ANNALES D'ÉCONOMIE ET DE STATISTIQUE. – N° 54 – 1999

Causalité de long terme et amélioration de la prévision : application aux courbes de taux d'intérêt

Catherine BRUNEAU, Éric JONDEAU

RÉSUMÉ. – Nous proposons une définition et une caractérisation de la propriété de causalité de long terme entre séries non-stationnaires, éventuellement cointégrées. Dans une représentation VAR, la condition de non-causalité de long terme peut être testée à partir d'un test de WALD standard, dont la statistique suit un chi-deux, une fois spécifiées les propriétés de long-terme de la dynamique. Cette méthodologie permet d'analyser les liens de causalité de long terme entre les taux d'intérêt d'une même courbe de taux, pour différents pays, entre 1983 et 1996. Contrairement aux prédictions de la théorie des anticipations, les différents taux ne jouent pas un rôle symétrique dans leur contribution à la tendance commune. Celle-ci apparaît en effet guidée principalement par les taux les plus longs, qui sont alors susceptibles de jouer le rôle d'indicateur des anticipations des marchés concernant la politique monétaire.

Long-Term Causality and Prediction Improvement with an Application to Interest Rate Term Structures

ABSTRACT. – We propose a definition and a characterization of long-run causality between non-stationary, possibly cointegrated, series. In a VAR framework, a WALD test can be performed to test for long-run non-causality, with the statistics distributed as a chi-square, conditionally on the cointegration rank. This methodology is used to study long-run causal links between the interest rates of euro-currency term structures, for different countries, between 1983 and 1996. Contrary to the implications of the expectations hypothesis, the interest rates of different maturities do not play a symmetrical role, when contributing to the common trend of the yield curve. The common trend appears to be mainly led by longest term rates, which can be used as an indicator of market participants expectations about future monetary policy.

---

* C. BRUNEAU, THEMA, Université de Paris X-Nanterre et Banque de France, Centre de Recherche. E. JONDEAU, Banque de France, Centre de Recherche.
Nous remercions les deux rapporteurs anonymes de la Revue pour leurs suggestions et commentaires. Nous restons bien sûr seuls responsables des erreurs qui peuvent subsister.

1 Introduction

Le contenu en information de la courbe des taux a été fréquemment étudié dans la littérature. De nombreuses études ont en effet montré que la pente des taux peut être utile pour anticiper l'évolution future de la croissance (HARVEY [1988], ESTRELLA et HARDOUVELIS [1991]), de l'inflation (FAMA [1990], MISHKIN [1990a], [1990b], [1991], JORION et MISHKIN [1991]) ou de la politique monétaire (FAMA et BLISS [1987], FAMA [1990], JORION et MISHKIN [1991]).

L'étude de la capacité prédictive de la pente des taux concernant l'évolution des taux d'intérêt repose en général directement sur les implications de la théorie des anticipations. Cette théorie, fondée sur l'hypothèse jointe d'absence d'opportunités d'arbitrage et de rationalité des anticipations, établit qu'un taux long est égal à la moyenne de taux courts futurs anticipés plus, éventuellement, une prime de risque constante dans le temps (SHILLER [1979]). Dans ce cadre d'analyse, on peut caractériser simplement le contenu en information de la pente des taux (CAMPBELL et SHILLER [1987], [1988]) : selon la théorie des anticipations, lorsque le taux court est non-stationnaire, le taux long doit l'être également, mais les deux taux doivent être cointégrés et la pente des taux caractérise la variable d'écart à l'équilibre dans la spécification « à correction d'erreur » de la dynamique (ENGLE et GRANGER [1987]). La pente des taux doit causer – au sens de GRANGER [1969] – les variations des deux taux. Autrement dit, la connaissance de la pente permet d'améliorer la prévision des variations futures des deux taux.

Le rejet empirique de la théorie des anticipations fait l'objet d'un large consensus sur données américaines (CAMPBELL et SHILLER [1991], HARDOUVELIS [1994], EVANS et LEWIS [1994], CAMPBELL [1995]). Les résultats obtenus pour d'autres pays, notamment européens, apparaissent toutefois plus favorables à la théorie (JORION et MISHKIN [1991], HARDOUVELIS [1994]). Dans tous les cas, elle constitue une référence privilégiée par la majorité des auteurs pour étudier la dynamique des taux d'intérêt. En particulier, cette hypothèse justifie le mode d'analyse adopté pour étudier la capacité prédictive de la pente des taux. Ainsi, certains auteurs, comme CAMPBELL et SHILLER [1991] et CAMPBELL [1995], tout en rejetant statistiquement la théorie, s'attachent à montrer que l'observation des mouvements de la pente des taux apporte une information utile pour prévoir les taux d'intérêt.

L'approche développée par CAMPBELL et SHILLER demeure toutefois fondamentalement bivariée. Cette démarche a été étendue au cas multivarié par HALL *et alii* [1992] puis ENGSTED et TANGGAARD [1994]. Dans un système de N taux d'intérêt représentant différentes maturités d'une même courbe des taux, la théorie des anticipations impose $(N-1)$ relations de cointégration, dont les résidus sont les $(N-1)$ pentes des taux vis-à-vis du taux le plus court. Il existe donc une seule tendance commune guidant l'ensemble de la courbe de taux. Cependant, comme le notent HALL *et alii* [1992] et ENGSTED et TANGGAARD [1994], contrairement aux implications de la théorie, les taux n'apparaissent pas, dans les applications empiriques, strictement substituables, puisque les effets des pentes des taux sur les variations de taux ne sont pas toujours conformes à la théorie. En particulier, HALL *et alii* [1992] montrent qu'aucune pente des taux n'influence la dynamique du taux le plus long.

Ce résultat est intéressant à considérer lorsqu'on étudie le contenu en information de la courbe des taux. En effet, d'après GRANGER [1988], elle s'interprète comme l'absence de causalité des taux les plus courts vers le taux le plus long, à un horizon de long terme, selon l'idée intuitive que les liens de causalité de long terme doivent être transmis par le mécanisme à correction d'erreur. Dès lors, les taux longs joueraient un rôle de leader dans la détermination de la tendance de la courbe des taux. La notion de causalité de long terme proposée par GRANGER [1988] est convaincante d'un point de vue intuitif, mais ne peut toutefois pas être interprétée en termes d'amélioration de la prévision, pour les variables en niveau et à l'horizon du long terme.

Nous proposons dans cet article une caractérisation explicite de la causalité de long terme comme une propriété d'amélioration de la prévision à un horizon infini et nous donnons une condition nécessaire et suffisante de non-causalité dans le cadre d'une dynamique VAR éventuellement cointégrée.

L'examen de cette condition montre que la non-significativité des forces de rappel n'est pas suffisante pour exclure tous liens de causalité à long terme.

Les liens de causalité de long terme entre les taux sur euro-devise sont étudiés pour les États-Unis, l'Allemagne, la France et le Royaume-Uni, sur la période 1983-96, en fréquence hebdomadaire et pour des maturités allant d'une semaine à un an. Cette analyse permet de conclure au rôle de leader du taux le plus long. En effet, le taux à un an (et à 3 mois pour les Etats-Unis et la France) n'est causé par aucun autre taux, mais cause tous les taux, à l'horizon du long terme. Les taux associés aux maturités longues des marchés monétaire ou euro-devise pourraient donc être utilisés par les Banques centrales comme des indicateurs des anticipations des marchés concernant la politique monétaire, tout particulièrement parce qu'ils sont susceptibles d'influencer les anticipations de long terme des agents.

Le plan de l'article est le suivant. Dans la section 2 sont rappelées les implications pour la courbe des taux, dans un cadre multivarié, de la théorie des anticipations de la structure par terme, qui justifient une représentation de la dynamique sous la forme d'un modèle à correction d'erreur (MCE). Dans la section 3, nous proposons une caractérisation de la propriété de causalité de long terme, énoncée en termes d'amélioration de la prévision à un horizon infini, lorsque la dynamique est caractérisée par un modèle VAR d'ordre fini, intégré, éventuellement cointégré. Nous décrivons une procédure de test de la non-causalité de long terme dans le cas où le modèle VAR est d'ordre et de rang de cointégration donnés. Dans la section 4, nous appliquons cette méthodologie pour analyser les liens de causalité de long terme entre les taux d'intérêt sur euro-devise des différents pays cités. La section 5 conclut l'article.

2 La théorie des anticipations de la structure par terme dans un cadre multivarié

La théorie des anticipations de la structure par terme établit que le rendement en t, noté $R_t^{(m_j)}$, d'un titre zéro-coupon de maturité m_j est égal au rendement moyen anticipé d'une succession d'investissements en t, $t+m_1,\dots,t+m_j-m_1$ en bons de court terme, de maturité m_1, plus une prime de risque constante dans le temps (SHILLER [1979]), soit :

$$R_t^{(m_j)} = \frac{m_1}{m_j} \sum_{i=0}^{\frac{m_j}{m_1}-1} E_t R_{t+im_1}^{(m_1)} + \varphi^{(m_1,m_j)} \qquad j=2,\dots,N \tag{1}$$

où $\frac{m_j}{m_1}$ est un entier et E_t représente l'espérance mathématique conditionnelle à l'information disponible à la date t. La prime $\varphi^{(m_1,m_j)}$ peut éventuellement dépendre des maturités m_1 et m_j mais doit être constante dans le temps.

Dans le cas bivarié, CAMPBELL et SHILLER [1987] ont montré que, si le taux court est non-stationnaire [1], alors le taux long doit également être non-stationnaire. Il doit toutefois exister une relation de cointégration entre ces deux taux, la pente des taux entre les maturités m_1 et m_j,

$$S_t^{(m_1,m_j)} = R_t^{(m_j)} - R_t^{(m_1)},$$

devant être stationnaire.

Ces implications de la théorie des anticipations ont été généralisées par HALL *et alii* [1992] au cas multivarié dans lequel on considère un ensemble de maturités $\{m_1,\dots,m_N\}$ avec $m_1 \leqslant \cdots \leqslant m_N$: si le taux le plus court $R_t^{(m_1)}$ est non-stationnaire, alors tous les autres taux doivent également être non-stationnaires, mais les différentes pentes des taux vis-à-vis du taux le plus court doivent être stationnaires. Il doit donc exister $(N-1)$ relations de cointégration dans un système de N taux d'intérêt et elles doivent satisfaire la contrainte selon laquelle la somme des coefficients de chaque vecteur de cointégration est nulle. Ainsi, si on note $X_t = \left(R_t^{(m_1)} \; R_t^{(m_2)} \; \cdots \; R_t^{(m_N)}\right)'$, le vecteur de cointégration s'écrit $Z_t = \beta' X_t$, de dimension $(N-1)$, où β est la matrice des paramètres de long terme, de dimension $(N, N-1)$. La théorie des anticipations implique que Z_t est stationnaire avec β tel que les sommes des différentes colonnes sont nulles. On vérifie aisément que cette contrainte sur les paramètres de β est équivalente à celle selon laquelle l'espace de cointégration est engendré par les colonnes de H défini par :

$$\underset{(N,N-1)}{H} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ -1 & 0 & \cdots & 0 \\ 0 & -1 & & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & -1 \end{bmatrix}. \tag{2}$$

1. Une telle hypothèse est généralement acceptée empiriquement pour les taux d'intérêt, bien que, en théorie, ils ne peuvent pas être intégrés, puisqu'ils sont bornés par zéro.

Lorsque la théorie des anticipations est valide, les pentes des taux représentent des relations d'équilibre entre les taux d'intérêt. Comme l'ont montré CAMPBELL et SHILLER [1987] dans le cas bivarié, chacun des taux d'intérêt doit être causé par le terme à correction d'erreur. De même, dans le cas multivarié, si un taux d'intérêt n'est causé par aucune des pentes, on doit rejeter la théorie. Or, en étudiant l'effet des termes à correction d'erreur dans la représentation MCE, HALL *et alii* [1992] concluent que la variation des taux les plus longs n'est causée par aucune pente des taux. L'une des implications de la théorie est donc clairement rejetée. De plus, selon GRANGER [1988], ce résultat implique que les différents taux d'intérêt ne jouent pas un rôle symétrique à long terme.

Le rejet de la théorie pourrait donc s'expliquer notamment par l'existence de relations causales unidirectionnelles à long terme entre les taux des différentes maturités. Autrement dit, il devrait exister un (ou plusieurs) taux *leader* à long terme, ce qui est contraire à la théorie des anticipations. C'est ce point que nous étudions dans la suite de l'article. Nous proposons en effet d'étudier la propriété de causalité de long terme comme une propriété explicite d'amélioration de la prévision, pour un horizon infini. En montrant alors que le taux long est causal à long terme, on peut en déduire qu'il possède une capacité prédictive effective lorsque l'on s'intéresse à des prévisions de long terme des taux d'intérêt, notamment de maturité courte.

3 Causalité de long terme

3.1. Définition et caractérisation

D'une manière générale, la causalité au sens de GRANGER est définie comme une propriété d'amélioration de la prévision. Ici, nous nous intéressons à la causalité de long terme et nous nous concentrons donc sur les propriétés d'amélioration des prévisions à un horizon de « long terme », selon la définition suivante :

DÉFINITION 1 : Etant donné un processus $X = (X_1, \ldots, X_N)'$, intégré d'ordre un, il n'existe pas de liens de causalité unidirectionnel de long terme de X_j vers X_k si et seulement si, à chaque date t, la connaissance des valeurs passées des variables $X_{j,t-l}, l \geqslant 0$, n'améliore pas la prévision linéaire de long terme de X_k, soit, $\forall t \geqslant 1$:

$$\lim_{h \to +\infty} EL\left(X_{k,t+h} \middle| \{\underline{X_{it}}; 1 \leqslant i \leqslant N\}\right) = \lim_{h \to +\infty} EL\left(X_{k,t+h} \middle| \{\underline{X_{it}}; 1 \leqslant i \leqslant N, i \neq j\}\right).$$

On recherche alors une condition testable de la non-causalité de long terme ainsi définie. On suppose que la dynamique d'un processus X_t, de dimension N, obéit à un modèle VAR en niveau, d'ordre fini p :

$$\Phi(L)\, X_t = \varepsilon_t, \forall t \geqslant 1 \tag{3}$$

où $\Phi(L) = I_N - \sum_{i=1}^{p} \Phi_i L^i$, $\Phi(0) = I_N$, et $\{\varepsilon_t\}_{t \geqslant 1}$ est un bruit blanc de matrice de variance régulière, notée Σ. Les racines de $\det(\Phi(x))$ sont toutes supposées être situées sur ou en dehors du disque unité, de sorte que $\{\varepsilon_t\}_{t \geqslant 1}$ est le processus des innovations canoniques de X. I_N est la matrice identité de dimension (N,N). Dans ces conditions, la propriété de non-causalité de long terme peut être caractérisée comme suit :

PROPOSITION 2 : Soit un processus de dimension N, $X = (X_1, \ldots, X_N)'$, dont la dynamique, intégrée d'ordre un, est décrite par le modèle (3). X_j ne cause pas X_k à long terme, dans l'univers $\{X_1, \ldots, X_N\}$, si et seulement la condition :

$$\left\{ \sum_i C_{ki}(1) \Phi_{ij}(L) = 0 \right\}$$

ou, de manière équivalente, la condition :

$$\left\{ C_{kj}(1) = 0 \text{ et } \sum_{i \neq j} C_{ki}(1) \Phi_{ij}(L) = 0 \right\}$$

est satisfaite, où $C(L)$ caractérise la décomposition de WOLD des différences premières :

(4) $$\Delta X_t = C(L) \varepsilon_t.$$

Preuve : Annexe 1.

3.2. Liens avec des notions usuelles

Il est intéressant de faire le lien entre la propriété de non-causalité de long terme précédente et des propriétés comparables étudiées dans la littérature. Il s'agit, d'une part, de la propriété de neutralité statistique introduite par STOCK et WATSON [1989] dans un cadre bivarié, intégré et non-cointégré et, d'autre part de la nullité des forces de rappel dans les modèles à correction d'erreur utilisés pour caractériser les dynamiques intégrées et cointégrées. Nous parlerons aussi brièvement de la propriété de causalité persistante étudiée par BRUNEAU et NICOLAÏ [1995].

3.2.1. Neutralité de STOCK et WATSON [1989]

La condition $\{C_{kj}(1) = 0\}$ qui intervient dans la proposition précédente correspond très précisément à la condition de neutralité statistique introduite par STOCK et WATSON [1989], dans un cadre bivarié, intégré et non-cointégré. Cette condition apparaît donc comme une condition nécessaire mais non suffisante de non-causalité de long terme, excepté dans le cas bivarié. Cependant dans le cas bivarié, il convient de remarquer que l'absence de causalité de long terme correspond alors simplement à l'absence de causalité (unidirectionnelle) au sens usuel de GRANGER puisqu'on a l'équivalence

$$\{C_{12}(1) = 0 \text{ et } C_{11}(1) \Phi_{12}(L) = 0\} \Longleftrightarrow \{\Phi_{12}(L) = 0\}.$$

La preuve est la suivante.

i) Condition nécessaire : en l'absence de causalité de long terme, $C_{12}(1)$ est nul et par conséquent $C_{11}(1)$ ne l'est pas, sinon X_1 serait stationnaire. $\Phi_{12}(L)$ doit donc être nul, ce qui exclut la causalité unidirectionnelle de X_2 vers X_1 au sens usuel de GRANGER [1969].

ii) Réciproquement, la nullité de $\Phi_{12}(L)$ implique la nullité de $C_{12}(L)$ (et donc celle de $C_{12}(1)$) parce que la dynamique est bivariée. En effet, quelle que soit la dimension N du système, si 1 est racine multiple d'ordre d de $\det(\Phi(L))$, alors $(1-L)^{d-1}$ divise chaque élément de l'adjointe $\operatorname{adj}(\Phi(L))$, que la dynamique soit ou non cointégrée, lorsque la dynamique est intégrée d'ordre 1 (ENGLE et GRANGER [1987]) :

$$\det(\Phi(L)) = (1-L)^d\,\varphi(L) \text{ et } \operatorname{adj}(\Phi(L)) = (1-L)^{d-1}\,\widetilde{\Phi}(L)$$

Par suite, les opérateurs $C(L)$ et $\Phi(L)$ sont liés selon :

$$C(L) = \frac{\operatorname{adj}(\Phi(L))}{\varphi(L)(1-L)^{d-1}}$$

et, par conséquent, dans le cas bivarié, on a toujours l'équivalence :

$$\{\Phi_{12}(L) = 0\} \Longleftrightarrow \{C_{12}(L) = 0\}.$$

3.2.2. **Causalité de long terme et significativité des « termes à correction d'erreur »**

Il convient de souligner en premier lieu que l'examen des termes à correction d'erreur dans la représentation MCE n'apporte des informations sur les liens de causalité de long terme que dans des cas très particuliers.

Dans le cas bivarié considéré par GRANGER [1988] et GRANGER et LIN [1995], les termes à correction d'erreur vérifient :

$$C(1)\,(\alpha_1,\alpha_2)' = (0,0)'.$$

Or, sous la condition de non-causalité de long terme, $C_{12}(1)$ est nul et donc $C_{11}(1)\,\alpha_1 = 0$. Mais $C_{11}(1)$ ne peut pas être nul, sinon X_1 serait stationnaire. Par conséquent, α_1 est nul et cette propriété, interprétée comme une indication de non-causalité de long terme dans GRANGER [1988] et GRANGER et LIN [1995], apparaît, de ce fait, comme une condition nécessaire mais non suffisante de non-causalité de long terme. Cependant, en utilisant les mêmes arguments que dans le cas non-cointégré, on peut montrer que l'absence de causalité de long terme est équivalente à l'absence de causalité au sens usuel de GRANGER.

Dans les systèmes de plus grande dimension, intégrés d'ordre 1 et cointégrés, l'examen des termes à correction d'erreur n'apporte généralement aucune information sur les liens de causalité de long terme. Toutefois, lorsque le système présente une seule tendance commune, la nullité de tous les termes à correction d'erreur dans l'équation de la variable causée X_k est équivalente à la « neutralité statistique » de tous les variables X_i $(i \neq k)$ vis-à-vis de la variable X_k :

$$\{\forall i \neq j,\ C_{ki}(1) = 0\} \Longleftrightarrow \{\forall l, 1 \leqslant l \leqslant N-1,\ \alpha_{kl} = 0\}$$

en notant α_{ij} les éléments de la matrice α, de dimension $(N, N-1)$, des termes à correction d'erreur. La preuve est la suivante.

i) Condition nécessaire : $C(1)$ est de rang 1 et s'écrit donc :

$$C(1) = (b_1, \ldots, b_N)' (d_1, \ldots, d_N)$$

où les b_i et d_i désignent des scalaires. Par suite, on a l'équivalence :

$$\begin{aligned}\{\forall i \neq j, C_{ki}(1) = 0\} &\Longleftrightarrow \{\forall i \neq k, b_k d_i = 0\} \\ &\Longleftrightarrow \{\forall i \neq k, d_i = 0\}\end{aligned}$$

parce que b_k ne peut pas être nul, sinon X_k serait stationnaire. $C(1)$ a alors une seule colonne non nulle, la *k*ème, et la propriété, systématiquement vérifiée, $C(1)\alpha = 0$, implique la nullité jointe des termes à correction d'erreur de la *k*ème équation :

$$\alpha_{k,1} = ... = \alpha_{k,N-1} = 0.$$

ii) Réciproquement, si tous les coefficients α_{kl} sont nuls, pour 1 allant de 1 à $N-1$, les coefficients d_i, $i \neq k$, satisfont le système de $N-1$ équations à $N-1$ inconnues :

$$\forall l, 1 \leqslant l \leqslant N-1, \sum_{i \neq k} d_i \alpha_{il} = 0.$$

Or le déterminant de la matrice $(\alpha_{il})_{i \neq k, 1 \leqslant l \leqslant N-1,}$ de dimension $(N-1, N-1)$ est non nul parce que la matrice α est de rang $N-1$. On en déduit la nullité jointe des coefficients d_i, $i \neq k$. La matrice $C(1)$ a alors une seule colonne non nulle, la *k*ème, ce qui implique la condition de neutralité annoncée $\{\forall i \neq j, C_{ki}(1) = 0\}$.

On notera toutefois, que même dans ce cas, la dernière condition est une condition nécessaire mais non suffisante de non-causalité de long terme, parce qu'il faut encore imposer la nullité de $\Phi_{kj}(L)$, c'est-à-dire l'absence de causalité au sens usuel de GRANGER de la variable X_j vers la variable X_k telle qu'elle est étudiée entre les variables en niveau dans TODA et PHILLIPS [1993], par exemple.

3.2.3. **Causalité persistante à la SIMS (BRUNEAU et NICOLAÏ [1995])**

BRUNEAU et NICOLAÏ [1995] s'intéressent à une caractérisation de la causalité persistante différente de celle qui est proposée ici. La causalité persistante, qui est une causalité de long terme, est ainsi associée à une propriété d'amélioration de la prévision à un horizon infini, mais le pouvoir prédictif en jeu concerne la variable causale courante, et non l'intégralité de son passé. La causalité est alors abordée selon les principes préconisés par SIMS [1980] : il s'agit d'étudier la contribution d'un choc sur la variable causale à la variance de l'erreur de prévision de la variable causée. Dans ce cas, la non-causalité de long terme est équivalente à la nullité d'un multiplicateur dynamique de long terme « structurel » associée à une orthogonalisation particulière des innovations :

$$\theta_{kj}(1) = 0 \quad \text{où} \quad \theta(L) = C(L)P$$

et $\omega_t = P^{-1}\varepsilon_t$ a pour variance l'identité.

3.3. Test de la non-causalité de long terme

La mise en œuvre du test de non-causalité de long terme repose sur l'estimation du MCE, représentation alternative du modèle VAR en niveau (3)

$$\Gamma(L)\,\Delta X_t = \Pi X_{t-1} + \varepsilon_t \tag{5}$$

où $\Gamma(L) = I_N - \sum_{i=1}^{p-1} \Gamma_i L^i,\ \Gamma_i = -\sum_{j=i+1}^{p} \Phi_j,\ i = 1,\ldots,p-1,$

et
$$\Pi = -\Big(I_N - \sum_{i=1}^{p} \Phi_i\Big).$$

S'il existe r relations de cointégration ($r \leqslant N$), la matrice Π s'écrit comme le produit $\alpha\beta'$ de deux matrices α et β de dimensions (N,r) de plein rang colonne r. Les colonnes de β définissent les vecteurs de cointégration. Comme le processus ΔX_t est stationnaire, ΠX_{t-1} doit aussi être stationnaire et s'écrire plus précisément comme la transformée linéaire de $Z_{t-1} = \beta' X_{t-1}$. L'équation (5) peut donc être récrite sous la forme :

$$\Gamma(L)\,\Delta X_t = \alpha Z_{t-1} + \varepsilon_t. \tag{6}$$

Les paramètres d'intérêt sont notés $\theta = [-\Gamma \ \ -\alpha]$, avec $\Gamma = \left[\Gamma_1,\ldots,\Gamma_{p-1}\right]$, mais nous utilisons également par la suite $\kappa = [I_N \ \ \theta]$. Les estimations de θ et de κ sont notées $\widehat{\theta} = \left[-\widehat{\Gamma} \ \ -\widehat{\alpha}\right]$ et $\widehat{\kappa} = \left[I_N \ \ \widehat{\theta}\right]$ respectivement.

L'annexe 2 détaille le passage des paramètres du MCE aux paramètres du VAR en niveau, $\Phi(L)$, et de la représentation de WOLD, $C(L)$. Elle précise également les distributions asymptotiques nécessaires pour établir la distribution asymptotique de la statistique de test. On montre en particulier que les paramètres du VAR en niveau et les multiplicateurs dynamiques de long terme s'écrivent sous la forme suivante, en fonction des paramètres du MCE :

$$\Phi = \left[I_N \ \ -\Phi_1 \ \ \cdots \ \ -\Phi_p\right] = [I_N \ \ -\Gamma \ \ -\alpha]\,D = \kappa D$$

et
$$C(1) = M^{-1} D(1)\,(\kappa G)^{-1}$$

avec $M = \begin{bmatrix} S_{N-r} \\ \beta' \end{bmatrix},\ D(1) = \begin{bmatrix} I_{N-r} & 0_{(N-r,r)} \\ 0_{(r,N-r)} & 0_{(r,r)} \end{bmatrix},$

$S_{N-r} = \left[I_{N-r} \ \ 0_{(N-r,r)}\right].$

D et G sont des matrices définies dans l'annexe 2.

Le test de l'hypothèse nulle selon laquelle X_j n'est pas une variable causale à long terme pour X_k nécessite l'estimation de l'expression : $\sum_{i=1}^{N} C_{ki}(1)\,\Phi_{ij}(L)$ pour j et k donnés, k. Cette expression peut être récrite sous la forme $u_k' C(1)\Phi v_j$. u_k est une matrice de sélection, de dimension $(N,1)$, dont la $k$ème ligne contient un 1 et les autres 0 ; v_j est une matrice, de dimension $(N(p+1),p)$, dont la $l$ème colonne contient un 1 dans la $(N(l-1)+j)$ ème ligne, pour $l = 1,\ldots,p$.

On définit la fonction :

$$g_{kj}(\theta) = vec\left(u_k' C(1)\,\Phi v_j\right) = \Lambda_{kj}\mathrm{vec}\left[(\kappa G)^{-1}\kappa D\right] \tag{7}$$

avec $\Lambda_{kj} = v_j' \otimes u_k' M^{-1} D(1)$. « vec » est l'opérateur de vectorialisation colonne. La distribution asymptotique de $g_{kj}(\theta)$ s'écrit alors sous la forme

$$\sqrt{T}\left(g_{kj}\left(\widehat{\theta}\right) - g_{kj}(\theta)\right) \xrightarrow{d} N\left(0, \Sigma_{g_{kj}}\right), \quad \forall k, j = 1, \ldots, N, k \neq j \tag{8}$$

avec $$\Sigma_{g_{kj}} = \frac{\partial g_{kj}}{\partial vec(\theta)'}\,\Sigma_\theta\,\frac{\partial g_{kj}'}{\partial vec(\theta)}$$

dont l'expression est précisée dans l'annexe 2.

Le test de la non-causalité de long terme, $H_0 : g_{kj}(\theta) = 0$, est finalement fondé sur la statistique de WALD usuelle :

$$\xi_{kj} = T g_{kj}'\left(\widehat{\theta}\right)\widehat{\Sigma}_{g_{kj}}^{-1} g_{kj}\left(\widehat{\theta}\right), \quad \forall k, j = 1, \ldots, N, k \neq j, \tag{9}$$

qui est distribuée comme un χ^2 à p degrés de liberté.

Une fois le MCE estimé, on peut aisément calculer la statistique de test ξ_{kj} et mettre en œuvre le test de non-causalité de long terme.

4 Application empirique

4.1. Propriétés des taux d'intérêt et cointégration

L'analyse des liens de causalité de long terme porte sur les taux sur euro-devise pour les États-Unis, l'Allemagne, la France et le Royaume-Uni.

L'échantillon couvre des données hebdomadaires (cours de clôture du vendredi) du 1er janvier 1983 au 27 décembre 1996, pour les maturités 7 jours, 1 mois, 3 mois et 1 an [2].

Les résultats des tests ADF pour l'hypothèse nulle de non-stationnarité sont présentés dans le tableau 1a. Ils indiquent que chaque série de taux d'intérêt peut être considérée comme stationnaire en différence première. Ces résultats sont confirmés par les tests multivariés réalisés dans le cadre du MCE estimé par maximum de vraisemblance, selon la méthodologie développée par JOHANSEN [1988] (voir également JOHANSEN et JUSELIUS [1992]) (tableau 1b).

Dans un système de N taux d'intérêt, la théorie des anticipations implique $(N-1)$ relations de cointégration. L'estimation de la dynamique du système est fondée sur la représentation MCE, à partir de la méthodologie proposée par JOHANSEN [1988]. Cette procédure du maximum de vraisemblance permet

2. Toutes les estimations ont été réalisées avec GAUSS. Les données proviennent de la base Datastream. Pour les États-Unis et la France, les données ont été corrigées pour exclure des variations trop importantes des taux d'intérêt. Pour les taux américains, nous avons corrigé certains taux de fin d'année (26 décembre 1986 ; 25 décembre 1987 ; 21 décembre 1990), sous la forme d'une moyenne du taux du vendredi précédent et du vendredi suivant. Pour les taux français, nous avons également corrigé les données correspondant aux crises de change de mars 1983 (4, 11 et 18 mars). Pour l'ensemble, 6 points ont été corrigés.

TABLEAU 1a

***Test ADF de non-stationnarité des taux sur euro-devise*.**

Pays	Maturité			
	7 jours	1 mois	3 mois	1 an
Niveau				
États-Unis	– 1,60	– 1,49	– 1,57	– 1,55
Allemagne	– 0,68	– 0,66	– 0,65	– 0,71
France	– 1,16	– 1,78	– 1,76	– 1,51
Roy.-Uni	– 1,10	– 1,08	– 1,12	– 1,15
Variation				
États-Unis	– 5,63 [a]	– 5,99 [a]	– 6,64 [a]	– 7,34 [a]
Allemagne	– 7,93 [a]	– 6,65 [a]	– 6,68 [a]	– 7,19 [a]
France	– 9,56 [a]	– 9,28 [a]	– 9,45 [a]	– 8,34 [a]
Roy.-Uni	– 8,33 [a]	– 7,59 [a]	– 7,82 [a]	– 7,85 [a]

TABLEAU 1b

***Test multivarié de non-stationnarité des taux sur euro-devise*.**

Pays	Maturité			
	7 jours	1 mois	3 mois	1 an
États-Unis	23,03 [a]	23,05 [a]	23,10 [a]	23,19 [a]
Allemagne	19,17 [a]	19,19 [a]	19,24 [a]	19,34 [a]
France	36,27 [a]	43,50 [a]	45,07 [a]	47,28 [a]
Roy.-Uni	20,57 [a]	20,60 [a]	20,68 [a]	20,89 [a]

Note : Les tests multivariés de non-stationnarité sont ceux suggérés par la méthodologie du maximum de vraisemblance (JOHANSEN et JUSELIUS [1992]). Ils sont fondés sur une constante cointégrée et $r = 3$ relations de cointégration (configuration retenue dans les estimations) et $p = 4$ retards (le choix du nombre de retards ne modifie pas les conclusions de ce test). Les statistiques de test sont distribuées comme un χ^2 à $(p - r)$ degrés de liberté. [a] indique que la statistique est significative au seuil de 1 %.

notamment de tester le rang de cointégration, d'estimer les vecteurs de cointégration et de tester des restrictions sur ces vecteurs de cointégration. Le test du rang de cointégration repose sur les statistiques du rapport des maxima de vraisemblance proposées par JOHANSEN [1988] et JOHANSEN et JUSELIUS [1990] : la valeur propre maximale $\lambda_{\max}$ teste l'hypothèse nulle qu'il existe au plus r relations de cointégration contre l'hypothèse alternative qu'il en existe au plus $r + 1$; la statistique de la trace λ_{trace} teste l'hypothèse nulle qu'il existe au plus r relations de cointégration contre l'hypothèse alternative qu'il en existe au plus N.

Pour chaque pays, le MCE est estimé, par maximum de vraisemblance, pour $X_t = \left(R_t^{(1)}, R_t^{(4)}, R_t^{(13)}, R_t^{(52)}\right)'$, associés aux taux à 7 jours, 1 mois, 3 mois et 1 an. Les tests des composantes déterministes pour le processus autorégressif ont conduit à sélectionner deux types de modèles : sans composante déterministe ou avec une constante contrainte à l'espace de cointégration. La principale différence entre ces deux types de modèles repose sur la nullité de la prime de risque. Comme ce point n'est pas au centre de notre problématique, nous n'avons pas imposé la nullité de la prime et avons donc retenu une constante contrainte à appartenir à l'espace de cointégration. Le choix du nombre de retards p est fondé sur le critère d'information HQ (HANNAN et QUINN [1979]) [3].

Le tableau 2 donne le nombre de retards sélectionné pour chaque système et les résultats des tests sur les vecteurs de cointégration. Les statistiques λ_{max} et λ_{trace} sont toutes significatives pour $r \leqslant 3$ et non significatives pour $r \leqslant 4$ (au moins à un seuil de risque de 5 %). Cette première implication de « long terme » de la théorie des anticipations est donc largement confirmée par les données et nous pouvons conclure qu'il existe une tendance commune unique guidant chacune des courbes de taux étudiées.

TABLEAU 2

Tests du rang de cointégration et des contraintes portant sur les vecteurs de cointégration.

Pays	Nombre de retards optimal (p)	hyp.	Test du rang de cointégration λ_{max}	λ_{trace}	Test des contraintes $LM(3)$
États-Unis	3	$r \leqslant 4$	2,41	2,41	6,242
		$r \leqslant 3$	25,45 [a]	27,86 [a]	
Allemagne	3	$r \leqslant 4$	1,13	1,13	19,558 [a]
		$r \leqslant 3$	19,18 [b]	20,32 [b]	
France	11	$r \leqslant 4$	7,17	7,17	8,751 [b]
		$r \leqslant 3$	15,51 [c]	22,67 [b]	
Roy.-Uni	2	$r \leqslant 4$	2,89	2,89	20,640 [a]
		$r \leqslant 3$	29,37 [a]	29,37 [a]	

Note : r est le rang de cointégration. λ_{max} et λ_{trace} sont les statistiques proposées par JOHANSEN [1988] pour le test du rang de cointégration. Les valeurs critiques de λ_{max} et λ_{trace} proviennent de OSTERWALD-LENUM (1992). $LM(3)$ est la statistique du rapport des maxima de vraisemblance pour le test selon lequel les vecteurs de cointégration s'écrivent sous la forme (10). La statistique de test suit asymptotiquement un χ^2 à 3 degrés de liberté. [a], [b] et [c] indiquent que la statistique est significative aux seuils de 1 %, 5 % et 10 % respectivement.

Les tests de restriction sur le vecteur de cointégration sont moins homogènes : tous les vecteurs de cointégration devraient s'écrire sous la forme d'une pente vis-à-vis du taux le plus court (ici, le taux à 7 jours), ce qui conduit à tester une contrainte du type

$$\beta = \begin{pmatrix} H \\ \varphi \end{pmatrix} \tag{10}$$

où H est défini dans (2) et $\varphi = [\varphi_1 \ \ \varphi_2 \ \ \varphi_3]$ est le vecteur des constantes cointégrées (les primes de risque constantes). Les résultats de ces tests sont présentés dans la dernière colonne du tableau 2 : l'hypothèse nulle d'égalité à $(1, -1, \varphi_i)$ de chacun des vecteurs de cointégration n'est pas rejetée dans le cas des États-Unis, mais est rejetée pour l'Allemagne, la France et le Royaume-Uni au seuil de 5 %.

3. Le critère d'information HQ pour un modèle i est défini par $HQ_i = \ln(\det(\Sigma_i)) + 2p_i \ln(\ln(T))/T$, où Σ_i la matrice de variance-covariance des résidus du modèle i, p_i est le nombre de paramètres estimés du modèle i, T est le nombre d'observations.

Il est important de souligner l'intérêt que présente la caractérisation de la causalité de long terme proposée dans les sections précédentes. En effet, les travaux antérieurs font généralement référence aux termes à correction d'erreur pour caractériser le rôle de *leader* de l'une ou l'autre des maturités (HALL *et alii* [1992], ENGSTED et TANGGAARD [1994]. Ici, excepté dans le cas américain, les pentes ne sont pas stationnaires et les liens de causalité de long terme, caractérisés selon GRANGER [1988], ne peuvent donc plus être analysés en considérant l'influence des pentes sur la dynamique des différents taux comme dans CAMPBELL et SHILLER [1987]. Il convient alors d'étudier les liens de causalité de long terme entre les taux eux-mêmes, ce que nous effectuons dans la section suivante.

4.2. **Tests de neutralité et de non-causalité de long terme**

Tout d'abord, l'estimation des MCE donne les configurations suivantes, en ce qui concerne les termes à correction d'erreur (tableau 3) : pour les États-Unis, l'Allemagne et le Royaume-Uni (à un seuil de 5 % dans ce dernier cas), un terme à correction d'erreur au moins intervient dans chacune des équations, à l'exception de l'équation de taux à 1 an. Pour la France, un terme à correction d'erreur au moins intervient dans chacune des équations (y compris l'équation de taux à 1 an), conformément à la théorie des anticipations, mais certains ont un signe contraire à celui prédit par la théorie. Ainsi, dans tous les cas, ces implications de la théorie des anticipations, relatives aux effets à correction d'erreur, doivent être rejetées. Il apparaît ainsi que les taux d'intérêt ne jouent pas un rôle symétrique dans leur contribution à la tendance de la courbe de taux. Ces résultats confirment ceux obtenus par HALL *et alii* [1992] et ENGSTED et TANGGAARD [1994].

TABLEAU 3
Statistiques de student associées aux forces de rappel pour les MCE.

Pays		7 jours	1 mois	3 mois	1 an
États-Unis	relation 1	– 6,54 [a]	7,03 [a]	1,62	0,79
	relation 2	– 4,32 [a]	– 1,80 c	3,15 [a]	0,40
	relation 3	– 3,30 [a]	– 2,92 [a]	– 2,27 [b]	0,37
Allemagne	relation 1	– 14,31 [a]	– 0,13	– 0,60	– 0,80
	relation 2	– 5,00 [a]	– 6,15 [a]	1,06	0,74
	relation 3	– 1,94 [b]	– 3,37 [a]	– 3,82 [a]	– 1,54
France	relation 1	1,64	1,41	– 2,75 [a]	– 0,79
	relation 2	– 0,34	– 4,35 [a]	– 4,33 [a]	– 2,89 [a]
	relation 3	3,15 [a]	2,43 [b]	1,90 [c]	0,16
Roy.-Uni	relation 1	– 7,35 [a]	1,59	0,20	0,94
	relation 2	– 10,19 [a]	– 7,18 [a]	– 0,88	– 0,89
	relation 3	2,56 [b]	4,36 [a]	4,59 [a]	1,87

Note : Il s'agit des statistiques de student associées aux paramètres α dans les MCE (6). [a, b] et [c] indiquent que la statistique est significative aux seuils de 1 %, 5 % et 10 % respectivement.

L'analyse de la neutralité et des liens de causalité de long terme, caractérisés en termes d'amélioration de la prévision à un horizon infini, permet de préciser ces résultats. Le tableau 4 présente deux statistiques de test : la statistique associée au test de neutralité ($C_{kj}(1) = 0$) ; la statistique pour le test de non-causalité de long terme ($(C(1)\Phi(L))_{kj} = 0$). Les distributions asymptotiques de ces deux statistiques sont détaillées dans l'annexe 2. Le premier test peut être considéré comme une étape préliminaire au second, puisque la neutralité est une condition nécessaire de non-causalité de long terme.

Les tests de neutralité permettent de valider l'hypothèse, suggérée par HALL *et alii* [1992], selon laquelle les taux les plus longs jouent un rôle de *leader* dans la détermination de la tendance commune de la courbe des taux. Il apparaît en effet que, pour chacun des pays, le taux à 1 an est le seul taux dont la contribution à la tendance commune est significativement différente de 0.

TABLEAU 4

Test de neutralité et test de non-causalité de long terme.

Pays		7 jours	1 mois	3 mois	1 an
États-Unis	neutralité	0,97	0,09	0,07	2,28
	non-caus. de l.t.	1,07	0,89	8,96	7,47 [c]
Allemagne	neutralité	0,72	0,68	1,24	2,37 [b]
	non-caus. de l.t.	2,40	4,42	3,93	6,96 [c]
France	neutralité	1,60	1,52	0,86	4,08 [a]
	non-caus. de l.t.	15,15	11,35	18,33 [c]	34,17 [a]
Roy.-Uni	neutralité	0,95	1,31	0,93	3,46 [a]
	non-caus. de l.t.	1,11	1,72	0,87	17,86 [a]

Note : Les statistiques des test de neutralité suivent asymptotiquement une loi normale. Les statistiques des tests de non-causalité de long terme suivent un χ^2 à *p* degrés de liberté. [a, b] et [c] indiquent que la statistique est significative aux seuils de 1 %, 5 % et 10 % respectivement.

Mais, comme on l'a indiqué précédemment, il n'est pas suffisant de considérer la neutralité pour étudier les liens de causalité à long terme, quand la dimension du système est supérieure à 2. L'analyse de la causalité de long terme permet en particulier de tester la capacité des taux longs à prévoir l'évolution des taux les plus courts, représentatifs de la politique monétaire. On obtient les résultats suivants en ce qui concerne la causalité de long terme : pour l'Allemagne et le Royaume-Uni, le taux à 1 an est le seul taux ayant un effet causal à long terme sur les autres maturités ; pour les États-Unis et la France, ce sont les deux taux les plus longs (taux à 3 mois et à 1 an) qui ont un effet causal sur les autres maturités.

Globalement, les tests de causalité de long terme conduisent à la même conclusion pour l'ensemble des devises. A long terme, les structures par terme sont principalement guidées par les taux longs. Les taux les plus courts (7 jours et 1 mois) n'apparaissent jamais comme variables causales à l'horizon du long terme. Ces résultats montrent que les taux les plus longs des marchés monétaire et euro-devise (typiquement, le taux à 1 an) sont de bons indicateurs pour les autorités monétaires, car ils reflètent les anticipations de long terme des agents concernant notamment l'évolution de la politique monétaire. De tels indicateurs permettent ainsi d'extraire la partie la plus structurelle (et donc la moins volatile) de ces anticipations.

5 Conclusion

Dans cet article, nous nous sommes intéressés au rejet de la théorie des anticipations et, plus particulièrement, aux résultats empiriques selon lesquels les taux longs semblent guider de façon persistante la courbe des taux, indiquant un rôle dissymétrique des différentes maturités dans la contribution à la tendance commune de la courbe des taux. Nous avons mis l'accent sur une contrepartie intéressante du rejet de la théorie, en montrant que les taux les plus longs ont, de fait, un rôle causal lorsque l'on s'intéresse à la causalité de long terme, caractérisée comme une propriété d'amélioration de la prévision à un horizon infini. Nous avons montré que la condition de non-causalité ainsi définie est une condition aisée à tester, par un test de WALD standard, lorsque la dynamique est caractérisée par un modèle VAR d'ordre fini, éventuellement cointégré, dès lors que le rang de cointégration est donné.

En considérant quatre des principales maturités des structures par terme sur euro-devise, de janvier 1983 à décembre 1996, nous avons montré que les implications de la théorie des anticipations en termes de rang de cointégration sont globalement respectées, confirmant ainsi les résultats obtenus par HALL *et alii* [1992]. Utilisant alors le modèle à tendance commune correspondant, nous avons mené l'analyse des liens de causalité de long terme entre les différents taux. Nous avons conclu au rôle causal des taux longs, les faisant ainsi apparaître comme des indicateurs utiles pour les autorités monétaires, soucieuses de prendre en compte les anticipations de long terme de leur politique par les participants au marché.

Preuve de la proposition 2

On se place sous les conditions d'initialisation usuelles : toutes les innovations sont nulles avant la date 0 et $c_0 = \left\{X_{-h} = x_{-h};\, 0 \leqslant h \leqslant p-1\right\}$. Les limites considérées, lorsque H tend vers l'infini, sont au sens de la norme L^2.

i) On suppose d'abord que la série X_j ne cause pas X_k à long terme. Par suite, on a, pour toute date $t \geqslant 1$, l'égalité des prévisions linéaires :

$$\lim_{H\to+\infty} EL\big(X_{k,t+H}\big|\{\underline{X_{it}};\, 1 \leqslant i \leqslant N\};\, c_0\big)$$
$$= \lim_{H\to+\infty} EL\big(X_{k,t+H}\big|\{\underline{X_{it}};\, 1 \leqslant i \leqslant N, i \neq j\};\, c_0\big).$$

Sous les conditions d'initialisation adoptées, la limite :

$$\lim_{H\to+\infty} EL\big(X_{k,t+H}\big|\{\underline{X_{it}};\, 1 \leqslant i \leqslant N\};\, c_0\big)$$

peut être évaluée comme suit. Le processus des différences premières ΔX_t peut être décomposé selon [4] :

$$\Delta X_t = \big(C(1) + C^*(L)(1-L)\big)\,\varepsilon_t,$$

ce qui donne, sous les conditions d'initialisation, l'expression de X_t :

$$X_t = c + C(1)\sum_{s=1}^{t}\varepsilon_s + C^*(L)\varepsilon_t$$

où c désigne une constante qui dépend des valeurs initiales des séries.

Par suite, la meilleure prévision linéaire :

$$\lim_{H\to+\infty} EL\big(X_{k,t+H}\big|\{\underline{X_{it}};\, 1 \leqslant i \leqslant N;\, c_0\}\big)$$
$$= \lim_{H\to+\infty} EL\big(X_{k,t+H}\big|\{\underline{\varepsilon_{it}};\, 1 \leqslant i \leqslant N;\, c_0\}\big)$$

peut être exprimée sous la forme :

$$c_k + \sum_{i=1}^{N} C_{ki}(1)\sum_{s=1}^{t}\varepsilon_{is}.$$

En effet, les innovations de différentes dates sont non-corrélées et les coefficients $C^*_{kj,H}$ tendent (exponentiellement) vers zéro, lorsque H tend vers l'infini, parce que le processus ΔX est stationnaire.

Enfin, à partir de la caractérisation VAR de la dynamique, pour tout $s \geqslant 1$ et tout i, $1 \leqslant i \leqslant N$, les innovations ε_{is} peuvent être exprimées en fonction des variables $X_{ls'}$, $s' \leqslant s$, $1 \leqslant l \leqslant N$:

$$\varepsilon_{is} = \sum_{l=1}^{N}\Phi_{il}(L)X_{ls}$$

4. On peut « inverser » le VAR : $\det(\Phi(L))\, X_t = \operatorname{adj}(\Phi(L))\,\varepsilon_t$, avec $\det(\Phi(L)) = (1-L)^d\varphi(L)$ et $\operatorname{adj}(\Phi(L)) = (1-L)^{d-1}\,\widetilde{\Phi}(L)$ si d est l'ordre de multiplicité de la racine unitaire.

de sorte que, pour tout $t \geqslant 1$, la meilleure prévision de long terme est donnée par :

$$\lim_{H\to+\infty} EL\left(X_{k,t+H}\middle|\{\underline{X_{it}};\, 1 \leqslant i \leqslant N, i \neq j;\, c_0\}\right) = c_k + \sum_{h=0}^{t-1}\sum_{i=1}^{N} C_{ki}(1) \sum_{h'\geqslant 0}\sum_{l=1}^{N} \Phi_{il,h} X_{l,t-h-h'}.$$

Par ailleurs, comme la limite $\lim_{H\to+\infty} EL\left(X_{k,t+H}\middle|\{\underline{X_{it}};\ 1 \leqslant i \leqslant N, i \neq j;\ c_0\}\right)$ est une fonction des seules variables $X_{is},\ i \neq j,\ s \leqslant t$, la condition de non-causalité implique l'existence de coefficients $P_{l,h}^{(t,H)}$ tels que $P_{j,h}^{(t,H)} = 0$, pour tous t, h, h' satisfaisant :

$$c_k + \sum_{h=0}^{t-1}\sum_{i=1}^{N} C_{ki}(1) \sum_{h'\geqslant 0} \Phi_{il,h} \sum_{l=1}^{N} X_{l,t-h-h'} = c_k + \sum_{h=0}^{t-1}\sum_{i=1}^{N} P_{l,h}^{(t)} X_{l,t-h}.$$

Or les variables $X_{l,t-h}$ sont linéairement indépendantes au sens où, pour toute date t :

$$\sum_{h=0}^{t-1}\sum_{i=1}^{N} \lambda_{i,h}^{(t)} X_{i,t-h} = 0 \Longrightarrow \forall i, \forall h, \lambda_{i,h}^{(t)} = 0$$

parce que la matrice de variance-covariance des innovations canoniques est régulière (BRUNEAU et NICOLAÏ [1994]).

La propriété de non-causalité implique donc la nullité des coefficients des variables $X_{j,t-h},\ h \geqslant 0$, soit :

$$\forall t, \sum_{h=0}^{t-1}\sum_{i=1}^{N} C_{ki}(1) \sum_{h'\geqslant 0} \Phi_{ij,h'} L^{h+h'}$$

et par conséquent la condition annoncée, obtenue pour $t = 1$:

$$\forall h, \sum_{i=1}^{N} C_{ki}(1)\Phi_{ij,h} = 0$$

ii) Réciproquement, si on suppose la dernière condition satisfaite, on peut affirmer que la prévision de long terme :

$$\lim_{H\to+\infty} EL\left(X_{k,t+H}\middle|\{\underline{X_{it}};\, 1 \leqslant i \leqslant N;\, c_0\}\right) = c_k + \sum_{s=1}^{t}\sum_{i=1}^{N} C_{ki}(1)\varepsilon_{is}\,,$$

peut être écrite sous la forme :

$$\lim_{H\to+\infty} EL\left(X_{k,t+H}\middle|\{\underline{X_{it}};\, 1 \leqslant i \leqslant N;\, c_0\}\right) = c_k + \sum_{s=1}^{t}\sum_{l\neq j}\sum_{i=1}^{N} C_{ki}(1)\Phi_{il}(L)X_{ls}.$$

Enfin, par suite de l'optimalité des prévisions, on a l'identité :

$$\lim_{H\to+\infty} EL\left(X_{k,t+H}\middle|\{\underline{X_{it}};\, 1 \leqslant i \leqslant N, i \neq j;\, c_0\}\right) = c_k + \sum_{s=1}^{t}\sum_{l\neq j}\sum_{i=1}^{N} C_{ki}(1)\Phi_{il}(L)X_{ls}, \forall t \geqslant 0$$

de sorte que les deux meilleures prévisions linéaires sont identiques.

Mise en œuvre des tests

A. **Notations**

On suppose qu'il existe N variables $X_t = (X_{1t},\ldots,X_{Nt})'$, pour $t = 1,\ldots,T$, avec p observations hors échantillon disponibles. On définit $\Delta X = (\Delta X_1,\ldots, \ \Delta X_T)'$, de dimension (T,N) et on suppose que la dynamique de processus X_t obéit à un modèle VAR en niveau, d'ordre fini p :

$$\Phi(L)X_t = \varepsilon_t, \forall t \geqslant 1 \tag{11}$$

où
$$\Phi(L) = I_N - \sum_{i=1}^{p} \Phi_i L^i,\ \Phi(0) = I_N, \text{ et } \{\varepsilon_t\}_{t\geqslant 1}$$

est un bruit blanc de matrice de variance régulière, notée Σ. Les racines de $\det(\Phi(x))$ sont toutes supposées sur ou en dehors du disque unité, de sorte que $\{\varepsilon_t\}_{t\geqslant 1}$ est le processus des innovations canoniques de X.

La représentation à correction d'erreur s'écrit sous la forme (Engle et Granger [1987]) :

$$\Gamma(L)\Delta X_t = \Pi X_{t-1} + \varepsilon_t \tag{12}$$

où
$$\Gamma(L) = I_N - \sum_{i=1}^{p-1} \Gamma_i L^i,\ \Gamma_i = -\sum_{j=i+1}^{p} \Phi_j,\ i = 1,\ldots,p-1,$$

et
$$\Pi = -\left(I_N - \sum_{i=1}^{p} \Phi_i\right).$$

S'il existe r relations de cointégration $(r \leqslant N)$, la matrice Π s'écrit comme le produit $\alpha\beta'$ de deux matrices α et β de dimensions (N,r) de plein rang colonne r. Les colonnes de β définissent les vecteurs de cointégration. Comme le processus ΔX_t est stationnaire, ΠX_{t-1} doit aussi être stationnaire et s'écrire plus précisément comme la transformée linéaire de $Z_{t-1} = \beta' X_{t-1}$. L'équation (12) peut donc être récrite sous la forme :

$$\Gamma(L)\Delta X_t = \alpha Z_{t-1} + \varepsilon_t. \tag{13}$$

Les paramètres d'intérêt du MCE sont $\theta = [-\Gamma \ \ -\alpha]$, avec $\Gamma = \left[\Gamma_1,\ldots,\Gamma_{p-1}\right]$, mais nous utilisons également par la suite $\kappa = [I_N \ \ \theta]$. Les estimations de θ et de κ sont notées $\widehat{\theta} = \left[-\widehat{\Gamma} \ \ -\widehat{\alpha}\right]$ et $\widehat{\kappa} = \left[I_N \ \ \widehat{\theta}\right]$ respectivement. Enfin, on définit la matrice, de dimension $(N(Np+r), N(N(p-1)+r))$,

$$\zeta = \frac{\partial \text{vec}(\kappa)}{\partial \text{vec}(\theta)'} = \begin{bmatrix} 0_{(N^2, N(p-1)+r)} \\ I_{(N(p-1)+r)} \end{bmatrix}.$$

« vec » est l'opérateur de vectorialisation colonne.

Le VAR contraint (RVAR) constitue une représentation alternative des modèles VAR en niveau et MCE (Campbell et Shiller [1988], Mellander *et alii*, 1990) et s'écrit sous la forme :

(14) $$B(L)Y_t = \eta_t$$

où : $$Y_t = D_\perp(L)MX_t,\ M = \begin{bmatrix} S_{N-r} \\ \beta' \end{bmatrix},\ \eta_t = M\varepsilon_t,$$

$$D(L) = \begin{bmatrix} I_{N-r} & 0 \\ 0 & (1-L)I_r \end{bmatrix},\ D_\perp(L) = \begin{bmatrix} (1-L)I_{N-r} & 0 \\ 0 & I_r \end{bmatrix}.$$

La matrice S_{N-r}, de dimension $(N-r, N)$, peut être choisie de telle sorte que : $S_{N-r} = \begin{bmatrix} I_{N-r} & 0_{(N-r,r)} \end{bmatrix}$. On montre aisément qu'il existe la relation suivante entre les paramètres des modèles MCE et RVAR :

(15) $$B(L) = M\left[\Gamma(L)M^{-1}D(L) - \alpha^* L\right]$$

avec $\underset{(N,N)}{\alpha^*} = \begin{bmatrix} 0_{(N,N-r)} & \alpha \end{bmatrix}$. A partir de la représentation RVAR, les multiplicateurs dynamiques de long terme s'obtiennent alors simplement en inversant la matrice de polynômes $B(L)$:

(16) $$C(L) = M^{-1}D(L)B(L)^{-1}M$$

avec $C(L) = \sum_{h=0}^{\infty} C_h L^h$ (Mellander et *alii* [1990]).

B. **L'estimation des multiplicateurs dynamiques de long terme**

L'estimation des multiplicateurs dynamiques de long terme est donc fondée sur la relation (16), qui indique que, une fois estimées les matrices *M* et $B(L)$, on peut déduire directement les multiplicateurs dynamiques de long terme et leur distribution asymptotique.

Dans une première étape, on doit tester le rang de cointégration *r* et estimer le vecteur de cointégration β. Pour cela, on peut mettre en œuvre la procédure du maximum de vraisemblance de Johansen [1988] pour estimer les paramètres du MCE. Dans une seconde étape, les paramètres du RVAR sont déduits de ceux du MCE à l'aide de la relation (15). Pour obtenir la matrice de variance-covariance asymptotique des paramètres du RVAR, il est nécessaire de recourir à l'hypothèse supplémentaire que les paramètres de long terme, contenu dans *M*, peuvent être considérés comme connus. Ce résultat vient de la propriété de super-convergence des paramètres de long terme (Stock [1987], et Lütkepohl et Reimers [1992]). On définit, pour alléger les notations :

(17) $$F(1) = M^{-1}B(1) = \left[\Gamma(1)M^{-1}D(1) - \alpha^*\right] = \kappa G$$

avec $$\underset{(Np+r,N)}{G} = \begin{bmatrix} \Lambda M^{-1}J & 0_{(Np,r)} \\ 0_{(r,N-r)} & I_r \end{bmatrix},$$

$$\underset{(Np,N)}{\Lambda} = \frac{\partial \mathrm{vec}(\Gamma(1))}{\partial \mathrm{vec}(\Gamma)'} = (\underbrace{1 \ \cdots \ 1}_{p})' \otimes I_N$$

et $$\underset{(N,N-r)}{J} = \begin{bmatrix} I_{N-r} \\ 0_{(r,N-r)} \end{bmatrix}.$$

De l'équation (16), les multiplicateurs de long terme $C(1)$ sont estimés par :

$$\widehat{C}(1) = \widehat{M}^{-1} D(1) \widehat{F}(1)^{-1} = \widehat{M}^{-1} D(1) \left(\widehat{\kappa} \widehat{G}\right)^{-1}. \tag{18}$$

Finalement, la distribution asymptotique des multiplicateurs de long terme est donnée par la proposition suivante, en notant $X_{-1} = \left(X'_{1-p} \cdots X'_{T-p}\right)'$ et $\widetilde{X} = \left(\widetilde{X}'_1 \cdots \widetilde{X}'_T\right)'$ où $\widetilde{X}_t = \left(\Delta X_{t-1} \cdots \Delta X_{t-p}\right)$:

PROPOSITION 3 : Si X_t est un processus gaussien défini par (11), la distribution asymptotique des multiplicateurs de long terme est donnée par :

$$\sqrt{T} \operatorname{vec}\left(\widehat{C}(1) - C(1)\right) \xrightarrow{d} N\left(0, \Sigma_{C(1)}\right) \tag{19}$$

où :

$$\Sigma_{C(1)} = \left(\left(F(1)^{-1}\right)' \otimes M^{-1} D(1) F(1)^{-1}\right) \cdot \Sigma_{F(1)} \cdot \left(\left(F(1)^{-1}\right)' \otimes M^{-1} D(1) F(1)^{-1}\right)'$$

$$\Sigma_{F(1)} = (G' \otimes I_N) \cdot \zeta \cdot \Sigma_\theta \cdot \zeta' \cdot (G' \otimes I_N)'$$

$\Sigma_\theta = \Omega^{-1} \otimes \Sigma$ est la matrice de variance-covariance de $vec(\widehat{\theta})$.

$$\Omega = \operatorname{plim} \frac{1}{T} \begin{bmatrix} \widetilde{X}'\widetilde{X} & \widetilde{X}' X_{-1}\beta \\ \beta' X'_{-1} \widetilde{X} & \beta' X'_{-1} X_{-1} \beta \end{bmatrix}.$$

Σ_θ est estimée, de façon convergente, par $\widehat{\Sigma}_\theta = \widehat{\Omega}^{-1} \otimes \widehat{\Sigma}$, avec $\widehat{\Sigma}$ donné dans JOHANSEN [1988] et $\widehat{\Omega} = \frac{1}{T} \begin{bmatrix} \widetilde{X}'\widetilde{X} & \widetilde{X}' X_{-1}\widehat{\beta} \\ \widehat{\beta}' X'_{-1} \widetilde{X} & \widehat{\beta}' X'_{-1} X_{-1} \widehat{\beta} \end{bmatrix}$.

Preuve : Se déduit directement de la relation (18).

C. **Test de non-causalité de long terme**

Le test de l'hypothèse nulle selon laquelle X_j n'est pas une variable causale de long terme pour X_k repose sur l'estimation de l'expression : $\sum_{i=1}^{N} C_{ki}(1)\Phi_{ij}(L)$, pour j et k donnés, sous la forme $u'_k C(1) \Phi v_j$. La matrice Φ, de dimension $(N, N(p+1))$, contient les paramètres du VAR en niveau et s'écrit sous la forme suivante, en fonction des paramètres du MCE

$$\Phi = \left[I_N \quad -\Phi_1 \cdots -\Phi_p\right] = \left[I_N \quad -\Gamma \quad -\alpha\right] D = \kappa D \tag{20}$$

avec : $$\underset{(Np+r, N(p+1))}{D} = \begin{bmatrix} I_N & -I_N & 0 & \cdots & 0 & 0 \\ 0 & I_N & -I_N & \ddots & \vdots & \vdots \\ 0 & 0 & I_N & \ddots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & -I_N & 0 \\ 0 & 0 & \cdots & 0 & I_N & -I_N \\ 0 & \beta' & 0 & \cdots & 0 & 0 \end{bmatrix}.$$

La matrice Φ est estimée par $\widehat{\Phi} = \widehat{\kappa} D$. u_k est une matrice de sélection, de dimension $(N,1)$, dont la $k$ème ligne contient un 1 et les autres 0, et v_j est une matrice, de dimension $(N(p+1),p)$, dont la $l$ème colonne contient un 1 dans la $(N(l-1)+j)$ème ligne, pour $l = 1,\ldots,p$. On note que v_j est une matrice de dimension $(N(p+1),p)$ (et non $(N(p+1),p+1)$), car l'une des contraintes est redondante parmi les $(p+1)$ sous-matrices (de dimension (N,N)) de $C(1)\Phi$. En effet, on a

$$C(1)\Phi = [C(1)\ \ C(1)\left(-I_N - \Gamma_1\right)\ \ C(1)\left(\Gamma_1 - \Gamma_2\right)$$
$$\cdots\ C(1)\left(\Gamma_{p-2} - \Gamma_{p-1}\right)\ \ C(1)\Gamma_{p-1}]$$

où il est clair que la somme des sous-matrices est nulle. Ce résultat est dû au fait que la deuxième colonne, qui devrait s'écrire $C(1)\left(-I_N - \Gamma_1 - \alpha\beta'\right)$, se simplifie en $C(1)\left(-I_N - \Gamma_1\right)$ (puisque la cointégration implique $C(1)\alpha = 0$). Donc seules les p premières contraintes non-redondantes sont testées.

On définit alors la fonction :

(21) $$g_{kj}(\theta) = \mathrm{vec}\left(u'_k C(1)\Phi v_j\right)$$
$$= \mathrm{vec}\left(u'_k\left[M^{-1}D(1)(\kappa G)^{-1}\kappa D\right]v_j\right) = \Lambda_{kj}\, vec\left[(\kappa G)^{-1}\kappa D\right]$$

avec : $\Lambda_{kj} = v_j' \otimes u'_k M^{-1} D(1)$.

Proposition 4 : Si X_t est un processus gaussien défini par (11), la distribution asymptotique de $g_{kj}(\widehat{\theta})$ est :

(22) $$\sqrt{T}\left(g_{kj}(\widehat{\theta}) - g_{kj}(\theta)\right) \xrightarrow{d} N\left(0,\Sigma_{g_{kj}}\right), \quad \forall k,j = 1,\ldots,N, k \neq j,$$

avec :
$$\Sigma_{g_{kj}} = \frac{\partial g_{kj}}{\partial vec\,(\theta)'}\Sigma_\theta \frac{\partial g_{kj}}{\partial vec\,(\theta)}'$$

$$\frac{\partial g_{kj}}{\partial vec\,(\theta)'} = \frac{\partial g_{kj}}{\partial vec\,(\kappa)'}\zeta$$

et
$$\frac{\partial g_{kj}}{\partial vec\,(\kappa)'} = \Lambda_{kj}\frac{\partial}{\partial vec\,(\kappa)'} vec\left[(\kappa G)^{-1}\kappa D\right]$$
$$= -\Lambda_{kj}\left((\kappa D)' \otimes I_N\right)\left((\kappa G)^{-1'} \otimes (\kappa G)^{-1}\right)\left(G' \otimes I_N\right)$$
$$+\Lambda_{kj}\left(I_{N(p+1)} \otimes (\kappa G)^{-1}\right)\left(D' \otimes I_N\right).$$

Preuve : Se déduit directement de (21).

Il est maintenant possible de proposer une statistique de test pour la non-causalité de long terme :

PROPOSITION 5 : Si X_t est un processus gaussien défini par (11), le test de $H_0 : g_{kj}(\theta) = 0$ est fondé sur la statistique :

$$\xi_{kj} = T g_{kj}(\widehat{\theta})' \widehat{\Sigma}_{g_{kj}}^{-1} g_{kj}(\widehat{\theta}), \quad \forall k, j = 1, \ldots, N, k \neq j, \tag{24}$$

qui est distribué comme un χ^2 à p degrés de liberté.

Preuve : Évident, puisque seules p contraintes non-redondantes sont testées.

• Références bibliographiques

BRUNEAU, C. (1996). - Analyse Econométrique de la Causalité, Thèse de Doctorat, Université de Paris IX-Dauphine.

BRUNEAU, C., NICOLAÏ, J.-P. (1994). – “Amélioration de la prévision et causalité entre deux séries d’un système multivarié autorégressif stationnaire”, *Annales d’Economie et de Statistique*, 36, pp. 1-22.

BRUNEAU, C., NICOLAÏ, J.-P. (1995). – “Causalité persistante entre séries non-stationnaires : application à l’étude comparée des politiques monétaires des pays du G5”, *Annales d’Economie et de Statistique*, 40, pp. 177-206.

CAMPBELL, J.Y. (1995). – “Some Lessons from the Yield Curve”, *Journal of Economic Perspectives*, 9(3), pp. 129-152.

CAMPBELL, J.Y., SHILLER, R.J. (1987). – “Cointegration and Tests of Present Value Models”, *Journal of Political Economy*, 95(5), pp. 1062-1088.

CAMPBELL, J.Y., SHILLER, R.J. (1988). – “Interpreting Cointegrated Models”, *Journal of Economic Dynamics and Control*, 12(2/3), pp. 505-522.

CAMPBELL, J.Y., SHILLER, R.J. (1991). – “Yield Spreads and Interest Rate Movements: A Bird’s Eye View”, *Review of Economic Studies*, 58(3), pp. 495-514.

ENGLE, R.F. GRANGER, C.W.J. (1987). – “Cointegration and Error Correction: Representation, Estimation, and Testing”, *Econometrica* 55(2), pp. 251-76.

ENGSTED, T., TANGGAARD, C. (1994). – “Cointegration and the US Term Structure”, *Journal of Banking and Finance*, 18(1), pp. 167-181.

ESTRELLA, A., HARDOUVELIS, G.A. (1991). – “The Term Structure as a Predictor of Real Economic Activity”, *Journal of Finance*, 46(2), pp. 555-576.

EVANS, M.D.D., LEWIS, K.K. (1994). – “Do Stationary Risk Premia Explain It All? Evidence from the Term Structure”, *Journal of Monetary Economics*, 33(2), pp. 285-318.

FAMA, E.F. (1990). – “Term-Structure Forecasts of Interest Rates, Inflation, and Real Returns”, *Journal of Monetary Economics*, 25(1), pp. 59-76.

FAMA, E.F., BLISS, R.R. (1987). – “The Information in Long-Maturity Forward Rates”, *American Economic Review*, 77(4), pp. 680-692.

FULLER, W.A. (1976). - Introduction to Statistical Time Series, New York, Wiley.

GRANGER, C.W.J. (1969). – “Investigating Causal Relations by Econometric Models and Cross-Spectral Methods”, *Econometrica*, 37(3), pp. 424-438.

GRANGER, C.W.J. (1988). – “Some Recent Developments in the Concept of Causality”, *Journal of Econometrics*, 39(1/2), pp. 199-211.

GRANGER, C.W.J, LIN, J.L. (1995). – “Causality in the Long Run”, *Econometric Theory*, 11, pp. 530-548.

HALL, A.D., ANDERSON, H.M. GRANGER, C.W.J. (1992). – “A Cointegration Analysis of Treasury Bill Yields”, *Review of Economics and Statistics*, 74(1), pp. 116-126.

HANNAN, E.J., QUINN, B.G. (1979). – “The Determination of the Order of an Autoregression”, *Journal of the Royal Statistical Society*, B, 41, pp. 190-195.

HARDOUVELIS, G.A. (1994). – “The Term Structure Spread and Future Changes in Long and Short Rates in the G7 Countries : Is There a Puzzle ?», *Journal of Monetary Economics*, 33(2), pp. 255-283.

HARVEY, C.R. (1988). – “The Real Term Structure and Consumption Growth”, *Journal of Financial Economics*, 22(2), pp. 305-333.

JOHANSEN, S. (1988). – “Statistical Analysis of Cointegration Vectors”, *Journal of Economic Dynamics and Control*, 12(1/2), pp. 231-254.

JOHANSEN, S., JUSELIUS, K. (1990). – “Maximum Likelihood Estimation and Inference on Cointegration – With Applications to the Demand for Money”, *Oxford Bulletin of Economics and Statistics*, 52(2), pp. 169-210.

JOHANSEN, S., JUSELIUS, K. (1992). – “Testing Structural Hypotheses in a Multivariate Cointegration Analysis of the PPP and the UIP for UK”, *Journal of Econometrics*, 53(1/3), pp. 211-244.

JORION, P., MISHKIN, F.C. (1991). – “A Multicountry Comparison of Term-Structure Forecasts at Long Horizons”, *Journal of Financial Economics*, 29(1), pp. 59-80.

LUTKEPOHL, H., REIMERS, H.R. (1992). – “Impulse Responses Analysis of Cointegrated Systems”, *Journal of Economic Dynamics and Control*, 16(1), pp. 53-78.

MELLANDER, E., VREDIN, A. WARNE, A. (1990). – “Stochastic Trends and Economic Fluctuations in a Small Open Economy”, *Journal of Applied Econometrics*, 7(4), pp. 369-394.

MISHKIN, F.C. (1990a). – “What Does the Term Structure Tell Us about Future Inflation ?”, *Journal of Monetary Economics*, 25(1), pp. 77-95.

MISHKIN, F.C. (1990b). – “The Information in the Longer-Maturity Term Structure about Future Inflation”, *Quarterly Journal of Economics*, 55(3), pp. 815-828.

MISHKIN, F.C. (1991). – “A Multi-Country Study of the Information in the Shorter Maturity Term Structure about Future Inflation”, *Journal of International Money and Finance*, 10(1), pp. 2-22.

OSTERWALD-LENUM, M. (1992). – “A Note with Quantiles of the Asymptotic Distribution of the Maximum Likelihood Cointegration Rank Test Statistics”, *Oxford Bulletin of Economics and Statistics*, 54(3), pp. 461-472.

SHILLER, R.J. (1979). – “The Volatility of Long-Term Interest Rates and Expectations Theories of the Term Structure”, *Journal of Political Economy*, 87(6), pp. 1190-1219.

STOCK J.H. (1987). – “Asymptotic Properties of Least Squares Estimators of Cointegrating Vectors”, *Econometrica*, 55(5), pp. 1035-1056.

STOCK, J.H., WATSON, M.W. (1989). – “Interpreting the Evidence on Money-Income Causality”, *Journal of Econometrics*, 40(1), pp. 161-181.

TODA, H.Y., PHILLIPS, M.W. (1993). – “Vector Autoregression and Causality”, *Econometrica*, 61, pp. 1367-1393.