ANNALES D'ÉCONOMIE ET DE STATISTIQUE. – N° 62 – 2001

La théorie des anticipations de la structure par terme permet-elle de rendre compte de l'évolution des taux d'intérêt sur euro-devise ?

Éric JONDEAU *

RÉSUMÉ. – À partir de l'estimation d'une représentation VAR contraint, il est possible de mesurer la capacité de la théorie des anticipations à rendre compte de l'évolution observée des taux d'intérêt sur euro-devise. Différents tests associés à cette représentation sont considérés dans cet article. Plusieurs problèmes méthodologiques rencontrés dans la mise en œuvre de ces tests sont examinés. Les estimations menées sur la période 1982-1997 indiquent que les tests de causalité et le test « formel » ne permettent pas de rejeter la théorie pour les taux sur euro-devise français et britanniques, mais la rejettent pour les taux américains et allemands. Les tests affaiblis indiquent que ce rejet provient essentiellement d'une sur-réaction des taux longs aux variations anticipées des taux courts.

Is the Expectations Hypothesis of the Term Structure Able to Explain Euro-Rate Movements?

ABSTRACT. – In this paper, I measure the ability of the expectations hypothesis (EH) to explain Euro-rate movements, in a restricted VAR framework. Several tests of the EH are performed. Some methodological issues in the design of these tests are addressed. Over the period 1982-1997, the EH is not rejected by the data for French and UK Euro-rates, but it is generally rejected for US and German Euro-rates. Comparing the actual spread with the theoretical spread, I find evidence that this rejection can be mainly explained by an over-reaction of long-term rates to expected changes in short-term rates.

---

* É. JONDEAU : Banque de France, Centre de recherche.
Je tiens à remercier les deux rapporteurs anonymes de la Revue ainsi que l'éditeur (Stéphane GREGOIR) pour leurs suggestions et commentaires. Je reste, bien sûr, seul responsable des erreurs qui peuvent subsister. Cet article ne reflète pas nécessairement la position de la Banque de France.

1 Introduction

L'une des implications les plus étudiées de la théorie des anticipations de la structure par terme établit que le taux long est une moyenne des taux courts futurs anticipés, à une prime de risque constante dans le temps près (SHILLER [1979]). De nombreux auteurs ont mis en cause la validité empirique de cette théorie (par exemple, SHILLER, CAMPBELL et SCHOENHOLTZ [1983], ou MANKIW [1986]). Mais d'autres ont utilisé ses implications pour mesurer le contenu en information des taux présents concernant les taux futurs. Par exemple, MISHKIN [1991] et JORION et MISHKIN [1991] se sont intéressés à l'information contenue dans la pente des taux ; FAMA et BLISS [1987] et FAMA [1990] ont considéré l'information contenue dans l'écart entre les taux à terme et au comptant.

Une approche alternative permettant de mesurer la capacité prédictive des taux d'intérêt repose sur la représentation autorégressive contrainte (*restricted* VAR, ou RVAR) proposée par CAMPBELL et SHILLER [1987 et 1988]. Fondée sur la dynamique jointe de la variation du taux court et de la pente des taux, cette représentation permet de déduire des prévisions du taux court à partir du RVAR lui-même et ainsi de déterminer un taux long compatible avec la théorie des anticipations et la dynamique jointe des taux d'intérêt. Cette approche offre un cadre d'analyse cohérent avec les propriétés statistiques des taux d'intérêt. Depuis CAMPBELL et SHILLER [1987] et HALL, ANDERSON et GRANGER [1992], de nombreux travaux ont, en effet, mis en évidence que les taux d'intérêt sont, en général, intégrés et cointégrés deux à deux.

Un autre intérêt du RVAR est qu'il permet de formuler plusieurs tests de la théorie des anticipations. La *causalité de la pente des taux* vis-à-vis de la variation des taux court et long constitue l'une des implications les plus directes de la théorie. La plupart des études ayant mis en œuvre le test de cette hypothèse ont conclu au rejet de l'hypothèse de non-causalité vis-à-vis de la variation du taux court, qu'il s'agisse de CAMPBELL et SHILLER [1987] ou DRIFFILL, PSARADAKIS et SOLA [1997] sur données américaines, de CUTHBERTSON [1996] ou CUTHBERTSON, HAYES et NITZSCHE [1996] sur données britanniques ou encore de COLLETAZ et GOURLAOUEN [1990] sur données françaises. En revanche, la causalité vis-à-vis de la variation du taux long n'est généralement pas étudiée dans le cadre d'un RVAR (HALL, ANDERSON et GRANGER [1992] et ENGSTED et TANGGAARD [1994] étudient cette relation dans le cadre d'un modèle à correction d'erreur).

Des travaux récents ont toutefois montré que, lorsque le VAR estimé résulte d'une séquence de tests d'hypothèses préalables (tels que le test du rang de cointégration), les statistiques de Wald ont des distributions asymptotiques non-standard, qui dépendent en outre de paramètres de nuisance (BRUNEAU et NICOLAÏ [1995], TODA et YAMAMOTO [1995], DOLADO et LÜTKEPOHL [1996]). Les deux derniers articles cités ont proposé l'ajout de retards superflus dans le VAR en niveau pour satisfaire les conditions de régularité des tests d'hypothèses. Ce principe est appliqué, dans cet article, au test de causalité de la théorie des anticipations.

Le *test formel* de la théorie des anticipations concerne les implications de la théorie sur les paramètres du RVAR : ceux-ci doivent assurer l'égalité entre la pente des taux observée et la pente des taux théorique telle qu'elle est évaluée à partir du RVAR (comme la moyenne sur longue période des variations du taux court futur anticipé). Ce test a conduit de nombreux auteurs à rejeter la théorie. C'est en particulier le cas, sur données américaines, de CAMPBELL et SHILLER [1987], KUGLER [1990] ou SHEA [1992] ou, sur données européennes, de COLLETAZ et GOURLAOUEN [1990], MACDONALD et SPEIGHT [1991], TAYLOR [1992], CUTHBERTSON [1996] ou CUTHBERTSON, HAYES et NITZSCHE [1996]. Néanmoins, des études empiriques récentes ont conduit à nuancer quelque peu ce diagnostic, notamment HARDOUVELIS [1994], HURN, MOODY et MUSCATELLI [1995], GERLACH [1996], DRIFFILL, PSARADAKIS et SOLA [1997].

Toutefois, certains auteurs considèrent que le test formel n'est pas totalement satisfaisant, car il conduirait à rejeter trop souvent la théorie des anticipations, sans que cela ne paraisse économiquement justifié. En effet, comme l'ont souligné notamment CAMPBELL et SHILLER [1991], la pente observée et la pente théorique (calculée en tenant compte des contraintes imposées par la théorie des anticipations) présentent souvent des évolutions très proches. La théorie conserverait donc toute sa pertinence pour expliquer le mode de formation des taux d'intérêt. De façon à s'affranchir des problèmes liés à l'utilisation de la distribution asymptotique, on estime également, par simulation de *bootstrapping* (comme décrit dans l'annexe 2), la distribution à distance finie de la statistique de test.

Enfin plusieurs *indicateurs affaiblis* permettent de mesurer la proximité entre les pentes des taux observée et théorique. Il s'agit du coefficient de régression, du coefficient de corrélation et du rapport des variances associés à ces deux variables. Ces indicateurs permettent d'évaluer la capacité de la théorie à rendre compte des mouvements observés de la pente des taux. Les tests empiriques récents apparaissent alors plus favorables envers la théorie des anticipations que le test formel. Ces différents *indicateurs affaiblis* ont été proposés par CAMPBELL et SHILLER [1987 et 1988], mais leurs propriétés asymptotiques n'ont pas encore fait l'objet d'un exposé détaillé dans la littérature. En outre, si leur estimation ne pose pas de problème technique, l'évaluation de leur distribution asymptotique est plus délicate et peut présenter des biais sur de petits échantillons. On procède donc également à une estimation des écarts-types fondée sur des simulations par *bootstrapping*.

L'enjeu de cette démarche, au-delà du simple test de la théorie (pour laquelle les abondants résultats empiriques ne permettent pas de dégager des conclusions non ambiguës), porte sur la capacité des taux courts futurs anticipés à rendre compte des taux longs observés. Tout l'intérêt de la représentation RVAR est de proposer un mode de formation des anticipations de taux courts futurs. Contrairement à l'approche standard fondée sur la régression des variations de taux sur la pente des taux (développée, par exemple, dans CAMPBELL et SHILLER [1991]), l'étude de la théorie des anticipations ne repose plus que sur une hypothèse de rationalité faible des agents, ceux-ci n'étant supposés connaître que le modèle représentant la dynamique des taux d'intérêt.

La suite de cet article est organisée de la façon suivante. La section 2 décrit les implications de la théorie des anticipations dans le cadre d'une représentation RVAR. Les tests de la théorie proposés dans ce cadre sont développés dans la section 3. L'accent est mis sur la démarche méthodologique adoptée et, plus particulièrement, sur les difficultés liées à l'existence d'une relation de cointégration. La section 4 présente les données et l'estimation préalable de la représentation RVAR. L'étude porte sur les taux d'intérêt sur euro-devise américains, allemands, français et britanniques, au cours de la période 1982-1997, en fréquence mensuelle. La section 5 présente les résultats des tests de la théorie des anticipations et leur interprétation. Les principales conclusions sont résumées dans la section 6.

2 La théorie des anticipations dans une représentation RVAR

2.1 Théorie des anticipations et pente des taux

La représentation vectorielle autorégressive (VAR) pour l'étude de la dynamique des taux d'intérêt a été développée initialement par SARGENT [1979] et adaptée par CAMPBELL et SHILLER [1987 et 1988] pour des variables stationnaires en différence et cointégrées. Ces derniers ont proposé une représentation sous forme RVAR pour modéliser la dynamique de la variation du taux court et de la pente des taux. On note $r_t^{(n)}$ est le rendement d'un titre zéro-coupon de maturité résiduelle n. La représentation adoptée dans cet article correspond au cas général, dans lequel les maturités des titres (court et long) sont quelconques : la maturité du titre court ne correspond pas nécessairement à la fréquence des données et la durée de vie résiduelle du titre long est *a priori* finie. La théorie des anticipations établit que le rendement en t d'un titre de maturité n est égal à la moyenne des rendements anticipés de placements aux dates t, $t+m$, ..., $t+n-m$ de titres de maturité m, plus une prime de risque $\varphi^{(m,n)}$:

$$r_t^{(n)} = \frac{m}{n} \sum_{i=0}^{\frac{n}{m}-1} E_t r_{t+im}^{(m)} + \varphi^{(m,n)}. \tag{1}$$

La prime de risque peut éventuellement dépendre de la maturité des titres, mais elle est constante au cours du temps. Le terme $E_t x_{t+i} = E\left(x_{t+i} \,|\Omega_t\right)$ représente la projection linéaire de la variable x_{t+i} sur l'information disponible à la date t, notée Ω_t.

Si on définit la pente des taux sous la forme $S_t^{(m,n)} = r_t^{(n)} - r_t^{(m)}$, on obtient l'expression suivante, en soustrayant $r_t^{(m)}$ aux deux termes de l'équation (1) et en regroupant les termes :

$$S_t^{(m,n)} = E_t S_t^{*(m,n)} + \varphi^{(m,n)} \tag{2}$$

où la pente des taux en prévision parfaite, $S_t^{*(m,n)}$, est définie par :

$$S_t^{*(m,n)} = \frac{m}{n} \sum_{i=1}^{\frac{n}{m}-1} \sum_{j=1}^{im} \Delta r_{t+j}^{(m)}. \tag{3}$$

La relation (2) est essentielle pour l'analyse de la théorie des anticipations, car elle établit un lien, proposé par CAMPBELL et SHILLER [1987 et 1988], avec la cointégration. En effet, si les taux sont courts non-stationnaires (et, plus précisément, $I(1)$), la pente des taux en prévision parfaite est stationnaire, puisqu'elle ne dépend que de variations futures du taux court (d'après l'équation (3)). Or, selon la théorie, son anticipation est égale, à une constante près, à la pente des taux observée (2)). Celle-ci doit donc être également stationnaire. Cela implique qu'il existe une relation de cointégration entre le taux court et le taux long et donc que ce dernier est non-stationnaire. Dans ce cas, il existe une combinaison linéaire stationnaire (la pente des taux) entre des taux d'intérêt non-stationnaires.

2.2 Les différentes représentations de la dynamique des taux d'intérêt

L'intérêt de l'interprétation en termes de cointégration est de conférer à la théorie des anticipations un cadre d'analyse statistique bien adapté. On suppose que la dynamique du processus $X_t = \left(r_t^{(m)} \;\; r_t^{(n)}\right)'$ obéit à un modèle VAR en niveau, d'ordre p, de la forme :

$$\Phi(L)\, X_t = \varepsilon_t \tag{4}$$

où $\Phi(L) = I_2 - \sum_{i=1}^{p} \Phi_i L^i$ est une matrice de polynômes de retard d'ordre p ; I_n est la matrice identité de dimension n ; $\varepsilon_t = \left(\varepsilon_{1,t} \;\; \varepsilon_{2,t}\right)'$ est le vecteur des perturbations, supposé bruit blanc multivarié. On note Φ la matrice, de dimension $(2 \times 2p)$, contenant les paramètres du VAR, $\Phi = \left(\Phi_1 \;\; \cdots \;\; \Phi_p\right)$, avec $\varphi = \text{vec}(\Phi)$, de dimension $(4p \times 1)$. S'il existe une relation de cointégration, la matrice de long terme, $\Phi(1) = I_2 - \sum_{i=1}^{p} \Phi_i$, s'écrit comme le produit $\Phi(1) = -\alpha\beta'$, où α et β sont deux vecteurs, de dimension (2×1), avec β le vecteur de cointégration. Lorsque les taux d'intérêt sont non-stationnaires mais que la pente des taux est stationnaire, le vecteur de cointégration s'écrit $\beta = (-1 \;\; 1)'$, de telle sorte que la pente joue le rôle de force de rappel, avec $\beta' X_t = r_t^{(n)} - r_t^{(m)} = S_t^{(m,n)}$. On note H_0^c l'hypothèse $\beta = (-1 \;\; 1)'$.

ENGLE et GRANGER [1987] ont montré que la dynamique d'un système de variables non-stationnaires et cointégrées peut également être étudiée sous la forme d'un modèle à correction d'erreur vectoriel (VECM). CAMPBELL et SHILLER [1987 et 1988] (dans le cas bivarié) et MELLANDER, VREDIN et WARNE [1992] (dans le cas multivarié) ont, en outre, montré l'équivalence entre les modèles VAR en niveau, VECM et RVAR. Le VECM s'écrit sous la forme :

$$\Gamma(L)\, \Delta X_t = \alpha\beta' X_{t-1} + \varepsilon_t \tag{5}$$

où $\Gamma(L) = I_2 - \sum_{i=1}^{p-1} \Gamma_i L^i$ est d'ordre $p-1$. La représentation RVAR décrit la dynamique de la variation de l'une des deux variables (ici, le taux court) et du résidu de la relation de cointégration, sous la forme :

$$B(L)\,Y_t = \eta_t \tag{6}$$

où $B(L) = I_2 - \sum_{i=1}^{p} B_i L^i$ est d'ordre p ; $Y_t = D_\perp(L) M X_t$; $\eta_t = \left(\eta_{1,t}\ \eta_{2,t}\right)'$ $= M\varepsilon_t$ est le vecteur des perturbations, avec $M = \begin{bmatrix} 1 & 0 \\ \beta_1 & \beta_2 \end{bmatrix}$ et

$$D_\perp(L) = \begin{bmatrix} (1-L) & 0 \\ 0 & 1 \end{bmatrix}.$$

On remarque que, sous l'hypothèse $H_0^c : \beta = (-1\ \ 1)'$, on a $M = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix}$ et donc $Y_t = \left(\Delta r_t^{(m)}\ \ S_t^{(m,n)}\right)'$. Le RVAR décrit alors la dynamique jointe de la variation du taux court et de la pente des taux.

Les relations suivantes établissent une relation entre les paramètres du VAR en niveau, du VECM et du RVAR (MELLANDER, VREDIN et WARNE [1992]) :

$$B(L) = M\left[\Gamma(L) M^{-1} D(L) - \alpha^* L\right] \tag{7}$$

et

$$\Phi(L) = M^{-1} B(L)\, D_\perp(L) M \tag{8}$$

avec $\alpha^* = \begin{bmatrix} 0 & \alpha_1 \\ 0 & \alpha_2 \end{bmatrix}$ et $D(L) = \begin{bmatrix} 1 & 0 \\ 0 & (1-L) \end{bmatrix}$.

On définit par ailleurs la matrice B, de dimension $(2 \times 2p)$, contenant les paramètres du RVAR, soit $B = \left(B_1\ \ \cdots\ \ B_p\right)$ avec $b = \mathrm{vec}(B)$ le vecteur colonne de dimension $(4p \times 1)$. On vérifie aisément que l'équivalence entre les écritures VECM et RVAR (7)) impose la contrainte $b_{11}^{(p)} = b_{21}^{(p)} = 0$, c'est-à-dire que les coefficients associés à $\Delta r_{t-p}^{(m)}$ dans le RVAR sont nuls par construction. Ce point a été établi, par exemple, par CAMPBELL et SHILLER [1987]. On définit donc la matrice B^c, de dimension $(2 \times 2p - 1)$, contenant les paramètres à estimer, soit

$$B^c = \left(\left| \begin{matrix} b_{11}^{(1)} & b_{12}^{(1)} \\ b_{21}^{(1)} & b_{22}^{(1)} \end{matrix} \right| \ldots \left| \begin{matrix} b_{11}^{(p-1)} & b_{12}^{(p-1)} \\ b_{21}^{(p-1)} & b_{22}^{(p-1)} \end{matrix} \right| \begin{matrix} b_{12}^{(p)} \\ b_{22}^{(p)} \end{matrix} \right| \right)$$

et on note $b^c = \mathrm{vec}\left(B^c\right)$, de dimension $(4p - 2 \times 1)$.

2.3 Les implications de la théorie des anticipations dans la représentation RVAR

Les représentations vectorielles (telles que le VAR en niveau, le VECM ou le RVAR) permettent la prévision des variables endogènes. Comme la théorie des anticipations implique que la pente des taux est égale à la moyenne des variations anticipées du taux court, elle impose certaines contraintes aux paramètres des dynamiques de pente et de variation du taux court. Ces contraintes trans-équations constituent, selon les termes de HANSEN et SARGENT [1981], « *the hallmark of rational expectations models* ». Le test de ces contraintes correspond, en fait, à un test joint de l'hypothèse de rationalité des anticipations et de constance des primes de risque[1].

Le choix du RVAR pour l'étude de la théorie des anticipations se justifie par le fait que ce modèle est particulièrement bien adapté au test de l'implication de la théorie donnée par la relation (2). Cette équation établit, en effet, une relation entre la pente des taux $S_t^{(m,n)}$ et les variations anticipées du taux court futur. Comme ces variables sont directement présentes dans le RVAR, celui-ci permet d'obtenir aisément des prévisions de la variation du taux court et donc une formulation très proche de l'équation (2). La démarche est la suivante.

Tout d'abord, sous l'hypothèse H_0^c, le modèle RVAR de dimension 2 et d'ordre p (relation (6)) peut s'écrire comme un VAR de dimension $2p$ et d'ordre 1, représentant la dynamique du vecteur

$$z_t = \left(\Delta r_t^{(m)}, S_t^{(m,n)}, \cdots, \Delta r_{t-p+1}^{(m)}, S_{t-p+1}^{(m,n)}\right)' :$$

$$\begin{pmatrix} \Delta r_t^{(m)} \\ S_t^{(m,n)} \\ \Delta r_{t-1}^{(m)} \\ S_{t-1}^{(m,n)} \\ \vdots \\ \Delta r_{t-p+1}^{(m)} \\ S_{t-p+1}^{(m,n)} \end{pmatrix}$$

$$= \begin{bmatrix} b_{11}^{(1)} & b_{12}^{(1)} & \cdots & b_{11}^{(p-1)} & b_{12}^{(p-1)} & b_{11}^{(p)} & b_{12}^{(p)} \\ b_{21}^{(1)} & b_{22}^{(1)} & \cdots & b_{21}^{(p-1)} & b_{22}^{(p-1)} & b_{21}^{(p)} & b_{22}^{(p)} \\ 1 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 & 0 & 0 \\ 0 & 0 & \cdots & 0 & 1 & 0 & 0 \end{bmatrix} \begin{pmatrix} \Delta r_{t-1}^{(m)} \\ S_{t-1}^{(m,n)} \\ \Delta r_{t-2}^{(m)} \\ S_{t-2}^{(m,n)} \\ \vdots \\ \Delta r_{t-p}^{(m)} \\ S_{t-p}^{(m,n)} \end{pmatrix} + \begin{pmatrix} \eta_{1,t} \\ \eta_{2,t} \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}$$

1. Ces modèles multivariés posent toutefois un problème dans la définition de l'ensemble d'information. D'un côté, ils supposent que les agents ont trop d'information par rapport à ce dont ils disposent en réalité : en effet, les coefficients du RVAR, par exemple, sont estimés sur l'ensemble de l'échantillon. Ainsi les prévisions faites en τ ($\tau < T$) à partir du RVAR utilisent de l'information mise à la disposition des agents après la date τ. À l'inverse, le RVAR suppose que les agents ont trop peu d'information : ils ont en fait accès à des informations qui dépassent largement le cadre des seuls taux d'intérêt. Donc, dans l'approche RVAR, l'ensemble de l'information utile aux agents est supposé correctement résumée par les taux d'intérêt et les pentes des taux.

ou, de façon plus compacte,

$$z_t = Az_{t-1} + v_t \tag{9}$$

avec $v_t = \left(\eta_{1,t}, \eta_{2,t}, 0, \cdots, 0\right)'$.

La matrice A est la matrice d'état du RVAR.

On suppose dans la suite que z_t est centré, pour se concentrer sur les relations entre les paramètres du VAR. Donc, plutôt que d'étudier la dynamique des primes de risque elles-mêmes, on privilégie, dans cette approche, l'effet de la théorie des anticipations sur la dynamique du RVAR. Ainsi, la question cruciale est la compatibilité entre les paramètres estimés du VAR (9) et ceux qui assurent la validité de la théorie.

L'intérêt de la réécriture (9) repose sur la propriété suivante : si z_t suit un modèle autorégressif d'ordre 1, alors la prévision à la date t de z_{t+j} s'écrit simplement

$$E\left[z_{t+j} \,|I_t\right] = A^j z_t \tag{10}$$

où $I_t = \left\{\Delta r_t^{(m)}, S_t^{(m,n)}, \cdots, \Delta r_{t-p+1}^{(m)}, S_{t-p+1}^{(m,n)}\right\}$ est l'ensemble d'information de l'économètre ($I_t \in \Omega_t$). On définit en outre h le vecteur de dimension $(2p \times 1)$ contenant 1 comme premier élément et 0 ailleurs et g le vecteur de dimension $(2p \times 1)$ contenant 1 comme deuxième élément et 0 ailleurs. La variation du taux court et la pente des taux à partir du vecteur z_t s'obtiennent ainsi de la façon suivante :

$$\Delta r_t^{(m)} = h' z_t \tag{11a}$$

$$S_t^{(m,n)} = g' z_t. \tag{11b}$$

La prévision de la variation du taux court futur $\Delta r_{t+j}^{(m)}$ à partir du VAR s'écrit alors sous la forme : $E\left[\Delta r_{t+j}^{(m)} \,|I_t\right] = h' A^j z_t$. Finalement, la prévision optimale de la pente des taux en prévision parfaite $S_t^{*(m,n)}$, notée $\tilde{S}_t^{(m,n)} = E\left[S_t^{*(m,n)} \,|I_t\right]$ a pour expression :

$$\tilde{S}_t^{(m,n)} = \frac{m}{n} \sum_{i=1}^{\frac{n}{m}-1} \sum_{j=1}^{im} E\left[\Delta r_{t+j}^{(m)} \,|I_t\right] = \frac{m}{n} \sum_{i=1}^{\frac{n}{m}-1} \sum_{j=1}^{im} h' A^j z_t \tag{12}$$

La variable $\tilde{S}_t^{(m,n)}$ est aussi appelée la pente des taux théorique. On définit le vecteur de paramètres :

$$\theta' = \frac{m}{n} \sum_{i=1}^{\frac{n}{m}-1} \sum_{j=1}^{im} h' A^j \tag{13}$$

de dimension $(2p \times 1)$, tel que $\tilde{S}_t^{(m,n)} = \theta' z_t$. La relation (12) met en évidence l'intérêt de la représentation RVAR bivariée. Une fois choisie cette représentation, on peut aisément en tirer des prévisions de la variation du taux court à un horizon quelconque. L'évaluation ainsi obtenue de la pente des taux est donc compatible à la fois avec la théorie des anticipations et avec le cadre statistique retenu.

3 Les tests de la théorie des anticipations

3.1 La causalité de la pente des taux dans l'approche bivariée

Sous l'hypothèse H_0^c, la pente des taux $S_t^{(m,n)}$ constitue une relation d'équilibre, stationnaire, entre les taux d'intérêt. Il s'agit d'une condition nécessaire, mais non suffisante, de la théorie des anticipations lorsque les taux d'intérêt sont non-stationnaires. Comme l'ont montré CAMPBELL et SHILLER [1988, *proposition p. 513*], dès que la pente ne s'exprime pas comme une fonction linéaire exacte des variations présentes et passées du taux court, chacun des taux doit être causé, au sens de GRANGER [1969], par la pente des taux[2]. Selon la théorie des anticipations, la pente des taux doit donc améliorer la prévision des variations futures de taux d'intérêt, conditionnellement aux variations passées de taux. Cette propriété de causalité constitue l'une des implications les plus directes de la théorie des anticipations.

Il est important de caractériser plus précisément cette forme de causalité de la pente des taux, en la positionnant par rapport à d'autre notions étudiées dans la littérature. Tout d'abord, dans la lignée de SHILLER, CAMPBELL et SCHOENHOLTZ [1983] et CAMPBELL et SHILLER [1991], la théorie des anticipations est couramment testée à partir des équations de régression suivante : la variation future du rendement du titre long dépend, à une erreur d'anticipation près, de la pente des taux, et la variation cumulée future des rendements du titre court est égale, à une erreur d'anticipation près, à la pente des taux. Ces formulations sont plus contraignantes que la précédente, car elles traduisent, dans le cadre d'une équation de régression, l'ensemble des implications de la théorie. En particulier, elles supposent la non-prévisibilité des erreurs d'anticipation (*cf.* JONDEAU et RICART [1999]).

D'autre part, la causalité de la pente des taux peut être mise en regard d'autres concepts de causalité définis dans le cadre d'un système cointégré. De façon à simplifier l'exposé, on considère le cas d'un VECM bivarié avec un seul retard ($p = 2$), dans lequel les deux variables X_1 et X_2 sont $I(1)$. L'hypothèse H_0^c est supposée vérifiée, ce qui donne $S_t = \beta' X_t = X_{2,t} - X_{1,t}$ comme relation de cointégration. Le VECM s'écrit donc sous la forme :

$$\begin{aligned}\Delta X_{1,t} &= \gamma_{11}\Delta X_{1,t-1} + \gamma_{12}\Delta X_{2,t-1} + \alpha_1 S_{t-1} + \varepsilon_{1,t}\\ \Delta X_{2,t} &= \gamma_{21}\Delta X_{1,t-1} + \gamma_{22}\Delta X_{2,t-1} + \alpha_2 S_{t-1} + \varepsilon_{2,t}.\end{aligned}$$

2. On rappelle qu'une série $X_{2,t}$ cause une autre série $X_{1,t}$ au sens de GRANGER [1969], si, et seulement si, le passé de $X_{2,t}$ permet d'améliorer la prévision de $X_{1,t}$ conditionnellement au passé de $X_{1,t}$.

En utilisant les relations (7) et (8), on en déduit la représentation RVAR :

$$\begin{aligned}
\Delta X_{1,t} &= (\gamma_{11} + \gamma_{12})\, \Delta X_{1,t-1} + (\gamma_{12} + \alpha_1)\, S_{t-1} - \gamma_{12} S_{t-2} + \varepsilon_{1,t} \\
S_t &= (\gamma_{21} - \gamma_{11} + \gamma_{22} - \gamma_{12})\, \Delta X_{1,t-1} + (1 - \alpha_1 + \alpha_2 + \gamma_{22} - \gamma_{12})\, S_{t-1} \\
&\quad - (\gamma_{22} - \gamma_{12})\, S_{t-2} + \left(\varepsilon_{2,t} - \varepsilon_{1,t}\right).
\end{aligned}$$

et le VAR en niveau :

$$\begin{aligned}
X_{1,t} = (1 + \gamma_{11} - \alpha_1)\, X_{1,t-1} - \gamma_{11} X_{1,t-2} + (\gamma_{12} + \alpha_1)\, X_{2,t-1} \\
-\gamma_{12} X_{2,t-2} + \varepsilon_{1,t}
\end{aligned}$$

$$\begin{aligned}
X_{2,t} = (\gamma_{21} - \alpha_2)\, X_{1,t-1} - \gamma_{21} X_{1,t-2} + (1 + \gamma_{22} + \alpha_2)\, X_{2,t-1} \\
-\gamma_{22} X_{2,t-2} + \varepsilon_{2,t}.
\end{aligned}$$

Le test de l'hypothèse nulle de non-causalité de la pente vis-à-vis de la variation du taux court dans le RVAR ($\alpha_1 = 0$ et $\gamma_{12} = 0$) est donc équivalent au test des deux hypothèses suivantes : (1) la non-causalité du taux long vis-à-vis du taux court dans le VAR en niveau ; et (2) la non-causalité de la variation du taux long vis-à-vis de la variation du taux court ($\gamma_{12} = 0$) et la nullité du coefficient de la force de rappel associée ($\alpha_1 = 0$) dans le VECM[3].

Le test envisagé ici est plus contraignant que le test de causalité de long terme caractérisé par GRANGER [1988] et GRANGER et LIN [1995] et mis en œuvre par HALL, ANDERSON et GRANGER [1992]. Ceux-ci étudient, en effet, la non-causalité à long terme du taux long vis-à-vis du taux court, à travers la non-significativité du coefficient de la force de rappel dans le VECM. Cette condition apparaît comme une condition nécessaire, mais non suffisante, de la non-causalité étudiée ici. On remarque en revanche que, dans un cadre bivarié, la non-causalité de la pente vis-à-vis de la variation du taux court est équivalente à la non-causalité de long terme du taux long vis-à-vis du taux court, définie par BRUNEAU et JONDEAU [1999a et b] et caractérisée par la propriété d'absence d'amélioration de la prévision à un horizon infini.

Dans la représentation (6), l'hypothèse nulle de non-causalité de la pente de taux vis-à-vis de la variation du taux court est définie par :

$$H_0^1 : b_{12}^{(k)} = 0, \forall k = 1, \ldots, p$$

ce qu'on peut réécrire sous la forme $H_0^1 : Rb^c = 0$, avec R la matrice, de dimension $(p \times 4p - 2)$, qui sélectionne les paramètres $b_{12}^{(k)}$, $k = 1, \ldots, p$. La statistique de Wald du test de H_0^1 s'écrit donc :

$$\xi = T \left(R\hat{b}^c\right)' \left(R \hat{\Sigma}_{b^c} R'\right)^{-1} \left(R\hat{b}^c\right)$$

3. De même, de façon évidente, le test de non-causalité de la pente des taux vis-à-vis de la variation du taux long peut être mené directement dans le RVAR associé au couple $\left(\Delta X_{2,t} \quad S_t\right)'$.

où $\hat{\Sigma}_{b^c}$ est un estimateur convergent de Σ_{b^c}, la matrice de variance-covariance de b^c. L'estimation de $\hat{\Sigma}_{b^c}$ est précisée dans la section 4. Quand $\left(R\hat{\Sigma}_{b^c}R'\right)$ est une matrice non-singulière, la statistique ξ suit un χ^2 à p degrés de liberté.

Ce test n'est toutefois pas sans poser de problème. En effet, SIMS, STOCK et WATSON [1990] ont montré que, dans un système non-stationnaire et cointégré, les statistiques de Wald n'ont pour distribution asymptotique un χ^2 que s'il est possible d'écrire le modèle de façon à ce que les restrictions dans le modèle transformé ne fassent intervenir que des coefficients de variables stationnaires. Or, comme le montre l'exemple précédent, le test de non-causalité fait intervenir le coefficient des forces de rappel S_{t-i} passées (que ce soit dans le VECM (5) ou le RVAR (6)), qui s'écrivent comme des combinaisons de variables non-stationnaires (puisque $S_t = \beta' X_t = X_{2,t} - X_{1,t}$). Préalablement à l'estimation du VECM ou du RVAR, différents tests doivent donc être menés pour s'assurer de la stationnarité de S_t. Il est ainsi nécessaire, dans une première étape, de valider l'hypothèse H_0^c, c'est-à-dire l'existence d'une relation de cointégration entre les taux court et long et l'égalité $\beta = (-1 \;\; 1)'$. Si on conclut à tort à l'existence d'une relation de cointégration et donc à la stationnarité de S_t, la distribution du paramètre α_1 est clairement non-standard, puisque ce coefficient est associé à une variable non-stationnaire. Donc, dès que l'hypothèse H_0^c n'est pas supposée *a priori* vérifiée, mais fait l'objet de tests préalables, les tests de non-causalité ne peuvent plus être menés de façon standard.

Ce problème a été étudié notamment par BRUNEAU et NICOLAÏ [1995], TODA et YAMAMOTO [1995] et DOLADO et LÜTKEPOHL [1996]. Dans les deux derniers articles cités[4], la démarche proposée pour surmonter ces difficultés consiste à estimer un VAR en niveau avec un $(p+1)$ème retard superflu. Le test de causalité mis en œuvre porte alors sur les p premiers retards seulement. Ainsi, la dernière matrice de retards permet d'« absorber » la singularité éventuelle de la distribution asymptotique des estimateurs, puisqu'il n'est plus nécessaire de tester préalablement le rang de cointégration et la forme du vecteur de cointégration.

On considère donc un modèle VAR en niveau d'ordre $p+1$, au lieu du VAR d'ordre p donné pour l'équation (4) :

$$X_t = \Phi_1 X_{t-1} + \cdots + \Phi_p X_{t-p} + \Phi_{p+1} X_{t-p-1} + \varepsilon_t.$$

On note φ_{p+1}^p le vecteur, de dimension $(4p \times 1)$, contenant les éléments des p premières matrices de retard, soit $\varphi_{p+1}^p = \mathrm{vec}\left(\Phi_1 \;\; \cdots \;\; \Phi_p\right)$. Le vecteur des paramètres estimés correspondant est noté $\hat{\varphi}_{p+1}^p$. Ce modèle VAR en niveau peut être réécrit sous la forme :

$$\left(X_t - X_{t-p-1}\right) = \sum_{i=1}^{p} \Phi_i \left(X_{t-i} - X_{t-p-1}\right) - \Pi X_{t-p-1} + \varepsilon_t$$

4. BRUNEAU et NICOLAÏ [1995] proposent un test du rang de la matrice $(R\hat{\Sigma}_{b^c}R')$. Si cette matrice est de rang égal au nombre de contraintes testées sous H_0, la distribution asymptotique de la statistique de Wald est un χ^2 standard.

où $\Pi = I_2 - \Phi_1 - \cdots - \Phi_p - \Phi_{p+1}$. Puisque, dans ce modèle, les matrices $\Phi_i,\ i = 1,...,p$, sont associées à des variables stationnaires, les estimateurs de $\Phi_i,\ i = 1,...,p$, ont une distribution asymptotique jointe non-singulière et normale, soit :

$$\sqrt{T}\left(\hat{\varphi}^p_{p+1} - \varphi^p_{p+1}\right) \xrightarrow{d} N\left(0, \Sigma^p_{p+1}\right)$$

où Σ^p_{p+1} est une matrice non-singulière de dimension $(4p \times 4p)$. Alors, pour une estimation convergente $\hat{\Sigma}^p_{p+1}$ de Σ^p_{p+1} et une matrice $\tilde{R}$, de dimension $(p \times 4p)$ et de rang p, la statistique de Wald du test de l'hypothèse nulle $H_0^{1\prime} : \tilde{R}\hat{\varphi}^p_{p+1} = 0$, équivalente à $H_0^1 : Rb^c = 0$, se réécrit, en fonction des paramètres du VAR en niveau, sous la forme :

$$\tilde{\xi} = T\left(\tilde{R}\hat{\varphi}^p_{p+1}\right)'\left(\tilde{R}\hat{\Sigma}^p_{p+1}\tilde{R}'\right)^{-1}\left(\tilde{R}\hat{\varphi}^p_{p+1}\right).$$

La statistique $\tilde{\xi}$ a comme distribution asymptotique un χ^2 à p degrés de liberté sous l'hypothèse $H_0^{1\prime}$ (théorème 1 de Dolado et Lütkepohl [1996]).

La différence entre les statistiques de Wald ξ et $\tilde{\xi}$ peut donc être résumée de la façon suivante : d'un côté, la statistique ξ est évaluée conditionnellement à l'hypothèse H_0^c. Si H_0^c est vraie, la statistique ξ est plus puissante que $\tilde{\xi}$. À l'inverse, quand H_0^c n'est pas vérifiée, la distribution asymptotique de ξ n'est plus un χ^2. D'un autre côté, comme $\tilde{\xi}$ est évaluée sans supposer que H_0^c est vraie, sa distribution asymptotique est un χ^2 quelles que soient les propriétés de cointégration du système, au prix d'une perte de puissance lorsque le système est réellement cointégré. Zapata et Rambaldi [1997] ont montré, à l'aide de simulations de Monte-Carlo, que $\tilde{\xi}$ présente de meilleures propriétés que ξ lorsque H_0^c n'est pas vérifiée par les données, sans pour autant souffrir d'une perte de puissance importante quand H_0^c est vraie.

3.2 Le test formel de la théorie des anticipations

Si les agents forment leurs anticipations à partir du modèle RVAR, la théorie des anticipations impose l'égalité entre la pente des taux observée et la pente des taux théorique :

$$H_0^2 : \quad S_t^{(m,n)} = \tilde{S}_t^{(m,n)}$$

soit, en égalisant (11b) et (12) :

$$g'z_t = \theta' z_t.$$

L'écriture du vecteur θ se simplifie de la façon suivante :

$$\theta' = \frac{m}{n}\sum_{i=1}^{\frac{n}{m}-1}\sum_{j=1}^{im} h'A^j = h'A\left[I - \frac{m}{n}(I - A^n)\left(I - A^m\right)^{-1}\right](I - A)^{-1}$$

d'où, finalement, l'hypothèse nulle :

$$(14) \qquad H_0^2 : g' = h^{'}A\left[I - \frac{m}{n}(I - A^n)\left(I - A^m\right)^{-1}\right](I - A)^{-1}.$$

Le vecteur de contraintes imposées par l'équation (14) peut s'écrire sous les deux formes alternatives suivantes :

$$(15) \quad r_1(b^c) \ = \ \text{vec}\left(g' - h'A\left[I - \frac{m}{n}(I - A^n)(I - A^m)^{-1}\right](I - A)^{-1}\right) = 0$$

$$(16) \quad r_2(b^c) \ = \ \text{vec}\left(g'(I - A) - h'A\left[I - \frac{m}{n}(I - A^n)(I - A^m)^{-1}\right]\right) = 0.$$

Les termes $r_i\left(b^c\right)$ dépendent, à travers la matrice A, des paramètres du RVAR (b^c). Même si elles sont algébriquement équivalentes, les expressions (15) et (16) peuvent conduire à des résultats contradictoires lors des tests empiriques, puisque la linéarisation (nécessaire pour le calcul de la distribution asymptotique de la statistique de test) porte sur des expressions différentes[5].

Sous l'hypothèse nulle donnée par l'égalité (14), les deux expressions $r_i(b^c)$ ont pour distributions asymptotiques respectives :

$$\sqrt{T}r_i\left(\hat{b}^c\right) \xrightarrow{d} N\left(0, \frac{\partial r_i\left(b^c\right)}{\partial b^{c\prime}}\Sigma_{b^c}\frac{\partial r_i\left(b^c\right)^{\prime}}{\partial b^c}\right) \qquad i = 1{,}2$$

où les dérivées sont évaluées au point $b^c = \hat{b}^c$. La statistique du test de Wald s'écrit alors :

$$(17) \qquad W_i = Tr_i^{'}(\hat{b}^c)(\hat{D}_i'\hat{\Sigma}_{b^c}\hat{D}_i)^{-1}r_i(\hat{b}^c) \qquad i = 1{,}2.$$

où $\hat{D}_i$ est un estimateur convergent de la matrice des dérivées partielles $D_i = \dfrac{\partial r_i(b^c)}{\partial b^{c\prime}}$, de dimension $(2p{,}4p - 2)$. Le calcul de D_i est détaillée dans l'annexe 1. Si $\Sigma_r = (D_i'\hat{\Sigma}_{b^c}D_i)$ est une matrice non-singulière, alors sous l'hypothèse nulle la statistique W_i a pour distribution asymptotique une loi du χ^2 dont le nombre de degrés de liberté est égal au nombre de contraintes testées dans $r_i(\hat{b}^c)$. Compte tenu de la contrainte $b_{11}^{(p)} = b_{21}^{(p)} = 0$, le nombre de degrés de liberté est égal à $(2p - 1)$ au lieu de $2p$. Dans la littérature, seul KUGLER [1990] a, semble-t-il, pris en compte cette contrainte dans le calcul du nombre de degrés de liberté[6].

---

5. Cette contradiction a été mise en évidence par GREGORY et VEALL [1985], à partir de simulations de Monte-Carlo, dans le cas général des tests de Wald de contraintes non-linéaires sur les paramètres.
6. Cette différence de degrés de liberté est susceptible d'influencer les résultats du test, lorsque le nombre de retards retenu dans le RVAR est réduit. Ce problème peut, par exemple, affecter les résultats de CUTHBERTSON [1996], GERLACH [1996] ou DRIFFILL, PSARADAKIS et SOLA [1997], qui retiennent entre 1 et 4 retards, et n'imposent pas la contrainte $b_{11}^{(p)} = b_{21}^{(p)} = 0$ lors de l'estimation de leur modèle. L'absence de prise en compte de cette contrainte conduit, en tout état de cause, à affaiblir la puissance des tests.

On remarque à nouveau que la mise en œuvre du test formel pose certaines difficultés. Tout d'abord, il est possible que la propriété de non-singularité, mentionnée précédemment, de la matrice Σ_r ne soit pas respectée[7]. En effet, la matrice des dérivées partielles D_i peut être de rang inférieur à $(2p-1)$ sous l'hypothèse nulle. Dans ce cas, la statistique de Wald W_i n'est plus distribuée asymptotiquement selon une loi du χ^2 sous l'hypothèse nulle. Ce type de problème est susceptible de survenir lorsque la statistique de test fait intervenir des produits croisés des éléments du vecteur des paramètres. De telles situations ont été mises en évidence, notamment, par BRUNEAU et NICOLAÏ [1995]. Quand D_i n'est pas de plein rang sous l'hypothèse nulle, la matrice Σ_r est singulière et $\hat{\Sigma}_r = \hat{D}_i' \hat{\Sigma}_{b^c} \hat{D}_i$ converge presque sûrement vers une matrice singulière. LÜTKEPOHL et BURDA [1997] ont proposé des modifications de la statistique de Wald standard, de façon à retrouver comme distribution asymptotique une loi du χ^2 sous l'hypothèse nulle. Dans le cas du test formel, le vecteur de contraintes $r_i\left(b^c\right)$ s'écrit comme une combinaison de produits croisés des éléments de b_c, puisque θ est défini comme une somme de puissances de A (*cf*. l'expression (13)). Les difficultés précédentes sont donc susceptibles de survenir. On montre toutefois, dans l'annexe 1, dans certaines configurations simples, que la matrice des dérivées partielles est de plein rang sous l'hypothèse nulle.

Une autre difficulté posée par le test formel est que, comme pour le test de causalité, la statistique de Wald associée au test formel a une distribution asymptotique non-standard lorsque les propriétés de cointégration des séries font l'objet de tests préalables, conformément au résultat de SIMS, STOCK et WATSON [1990]. La présentation du test formel qui précède admet que l'hypothèse H_0^c est vérifiée par les données, de sorte que la pente des taux est directement introduite dans le RVAR comme résidu de la relation de cointégration (le vecteur de cointégration est supposé égal à $\beta = (-1\ \ 1)'$). Comme indiqué précédemment, les tests de causalité sont menés à la fois dans le RVAR et dans le VAR en niveau, de façon à se prémunir contre le risque d'avoir accepté à tort H_0^c. Pour les autres tests en revanche, l'hypothèse H_0^c est supposée vérifiée, de façon à se concentrer sur la représentation RVAR.

D'autre part, de nombreuses études empiriques ont mis en évidence que le test formel conduit à rejeter trop souvent la théorie, alors que la comparaison entre la pente des taux observée $S_t^{(m,n)}$ et la pente des taux théorique $\tilde{S}_t^{(m,n)}$ semble souvent plus favorable envers la théorie (CAMPBELL et SHILLER [1987], CUTHBERTSON [1996]). Plusieurs explications ont été avancées pour rendre compte de ce paradoxe. Le principal argument repose sur la sensibilité du test formel à la spécification retenue. Ainsi, TAYLOR [1992] suggère que des imperfections « mineures » sur les données pourraient conduire au rejet de la théorie, alors même que les deux pentes des taux seraient très proches. Selon CUTHBERTSON [1996], le rejet de la théorie des anticipations peut être dû à un biais dans l'estimation des paramètres du RVAR causé par une petite prime de risque variant dans le temps et corrélée avec la pente des taux. Cette prime n'aurait, en revanche, que peu d'effet sur la corrélation entre $S_t^{(m,n)}$ et $\tilde{S}_t^{(m,n)}$.

7. Je suis reconnaissant à l'un des rapporteurs anonymes de la revue pour avoir soulevé ce problème.

De façon à éliminer les biais possibles liés à l'estimation de la distribution asymptotique, on évalue également la distribution à distance finie de la statistique de test, à partir de simulation de *bootstrapping*. Les méthodes de simulation constituent en effet une approche alternative, permettant de résoudre le problème de l'adéquation entre la distribution asymptotique et la distribution à distance finie. L'annexe 2 présente l'approche utilisée pour estimer par simulation la distribution de la statistique du test formel et des statistiques de test affaiblies.

3.3 Les tests affaiblis

Les critiques présentées ci-dessus à l'adresse du test formel ont conduit certains auteurs, dans la suite de CAMPBELL et SHILLER [1987], à ne pas considérer uniquement ce test dans l'analyse du RVAR, mais à rechercher également une mesure de l'écart entre les pentes observée et théorique. Cette approche ne s'intéresse plus aux contraintes portant sur les paramètres du modèle, mais directement aux dynamiques des pentes des taux observée et théorique. Ces procédures de test « affaiblies » reposent alors sur la régression suivante :

$$\tilde{S}_t^{(m,n)} = \alpha^{(m,n)} + \beta^{(m,n)} S_t^{(m,n)} + \varepsilon_t^{(m,n)}. \tag{18}$$

Le test est fondé sur l'égalité à 1 du paramètre $\beta^{(m,n)}$. L'estimation de la statistique de test associée est toutefois relativement complexe. En effet, la variable $\tilde{S}_t^{(m,n)}$ n'est pas une série directement observée, mais calculée avec incertitude (puisque son évaluation dépend des paramètres estimés du RVAR). Ainsi, $\hat{\beta}^{(m,n)}$ peut être obtenu simplement par l'estimation de l'équation (18), mais sa distribution asymptotique dépend de la matrice de variance-covariance des estimateurs du RVAR[8].

Comme indiqué plus haut, les pentes des taux se calculent respectivement sous la forme : $S_t^{(m,n)} = g' z_t$ et $\tilde{S}_t^{(m,n)} = \theta' z_t$. Comme z_t a été préalablement centré, $S_t^{(m,n)}$ et $\tilde{S}_t^{(m,n)}$ le sont également. Le paramètre estimé $\hat{\beta}^{(m,n)}$ de la relation (18) s'exprime directement en fonction de la corrélation entre $S_t^{(m,n)}$ et $\tilde{S}_t^{(m,n)}$ et du rapport de leur variance, sous la forme :

$$\hat{\beta}_\theta^{(m,n)} = \operatorname{corr}\left(\tilde{S}_t^{(m,n)}, S_t^{(m,n)}\right) \sqrt{\frac{\operatorname{var}\left(\tilde{S}_t^{(m,n)}\right)}{\operatorname{var}\left(S_t^{(m,n)}\right)}} = \rho_\theta^{(m,n)} \sqrt{V_\theta^{(m,n)}} \tag{19}$$

où l'indice θ précise que la statistique dépend des paramètres du RVAR à travers θ. L'hypothèse nulle $H_0^3 : \tilde{S}_t^{(m,n)} = S_t^{(m,n)}$ peut donc être étudiée à

8. De fait, certains auteurs se sont intéressés aux statistiques de test affaiblies, sans donner d'indication sur leur écart-type (par exemple, MACDONALD et SPEIGHT [1991], ou GERLACH [1996]).

partir des trois égalités suivantes[9] : (1) le coefficient de régression est égal à 1 : $\beta_\theta^{(m,n)} = 1$ (le paramètre $\alpha_\theta^{(m,n)}$ est nul par définition, puisque les pentes ont été préalablement centrées) ; (2) le coefficient de corrélation est égal à 1 : $\rho_\theta^{(m,n)} = 1$[10] ; (3) les variances de la pente observée et de la pente théorique sont égales : $V_\theta^{(m,n)} = 1$.

En notant $\Omega_T = zz'/T$ la matrice de variance-covariance de $z = (z_1 \cdots z_T)$, on définit :

$$a(\theta) = \theta'\Omega_T g = \widehat{cov}\left(\tilde{S}_t^{(m,n)}, S_t^{(m,n)}\right),$$

$$b(\theta) = \theta'\Omega_T\theta = \widehat{var}\left(\tilde{S}_t^{(m,n)}\right)$$

et $$c = g'\Omega_T g = \widehat{var}\left(S_t^{(m,n)}\right).$$

Le coefficient de régression, le coefficient de corrélation et le rapport des variances s'écrivent en fonction des moments du second ordre des pentes des taux observée et théorique :

$$\begin{aligned}\hat{\beta}_\theta^{(m,n)} &= \frac{g'zz'\theta}{g'zz'g} = \frac{a(\theta)}{c}\\ \rho_\theta^{(m,n)} &= \frac{\theta'zz'g}{\sqrt{(g'zz'g)(\theta'zz'\theta)}} = \frac{a(\theta)}{\sqrt{b(\theta)c}}\\ V_\theta^{(m,n)} &= \frac{\theta'zz'\theta}{g'zz'g} = \frac{b(\theta)}{c}.\end{aligned}$$

Chacune de ces statistiques dépend donc, de façon évidente (à travers θ), des paramètres du RVAR.

Dans la suite de CAMPBELL et SHILLER [1991], de HARDOUVELIS [1994] ou de TZAVALIS et WICKENS [1998], on peut interpréter le rejet de l'égalité à 1 de $\beta_\theta^{(m,n)}$ en termes de prime de risque variable et/ou de sur-réaction du taux long au taux court. Deux types de situations peuvent être distinguées. D'une part, si le rapport des variances est inférieur à 1, mais la corrélation proche de 1, on peut en déduire qu'il existe une relation très forte entre les pentes observée et théorique, mais la pente observée réagit plus fortement que la pente théorique aux variations anticipées du taux court. Autrement dit, le taux long sur-réagit aux variations anticipées du taux court. La non-rationalité des anticipations des agents apparaît alors comme une explication suffisante du rejet de l'hypothèse $\beta_\theta^{(m,n)} = 1$. D'autre part, si le rapport des variances et la corrélation sont inférieurs à 1, non seulement il y a sur-réaction des taux longs

9. CAMPBELL et SHILLER [1987 et 1988] se sont intéressés à une autre statistique, le rapport des variances des innovations, observée et théorique, portant sur le taux long. Lorsque la maturité du titre long est infinie, ces innovations s'interprètent également comme l'excès de rendement de portage. Toutefois, lorsque la maturité du titre long est finie, on ne peut pas déduire de la théorie des anticipations des implications testables concernant le rapport des variances des innovations du taux long.

10. Ce test est équivalent au test $R^2 = 1$, où R^2 est le coefficient de détermination de la régression (18), mis en œuvre, par exemple, par HURN, MOODY et MUSCATELLI [1995] ou CUTHBERTSON [1996].

aux variations anticipées des taux courts, mais de plus la pente observée évolue, au moins partiellement, indépendamment de la pente théorique. Ce résultat peut traduire la présence d'une prime de risque variable, et la surréaction des taux longs s'explique alors par le fait que la prime de risque est corrélée aux variations du taux court.

Le calcul analytique de la distribution asymptotique de $\hat{\beta}_{\theta}^{(m,n)}$, $\rho_{\theta}^{(m,n)}$ et $V_{\theta}^{(m,n)}$ repose sur la dépendance de ces statistiques vis-à-vis des paramètres estimés $\hat{b}^{c}$ du RVAR. Il s'agit d'exprimer la statistique de test en fonction des paramètres du RVAR et d'en déduire l'expression de la variance associée. La distribution asymptotique des différentes statistiques de test est donc évaluée à partir d'un développement de Taylor de la fonction, autour de la valeur estimée de $\hat{b}^{c}$. Si $x_{\theta}^{(m,n)}$ représente alternativement $\hat{\beta}_{\theta}^{(m,n)}$, $\rho_{\theta}^{(m,n)}$ ou $V_{\theta}^{(m,n)}$, les distributions asymptotiques s'écrivent alors de la façon suivante :

$$\sqrt{T}\left(\hat{x}_{\theta}^{(m,n)} - x_{\theta}^{(m,n)}\right) \xrightarrow{d} N\left(0, \frac{\partial x_{\theta}^{(m,n)}}{\partial b^{c\prime}} \Sigma_{b^c} \frac{\partial x_{\theta}^{(m,n)}}{\partial b^{c}}^{\prime}\right) \tag{20}$$

avec

$$\frac{\partial x_{\theta}^{(m,n)}}{\partial b^{c\prime}} = \frac{\partial x_{\theta}^{(m,n)}}{\partial \theta^{\prime}} \frac{\partial \theta}{\partial b^{c\prime}}$$

où

$$\begin{aligned}
\frac{\partial \hat{\beta}_{\theta}^{(m,n)}}{\partial \theta^{\prime}} &= \frac{g^{\prime} \Omega_T}{c} \\
\frac{\partial \rho_{\theta}^{(m,n)}}{\partial \theta^{\prime}} &= \frac{1}{\sqrt{b(\theta) c}} \left(g - \frac{a(\theta)}{b(\theta)} \theta\right)^{\prime} \Omega_T \\
\frac{\partial V_{\theta}^{(m,n)}}{\partial \theta^{\prime}} &= \tfrac{2}{c} \theta^{\prime} \Omega_T .
\end{aligned}$$

L'expression $\frac{\partial \theta}{\partial b^{c\prime}} = -\frac{\partial r_1\left(b^c\right)}{\partial b^{c\prime}} = -D_1$ est définie dans l'annexe 1. Les dérivées sont évaluées au point $\theta = \hat{\theta}$, correspondant aux paramètres estimés $\hat{b}^{c}$ du RVAR.

Comme l'ont souligné STAMBAUGH [1986] puis CAMPBELL et SHILLER [1991], il peut exister des biais de petit échantillon dans l'estimation de la matrice de variance-covariance, lorsque le modèle inclut des variables prédéterminées mais non exogènes. Dans ce cas, les propriétés à distance finie peuvent différer sensiblement des distributions asymptotiques, et la mise en œuvre de la méthode analytique est susceptible de conduire à des conclusions erronées. En outre, la vraie distribution de l'estimateur du coefficient de corrélation est non-normale, puisque le coefficient est borné par -1 et 1. L'utilisation de la distribution asymptotique (normale) n'est donc pas adaptée à proximité des bornes. Comme pour le test formel, la distribution à distance finie des estimateurs est évaluée à partir de simulations de *bootstrapping*.

TABLEAU 1

Test de stationnarité des taux d'intérêt et des pentes des taux d'intérêt

Le test ADF est fondé sur l'estimation de la relation suivante : $\Delta x_t = \mu + \varphi x_{t-1} + \sum_{i=1}^{l} \theta_i \Delta x_{t-i} + u_t$, où x_t est le taux d'intérêt et u_t le terme d'erreur. L'ordre du processus autorégressif, l, est sélectionné de façon à blanchir les résidus (le nombre de retards est incrémenté jusqu'à ce que la statistique de Ljung-Box ne permette plus de rejeter l'hypothèse nulle d'absence d'auto-corrélation des résidus, au seuil de 5 %). La statistique de test, τ_φ, est la statistique de Student associée au paramètre φ. Les valeurs critiques sont celles de FULLER [1976]. Le test de SCHMIDT et PHILLIPS [1992] est fondé sur l'estimation de la relation suivante : $\Delta x_t = \tilde{\mu} + \tilde{\varphi}\tilde{S}_{t-1} + \tilde{u}_t$, où $\tilde{S}_t$ est le résidu de la régression $x_t = \psi + \xi t + S_t$. La statistique de test $\tau_{\tilde{\varphi}}$ est la statistique de Student associée au paramètre $\tilde{\varphi}$. Les statistiques de test sont corrigées de façon à tenir compte de l'auto-corrélation des résidus selon la procédure développée par PHILLIPS [1987]. Les valeurs critiques sont celles de SCHMIDT et PHILLIPS [1992]. [a] et [b] indiquent que la statistique de test est significative à un seuil de 1 % et 5 % respectivement.

Variable	Stat. ADF τ_φ	Stat. de SCHMIDT-PHILLIPS $\tau_{\tilde{\varphi}}$	Stat. de SCHMIDT-PHILLIPS $T\tilde{\varphi}$	Stat. ADF τ_φ	Stat. de SCHMIDT-PHILLIPS $\tau_{\tilde{\varphi}}$	Stat. de SCHMIDT-PHILLIPS $T\tilde{\varphi}$
		États-Unis			**Allemagne**	
$r_t^{(1)}$	-1,554	-2,032	-8,180	-0,985	-1,256	-3,159
$r_t^{(3)}$	-1,592	-2,089	-8,647	-0,875	-1,274	-3,252
$r_t^{(6)}$	-1,667	-2,189	-9,469	-1,200	-1,308	-3,425
$r_t^{(12)}$	-1,782	-2,349	-10,870	-1,107	-1,389	-3,859
$\Delta r_t^{(1)}$	-13,453 [a]	-14,905 [a]	-235,605 [a]	-6,304 [a]	-16,231 [a]	-291,049 [a]
$\Delta r_t^{(3)}$	-7,850 [a]	-13,283 [a]	-197,905 [a]	-9,033 [a]	-15,017 [a]	-255,809 [a]
$\Delta r_t^{(6)}$	-8,217 [a]	-12,862 [a]	-183,245 [a]	-3,837 [a]	-13,407 [a]	-210,614 [a]
$\Delta r_t^{(12)}$	-8,467 [a]	-12,115 [a]	-163,764 [a]	-4,544 [a]	-12,256 [a]	-182,106 [a]
$S_t^{(1,3)}$	-8,174 [a]	-7,977 [a]	-106,160 [a]	-3,728 [a]	-7,800 [a]	-101,450 [a]
$S_t^{(1,6)}$	-3,755 [a]	-5,679 [a]	-56,913 [a]	-3,236 [a]	-4,304 [a]	-34,164 [a]
$S_t^{(1,12)}$	-4,155 [a]	-4,536 [a]	-37,449 [a]	-2,967 [a]	-3,036 [a]	-17,694 [a]
		France			**Royaume-Uni**	
$r_t^{(1)}$	2,154	-2,904	-16,118	1,389	-1,795	-6,396
$r_t^{(3)}$	-1,735	-2,923	-16,332	-1,380	-1,780	-6,288
$r_t^{(6)}$	-1,835	-2,454	-11,709	-1,401	-1,830	-6,642
$r_t^{(12)}$	-2,185	-2,120	-8,817	-1,490	-1,917	-7,272
$\Delta r_t^{(1)}$	-8,924 [a]	-7,154 [a]	-58,365 [a]	-9,013 [a]	-12,279 [a]	-168,909 [a]
$\Delta r_t^{(3)}$	-7,444 [a]	-6,612 [a]	-52,888 [a]	-8,982 [a]	-11,970 [a]	-163,303 [a]
$\Delta r_t^{(6)}$	-9,341 [a]	-7,778 [a]	-73,570 [a]	-9,562 [a]	-12,733 [a]	-173,024 [a]
$\Delta r_t^{(12)}$	-7,186 [a]	-7,975 [a]	-84,800 [a]	-10,149 [a]	-11,626 [a]	-150,429 [a]
$S_t^{(1,3)}$	-5,491 [a]	-7,911 [a]	-100,209 [a]	-5,541 [a]	-7,978 [a]	-102,009 [a]
$S_t^{(1,6)}$	-3,811 [a]	-4,379 [a]	-34,472 [a]	-4,994 [a]	-5,842 [a]	-59,206 [a]
$S_t^{(1,12)}$	-4,196 [a]	-3,258 [b]	-19,787 [b]	-4,215 [a]	-4,331 [a]	-34,458 [a]

4 Les données et l'estimation de la représentation RVAR

L'étude empirique porte sur les taux sur euro-devise pour le Dollar (*USD*), le Deutsche Mark (*DEM*), le Franc français (*FRF*) et la Livre britannique (*GBP*). L'échantillon couvre des données fin de mois, d'octobre 1982 à décembre 1997, pour les maturités 1 mois, 3 mois, 6 mois et 12 mois. m est égal à 1 mois et n est compris entre 3 et 12 mois[11]. Le choix de cette période d'estimation permet d'éviter les perturbations liées au changement de politique monétaire par la *Federal Reserve Bank* (entre octobre 1979 et octobre 1982) et à l'adoption d'une nouvelle cible opérationnelle à court terme[12]. Le choix d'une fréquence mensuelle correspond à l'horizon du taux le plus court retenu dans cette étude, de façon à éviter le chevauchement des données. De plus, cette fréquence atténue les problèmes liés aux propriétés statistiques des données (notamment l'hétéroscédasticité résiduelle). L'utilisation des taux sur euro-devise vise un double objectif. D'une part, pour chaque devise, les titres échangés et les intervenants sur ce segment sont homogènes (ce qui est très rarement le cas, à l'inverse, quand on considère un spectre de maturités plus large, comme le couple de maturités 3 mois-10 ans). D'autre part, les comparaisons internationales sont facilitées par le fait que les marchés d'euro-devise sont harmonisés (mêmes conventions de calcul, même fiscalité, ...).

Avant de considérer les différents tests de la théorie des anticipations, il est nécessaire de valider au préalable les hypothèses sous-jacentes à la représentation RVAR, à savoir d'une part, de la non-stationnarité des taux d'intérêt et, d'autre part, de la stationnarité des pentes des taux. Le tableau 1 présente les résultats de différents tests de stationnarité menés sur les taux, les variations de taux et les pentes de taux. Il s'agit d'une part de la statistique ADF standard et d'autre part des statistiques de SCHMIDT et PHILLIPS [1992]. Ces dernières, en traitant préalablement les composantes déterministes du processus, permettent de clarifier les hypothèses testées et de simplifier la mise en œuvre du test de stationnarité[13]. Les différents tests menés ne permettent jamais de rejeter la non-stationnarité des taux d'intérêt sur euro-devise au

11. Les données proviennent de la base Datastream. Dans le cas de l'euro-franc, les estimations débutent en mars 1983, de façon à exclure la période de forte volatilité des taux courts français, lors de la dévaluation du franc survenue en mars 1983. Tous les calculs ont été réalisés avec le logiciel GAUSS.

12. La réunification allemande ne paraît avoir perturbé significativement la dynamique des taux d'intérêt allemands. L'estimation du modèle sur les deux sous-périodes (avant et après la réunification) conduit aux mêmes conclusions que sur l'ensemble de la période : le test formel rejette la théorie des anticipations et les tests affaiblis indiquent que le rejet provient du rapport des variances.

13. On ne considère pas ici d'hypothèses alternatives plus générales que la stationnarité (éventuellement autour d'une tendance déterministe). En particulier, on exclut l'éventualité de ruptures déterministes, suivant en cela les résultats obtenus par PERRON [1989] pour les taux d'intérêt américains. L'hypothèse de non-stationnarité est donc supposée être la représentation la plus parcimonieuse des taux d'intérêt.

TABLEAU 2
Estimation du RVAR

Le tableau présente les principales statistiques associées à l'estimation du RVAR : le nombre de retards optimal p (sélectionné par la procédure HQ) ; pour chaque équation du RVAR, l'écart-type estimé des résidus de la régression σ_η, le coefficient de détermination associé $\bar{R}^2$; la statistique de Ljung-Box, corrigée pour l'éventuelle hétéroscédasticité, $LB_c\,(24)$, associée au test de l'hypothèse nulle d'absence d'auto-corrélation des résidus ; enfin, la statistique du multiplicateur de Lagrange, $ARCH(24)$, associée au test de l'hypothèse nulle d'homoscédasticité des résidus. Ces deux statistiques de test suivent, sous leur hypothèse nulle respective, un χ^2 à 24 degrés de liberté. [a] et [b] indiquent que la statistique de test est significative à un seuil de 1 % et 5 % respectivement.

$m-n$	p	éq.	σ_η	$\bar{R}^2$	LB_c (24)	$ARCH$ (24)
États-Unis						
$1-3$	4	Δr	0,345	0,194	25,848	36,525 [b]
		S	0,141	0,306	42,724 [b]	29,061
$1-6$	4	Δr	0,358	0,131	24,922	24,178
		S	0,211	0,513	29,174	22,709
$1-12$	4	Δr	0,362	0,108	28,375	27,562
		S	0,303	0,623	27,060	16,073
Allemagne						
$1-3$	2	Δr	0,268	0,134	21,701	35,585
		S	0,124	0,417	30,833	24,143
$1-6$	2	Δr	0,271	0,111	23,381	33,385
		S	0,180	0,666	20,895	12,297
$1-12$	2	Δr	0,274	0,093	25,711	32,029
		S	0,230	0,796	16,817	18,779
France						
$1-3$	5	Δr	1,122	0,284	13,117	29,101
		S	0,519	0,194	20,333	2,049
$1-6$	4	Δr	1,164	0,225	19,400	20,705
		S	0,730	0,417	21,606	10,713
$1-12$	4	Δr	1,184	0,199	18,723	20,449
		S	0,927	0,501	18,329	28,123
Royaume-Uni						
$1-3$	4	Δr	0,429	0,175	11,366	33,024
		S	0,173	0,415	17,364	27,489
$1-6$	4	Δr	0,435	0,154	11,351	34,162
		S	0,310	0,546	17,574	41,285 [b]
$1-12$	3	Δr	0,435	0,147	11,524	36,056
		S	0,412	0,671	18,821	32,629

seuil de 5 %. En ce qui concerne les pentes des taux, elles peuvent toutes être considérées comme stationnaires au seuil de 5 %. L'hypothèse H_0^c ne semble donc pas être rejetée sur les données étudiées. Dans la suite, les conditions de validité du modèle RVAR, concernant les propriétés de cointégration, sont supposées satisfaites. On reviendra toutefois sur cette hypothèse dans le cadre des tests de non-causalité.

On considère maintenant les résultats des estimations portant sur la représentation RVAR suivante :

$$
(21) \quad \begin{pmatrix} \Delta r_t^{(m)} \\ S_t^{(m,n)} \end{pmatrix} = \begin{bmatrix} b_{11}^{(1)} & b_{12}^{(1)} \\ b_{21}^{(1)} & b_{22}^{(1)} \end{bmatrix} \begin{pmatrix} \Delta r_{t-1}^{(m)} \\ S_{t-1}^{(m,n)} \end{pmatrix} + \ldots + \begin{bmatrix} b_{11}^{(p-1)} & b_{12}^{(p-1)} \\ b_{21}^{(p-1)} & b_{22}^{(p-1)} \end{bmatrix} \begin{pmatrix} \Delta r_{t-p+1}^{(m)} \\ S_{t-p+1}^{(m,n)} \end{pmatrix} + \begin{bmatrix} 0 & b_{12}^{(p)} \\ 0 & b_{22}^{(p)} \end{bmatrix} \begin{pmatrix} \Delta r_{t-p}^{(m)} \\ S_{t-p}^{(m,n)} \end{pmatrix} + \eta_t .
$$

Le nombre de retards p est sélectionné par la procédure de HQ avec $p_{\max} = 24$ (HANNAN et QUINN [1979]). La matrice de variance-covariance des paramètres du RVAR (Σ_{b^c}) est estimée de la façon suivante : elle est corrigée pour l'auto-corrélation et l'hétéroscédasticité éventuelle des innovations selon la démarche de HANSEN et HODRICK [1980] et WHITE [1980]. La semi-positivité de la matrice est assurée par la correction de NEWEY et WEST [1987]. En l'absence d'information sur le degré d'auto-corrélation et d'hétéroscédasticité des innovations, le nombre optimal d'auto-corrélations calculées pour la correction de NEWEY et WEST est déterminé selon l'approche développée, notamment, dans DEN HAAN et LEVIN [1996][14].

Le tableau 2 présente les principales caractéristiques des modèles RVAR estimés pour chaque marché et chaque couple de maturités. Il s'agit du nombre de retard optimal et, pour chacune des deux équations du modèle, de l'écart-type résiduel, du $\bar{R}^2$, de la statistique de Ljung-Box LB_c (corrigée de l'hétéroscédasticité), associée au test de nullité des 24 premières auto-corrélations des résidus, et de la statistique du multiplicateur de Lagrange $ARCH$, associée à l'hypothèse nulle d'homoscédasticité des résidus.

On note tout d'abord que la qualité des modèles estimés pour les variations de taux se dégrade, de façon sensible, pour tous les marchés, avec la maturité des placements. En effet, le $\bar{R}^2$ de la régression est plus fort pour les maturités les plus courtes que pour les maturités longues. Ainsi pour le couple (1-3) mois, le $\bar{R}^2$ est compris entre 0,13 pour le taux sur l'euro-mark et 0,28 pour les taux sur l'euro-franc. Pour le couple (1-12) mois, en revanche, le $\bar{R}^2$ est compris entre 0,09 et 0,20. À l'inverse, en ce qui concerne les pentes des taux, le $\bar{R}^2$ croît sensiblement avec les maturités : il est compris entre 0,19 et

14. Le nombre optimal d'auto-corrélations est déterminé, selon ANDREWS [1991] et ANDREWS et MONAHAN [1992], par

$$
l^* = 1{,}1447 \left(T\hat{\alpha}\right)^{1/3} \qquad \text{avec} \qquad \hat{\alpha} = \frac{\sum_{i=1}^{N} \frac{4\hat{\rho}_i^2 \hat{\sigma}_i^4}{(1-\hat{\rho}_i)^6 (1+\hat{\rho}_i)^2}}{\sum_{i=1}^{N} \frac{\hat{\sigma}_i^4}{(1+\hat{\rho}_i)^4}}
$$

avec $\hat{\sigma}_i^2$ et $\hat{\rho}_i$ la variance et l'auto-corrélation d'ordre 1 du terme $\left(X_t \hat{\eta}_t\right)$, où X_t est le vecteur des N variables exogènes et $\hat{\eta}_t$ le vecteur des résidus estimés.

TABLEAU 3

Tests de causalité

Le tableau présente les résultats des tests de causalité associés à la théorie des anticipations. Les colonnes 1 et 3 reportent le nombre de retards optimal p_1 et p_2 identifiés par la procédure HQ dans les RVAR associés aux systèmes $\left(\Delta r_t^{(m)}, S_t^{(m,n)}\right)'$ et $\left(\Delta r_t^{(n)}, S_t^{(m,n)}\right)'$. Les colonnes 2 et 4 indiquent les statistiques du test de nullité jointe des paramètres associés à la pente des taux dans l'équation de variation du taux court et de variation du taux long respectivement. Ces statistiques suivent, sous l'hypothèse nulle de non-causalité, un χ^2 à p_i degrés de liberté, $i = 1,\ 2$. Entre crochets est indiqué le niveau de significativité du test obtenu à partir de l'écart-type asymptotique. La colonne 5 reporte le nombre de retards optimal pour le VAR en niveau $(\tilde{p}+1)$, où $\tilde{p}$ est le nombre optimal déterminé par la procédure HQ. Les colonnes 6 et 7 indiquent les statistiques du test de nullité jointe des paramètres associés aux $\tilde{p}$ premiers retards du taux long dans l'équation de taux court et du taux court dans l'équation de taux long respectivement. Ces statistiques suivent, sous l'hypothèse nulle de non-causalité, un χ^2 à $\tilde{p}$ degrés de liberté.

	Représentation RVAR				Représentation VAR en niveau		
$m-n$	p_1	$S^{(m,n)} \to \Delta r^{(m)}$	p_2	$S^{(m,n)} \to \Delta r^{(n)}$	$\tilde{p}+1$	$r^{(n)} \to r^{(m)}$	$r^{(m)} \to r^{(n)}$
États-Unis							
$1-3$	4	47,650	4	11,566	4	37,697	11,172
		[0,000]		[0,021]		[0,000]	[0,011]
$1-6$	4	28,913	4	4,438	5	22,163	3,432
		[0,000]		[0,350]		[0,000]	[0,488]
$1-12$	4	27,104	2	0,673	5	21,942	0,792
		[0,000]		[0,714]		[0,000]	[0,940]
Allemagne							
$1-3$	2	29,975	3	5,790	4	14,785	1,634
		[0,000]		[0,122]		[0,002]	[0,652]
$1-6$	2	34,392	2	4,057	4	8,790	0,706
		[0,000]		[0,132]		[0,032]	[0,872]
$1-12$	2	23,055	2	2,087	4	6,743	0,875
		[0,000]		[0,352]		[0,081]	[0,831]
France							
$1-3$	5	33,147	5	8,631	7	48,858	10,836
		[0,000]		[0,125]		[0,000]	[0,094]
$1-6$	4	12,942	5	9,569	7	26,794	17,239
		[0,012]		[0,088]		[0,000]	[0,008]
$1-12$	4	10,758	4	4,532	6	15,521	8,809
		[0,029]		[0,339]		[0,008]	[0,117]
Royaume-Uni							
$1-3$	4	31,760	2	5,227	2	14,980	7,769
		[0,000]		[0,073]		[0,000]	[0,005]
$1-6$	4	33,290	2	3,115	2	16,263	4,479
		[0,000]		[0,211]		[0,000]	[0,034]
$1-12$	3	29,167	2	3,682	2	21,514	4,368
		[0,000]		[0,159]		[0,000]	[0,037]

0,42 pour le couple de maturités (1-3) mois et entre 0,50 et 0,80 pour le couple (1-12) mois.

Les statistiques de Ljung-Box (corrigées pour l'éventuelle hétéroscédasticité résiduelle) montrent que les résidus ne présentent pas d'auto-corrélation, à l'exception de l'équation de pente des taux pour le couple de maturités (1-3) mois aux États-Unis. Enfin, les statistiques *ARCH* ne mettent en évidence une hétéroscédasticité résiduelle que pour deux des 24 équations estimées.

Pour l'essentiel, les différents modèles estimés présentent donc des propriétés statistiques satisfaisantes (absence d'auto-corrélation, homoscédasticité). Elles assurent en particulier la validité des exercices de simulation menés dans la section suivante.

5 Les tests de la théorie des anticipations

5.1 Les tests de causalité

Le tableau 3 présente les résultats des tests de non-causalité. Deux types de test sont reportés : d'une part, l'hypothèse de non-causalité de la pente vis-à-vis des variations de taux est évaluée dans le cadre du RVAR. De façon à montrer le sens de causalité entre les deux taux d'intérêt, on considère à la fois la non-causalité de la pente vis-à-vis de la variation du taux court et vis-à-vis de la variation du taux long. On estime donc, pour chaque couple de maturités (m,n), deux modèles RVAR, associés aux systèmes $\left(\Delta r_t^{(m)} \;\; S_t^{(m,n)}\right)'$ et $\left(\Delta r_t^{(n)} \;\; S_t^{(m,n)}\right)'$. D'autre part, le test de non-causalité entre les taux d'intérêt eux-mêmes est mené dans le VAR en niveau, associé au système $\left(r_t^{(m)} \;\; r_t^{(n)}\right)'$, selon l'approche de Dolado et Lütkepohl [1996]. Dans tous les cas, le nombre de retards optimal est identifié par la procédure de HQ, tout en s'assurant de la significativité jointe des paramètres associés à la dernière matrice de retards.

Le tableau 3 montre que l'hypothèse de non-causalité de la pente des taux vis-à-vis de la variation du taux court est rejetée pour tous les couples de maturités et tous les marchés, confirmant l'implication de la théorie des anticipations en termes de causalité au sein du RVAR (6). Ce résultat confirme ainsi très largement les résultats obtenus par des études antérieures, sur données américaines (Campbell et Shiller [1987 et 1988]), allemandes (MacDonald et Speight [1991]), françaises (Colletaz et Gourlaouen [1990]) ou britanniques (Cuthbertson [1996]). En revanche, on ne met pas en évidence de causalité marquée de la pente des taux vis-à-vis du taux long. En effet, la non-causalité n'est rejetée que pour le couple (1-3) mois aux États-Unis, au seuil de 5 %, et pour les couples (1-6) mois en France et (1-3) mois au Royaume-Uni, au seuil de 10 %.

TABLEAU 4
Tests formels

Le tableau présente les résultats des tests formels associés à la théorie des anticipations. Deux colonnes sont consacrées à chacune des statistiques de test $r_1(b^c)$ et $r_2(b^c)$. Dans la première colonne, on trouve la statistique $r_1(b^c)$ ou $r_2(b^c)$ définie dans les équations (15) et (16). Sous l'hypothèse nulle, ces statistiques suivent un χ^2 à $(2p-1)$ degrés de liberté. La seconde colonne indique les niveaux de significativité : d'une part, le niveau de significativité du test obtenu à partir de l'écart-type asymptotique ; d'autre part, le niveau de significativité du test obtenu par *bootstrapping* à partir de 10 000 simulations du RVAR estimé.

	$r_1(b^c)=0$		$r_2(b^c)=0$	
$m-n$	Statistique	Niveau de significativité	Statistique	Niveau de significativité
États-Unis				
$1-3$	22,216	[0,002]	27,724	[0,000]
		[0,017]		[0,008]
$1-6$	13,064	[0,071]	15,882	[0,026]
		[0,100]		[0,041]
$1-12$	14,744	[0,039]	12,237	[0,093]
		[0,071]		[0,092]
Allemagne				
$1-3$	19,052	[0,000]	13,596	[0,004]
		[0,005]		[0,021]
$1-6$	17,487	[0,001]	10,179	[0,017]
		[0,006]		[0,039]
$1-12$	10,259	[0,017]	10,735	[0,013]
		[0,053]		[0,031]
France				
$1-3$	14,705	[0,099]	10,900	[0,283]
		[0,216]		[0,364]
$1-6$	9,585	[0,213]	7,605	[0,369]
		[0,185]		[0,251]
$1-12$	6,956	[0,434]	6,517	[0,481]
		[0,266]		[0,278]
Royaume-Uni				
$1-3$	4,438	[0,728]	4,067	[0,772]
		[0,354]		[0,383]
$1-6$	7,315	[0,397]	5,757	[0,568]
		[0,241]		[0,326]
$1-12$	9,631	[0,086]	6,731	[0,241]
		[0,147]		[0,273]

Comme on l'a vu précédemment, ces tests de non-causalité peuvent conduire à des conclusions erronées lorsque l'hypothèse H_0^c n'est pas valide. Pour se prémunir contre le risque d'avoir conclu à tort en faveur de H_0^c, on teste également l'hypothèse de causalité dans le VAR en niveau. Le VAR est alors estimé avec $(\tilde{p}+1)$ retards, où $\tilde{p}$ est le nombre de retards optimal du VAR en niveau. Au seuil de risque de 5 %, la plupart des conclusions obtenues avec le RVAR sont confirmées avec le VAR en niveau. D'une part, on rejette la non-causalité du taux long vis-à-vis du taux court. D'autre part, on ne rejette pas en général la non-causalité du taux court vis-à-vis du taux long aux États-Unis et en Allemagne. En revanche, on rejette systématiquement la non-causalité pour les données britanniques. Dans le cas de la France, la non-causalité est rejetée pour les trois couples de maturités si on adopte un seuil de risque de 12 %.

Le théorème de représentation de ENGLE et GRANGER [1987] établit qu'en cas de cointégration la force de rappel doit avoir un effet significatif dans au moins l'une des deux équations du VECM. Les estimations menées mettent bien en évidence ce résultat, puisque la pente des taux a toujours un effet causal significatif sur la variation du taux court. L'hypothèse H_0^c est donc bien confirmée pour tous les couples de maturités. Pour cette raison, on poursuit la démarche en supposant H_0^c vérifiée. En revanche, selon la théorie des anticipations, la pente des taux devrait également avoir un effet causal sur la variation du taux long ou, de façon alternative, le taux court devrait avoir un effet causal sur le taux long. On n'obtient un tel résultat que pour certains couples de maturités pour les taux britanniques et français. La difficulté à obtenir un effet causal du taux court sur le taux long confirme les résultats obtenus, dans un cadre multivarié, par BRUNEAU et JONDEAU [1999a].

5.2 Les tests formels

Les résultats des tests formels sont regroupés dans le tableau 4. Pour chacune des deux statistiques de test $r_1\left(b^c\right)$ et $r_2\left(b^c\right)$ sont indiqués la statistique de test, le niveau de significativité obtenu à partir de la distribution asymptotique et enfin le niveau de significativité obtenu à partir des simulations par *bootstrapping*.

Les tests formels permettent à nouveau de distinguer deux groupes de pays : dans le premier groupe (France et Royaume-Uni), l'hypothèse nulle $H_0^2 : S_t^{(m,n)} = \tilde{S}_t^{(m,n)}$ n'est rejetée pour aucun couple de maturités. Ce résultat est obtenu quelle que soit la statistique de test ($r_1\left(b^c\right)$ ou $r_2\left(b^c\right)$) et quelle que soit la distribution (asymptotique ou à distance finie) utilisée pour le calcul de la statistique de Wald.

Dans le second groupe (États-Unis et Allemagne), la théorie des anticipations est en général rejetée aux seuils de risque usuels. Le rejet est systématique en Allemagne, puisque tous les niveaux de significativité obtenus sont inférieurs à 5,3 %. Dans le cas des États-Unis, les résultats sont moins nets, mais les niveaux de significativité du test formel sont tous inférieurs à 10 %.

Pour positionner ces résultats par rapport à la littérature empirique antérieure, on note tout d'abord que de nombreuses études ont porté sur des maturités plus longues (généralement des taux à 3 mois et à 10 ans). MACDONALD et SPEIGHT [1991] rejettent la théorie des anticipations pour les États-Unis et l'Allemagne, mais pas pour le Royaume-Uni. GERLACH [1996] trouve que le test formel ne conduit pas à un rejet de la théorie des anticipations pour l'Allemagne, la France et le Royaume-Uni. Pour les États-Unis, en revanche, le rejet est très net. À partir du test formel, HARDOUVELIS [1994] ne rejette jamais la théorie pour les quatre pays étudiés ici. Concernant les marchés de plus court terme (typiquement, à moins d'un an), les résultats obtenus jusqu'à présent sont également ambigus. Ainsi sur données britan-

TABLEAU 5
Statistiques de test affaiblies

Le tableau présente les statistiques de test affaiblies associées à la théorie des anticipations. Deux colonnes sont consacrées à chacune des statistiques d'intérêt (β, V et ρ). Dans la première colonne, on trouve la statistique de test et, entre parenthèses, l'écart-type asymptotique. La seconde colonne indique les niveaux de significativité S_1 et S_2 associés à l'hypothèse d'égalité à 1 de la statistique : S_1 est la proportion d'échantillons simulés qui ont produit une statistique de test (β^s, V^s ou ρ^s) plus éloignée de 1 (dans la même direction) que la statistique de test obtenue à partir des observations ; S_2 est la proportion d'échantillons simulés qui ont produit une statistique de test dont la statistique de Student (évaluée à partir de l'écart-type asymptotique) rejette l'hypothèse nulle plus fortement (quelle que soit la direction) que la statistique de Student obtenue à partir des observations. Les niveaux de significativité sont obtenus par *bootstrapping*, à partir de 10 000 simulations du RVAR estimé.

	β		V		ρ	
$m-n$	Statistique (écart-type)	Niveau de signif.	Statistique (écart-type)	Niveau de signif.	Statistique (écart-type)	Niveau de signif.
États-Unis						
1 – 3	0,671	[0,015]	0,636	[0,049]	0,841	[0,019]
	(0,172)	[0,076]	(0,227)	[0,128]	(0,113)	[0,184]
1 – 6	0,399	[0,007]	0,273	[0,004]	0,763	[0,046]
	(0,237)	[0,023]	(0,163)	[0,008]	(0,339)	[0,650]
1 – 12	0,211	[0,003]	0,130	[0,001]	0,586	[0,016]
	(0,301)	[0,026]	(0,090)	[0,000]	(0,745)	[0,769]
Allemagne						
1 – 3	0,604	[0,021]	0,372	[0,015]	0,991	[0,584]
	(0,110)	[0,001]	(0,137)	[0,003]	(0,016)	[0,556]
1 – 6	0,640	[0,069]	0,412	[0,062]	0,997	[0,675]
	(0,112)	[0,002]	(0,145)	[0,011]	(0,007)	[0,665]
1 – 12	0,682	[0,156]	0,471	[0,152]	0,995	[0,331]
	(0,149)	[0,052]	(0,206)	[0,069]	(0,007)	[0,341]
France						
1 – 3	0,737	[0,010]	0,899	[0,182]	0,777	[0,001]
	(0,138)	[0,088]	(0,302)	[0,775]	(0,130)	[0,100]
1 – 6	0,688	[0,009]	0,550	[0,011]	0,927	[0,021]
	(0,202)	[0,202]	(0,244)	[0,125]	(0,083)	[0,360]
1 – 12	0,728	[0,014]	0,560	[0,014]	0,973	[0,030]
	(0,206)	[0,262]	(0,277)	[0,169]	(0,040)	[0,499]
Royaume-Uni						
1 – 3	0,723	[0,060]	0,564	[0,056]	0,963	[0,218]
	(0,229)	[0,254]	(0,311)	[0,178]	(0,052)	[0,455]
1 – 6	0,635	[0,066]	0,418	[0,046]	0,982	[0,430]
	(0,206)	[0,090]	(0,258)	[0,054]	(0,021)	[0,370]
1 – 12	0,682	[0,155]	0,480	[0,144]	0,985	[0,341]
	(0,221)	[0,169]	(0,303)	[0,130]	(0,016)	[0,231]

niques, HURN, MOODY et MUSCATELLI [1995] et DRIFFILL, PSARADAKIS et SOLA [1997] ne rejettent pas la théorie, alors que CUTHBERTSON, HAYES et NITZSCHE [1996] rejettent clairement la théorie. De même, sur données américaines, KUGLER [1990] et DRIFFILL, PSARADAKIS et SOLA [1997] obtiennent des conclusions contradictoires. Pour les quatre pays étudiés, les résultats obtenus sont, en revanche, conformes à ceux trouvés par JONDEAU [1998] sur une période sensiblement plus longue.

5.3 Les tests affaiblis

Le tableau 5 contient les différentes statistiques de test affaiblies : le coefficient de la régression de la pente théorique sur la pente observée (β), le rapport des variances (V) et le coefficient de corrélation (ρ). Pour chacune de ces statistiques, sont indiqués la statistique de test, l'écart-type asymptotique entre parenthèses et, entre crochets, les deux niveaux de significativité obtenus par *bootstrapping*, notés S_1 et S_2. Les statistiques S_1 et S_2 donnent la probabilité d'obtenir la valeur estimée de la statistique de test (β^s, V^s ou ρ^s) lorsque la dynamique des rendements est contrainte de façon à satisfaire la théorie des anticipations, selon l'approche de CAMPBELL et SHILLER [1991]. S_1 est la proportion d'échantillons simulés parmi 10 000 expériences qui ont produit une statistique de test (β, V ou ρ) plus éloignée de 1 (dans la même direction) que la statistique de test obtenue à partir des observations. S_2 est à la proportion d'échantillons simulés parmi les 10 000 expériences qui ont produit une statistique de test dont la statistique de Student (évaluée à partir de l'écart-type asymptotique) rejette l'hypothèse nulle plus fortement (quelle que soit la direction) que la statistique de Student obtenue à partir des observations. La première expérience correspond donc à un test unilatéral, alors que la seconde correspond à un test bilatéral. Le mode de calcul des niveaux de significativité S_1 et S_2 est détaillé dans l'annexe 2[15].

Les statistiques de test affaiblies permettent de compléter les conclusions des tests formels. Tout d'abord, en considérant les écarts-types asymptotiques, on retrouve bien la distinction entre taux français et britanniques d'une part, et taux américains et allemands d'autre part. Pour les premiers, on ne rejette jamais l'égalité à 1 du coefficient de régression $\beta_\theta^{(m,n)}$. Pour les seconds, le rejet est systématique au seuil de 5 %.

Dans le cas des taux sur l'euro-franc et, dans une moindre mesure, sur l'euro-livre, on note une certaine proximité de $\beta_\theta^{(m,n)}$ par rapport à 1 : l'égalité à 1 n'est jamais rejetée lorsque le test est fondé sur l'écart-type asymptotique. Le paramètre $\beta_\theta^{(m,n)}$ est compris entre 0,68 et 0,74 pour la France, et entre 0,63 et 0,73 pour le Royaume-Uni. Pour l'euro-mark, les coefficients sont relativement élevées (entre 0,60 et 0,68), mais la précision des estimations est suffisante pour rejeter de l'égalité à 1. Enfin, l'égalité $\beta_\theta^{(m,n)} = 1$ est systématiquement rejetée pour l'euro-dollar.

On considère maintenant les niveaux de significativité obtenus par *bootstrapping*. On note tout d'abord que, sauf pour l'Allemagne, le test fondé sur la statistique de Student (S_2) conduit à un moindre rejet de l'hypothèse nulle que le test fondé sur la statistique de test elle-même (S_1). Ce résultat s'explique par le fait que les données simulées sont, en général, associées à des écarts-types asymptotiques des statistiques de test plus petits que les données

15. Dans le tableau 5, seuls les résultats obtenus par *bootstrapping* sont présentés. L'évaluation des niveaux de significativité par simulation de Monte-Carlo ne modifie pas l'essentiel des résultats.

observées[16]. Pour expliquer ce constat, CAMPBELL et SHILLER [1991] ont suggéré qu'il pourrait provenir de la présence d'hétéroscédasticité dans les données. Toutefois, on obtient ici des résultats opposés pour les données allemandes et françaises, qui sont pourtant caractérisées par des résidus non auto-corrélés et homoscédastiques. On privilégie dans la suite la statistique S_2, fondée sur la statistique de Student, car elle permet de tenir compte de la précision des statistiques obtenues par simulation.

Pour les données américaines et allemandes, les deux statistiques de test S_1 et S_2 conduisent au rejet de l'hypothèse $\beta_\theta^{(m,n)} = 1$, sauf pour le couple (1-12) mois en Allemagne. Dans le cas français, cette hypothèse est rejetée quand le test est mené avec la statistique S_1, mais pas lorsqu'il est mené avec la statistique S_2. Pour le Royaume-Uni, les deux tests sont favorables à l'hypothèse nulle, puisqu'aucun niveau de significativité n'est inférieur à 6 %.

La décomposition du coefficient de régression entre le coefficient de corrélation et le rapport des variances, donnée par la relation (19), permet de préciser ces résultats. Les États-Unis se distinguent nettement des pays européens : d'une part, les corrélations ρ sont assez faibles (de 0,58 à 0,84), quoique non-significativement différentes de 1 selon la statistique S_2 ; et surtout le rapport des variances est significativement inférieur à 1 (V est compris entre 0,13 et 0,64), même en considérant les niveaux de significativité obtenus par *bootstrapping*. La variance de la pente théorique est donc nettement plus faible que celle de la pente observée.

Sur données européennes, les coefficients de corrélation entre les pentes observée et théorique sont très élevés. Ils sont en effet tous supérieurs à 0,92 – à l'exception du couple (1-3) mois pour la France. Les corrélations ne sont pas significativement différentes de 1 pour l'Allemagne et le Royaume-Uni. Pour la France, le test asymptotique et la statistique S_2 ne permettent pas de rejeter l'égalité à 1. Pour ce qui concerne le rapport des variances, la variance de la pente théorique est de l'ordre de 0,4 à 0,6 fois la variance de la pente observée. En Allemagne, les rapports des variances sont les plus faibles et en général significativement différents de 1. Au Royaume-Uni et surtout en France, l'égalité à 1 est rarement rejetée. Avec la statistique S_2, on ne rejette pas l'hypothèse $V_\theta^{(m,n)} = 1$ au seuil de 5 % pour ces deux pays.

Les résultats obtenus pour l'Allemagne et les États-Unis peuvent être directement reliés à l'hypothèse de sur-réaction des taux longs aux mouvements observés des taux courts. Cette explication a été mise en avant notamment par SHILLER [1979] et, plus récemment, par HARDOUVELIS [1994], pour rendre compte des résultats obtenus sur données américaines. HARDOUVELIS obtient, en particulier, des corrélations et des rapports de variance très proches de l'unité pour l'Allemagne, la France et le Royaume-Uni, mais sensiblement plus faibles pour les États-Unis.

L'analyse des statistiques de test affaiblies permet finalement de dégager les résultats suivants : pour les taux français et britanniques, on ne rejette pas la

16. Pour comprendre pourquoi la statistique de Student de $(\beta - 1)$ est plus favorable envers l'hypothèse nulle en France qu'en Allemagne, on peut prendre l'exemple du coefficient de régression β pour le couple (1-12) mois. Dans le cas de la France, l'écart-type asymptotique est de 0,21 sur les données observées et la médiane des écarts-types obtenus par simulation est de 0,11. Pour l'Allemagne, on obtient 0,15 et 0,25 respectivement.

théorie des anticipations. La situation allemande apparaît en quelque sorte intermédiaire, puisque les corrélations sont très proches de 1 mais les rapports des variances sont faibles et en général significativement différents de 1. L'hypothèse de sur-réaction semble donc suffisante pour rendre compte du rejet de la théorie des anticipations. Sur données américaines, enfin, cette hypothèse apparaît suffisante pour rendre compte du rejet de la théorie des anticipations. Mais le faible niveau des corrélations, même si l'égalité à 1 n'est pas toujours rejetée, indique que l'hypothèse d'une prime de risque variable serait peut-être plus adaptée pour expliquer le rejet de la théorie des anticipations.

6 Conclusion

Cet article permet de préciser certains éléments méthodologiques du test de la théorie des anticipations, dans le cadre d'une représentation explicite des anticipations des agents, à partir d'un modèle RVAR. Dans ce cadre, il est possible d'évaluer la théorie, ou bien à travers ses conséquences en termes de causalité entre les variables, ou bien à travers ses implications sur les paramètres du modèle (test formel). Cette approche permet en outre d'obtenir des prévisions des variations du taux court futur, donnant ainsi une évaluation de la pente des taux théorique. La comparaison des pentes observée et théorique, à travers les statistiques de test affaiblies, permet de préciser les causes de l'éventuel rejet de la théorie des anticipations.

La mise en œuvre des différentes formes de test pose certaines difficultés. En particulier, dans un système de variables non-stationnaires et cointégrées, la distribution asymptotique des statistiques de Wald n'est pas un χ^2 standard. La procédure de test proposée par TODA et YAMAMOTO [1995] et DOLADO et LÜTKEPOHL [1996] est alors appliquée aux tests de causalité. Elle ne remet pas en cause les résultats obtenus par l'approche usuelle. En ce qui concerne le test formel et les tests affaiblis, ces tests sont menés à partir de la distribution asymptotique de la statistique de test, mais également à partir de sa distribution à distance finie, obtenue par une procédure de simulation par *bootstrapping*.

Les résultats de ces tests sur les différents marchés sur euro-devise étudiés mettent en évidence l'opposition entre la France et le Royaume-Uni d'une part, pour lesquels la théorie n'est pas rejetée, et les États-Unis et l'Allemagne d'autre part, pour lesquels le rejet est *quasi* systématique. L'étude des statistiques de test affaiblies permet de préciser ce résultat. L'hypothèse de sur-réaction semble suffisante pour rendre compte du rejet de la théorie pour les données allemandes et américaines, même si, pour les États-Unis, l'hypothèse d'une prime de risque variable pourrait s'avérer plus adaptée. ■

• Références bibliographiques

ANDREWS D.W.K. (1991). – « Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation », *Econometrica*, 59, 3, pp. 817-858.

ANDREWS D.W.K., MONAHAN J.C. (1992). – « An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation », *Econometrica*, 60, 4, pp. 953-966.

BRUNEAU C., JONDEAU É. (1999a). – « Causalité de long terme et amélioration de la prévision : application aux courbes de taux d'intérêt », *Annales d'Économie et de Statistique*, 54, pp. 23-45.

BRUNEAU C., JONDEAU É. (1999b). – « Long-Run Causality, with an Application to International Links between Long-Term Interest Rates », *Oxford Bulletin of Economics and Statistics*, 61, 4, pp. 545-568.

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'Économie et de Statistique*, 40, pp. 177-206.

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.

COLLETAZ G., GOURLAOUEN J.-P. (1990). – « Cointégration et structure par terme des taux d'intérêt », *Revue Économique*, 41, 4, pp. 687-712.

CUTHBERTSON K. (1996). – « The Expectations Hypothesis of the Term Structure: The UK Interbank Market », *Economic Journal*, 106, 436, pp. 578-592.

CUTHBERTSON K., HAYES S., NITZSCHE D. (1996). – « The Behaviour of Certificate of Deposit Rates in the UK », *Oxford Economic Papers*, 48, 3, pp. 397-414.

DEN HAAN W.J., LEVIN A. (1996). – « A Practitioner's Guide to Robust Covariance Matrix Estimation », NBER Technical *Working Paper* 197.

DOLADO J.J., LÜTKEPOHL H. (1996). – « Making Wald Tests Work for Cointegrated VAR Systems », *Econometric Reviews*, 15, 4, pp. 369-386.

DRIFFILL J., PSARADAKIS Z., SOLA M. (1997). – « A Reconciliation of Some Paradoxical Empirical Results on the Expectations Model of the Term Structure », *Oxford Bulletin of Economics and Statistics*, 59, 1, pp. 29-42.

ENGLE R., GRANGER C.W.J. (1987). – « Cointegration and Error Correction: Representation, Estimation, and Testing », *Econometrica*, 55, 2, pp. 251-276.

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

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.

GERLACH S. (1996). – « Monetary Policy and the Behaviour of Interest Rates: Are Long Rates Excessively Volatile? », BIS *Working Paper* n° 34.

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, 3, pp. 530-548.

GREGORY A.W., VEALL M.R. (1985). – « Formulating Wald Tests of Nonlinear Restrictions », *Econometrica*, 53, 6, pp. 1465-1468.

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 Determinant of the Order of an Autoregression », *Journal of the Royal Statistical Society*, série B, n° 41, pp. 190-195.

HANSEN L.P., HODRICK R.J. (1980). – « Forward Rates as Optimal Predictors of Future Spot Rates », *Journal of Political Economy*, 88, 5, pp. 829-853.

HANSEN L.P., SARGENT T.S. (1981). – « Linear Rational Expectations Models for Dynamically Interrelated Variables », in *Rational Expectations and Econometric Practice*, Lucas, R.E., et T.J. Sargent (éd.), Minneapolis, University of Minnesota Press.

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

HURN A.S., MOODY T., MUSCATELLI V.A. (1995). – « The Term Structure of Interest Rates in the London Interbank Market », *Oxford Economic Papers*, 47, 3, pp. 418-436.

JONDEAU É. (1998). – « Représentation VAR et test de la théorie des anticipations de la structure par terme », *Journal de la Société de Statistique de Paris*, 139, 1, pp. 49-71.

JONDEAU É., RICART R. (1999). – « The Expectations Theory: Tests on US, German, French, and UK Euro-Rates », *Journal of International Money and Finance*, 18, 5, pp. 725-750.

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.

KUGLER P. (1990). – « The Term Structure of Euro Interest Rates and Rational Expectations », *Journal of International Money and Finance*, 9, 2, pp. 234-244.

LÜTKEPOHL H., BURDA M.M. (1997). – « Modifed Wald Tests under Nonregular Conditions », *Journal of Econometrics*, 78, 2, pp. 315-332.

MACDONALD R., SPEIGHT A.E. (1991). – « The Term Structure of Interest Rates under Rational Expectations: Some International Evidence », *Applied Financial Economics*, 1, 4, pp. 211-221.

MANKIW N.G. (1986). – « The Term Structure of Interest Rates Revisited », *Brookings Papers on Economic Activity*, 1, pp. 61-96.

MELLANDER E., VREDIN A., WARNE A. (1992). – « Stochastic Trends and Economic Fluctuations in a Small Open Economy », *Journal of Applied Econometrics*, 7, 4, pp. 369-394.

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.

NEWEY W.K., WEST K.D. (1987). – « A Simple, Positive Definite, Heteroscedasticity and Autocorrelation Consistent Covariance Matrix », *Econometrica*, 55, 3, 703-708.

PERRON P. (1989). – « The Great Crash, the Oil Price Shock, and the Unit Root Hypothesis », *Econometrica*, 57, 6, pp. 1361-1401.

PHILLIPS P.C.B. (1987). – « Time Series Regression with a Unit Root », *Econometrica*, 55, 2, pp. 277-301.

SARGENT T.J. (1979). – « A Note on Maximum Likelihood Estimation of the Rational Expectations Model of the Term Structure », *Journal of Monetary Economics*, 5, 1, pp. 133-143.

SCHMIDT P., PHILLIPS P.C.B. (1992). – « LM Tests for a Unit Root in the Presence of Deterministic Trends », *Oxford Bulletin of Economics and Statistics*, 54, 3, pp. 257-287.

SHEA G.S. (1992). – « Benchmarking the Expectations Hypothesis of the Interest-Rate Term Structure: An Analysis of Cointegration Vectors », *Journal of Business and Economic Statistics*, 10, 3, pp. 347-366.

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.

SHILLER R.J., CAMPBELL J.Y., SCHOENHOLTZ K.L. (1983). – « Forward Rates and Future Policy: Interpreting the Term Structure of Interest Rates », *Brookings Papers on Economic Activity*, 1, pp. 173-217.

SIMS C.A., STOCK J.H., WATSON M.W. (1990). – « Inference in Linear Time Series Models with Some Unit Roots », *Econometrica*, 58, 1, pp. 113-144.

STAMBAUGH R.F. (1986). – « Bias in Regressions with Lagged Stochastic Regressors », CRSP *Working Paper* 156, University of Chicago.

TAYLOR M.P. (1992). – « Modelling the Yield Curve », *Economic Journal*, 102, 412, pp. 524-537.

TODA H.Y., YAMAMOTO T. (1995). – « Statistical Inference in Vector Autoregressions with Possibly Integrated Processes », *Journal of Econometrics*, 66, 1-2, pp. 225-250.

TZAVALIS E., WICKENS M.R. (1998). – « A Re-Examination of the Rational Expectations Hypothesis of the Term Structure: Reconciling the Evidence of Long-Run and Short-Run Tests », *International Journal of Finance and Economics*, 3, 3, pp. 229-239.

WHITE H. (1980). – « A Heteroscedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroscedasticity », *Econometrica*, 48, 4, pp. 817-838.

ZAPATA H.O., RAMBALDI A.N. (1997). – « Monte-Carlo Evidence on Cointegration and Causation », *Oxford Bulletin of Economics and Statistics*, 59, 2, pp. 285-298.

ANNEXES

1. Expression et rang de la matrice D_i

Expression de la matrice D_i

Dans le cas de m et n quelconques, la matrice D_i s'exprime sous la forme :

$$\underset{(2p,4p-2)}{D_i} = \frac{\partial r_i(b^c)}{\partial b^{c\prime}} = \frac{\partial r_i(b^c)}{\partial \mathrm{vec}(A)'} \frac{\partial \mathrm{vec}(A)}{\partial b'} \frac{\partial b}{\partial b^{c\prime}} \qquad i = 1{,}2.$$

Le premier terme de cette expression s'écrit :

$$\begin{aligned}
\frac{\partial r_1(b^c)}{\partial \mathrm{vec}(A)'} &= -\left((I-A)^{-1\prime} \otimes h'I\right) - \left((I-A)^{-1\prime} \otimes h'A(I-A)^{-1}\right) \\
&\quad + \frac{m}{n}\left(((I-A^n)(I-A^m)^{-1}(I-A)^{-1})' \otimes h'I\right) \\
&\quad - \frac{m}{n}\left(((I-A^m)^{-1}(I-A)^{-1})' \otimes h'A\right)\left(\sum_{j=0}^{n-1} A'^{n-1-j} \otimes A^j\right) \\
&\quad + \frac{m}{n}\left(((I-A^m)^{-1}(I-A)^{-1})' \otimes h'A(I-A^n)(I-A^m)^{-1}\right) \\
&\qquad \left(\sum_{j=0}^{m-1} A'^{m-1-j} \otimes A^j\right) \\
&\quad + \frac{m}{n}\left((I-A)^{-1\prime} \otimes h'A(I-A^n)(I-A^m)^{-1}(I-A)^{-1}\right)
\end{aligned}$$

et

$$\begin{aligned}
\frac{\partial r_2(b^c)}{\partial \mathrm{vec}(A)'} &= -\left(I \otimes (g+h)'I\right) + \frac{m}{n}\left(((I-A^n)(I-A^m)^{-1})' \otimes h'I\right) \\
&\quad - \frac{m}{n}\left((I-A^m)^{-1\prime} \otimes h'A\right)\left(\sum_{j=0}^{n-1} A'^{n-1-j} \otimes A^j\right) \\
&\quad + \frac{m}{n}\left((I-A^m)^{-1\prime} \otimes h'A(I-A^n)(I-A^m)^{-1}\right) \\
&\qquad \left(\sum_{j=0}^{m-1} A'^{m-1-j} \otimes A^j\right).
\end{aligned}$$

Le second terme de l'expression, $\frac{\partial \text{vec}(A)}{\partial b'}$, de dimension $\left(4p^2,4p\right)$, s'écrit :

$$\frac{\partial \text{vec}(A)}{\partial b'} = I_{2p} \otimes \begin{bmatrix} I_2 \\ 0_{(2(p-1),2)} \end{bmatrix}.$$

Enfin, le terme $\frac{\partial b}{\partial b^{c\prime}}$, de dimension $(4p,4p-2)$, s'écrit :

$$\frac{\partial b}{\partial b^{c\prime}} = \begin{bmatrix} I_{4(p-1)} & 0_{(2(p-1),2)} \\ 0_{(2,4(p-1))} & 0_{(2,2)} \\ 0_{(2,4(p-1))} & I_2 \end{bmatrix}.$$

Rang de la matrice D_i

Montrer la non-singularité de $\left(D_i' \Sigma_{b^c} D_i\right)$ n'est pas une tâche aisée, du fait de l'expression complexe de D_i. De façon à pouvoir mener certains calculs et certaines simplifications sur cette expression, on se concentre ici sur le cas $m = 1$ (cas retenu dans l'application empirique) et $p = 1$. La matrice des paramètres du RVAR s'écrit alors

$$A = \begin{pmatrix} 0 & b_{12} \\ 0 & b_{22} \end{pmatrix}.$$

Dans ce cas, θ est de dimension (2×1). Le premier élément vaut 0 par construction et le second s'écrit simplement sous la forme

$$\begin{aligned} \theta_2 &= \left(\frac{1}{n} \sum_{i=1}^{n-1} \sum_{j=1}^{i} h' A^j \right)_2 \\ &= \frac{1}{n} \left[(n-1)\, b_{12} + (n-2)\, b_{12} b_{22} + ... + b_{12} \left(b_{22}\right)^{n-1} \right]. \end{aligned}$$

On ne s'intéresse dans la suite qu'à la seconde contrainte de $r_1\left(b^c\right)$ (la première est toujours respectée), qui s'écrit sous la forme :

$$1 = \theta_2 = \frac{b_{12}}{n} \left[(n-1) + (n-2)\, b_{22} + ... + \left(b_{22}\right)^{n-1} \right].$$

Le vecteur des dérivées partielles (D) associé à cette contrainte a alors comme expression

$$D = -\frac{1}{n} \begin{pmatrix} (n-1) + (n-2)\, b_{22} + ... + \left(b_{22}\right)^{n-1} \\ b_{12} \left[(n-2) + 2\,(n-3)\, b_{22} + ... + (n-1)\, \left(b_{22}\right)^{n-2} \right] \end{pmatrix}.$$

Sous l'hypothèse nulle, ce vecteur s'écrit

$$D = -\frac{1}{n} \begin{pmatrix} n / b_{12} \\ b_{12} \left[(n-2) + 2\,(n-3)\, b_{22} + ... + (n-1)\, \left(b_{22}\right)^{n-2} \right] \end{pmatrix}.$$

Donc, le premier élément du vecteur D est non nul, ce qui assure que le rang de D est toujours égal à 1 sous l'hypothèse nulle[17].

17. On remarque en outre que le terme b_{12} ne peut pas être nul sous l'hypothèse nulle.

2. L'estimation par simulation des distributions des statistiques de test

Le principe de l'estimation par simulation des distributions est de simuler un grand nombre d'échantillons ayant les mêmes caractéristiques que les séries initiales sous l'hypothèse nulle, de procéder à l'estimation du modèle à partir des échantillons simulés, d'en déduire un ensemble de statistiques de test et finalement de vérifier si la valeur testée de la statistique de test est compatible avec la distribution simulée (issue des échantillons simulés).

Première étape : la phase d'estimation

Dans un premier temps, la dynamique des taux d'intérêt est estimée à partir des séries observées, de façon à obtenir, d'une part, une estimation des paramètres du modèle à simuler et, d'autre part, une estimation des statistiques d'intérêt x_θ (statistique de Wald, coefficient de régression, coefficient de corrélation, rapport des variances).

– Après avoir identifié le nombre de retards (p), on estime les paramètres du modèle RVAR :

$$\begin{pmatrix} \Delta r_t^{(m)} \\ S_t^{(m,n)} \end{pmatrix} = \sum_{i=1}^{p} B_i \begin{pmatrix} \Delta r_{t-i}^{(m)} \\ S_{t-i}^{(m,n)} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix} \qquad t = 1,\ldots,T \tag{22}$$

ainsi que leur loi asymptotique.

– On en déduit la représentation VAR (9)), d'où l'on tire le vecteur θ, représentatif des contraintes imposées par la théorie des anticipations, et les différentes statistiques d'intérêt.

Deuxième étape : la phase de simulation

Le principe des méthodes de simulation consiste à construire un grand nombre d'échantillons de séries ayant des caractéristiques similaires aux variables initiales de taux d'intérêt. Il suffit pour cela de simuler des séries de perturbations, notées par exemple $\varepsilon_t^s = \left(\varepsilon_{1,t}^s \;\; \varepsilon_{2,t}^s\right)'$, et d'en déduire des séries de taux d'intérêt à partir du RVAR (22) et des coefficients $\left(\hat{B}_1 \;\; \cdots \;\; \hat{B}_p\right)$ estimés lors de la première étape.

Dans le cas de simulations de Monte-Carlo, la distribution à distance finie des perturbations est supposée normale et les perturbations simulées sont donc simplement obtenues en créant une série constituée de T réalisations tirées d'une loi normale. Pour le *bootstrapping*, la distribution des perturbations est supposée bien approchée par la distribution empirique des résidus estimés. Les perturbations simulées sont donc obtenues en créant une série constituée de T réalisations tirées, de façon aléatoire, dans la série des résidus estimés.

La deuxième étape consiste donc à estimer I statistiques de test (dans l'application numérique, on retient $I = 10\,000$). Celles-ci sont obtenues en deux temps, de la façon suivante : I échantillons des résidus du RVAR sont simulés (à partir d'une loi normale, pour la méthode de Monte-Carlo, ou à partir de la distribution empirique des résidus estimés lors de la première étape, pour le *bootstrapping*). Des séries simulées $\Delta r_t^{s(m)}$ et $S_t^{s(m,n)}$ sont ainsi obtenues à partir des paramètres du RVAR estimés lors de la première étape et des résidus simulés, par la relation :

$$\begin{pmatrix} \Delta r_t^{s(m)} \\ S_t^{s(m,n)} \end{pmatrix} = \sum_{i=1}^{p} \hat{B}_i \begin{pmatrix} \Delta r_{t-i}^{s(m)} \\ S_{t-i}^{s(m,n)} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1,t}^s \\ \varepsilon_{2,t}^s \end{pmatrix} \qquad t = 1,...,T. \tag{23}$$

Les p conditions initiales $\left(\Delta r_1^{s(m)}, S_1^{s(m,n)} ..., \Delta r_p^{s(m)}, S_p^{s(m,n)}\right)$ correspondent aux p premières valeurs observées de l'échantillon. Les 50 premières observations simulées sont, en outre, supprimées de façon à atténuer l'effet des conditions initiales. La taille de l'échantillon final, T, correspond à celle des observations.

Une fois obtenues des séries temporelles simulées pour $\Delta r_1^{s(m)}$ et $S_1^{s(m,n)}$ à partir du modèle (23), la pente des taux simulée $S_t^{s(m,n)}$ est remplacée par la pente des taux théorique construite en utilisant les paramètres estimés à partir des observations, soit $\tilde{S}_t^{s(m,n)} = \theta' z_t^s$, avec

$$z_t^s = \left(\Delta r_t^{s(m)}, S_t^{s(m,n)}, \cdots, \Delta r_{t-p+1}^{s(m)}, S_{t-p+1}^{s(m,n)}\right)'.$$

Les données vérifient donc par construction la théorie des anticipations, mais les coefficients du modèle ne sont pas nécessairement compatibles avec cette théorie (Campbell et Shiller [1991]).

Troisième étape : le calcul des niveaux de significativité

Les opérations de la première étape sont alors de nouveau mises en œuvre : le nombre de retards optimal p est identifié par la procédure HQ, les paramètres du RVAR

$$\begin{pmatrix} \Delta r_t^{s(m)} \\ \tilde{S}_t^{s(m,n)} \end{pmatrix} = \sum_{i=1}^{p} B_i \begin{pmatrix} \Delta r_{t-i}^{s(m)} \\ \tilde{S}_{t-i}^{s(m,n)} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix}.$$

et leur distribution asymptotique sont estimés, le vecteur θ^s et les différentes statistiques d'intérêt ($r_i^s\ (b_c)$, $i = 1,2$, β^s, V^s et ρ^s) sont calculés.

Campbell et Shiller définissent deux niveaux de significativité associés aux tests affaiblis : S_1 est la proportion d'échantillons (parmi les I simulés) qui ont produit une statistique de test (β^s, V^s ou ρ^s) plus éloignée de 1 (dans la même direction) que la statistique de test obtenue à partir des observations. Il s'agit donc d'un test unilatéral. S_2 est la proportion d'échantillons qui ont produit une statistique de test dont la statistique de Student (évaluée à partir de l'écart-type asymptotique) rejette l'hypothèse nulle plus fortement (quelle que soit la direction) que la statistique de Student obtenue à partir des observations. Il s'agit donc d'un test bilatéral.