

Modélisation bayésienne non linéaire du taux d'intérêt de court terme américain : l'aide des outils non paramétriques

Michel Lubrano

Volume 80, Number 2-3, juin–septembre 2004

Hommage à Marcel Dagenais

URI: <https://id.erudit.org/iderudit/011396ar>

DOI: <https://doi.org/10.7202/011396ar>

[See table of contents](#)

Publisher(s)

HEC Montréal

ISSN

0001-771X (print)

1710-3991 (digital)

[Explore this journal](#)

Cite this article

Lubrano, M. (2004). Modélisation bayésienne non linéaire du taux d'intérêt de court terme américain : l'aide des outils non paramétriques. *L'Actualité économique*, 80(2-3), 465–499. <https://doi.org/10.7202/011396ar>

Article abstract

This paper investigates empirical models of the US short term interest rate. It make use of a combination of classical non-parametric methods and of parametric bayesian methods. In a first step, it investigates the shape of drift and volatility functions using non parametric tools. The paper then develops a bayesian approach to model selection based on the minimisation of the Hellinger distance between the posterior predictive density of a discretised model and a non-parametric estimation of the data density. A discretisation of various parametric formulations are then estimated, ranging between constant elasticity of variance to switching regimes.

MODÉLISATION BAYÉSIENNE NON LINÉAIRE DU TAUX D'INTÉRÊT DE COURT TERME AMÉRICAIN : L' AIDE DES OUTILS NON PARAMÉTRIQUES*

Michel LUBRANO
GREQAM-CNRS

RÉSUMÉ – Cet article a pour objet l'investigation des modèles empiriques de taux d'intérêt de court terme sur données américaines. Il utilise une combinaison de méthodes classiques non paramétriques et de méthodes bayésiennes paramétriques. La forme des fonctions de dérive et de volatilité du modèle discrétisé est tout d'abord examinée au moyen d'une analyse non paramétrique préliminaire. Le texte développe ensuite une méthode bayésienne de choix de modèles qui est basée sur la capacité d'un modèle à minimiser la distance de Hellinger entre la densité prédictive *a posteriori* du modèle discrétisé et la densité de l'échantillon observé. Une discrétisation du modèle en temps continu est estimée en utilisant différentes variantes paramétriques allant du modèle à variance constante jusqu'à différents types de modèles de *switching* suggérés par l'analyse non paramétrique préliminaire.

ABSTRACT – This paper investigates empirical models of the US short term interest rate. It make use of a combination of classical non-parametric methods and of parametric bayesian methods. In a first step, it investigates the shape of drift and volatility functions using non parametric tools. The paper then develops a bayesian approach to model selection based on the minimisation of the Hellinger distance between the posterior predictive density of a discretised model and a non-parametric estimation of the data density. A discretisation of various parametric formulations are then estimated, ranging between constant elasticity of variance to switching regimes.

INTRODUCTION

Dans un modèle d'évaluation d'obligations, les mouvements du taux d'intérêt de court terme r sont en général décrits au moyen d'une équation de diffusion du type

* De nombreuses discussions menées avec Luc Bauwens, Laurence Broze, Christian Hafner, Frédéric Jouneau, Jean-François Richard et Peter Schotman ont beaucoup aidé à l'écriture de ce papier. Sa révision a grandement bénéficié des remarques d'un arbitre anonyme et de la lecture attentive de Ludovic Giet. Mais l'auteur reste seul responsable des erreurs restantes. La présente version de ce papier est dédiée à la mémoire de Marcel Dagenais. Marcel fut un visiteur régulier de l'Université d'Aix-Marseille. La première fois que je l'y avais rencontré, je n'étais qu'un étudiant de Ph.D. Il m'avait alors fortement motivé à poursuivre mes recherches dans le domaine de l'économétrie bayésienne.

$$dr = \mu(r) dt + \sigma(r) dW \quad (1)$$

où W est un mouvement brownien standard, $\mu(r)$ et $\sigma(r)$ désignent respectivement les fonctions de dérive et de volatilité. Toutes deux peuvent dépendre de r , mais sont supposées indépendantes de t . L'estimation économétrique de ce type d'équation de diffusion est rendue difficile par le fait que premièrement l'on ne dispose que d'observations discrètes et deuxièmement qu'il n'existe de discrétisations exactes que pour des spécifications très restrictives de $\mu(r)$ et $\sigma(r)$. Par exemple, Vasicek (1977) utilise un processus d'Ornstein-Uhlenbeck stationnaire, Dothan (1978) un brownien géométrique qui permet une volatilité non constante mais n'est pas stationnaire et Cox, Ingersoll et Ross (1985) un modèle stationnaire avec une volatilité qui est fonction de \sqrt{r} . Ces trois modèles possèdent une discrétisation exacte qui permet d'écrire les densités marginales et de transition des observations discrètes. Mais toute généralisation nous fait sortir de ce cadre analytique et impose de passer par une approximation pour conduire l'inférence. L'approximation la plus simple consiste à linéariser l'équation de diffusion (1) au moyen du schéma d'Euler. On peut se contenter d'estimer le modèle discrétisé, mais alors les estimations seront biaisés. La discrétisation d'Euler permet également de simuler la trajectoire du modèle pour un pas plus faible que celui des observations et ouvre alors la voie à toute une classe de méthodes qui corrigent le biais de discrétisation : l'inférence indirecte exploitée dans Broze, Scaillet et Zakoian (1995), le maximum de vraisemblance simulé développé entre autres par Pedersen (1995) ainsi que l'approche bayésienne par augmentation des données comme dans Elerian, Chib et Shepard (2001) ou Eraker (2001).

Les modèles empiriques que comparent Chan, Karolyi, Longstaff et Sanders (1992) (CKLS) sont des discrétisations d'Euler de l'équation de diffusion générale (1) pour divers choix paramétriques de $\mu(r)$ et $\sigma(r)$. Ces modèles partagent tous la caractéristique commune de présenter une fonction de dérive $\mu(r) = (\alpha + \beta r_t)$ qui est linéaire en r . Ils se différencient au niveau de l'expression de leur fonction de volatilité $\sigma(r)$ qui peut être constante (Vasicek, 1977), peut dépendre de la racine carrée de r (Cox, Ingersoll et Ross, 1985) ou finalement peut être une fonction de r à élasticité constante avec $\sigma(r) = \sigma r^\gamma$. CKLS (1992) montrent que les modèles autorisant $\gamma > 1$ capturent mieux la dynamique de court terme du taux d'intérêt que ceux qui imposent $\gamma < 1$. Cependant, Broze, Scaillet et Zakoian (1995) ont montré que la discrétisation du modèle favorisé par CKLS (1992) n'est ergodique et stationnaire au second ordre que tant que $\gamma \leq 1$. En conséquence, le fait que γ soit plus grand que un doit être pris comme une indication de mauvaise spécification. Voir à ce propos la discussion dans Gouriéroux et Monfort (1996, chap. 6 : 127-129)¹. En effet, le modèle doit avoir une propriété de retour à la moyenne et ainsi doit impliquer des taux d'intérêt positifs et finis qui retournent à leur valeur d'équilibre à la suite d'un choc positif ou négatif. Si les taux d'intérêt sont modélisés par exemple au moyen d'une marche aléatoire, comme les tests statistiques

1. Remarquons cependant que Conley, Hansen, Luttmer et Scheinkman (1997) donnent des conditions de stationnarité différentes pour le modèle CEV en temps continu.

de racine unité auraient tendance à l'indiquer pour la plupart des séries empiriques, on arrive à des incohérences pour la structure à terme. Voir par exemple Ingersoll (1987, chap. 18 : p. 396).

Ait-Sahalia (1996b) argumente le fait que la linéarité de la fonction de dérive est sans doute la source principale de mauvaise spécification pour les modèles existants. Une estimation non paramétrique de la dérive montre que celle-ci est pratiquement constante et nulle pour des r petits et qu'elle devient fortement négative pour des valeurs plus importantes de r exerçant alors une forte force de rappel. Ait-Sahalia (1996b) suggère donc de remplacer la dérive linéaire par une fonction non linéaire de la forme $\mu(r) = \alpha_0 + \alpha_1 r + \alpha_2 r^2 + \alpha_3 / r$. Un α_2 négatif force les taux élevés à retourner vers la région centrale tandis qu'un α_3 positif fait de même pour des taux trop proches de zéro. Cette dérive non linéaire avec des coefficients correctement signés garantit l'ergodicité et la stationnarité au second degré du modèle paramétrique en temps discret même lorsque $\gamma > 1^2$. La recherche d'une formulation paramétrique correcte pour les fonctions de dérive et de volatilité dans une équation de diffusion peut être sans fin en l'absence d'outils appropriés. Dans cet article, nous avons décidé de guider notre choix de paramétrisation au moyen d'une analyse préliminaire non paramétrique classique basée sur une discrétisation d'Euler de l'équation de diffusion qui permettra de donner une idée sur les formes à atteindre. Une fois que l'on aura choisi une formulation paramétrique particulière, on pourra alors vérifier comment le modèle paramétrique en temps discret se comporte, sa capacité à reproduire les caractéristiques des données. Une approche particulière est développée dans cet article pour comparer les modèles. Notre procédure pour le choix de modèle est basée sur la comparaison entre la densité prédictive *a posteriori* du modèle en temps discret à évaluer et un estimateur non paramétrique de la densité des observations qui lui ne repose pas sur une hypothèse de discrétisation.

Indépendamment de l'analyse non paramétrique préliminaire, l'approche statistique suivie dans cet article est bayésienne. Ce cadre a été appliqué avec succès dans Lubrano (2001) pour les indices boursiers, par Pfann, Schotman et Tschernig (1996) et Pai et Pedersen (1998) pour les taux d'intérêt par exemple. Elle a connu des développements récents avec Elerian, Chib et Shepard (2001) et Eraker (2001). L'approche bayésienne est plus souple et plus robuste que l'approche classique toutes les fois qu'il s'agit de traiter un modèle non linéaire et particulièrement un modèle à seuil parce qu'il est plus facile de calculer une moyenne que d'essayer de maximiser une fonction de vraisemblance possédant probablement plusieurs maxima. (Voir Lubrano, 2000 pour une revue.) Une meilleure compréhension de la fonction de vraisemblance est également fournie par l'approche bayésienne.

L'article est organisé comme suit. Dans la première section, nous montrons comment une discrétisation d'Euler du modèle de diffusion peut être estimée de façon non paramétrique ainsi que la façon dont cette méthode est reliée aux résul-

2. Voir l'hypothèse A2' dans Ait-Sahalia (1996b).

tats récents de la littérature. Dans la deuxième section, nous expliquons comment l'inférence bayésienne peut être conduite sur ce même modèle discrétisé. Nous montrons également comment le biais de discrétisation inhérent à la méthode pourrait être réduit et à quel prix. Dans la troisième section, nous développons une stratégie bayésienne qui permette d'évaluer les modèles estimés sur la base d'une discrétisation selon un critère qui soit indépendant du schéma de discrétisation. La quatrième section présente les données et donne une évaluation non paramétrique des fonctions de dérive et de volatilité. La cinquième section est consacrée à l'analyse bayésienne de divers modèles non linéaires de taux d'intérêt tandis que la sixième section analyse l'impact d'une modélisation à deux régimes. La dernière section conclut.

1. INFÉRENCE NON PARAMÉTRIQUE DANS LES ÉQUATIONS DE DIFFUSION

L'inférence non paramétrique dans les équations de diffusion peut être abordée de plusieurs façons selon le degré de complexité que l'on accepte. Si l'on recherche un outil de recherche de spécification, comme c'est le cas dans cet article, il suffit de discrétiser l'équation de diffusion au moyen du schéma d'Euler :

$$r_{n\Delta} - r_{(n-1)\Delta} = \Delta \mu(r_{(n-1)\Delta}) + \sqrt{\Delta} \sigma(r_{(n-1)\Delta}) \varepsilon_n \quad (2)$$

où ε_n est un bruit blanc gaussien de variance unitaire, Δ l'unité de temps entre deux observations (ou pas de discrétisation avec $1/\Delta$ entier) et n l'index des observations $n = 1, \dots, N$. Si on observe le processus dans l'intervalle de temps $[0, T]$, alors $\Delta = T/N$. Quand $\Delta \rightarrow 0$, c'est-à-dire quand $N \rightarrow \infty$, le processus tend vers le processus en temps continu. Quand Δ est strictement positif, il y a en général un biais de discrétisation parce que la dynamique du processus peut être beaucoup plus rapide que la fréquence d'observation. Ce biais est faible tant que la période d'observation est élevée (données journalières). Mais il augmente quand cette fréquence baisse. L'idée de Stanton (1997), pour corriger ce biais, consiste à utiliser un développement de Taylor d'ordre supérieur sur le générateur infinitésimal. Enfin, Ait-Sahalia (1996a) (voir aussi Jiang et Knight, 1997) utilise les équations de Kolmogorov qui donnent une relation entre la densité marginale du processus, la dérive et la volatilité pour trouver une estimation non paramétrique de la volatilité.

1.1 Régression non paramétrique

La discrétisation de (2) nous ramène à un modèle de série temporelle dont le traitement est rappelé dans la recension de Härdle, Lütkepohl et Chen (1997). Considérons $\Delta = 1$ et notons le modèle comme

$$\Delta r_t = m(r_{t-1}) + s(r_{t-1}) \varepsilon_t \quad (3)$$

Nous avons donc un modèle de régression hétéroscédastique où $\Delta r_t = r_t - r_{t-1}$ est la variable endogène et r_{t-1} la variable prédéterminée. L'espérance conditionnelle

est $m(r_{t-1})$ et la fonction scédastique est $s(r_{t-1})$. Comme l'ont montré Auestadt et Tjøstheim (1990), Tjøstheim et Auestadt (1994a, b), la fonction $m(r_{t-1})$ peut être estimée en utilisant l'estimateur à noyau de Nadaraya-Watson

$$\hat{m}_h(r) = \frac{\sum_t K[(r_{t-1} - r) / h] \Delta r_t}{\sum_t K[(r_{t-1} - r) / h]} \tag{4}$$

où K est un noyau de lissage (le noyau normal sera choisi ici) et h le paramètre de lissage ou la taille de la fenêtre. La fonction scédastique est estimée pareillement au moyen de

$$\hat{s}_h(r)^2 = \frac{\sum K[(r_{t-1} - r) / h] \Delta r_t^2}{\sum K[(r_{t-1} - r) / h]} - \hat{m}_h^2(r). \tag{5}$$

La fonction $\hat{s}_h(r)^2$ peut être également considérée comme la variance de l'estimateur non paramétrique de l'espérance conditionnelle $\hat{m}_h(r)$. Il est alors possible de l'employer pour construire une bande uniforme de confiance pour la fonction de régression $\hat{m}_h(r)$. Dans le cas IID, la quantité $\sqrt{Th} [\hat{m}_h(r) - m(r)]$ peut être approchée par un processus gaussien $G(r)$. La distribution asymptotique du supremum de $|g(r)|$ a été dérivée par Bickel et Rosenblatt (1973). Härdle (1990, chap. 4) montre comment ce résultat peut être employé pour approximer une bande de confiance uniforme pour une valeur donnée de r . Pour le noyau normal, nous pouvons définir d_T comme

$$d_T = (2\delta \log T)^{1/2} + (2\delta \log T)^{-1/2} \log \left(\frac{0,25}{2\pi^2} \right)^{1/2} \tag{6}$$

où $1/5 < \delta < 1/3$. Une bande de confiance de niveau $(1 - \alpha)$ est définie par

$$\hat{m}_h(r) \pm \left(\frac{c_\alpha}{(2\delta \log T)^{1/2}} + d_T \right) \left(\frac{1}{2Th\sqrt{\pi}} \frac{\hat{s}_h^2}{\hat{f}_h(r)} \right)^{1/2} \tag{7}$$

où $\hat{f}_h(r)$ est un estimateur non paramétrique de la densité marginale de r ; c_α est défini de telle sorte que $\exp(-2 \exp(-c_\alpha)) = 1 - \alpha$. Pour un intervalle de confiance de 95 %, $c_{0,05} = 3,66$.

La valeur de la fenêtre optimale peut être approchée en minimisant l'erreur estimée de prévision

$$\frac{1}{T} \sum_t [\Delta r_t - \hat{m}_h(r_t)]^2 \tag{8}$$

où $\hat{m}_h(r_t)$ est calculé comme l'estimateur « *leave-one-out* » obtenu en supprimant la $t^{\text{ième}}$ observation dans les sommes. Cette méthode automatique est asymptotiquement optimale, mais demande des calculs intensifs et elle ne converge que très lentement. La convergence peut être améliorée en multipliant (8) par une fonction

pour pénaliser les valeurs trop petites de h . La méthode est décrite dans Rice (1984) qui propose par exemple $(1 - 2T^{-1} h^{-1} K(0))^{-1}$ comme fonction de pénalité. Une méthode plus rapide consiste à choisir graphiquement le h de manière à obtenir une courbe de régression visuellement satisfaisante.

1.2 Comparaison avec les autres approches

Une régression non paramétrique conduite sur la discrétisation d'Euler du processus conduit à des résultats qui semblent relativement proches de ceux proposés dans la littérature non paramétrique. L'estimateur non paramétrique usuel de la volatilité dans une équation de diffusion est celui de Florens-Zmirou (1993). Il se note

$$\hat{\sigma}^2(r) = \frac{\sum_{i=1}^n n K[(r_i - r) / h] (r_{i+1} - r_i)^2}{\sum_{i=1}^n T K[(r_i - r) / h]} \tag{9}$$

où n est le nombre d'observations équidistantes et T la période de temps d'observation. Quand $n = T$, cet estimateur est équivalent à (5), excepté le terme de correction pour la dérive. On notera toutefois que Florens-Zmirou (1993) renonce à estimer la dérive, car celle-ci est en général mal identifiée.

Considérons maintenant une fonction f de r où r suit la diffusion (1). On définit alors le générateur infinitésimal \mathcal{L} de r par

$$\begin{aligned} \mathcal{L} f(r, t) &= \lim_{\Delta \rightarrow 0} \frac{E[f(r_{t+\Delta} | r_t = r) - f(r, t)]}{\Delta} \\ &= \frac{\partial f(r, t)}{\partial t} + \frac{\partial f(r, t)}{\partial r} \mu(r) + \frac{1}{2} \frac{\partial^2 f(r, t)}{\partial r^2} \sigma^2(r). \end{aligned} \tag{10}$$

Pour des choix adéquats de la fonction f , $\mathcal{L} f(r, t)$ permettra de générer les différents moments de r . En particulier, si l'on pose $f(r, t) = r$, on aura $\mathcal{L} f(r, t) = \mu(r)$; et pour $f(r, t) = (r - r_t)^2$, l'on obtiendra $\mathcal{L} f(r, t) = 2(r - r_t) \mu(r) + \sigma^2(r)$ et finalement $\mathcal{L} f(r_t, t) = \sigma^2(r_t)$. Stanton (1997) utilise ce générateur infinitésimal pour calculer un développement de Taylor de l'espérance conditionnelle $E_t(r_{t+\Delta})$. En réordonnant les termes de ce développement, on arrive pour l'approximation au premier ordre à l'expression suivante :

$$\mathcal{L} f(r_t, t) = \frac{1}{\Delta} E_t[f(r_{t+\Delta}, t + \Delta) - f(r_t, t)] + O(\Delta) \tag{11}$$

et au second ordre

$$\begin{aligned} \mathcal{L} f(r_t, t) &= \frac{1}{2\Delta} \{4E_t[f(r_{t+\Delta}, t + \Delta) - f(r_t, t)] - E_t[f(r_{t+2\Delta}, t + 2\Delta) - f(r_t, t)]\} \\ &\quad + O(\Delta^2). \end{aligned} \tag{12}$$

On peut alors trouver une approximation à l'ordre un, deux ou plus pour la dérive $\mu(r)$ et pour la volatilité $\sigma^2(r)$ en choisissant de manière adéquate la fonction f . L'approximation au premier ordre de la dérive est

$$\mu_1(r_t) = \frac{1}{\Delta} E_t[r_{t+\Delta} - r_t] + O(\Delta) \tag{13}$$

et l'approximation correspondante de la volatilité

$$\sigma_1^2(r_t) = \frac{1}{\Delta} E_t[(r_{t+\Delta} - r_t)^2] + O(\Delta) . \tag{14}$$

Pour implémenter ces deux approximations, il suffit alors de trouver un estimateur non paramétrique de l'espérance conditionnelle $E_t(\cdot)$ qui n'est rien d'autre qu'une fonction de régression. On peut donc utiliser l'estimateur à noyau de Nadaraya-Watson que l'on a détaillé plus haut. En suivant Stanton (1997), on arrive à la formule suivante pour l'approximation au premier ordre de la dérive

$$\widehat{\mu_1}(r) = \frac{1}{\Delta} \frac{\sum (r_{t+\Delta} - r_t) K([(r - r_t) / h]}{\sum K([(r - r_t) / h]} \tag{15}$$

et pour l'approximation au premier ordre de la volatilité

$$\widehat{\sigma_1^2}(r) = \frac{1}{\Delta} \frac{\sum (r_{t+\Delta} - r_t)^2 K([(r - r_t) / h]}{\sum K([(r - r_t) / h]} . \tag{16}$$

On constate premièrement que l'estimateur (15) est identique à l'estimateur non paramétrique que nous avons proposé en (4) pour un pas de discrétisation de un et que l'estimateur (16) est identique à l'estimateur de Florens-Zmirou (1993).

Stanton (1997) étudie, pour le modèle CIR, la qualité de ces approximations. Il trouve qu'elle est liée à la fréquence des observations. Pour des données journalières, les approximations à l'ordre 1, 2 ou 3 ne peuvent pas être distinguées de la vraie courbe. Pour des données hebdomadaires ou même mensuelles, les approximations au deuxième ou troisième ordre ne sont pas distinguables de l'estimateur exact pour des valeurs pas trop fortes de r .

1.3 L'utilisation des équations de Kolmogorov

Si nous supposons que le processus de diffusion est strictement stationnaire, nous pouvons utiliser une propriété des équations différentielles stochastiques pour identifier les fonctions de dérive et de volatilité, propriété qui fait intervenir les probabilités marginales et de transition de r . Cette approche a été développée par Ait-Sahalia (1996a, b) et par Jiang et Knight (1997). Pour donner une certaine intuition sur la façon dont ceci fonctionne, notons qu'un processus gaussien se caractérise entièrement par sa moyenne et sa variance. Puisque ses incréments sont gaussiens, les distributions (marginales et les probabilités de transition) d'un processus de diffusion d'Ito se caractérisent entièrement par leurs fonctions de

dérive et de volatilité sous certaines conditions de régularité. Ceci peut être montré (voir par exemple Jiang et Knight, 1997) en résolvant les équations de Kolmogorov *backward* et *forward*. On obtient alors une relation exacte entre l'expression de la densité marginale des observations $f(r_t)$ et les fonctions de dérive et de volatilité :

$$f(r_t) = \frac{A}{\sigma^2(r)} \exp \left\{ 2 \int_{r_0}^{r_t} \frac{\mu(u)}{\sigma^2(u)} du \right\} \quad (17)$$

où A est une constante de normalisation et r_0 une borne inférieure arbitraire.

L'expression (17) permet une gamme de variantes qui vont toutes impliquer que l'on estime de manière non paramétrique la densité marginale des observations $f(r_t)$. On peut souligner au passage que la procédure d'estimation non paramétrique de la densité marginale de r est totalement indépendante du fait que l'on soit en temps discret ou en temps continu.

On peut tout d'abord spécifier une forme paramétrique pour la dérive et la volatilité indexée par le paramètre θ . On va ensuite chercher à minimiser en θ une certaine distance entre l'estimateur non paramétrique $\hat{f}(r_t)$ et sa contrepartie paramétrique évaluée avec (17). Ait-Sahalia (1996b) utilise cette approche pour construire un test de mauvaise spécification. Il teste ainsi les différents modèles usuels du taux d'intérêt de court terme et les rejette tous à l'exception de son propre modèle.

Dans un autre papier, Ait-Sahalia (1996a) impose seulement une forme paramétrique linéaire pour la dérive en choisissant $\mu(r, \theta)$. Après quelques manipulations, il obtient l'estimateur non paramétrique suivant pour $\sigma^2(r)$:

$$\hat{\sigma}^2(r) = \frac{2}{\hat{f}(r)} \int_0^r \mu(u, \hat{\theta}) \hat{f}(u) du \quad (18)$$

qui nécessite une estimation séparée du paramètre de la dérive θ . Si la fonction de dérive est paramétrisée afin d'obtenir un retour à la moyenne, par exemple $\mu(r, \theta) = \beta(\alpha - r)$ dans le cas linéaire, les propriétés de la distribution transitoire impliquent $E(r_t | r_{t-1}) = \alpha + e^{-\beta}(r_{t-1} - \alpha)$. Alors $\hat{\theta}$ peut s'obtenir par régression simple³. Ce cadre implique que le pas d'échantillonnage Δ soit égal à 1 et que toute nouvelle information soit obtenue par l'intermédiaire d'observations futures avec un pas fixe d'échantillonnage.

Ait-Sahalia (1996a) se concentre sur une formulation paramétrique de la dérive parce qu'il veut évaluer des actifs dérivés sur les taux d'intérêt. Il doit représenter le prix du risque au moyen d'un paramètre et pour cela il a besoin d'une fonction paramétrique pour la dérive. Jiang et Knight (1997) ont des centres d'intérêt différents. Ils critiquent les résultats d'estimation d'Ait-Sahalia (1996a)

3. Une fonction de dérive non linéaire telle que $\mu(r, \theta) = \alpha_0 + \alpha_1 r + \alpha_2 r^2 + \alpha^3 / r$ est aussi possible. Elle implique simplement un recours aux moindres carrés non linéaires.

pour leur sensibilité à la formulation paramétrique choisie pour la fonction de dérive. Ils préfèrent employer une évaluation non paramétrique de $\sigma(r)$ inspirée de Florens-Zmirou (1993) et donnée en (9). Un nouvel estimateur pour la fonction de dérive peut alors être construit en résolvant (17) en $\mu(r)$

$$\mu(r) = \frac{1}{2} \left[\frac{d\sigma^2(r)}{dr} + \sigma^2(r) \frac{f'(r)}{f(r)} \right]. \tag{19}$$

Nous pouvons maintenant remplacer dans (19) $\sigma^2(r)$ par son estimateur $\hat{\sigma}^2(r)$. Celui-ci est différentiable pourvu que le noyau utilisé le soit. On approximera ensuite $f'(r) / f(r)$ par l'estimateur non paramétrique

$$\frac{\hat{f}'(r)}{\hat{f}(r)} = \frac{\sum \frac{1}{h} K' \left(\frac{r_n - r}{h} \right)}{\sum K \left(\frac{r_n - r}{h} \right)}. \tag{20}$$

Jiang et Knight (1997) montrent que $\hat{\mu}(r)$ et $\hat{\sigma}^2(r)$ sont consistants quand T est fixé et que le pas d'échantillonnage Δ tend vers zéro.

2. INFÉRENCE BAYÉSIENNE DANS LES ÉQUATIONS DE DIFFUSION

Dans cette section, nous allons proposer une méthode d'inférence bayésienne relativement simple qui est basée sur une discrétisation au premier ordre de l'équation de diffusion. Elle correspond en fait à conduire l'inférence sur le même type de modèle de régression que celui examiné dans la section 4.6 de Ait-Sahalia (1996b). Nous suggérerons ensuite comment cette approche pourrait être améliorée et enfin quel est l'état récent de la littérature en ce domaine.

2.1 Inférence bayésienne pour le modèle discrétisé

Une analyse non paramétrique est utile pour suggérer des paramétrisations particulières pour $m(r)$ et $s(r)$. Dans le cas général, nous proposerons une fonction paramétrique pour la dérive $m(r_{t-1}, \alpha)$ qui sera linéaire en α (mais non linéaire en r) tandis que pour la fonction scédastique $s(r_{t-1}, \gamma)$ nous considérerons des fonctions fortement non linéaires en γ . En conséquence, le modèle envisagé est un modèle de régression dynamique avec des erreurs hétéroscedastiques. Conditionnellement aux paramètres non linéaires γ de la fonction scédastique, ce modèle est un modèle linéaire. L'inférence bayésienne reste donc simple pourvu que nous restions dans un cadre naturel conjugué pour les paramètres de régression α et pour σ^2 qui représente la partie constante de la variance du terme d'erreurs. La taille de l'intégration numérique correspondra à la taille du paramètre γ . Pour clarifier les notations, appelons y_t la variable endogène Δr_t et x_t l'ensemble de variables prédéterminées et définissons les transformations suivantes

$$\begin{aligned} y_t(\gamma) &= y_t / s_t(r_{t-1}, \gamma), \\ x_t(\gamma) &= x_t / s_t(r_{t-1}, \gamma) \end{aligned} \quad (21)$$

de telle sorte que notre modèle dynamique de régression linéaire avec erreurs hétéroscédastiques puisse être écrit sous la forme

$$y_t(\gamma) = x_t'(\gamma) \alpha + v_t \quad (22)$$

où v_t est $N(0, \sigma^2)$. La fonction de vraisemblance de ce modèle est

$$L(\alpha, \sigma^2, \gamma; y) \propto \sigma^{-T} [\prod s_t(\gamma)]^{-1/2} \exp - \frac{1}{2\sigma^2} \sum [y_t(\gamma) - x_t'(\gamma) \alpha]^2. \quad (23)$$

Nous considérerons un *a priori* uniforme sur α et $\log(\sigma^2)$. L'*a priori* sur γ est laissée non spécifiée pour le moment et est notée $\varphi(\gamma)$. La densité *a posteriori* de ce modèle peut être analysée en utilisant un échantillonnage partiel de Gibbs. Plus précisément, le paramètre σ^2 peut d'abord être intégré analytiquement. En second lieu, la densité *a posteriori* conditionnelle de α étant donné γ est Student⁴ avec :

$$\varphi(\alpha \mid \gamma, y) \propto f_t(\alpha \mid \alpha_*(\gamma), M_*^{-1}(\gamma), s_*(\gamma), T) \quad (24)$$

où

$$\begin{cases} M_*(\gamma) = \sum_t x_t(\gamma) x_t'(\gamma) \\ \alpha_*(\gamma) = M_*^{-1}(\gamma) [\sum_t x_t'(\gamma) y_t(\gamma)] \\ s_*(\gamma) = \sum_t y_t'(\gamma) y_t(\gamma) - \alpha_*(\gamma) M_*(\gamma) \alpha_*(\gamma). \end{cases} \quad (25)$$

Enfin, la densité *a posteriori* marginale de γ a la forme suivante :

$$\varphi(\gamma \mid y) \propto [\prod s_t(\gamma)]^{-1} [s_*(\gamma)]^{-T/2} |M_*(\gamma)|^{-1/2} \varphi(\gamma). \quad (26)$$

L'algorithme Griddy Gibbs⁵ de Bauwens et Lubrano (1998) est employé pour obtenir des tirages de γ . Conditionnellement à chaque tirage de γ , un tirage de α est obtenu au moyen de (24). Une technique de réduction de variance pour obtenir les moments marginaux de α et de σ^2 consiste à faire la moyenne des moments conditionnels analytiques selon les principes expliqués en Bauwens, Lubrano et Richard (1999, chap. 3).

4. Bauwens, Lubrano et Richard (1999, annexe A) pour les notations de la Student.

5. Une évaluation directe de la densité *a posteriori* de γ est possible avec une règle de Simpson quand la dimension de γ n'excède pas 2. Mais cette approche n'est pas commode pour la comparaison de modèles.

2.2 La discrétisation de Milstein

De même que Stanton (1997) examine l'utilisation de développements de Taylor d'ordre supérieur pour améliorer la qualité de son approximation, de même il est possible de remplacer le schéma de discrétisation d'Euler par celui de Milstein (1978) pour pouvoir écrire la fonction de vraisemblance. L'approximation de Milstein peut en fait se déduire de celle d'Euler en développant au premier ordre son reste. Il en résulte que la discrétisation d'Euler est un développement d'Ito-Taylor au premier ordre et celle de Milstein un développement d'Ito-Taylor au deuxième ordre qui prend la forme

$$r_{n\Delta} - r_{(n-1)\Delta} = \Delta \mu(r_{(n-1)\Delta}, \theta) + \sqrt{\Delta} \sigma(r_{(n-1)\Delta}) \varepsilon_n + \frac{\Delta}{2} \sigma(r_{(n-1)\Delta}) \sigma'(r_{(n-1)\Delta}) (\varepsilon_n^2 - 1).$$

Elerian (1998) a montré que la densité transitoire associée à ce modèle est une χ^2 non centrée à un degré de liberté, ce qui rend relativement complexe l'évaluation de la fonction de vraisemblance. On peut alors se contenter d'une pseudo-vraisemblance et travailler sur une normale de moyenne $\Delta \mu(r_{n-1}, \theta)$ identique à celle fournie par le schéma d'Euler et de variance $\Delta \sigma^2(r_{n-1}) + 0,5 \Delta^2 [\sigma(r_{n-1}) \sigma'(r_{n-1})]^2$. Il est alors facile de conduire une inférence bayésienne. L'*a posteriori* ne change pas de forme par rapport à (24-26). Il suffit de changer la définition (21) en conséquence. On remarquera que Kessler (1997) suggère aussi d'approximer la fonction de transition par une densité gaussienne, mais avec une moyenne et une variance obtenues avec le générateur infinitésimal (10), ce qui devrait donner de meilleurs résultats.

2.3 Approximations non gaussiennes de la vraisemblance

Les approximations précédentes reposent sur l'hypothèse gaussienne, alors que comme le montre la discrétisation de Miltstein, on peut être loin de la normalité dès que la volatilité n'est plus constante. Une idée défendue dans la littérature consiste à transformer l'équation de diffusion par un changement de variable de telle sorte que la diffusion transformée soit de volatilité constante. Posons $y = G(r)$ avec $G'(r) = 1 / \sigma(r)$. Alors par application du lemme d'Ito, la diffusion (1) devient

$$dy = \left[\frac{\mu [G^{-1}(y)]}{\sigma [G^{-1}(y)]} - \frac{1}{2} \sigma' [G^{-1}(y)] \right] dt + dW.$$

Durham et Gallant (2002) soulignent par exemple que l'approximation de Kessler (1997) a une performance bien supérieure sur le modèle transformé que sur le modèle initial. Ait-Sahalia (2002) utilise cette transformation pour se rapprocher d'un processus gaussien et pouvoir ensuite utiliser l'expansion d'Hermite de la densité de transition autour de la densité normale. Appelons $\tilde{p}_y(y | y_0, \theta)$ la densité de transition ainsi approchée. Elle se décompose en

$$\tilde{p}_y(y | y_0, \theta) = \Delta^{-1/2} \phi\left(\frac{y - y_0}{\Delta^{-1/2}}\right) \exp\left(\int_{y_0}^y \mu_y(w, \theta) dw\right) \sum_{k=0}^K c_k(y | y_0, \theta) \frac{\Delta^k}{k!}.$$

La densité de transition approchée $\tilde{p}_y(y | y_0, \theta)$ se compose d'un premier terme qui est un noyau gaussien $\phi(\cdot)$ dont la variance est le pas d'observation Δ , suivi par un terme de correction venant de la dérive et des K termes de l'expansion. Pour $k = 0$, on aura $c_0(y | y_0, \theta) = 1$ et pour les autres valeurs de k on a

$$c_k(y | y_0, \theta) = k(y - y_0)^{-k} \int_{y_0}^y (w - y_0)^{k-1} \times \left(\lambda_y(w, \theta) c_{k-1}(w | y_0, \theta) + 0,5 \frac{\partial^2 c_{k-1}(w | y_0, \theta)}{\partial w^2} \right) dw$$

avec $\lambda_y(w, \theta) = -(\mu_y^2(w, \theta) + \partial \mu_y(w, \theta) / \partial y) / 2$. Ait-Sahalia (2002) montre qu'il suffit en général de peu de termes ($K = 2, 3$) pour obtenir une excellente approximation dans les modèles usuels. Nous disposons donc ainsi d'une approximation analytique de la fonction de vraisemblance du modèle de diffusion en y . Mais comme l'objet d'intérêt est la densité de transition du processus initial en r , il nous faut écrire la relation entre ces deux densités :

$$\tilde{p}_r(r | r_0, \theta) = \frac{1}{\sigma(r, \theta)} \tilde{p}_y(G(r, \theta) | G(r_0, \theta), \theta).$$

Il est alors possible de conduire l'inférence bayésienne sur les paramètres des fonctions de dérive $\mu(r, \theta)$ et de volatilité $\sigma(r, \theta)$. On se trouvera dans un modèle totalement non linéaire et l'*a posteriori* devra être intégrée directement au moyen par exemple d'un algorithme de Metropolis.

Cette approche est très séduisante, mais elle implique de se limiter à des modèles relativement simples pour lesquels la transformation G est inversible. L'analyse empirique qui va suivre montre que la volatilité peut parfois présenter des formes non triviales. Ensuite, même pour des modèles relativement simples comme celui de Chan, Karolyi, Longstaff et Sanders (1992) à élasticité constante, les formules deviennent rapidement d'une longueur trop importante pour se prêter à une procédure de recherche de spécification.

2.4 La complétion des données

L'inférence dans les équations de diffusion serait beaucoup plus simple si le processus était observé de manière continue. Le processus n'est malheureusement observé qu'à des dates fixes que l'on désigne au moyen de l'indice entier $t = 1, \dots, T$.

On peut alors supposer que l'on se trouve face à un problème de données manquantes auquel on peut appliquer le principe de complétion des données initié par Tanner et Wong (1987). Il va s'agir de simuler le processus entre deux observations, r_t et r_{t+1} . On utilisera la discrétisation d'Euler avec un pas de discrétisation $\Delta = 1/M$ pour générer $\tilde{r}_{t+k\Delta}$, $k = 0, \dots, M$ avec la convention que pour $k = 0$ et $k = M$, on retrouve respectivement les observations r_t et r_{t+1} . Pour un M suffisamment grand, la discrétisation d'Euler va permettre d'écrire une approximation gaussienne relativement bonne de la fonction de vraisemblance. On va donc pouvoir construire un algorithme de Gibbs dans lequel on génère alternativement les observations manquantes $\tilde{r}_{t+k\Delta}$ conditionnellement aux paramètres et à l'échantillon observé et les paramètres conditionnellement aux données simulées et observées. Après une période d'initialisation, cet algorithme va converger vers la distribution *a posteriori* des paramètres θ .

Malgré la simplicité de son principe, cet algorithme ne va pas sans poser un certain nombre de problèmes comme l'ont souligné les différents auteurs qui s'y sont attaqués : Jones (1998), Eraker (2001), Elerian, Chib et Shepard (2001), Roberts et Stramer (2001). Les problèmes sont de trois ordres :

- Il s'agit de simuler le processus entre deux points déjà observés. On s'intéresse donc non pas à la densité $f(r_{t+(k+1)\Delta} \mid r_{t+k\Delta}, r_t, \theta)$ qui est une Normale dès que l'on utilise la discrétisation d'Euler, mais à la densité $f(r_{t+(k+1)\Delta} \mid r_{t+k\Delta}, r_t, r_{t+1}, \theta)$ dont on ne connaît pas la forme. Eraker (2001) simule $r_{t+(k+1)\Delta}$ en utilisant un algorithme de Metropolis. Elerian, Chib et Shepard (2001) choisissent de simuler tout un bloc d'inobservables.
- L'*a posteriori* de θ dans le modèle simulé n'a une forme simple que pour une catégorie très réduite de diffusions. Les différents auteurs cités proposent d'utiliser un algorithme de Metropolis pour tirer dans cette densité. Mais il serait aussi possible d'appliquer ici l'analyse que nous avons conduite en début de section. On retombe en fait sur les formules (24-26) qui permettent d'exploiter la linéarité partielle du modèle discrétisé et ainsi d'obtenir une procédure de simulation plus efficace.
- Enfin, il est connu (voir Sahu et Roberts, 1999) que l'algorithme de Gibbs converge mal quand la proportion des données simulées par rapport aux données observées est trop importante. On est alors en face du paradoxe suivant. L'approximation d'Euler sera d'autant meilleure que M sera grand. Mais un M trop grand empêche l'algorithme de Gibbs de converger. Roberts et Stramer (2001) proposent une reparamétrisation du processus pour contourner ce problème.

Pour conclure cette section, on remarquera avec Durham et Gallant (2002) que toute procédure d'estimation des équations de diffusion basée sur des simulations conduit à des temps de calcul beaucoup plus longs que les méthodes alternatives, en particulier celles basées sur une approximation de la fonction de vraisemblance.

3. CHOIX DE MODÈLES DANS UN CADRE BAYÉSIEN

CKLS (1992) furent les premiers à chercher à évaluer les différents modèles empiriques du taux d'intérêt de court terme. Ils partent d'un modèle en temps continu à dérive linéaire dont ils estiment une version discrétisée par la méthode des moments. Ils comparent les restrictions paramétriques impliquées par les différents modèles emboîtés au moyen d'un test du χ^2 basé sur la fonction objectif de la méthode d'estimation. Mais Brenner, Harjes et Kroner (1996) ont fait remarquer que la méthode des moments requiert la stationnarité du modèle discrétisé, hypothèse qui est violée par le modèle le plus général de CKLS (1992). Ait-Sahalia (1996b) utilise une méthode de comparaison de modèles qui est directement basée sur une estimation du modèle en temps continu. Son test est basé sur la distance entre une estimation non paramétrique de la densité marginale des observations et son expression paramétrique équivalente donnée en (17). Cette méthode ne souffre pas des mêmes inconvénients que la méthode de CKLS (1992). Conley, Hansen, Luttmer et Scheinkman (1997) considèrent un modèle général à la fois plus simple que celui de Ait-Sahalia (1996b) (leur fonction de volatilité est contrainte à être du type CEV $\sigma(r) = \sigma r^\gamma$) et plus compliqué car la relation entre temps discret et temps continu repose sur un processus auxiliaire non observé. Leur test suppose une estimation par la méthode des moments, mais ceux-ci sont calculés au moyen du générateur infinitésimal (10). Leur estimateur converge même pour des $\gamma > 1$ car les conditions de stationnarité en temps discret et en temps continu ne sont pas les mêmes.

Nous avons choisi d'estimer en bayésien une version discrétisée de différents modèles dont la forme paramétrique sera inspirée par les résultats d'une analyse non paramétrique préliminaire. Comme la méthode n'est pas exempte d'un biais de discrétisation, il n'est pas recommandé de comparer les modèles sur la seule base de leur ajustement comme le ferait la procédure traditionnelle des *posterior odds*. Nous avons donc privilégié une méthode alternative, issue, mais adaptée de la littérature bayésienne sur les choix de modèles. Cette méthode peut permettre de contourner les inconvénients liés à la discrétisation.

3.1 Cadre général

L'évaluation statistique d'un modèle consiste à vérifier si ce modèle, estimé sur un échantillon donné, parvient à reproduire correctement certaines des caractéristiques de l'échantillon. Ce point de vue a été examiné dans plusieurs articles de la littérature bayésienne. Les deux articles fondateurs de Box (1980) et de Rubin (1984) considèrent les propriétés d'échantillonnage des données simulées au moyen de la densité prédictive du modèle. La distribution prédictive attachée au modèle M

$$p(y | M) = \int f(y | \theta, M) \varphi(\theta | M) d\theta \quad (27)$$

représente la distribution de tous les échantillons possibles y qui se produiraient si le modèle M était le vrai modèle. Quand on observe un échantillon yd , celui-ci est utilisé pour faire de l'inférence (réviser l'*a priori* au moyen du théorème de Bayes), mais aussi pour vérifier l'adéquation du modèle aux données en calculant la probabilité que l'échantillon ait réellement été généré par le modèle M . Si $g(yd)$ est une fonction de l'échantillon observé, Box (1980) calcule la P -valeur prédictive de cette transformation au moyen de la distribution prédictive $p(g(y) | M)$. Une valeur faible remet en cause la validité du modèle. Mais naturellement, cette approche ne peut être suivie qu'avec un *a priori* $\varphi(\theta | M)$ informatif.

Rubin (1984) considère des simulations hypothétiques de l'échantillon y et vérifie si elles sont assez proches de l'échantillon original yd . Il considère les mêmes transformations $g(yd)$ que celles de Box (1980) mais préfère les évaluer utilisant certes $f(g(y) | \theta, M)$ mais en prenant la moyenne non par rapport à l'*a priori* en θ , mais par rapport à l'*a posteriori* de θ . C'est ce qu'il appelle la densité prédictive *a posteriori*. Si l'inférence a été faite en utilisant une méthode de Monte-Carlo par chaîne de Markov (MCMC), cette approche est commode parce que les tirages de la densité *a posteriori* de θ sont disponibles et par conséquent la densité prédictive *a posteriori* est obtenue comme sous-produit de l'inférence. Supposons que nous avons observé un échantillon yd de taille T et qu'une méthode MCMC ait produit N tirages pour θ . Pour chaque tirage θ_i , nous pouvons générer un nouvel échantillon y_i de taille T de $f(y | \theta_i)$ qui représente la densité d'échantillonnage de notre modèle à évaluer et ainsi calculer des « statistiques » à partir de $g(y_i)$. Avec les N tirages, nous pouvons alors estimer la densité prédictive *a posteriori* de $g(\cdot)$ et voir dans quel quantile se situe la statistique équivalente $g(yd)$ obtenue à partir de l'échantillon observé yd . Pour des références plus récentes sur le sujet, voir par exemple Meng (1994).

La principale difficulté de l'analyse réside dans le choix de la « bonne » transformation $g(\cdot)$ à faire. Par exemple, dans le problème classique d'échantillonnage normal avec moyenne et variance inconnues, Rubin note que les échantillons observés et simulés peuvent avoir leurs deux premiers moments identiques, mais que les écarts entre les statistiques d'ordre seront différents si le vrai processus de génération des données est Cauchy. L'idée que nous voudrions développer ici est qu'au lieu de chercher à trouver des bonnes statistiques, il vaut mieux considérer la distribution prédictive *a posteriori* complète et la comparer à une estimation non paramétrique de la distribution empirique de l'échantillon. Nous devons ainsi comparer la densité de l'échantillon observé à la densité d'un échantillon simulé au moyen de la densité prédictive. Ce choix est d'autant plus motivé dans notre contexte que l'estimateur non paramétrique de la densité des observations $\hat{f}(r)$ est indépendant du fait que l'on soit ou non en temps continu. Si notre modèle discrétisé peut présenter un biais de discrétisation, il sera jugé sur sa proximité par rapport à $\hat{f}(r)$, ce qui va pénaliser le biais de discrétisation par rapport à un critère usuel d'ajustement type *posterior odds*. De ce fait, notre procédure peut être vue en partie comme une extension de la procédure de test d'Ait-Sahalia (1996b).

3.2 Comment mesurer la distance entre deux densités

Nous sommes donc confrontés au problème de comparer deux fonctions de distribution inconnues. Ce problème a été traité dans la littérature non paramétrique, voir les références citées par exemple dans Pagan et Ullah (1999, section 2.9). Fondamentalement, nous devons d'abord définir une distance entre deux densités f et g . La plupart des tests non paramétriques considèrent l'erreur quadratique intégrée

$$ISE = \int [f(x) - g(x)]^2 dx = \int f(x) dF(x) + \int g(x) dG(x) - 2 \int g(x) dF(x)$$

où nous avons développé les carrés et utilisé la définition de l'intégrale de Stieltjes. L'évaluation de Monte-Carlo de $\int f(x) dF(x)$ donne $\sum f(x_i) / n$ quand les x_i sont distribués selon $F(x)$ tandis que l'évaluation par noyau de $f(x)$ est $\sum K((x_i - x) / h) / nh$. En conséquence si y_d et y_s sont les données observées et les données simulées avec des distributions respectives F et G , l'erreur quadratique intégrée peut être estimée au moyen de la double somme suivante :

$$\widehat{ISE}_T = \frac{1}{T^2 h} \sum_i \sum_j \left\{ K\left(\frac{y_{d_i} - y_{d_j}}{h}\right) + K\left(\frac{y_{s_i} - y_{s_j}}{h}\right) - 2K\left(\frac{y_{d_i} - y_{s_j}}{h}\right) \right\}.$$

En utilisant un argument asymptotique, Li (1996) construit un test classique basé sur

$$\widetilde{ISE}_T = \frac{1}{T^2 h} \sum_i \sum_{j \neq i} \left\{ K\left(\frac{y_{d_i} - y_{d_j}}{h}\right) + K\left(\frac{y_{s_i} - y_{s_j}}{h}\right) - K\left(\frac{y_{d_i} - y_{s_j}}{h}\right) - K\left(\frac{y_{s_i} - y_{d_j}}{h}\right) \right\}.$$

Quand $h \rightarrow 0$ et $Th \rightarrow \infty$,

$$T\sqrt{h} \widetilde{ISE}_T \rightarrow N\left(0, \sigma_{ISE}^2\right)$$

où $\sigma_{ISE}^2 = 2\left\{\int [f(x) + g(x)]^2 dx\right\} \left\{\int K^2(u) du\right\}$. Ce test, dont on dit qu'il est performant, a l'inconvénient pour nous de reposer sur deux approximations asymptotiques : une pour l'évaluation de l'intégrale par un estimateur de Monte-Carlo, une pour l'évaluation de sa distribution. Nous notons également que d'autres tests existants qui sont basés sur différentes mesures de contraste. Par exemple Ait-Sahalia (1996b) emploie l'erreur quadratique moyenne intégrée. D'autres exemples sont donnés dans Pagan et Ullah (1999). Dans cet article, nous suivons naturellement un itinéraire différent. Nous considérons tout d'abord un autre type de mesure de contraste qui est la ϕ divergence définie par :

$$D_\phi = \int \phi\left(\frac{f(x)}{g(x)}\right) g(x) dx. \quad (28)$$

Pour différents choix de ϕ , nous retrouvons les mesures habituelles de distance telles que la distance du χ^2 ($\phi(z) = (z - 1)^2$), l'entropie négative ou la divergence de Kullback-Leibler ($\phi(z) = z \log(z)$), et la distance de Hellinger à laquelle nous consacrerons notre attention. Elle est obtenue pour $\phi(z) = (\sqrt{z} - 1)^2$. Cette distance a deux avantages. Elle est symétrique et elle est à valeur entre 0 et 2. Nous avons

$$D_H^2 = 2 \left(1 - \int \sqrt{f(x)g(x)} \, dx \right) = 2 \left(1 - \int \frac{\sqrt{f(x)g(x)}}{f(x)} \, dF(x) \right). \tag{29}$$

Nous avons deux manières d'évaluer cette distance dans notre cas. Si nous suivons la démarche classique, nous obtenons

$$\tilde{D}_H^2 = 2 \left(1 - \frac{\frac{1}{T} \sum_i \sqrt{\sum_j K\left(\frac{yd_i - yd_j}{h}\right) \sum_j K\left(\frac{yd_i - yds_j}{h}\right)}}{\sum_j K\left(\frac{yd_i - yd_j}{h}\right)} \right). \tag{30}$$

Mais cette évaluation repose sur une approximation comme déjà mentionné plus haut. Une méthode alternative nous permet de réduire la variance de cet estimateur en remplaçant l'évaluation Monte-Carlo de l'intégrale par une règle de Simpson ce qui donne

$$\hat{D}_H^2 = 2 \left(1 - \frac{d}{3} \sum_i \sqrt{\hat{f}(y_i) \hat{g}(y_i)} w_i \right) \tag{31}$$

où d est la taille de l'incrément de y , w_i un poids valant 1 pour les points extrêmes, 2 pour les points pairs et 4 pour les points impairs. L'évaluation par noyau est conservée pour f et g :

$$\hat{f}(y) = \frac{1}{Th} \sum K\left(\frac{yd_i - y}{h}\right) \quad \hat{g}(y) = \frac{1}{Th} \sum K\left(\frac{yds_i - y}{h}\right). \tag{32}$$

Il nous reste alors à évaluer la distribution empirique de (31) en utilisant les valeurs simulées des paramètres et l'algorithme que nous allons maintenant décrire.

3.3 Tests utilisant une estimation de la distance de Hellinger

Nous emploierons la distance de Hellinger estimée pour comparer nos modèles. Le modèle produisant la plus petite distance à la distribution empirique des observations sera le modèle préféré. Un modèle produisant une « petite » distance estimée (suffisamment près de zéro) sera qualifié de correctement spécifié. Comme la distance de Hellinger est à valeur entre 0 et 2, il est relativement facile de dire

si elle est petite ou pas. Nous avons juste besoin de calculer la probabilité que \hat{D}_H^2 soit inférieure à une taille donnée a , choisie relativement près de zéro, en général 0,01, 0,05 ou 0,1. Cette probabilité correspond à une P -valeur prédictive *a posteriori* pour a . Nous choisissons le modèle qui produit la probabilité la plus élevée, qui signifie choisir le modèle qui a la probabilité la plus élevée de produire une D_H^2 proche de zéro. Afin de calculer cette probabilité, nous devons simuler la distribution empirique de \hat{D}_H^2 , c'est-à-dire de spécifier un algorithme de simulation pour la densité prédictive.

3.4 Comment simuler la densité prédictive

Nous sommes typiquement dans le contexte d'un modèle dynamique avec conditions initiales. La densité marginale de la première observation est obtenue comme

$$f(r_1) = \int f(r_1 | r_0, \theta) \hat{f}(r_0) \varphi(\theta | yd) dr_0 d\theta. \quad (33)$$

La condition initiale r_0 peut être tirée de la distribution empirique de l'échantillon observé yd et θ est tiré dans la densité *a posteriori* des paramètres. Conditionnellement à ces deux tirages, on peut obtenir un tirage dans $f(r_1 | r_{0i}, \theta_i)$. Si nous répétons T fois cette opération (un tirage dans $\hat{f}(r_0)$, un tirage dans $\varphi(\theta | yd)$ et un tirage correspondant dans $f(r_1 | r_{0i}, \theta_i)$), nous obtenons un échantillon de T tirages de la distribution prédictive *a posteriori* de r_1 . Nous calculons la distance de Hellinger entre ce premier échantillon simulé et l'échantillon observé yd . Nous appelons D_{H1}^2 cette valeur. Une fois que nous avons obtenu ce premier échantillon, nous pouvons poursuivre le processus avec

$$f(r_2) = \int f(r_2 | r_1, \theta) \hat{f}(r_1) \varphi(\theta | yd) dr_1 d\theta \quad (34)$$

ce qui signifie que cette fois nous considérons des tirages de r_2 , étant donné le tirage précédent de r_1 . Nous calculons une deuxième distance de Hellinger. Si nous continuons le processus, nous obtenons des tirages de la distribution empirique de \hat{D}_H^2 .

3.5 Discussion

Le point fondamental de cette méthode, et plus généralement de l'approche entreprise dans ce papier, concerne la relation entre modèle en temps continu et modèle en temps discret. La discrétisation d'Euler approche le modèle en temps continu, seulement pour $\Delta \rightarrow 0$. Pour un Δ strictement positif, il n'est pas évident de voir la relation entre les fonctions $\mu(r)$ et $\sigma(r)$ du modèle en temps continu et leur équivalent en temps discret $m(r)$ et $s(r)$. Par exemple, sauf à supposer que $\mu(r)$ est linéaire, l'espérance conditionnelle de $r_t | r_{t-1}$, donc $m(r)$, est une fonction à la fois de $\mu(r)$ et de $\sigma(r)$. Ait-Sahalia (1996b) utilise cet exemple pour justifier son test classique, basé sur le temps continu. En utilisant des données journalières, le même auteur, en section 4.6, estime par GLS une discrétisation d'Euler de son modèle en temps continu dans un but de comparer avec l'article de CKLS (1992).

Les graphiques induits pour $m(r)$ et $s(r)$ ne semblent pas très différents de ceux obtenus pour $\mu(r)$ et $\sigma(r)$ basés sur une estimation en temps continu (voir ses figures 5b-5c et 7a-7b). Étant donné que notre test est basé sur une comparaison entre la distribution marginale des observations et la même distribution simulée par le modèle, on peut également supposer que notre procédure ne sera pas trop pénalisée. On peut souligner également que l'approche d'inférence bayésienne dominante utilisée dans la littérature procède par complétion des données au moyen d'une approximation d'Euler avec un pas de discrétisation qui demeure strictement positif.

4. ÉVIDENCES NON PARAMÉTRIQUES POUR LES BONS DU TRÉSOR AMÉRICAIN

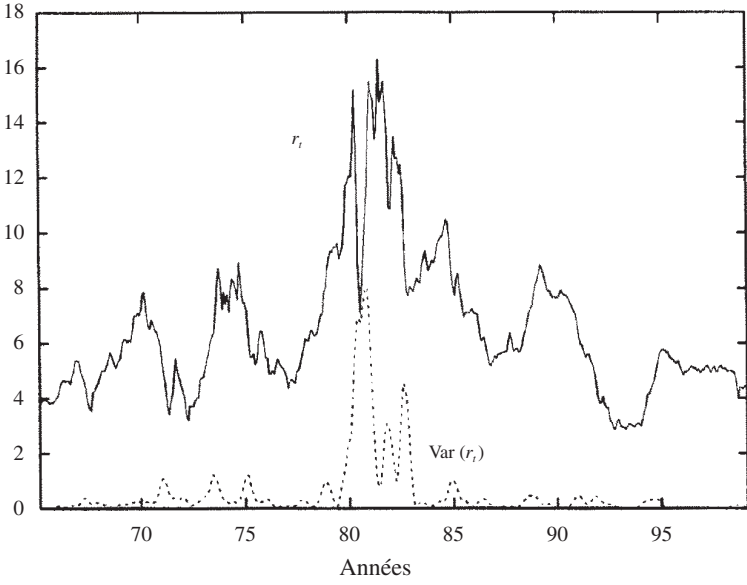
Pour le marché des obligations aux États-Unis, le taux des bons du Trésor à trois mois est souvent pris comme une *proxy* pour le taux sans risque, (voir par exemple Pfann, Schotman et Tschernig, 1996 ou Jiang, 1998). Les données utilisées sont extraites du site web de la Réserve Fédérale de St-Louis. Les observations mensuelles couvrent une période très longue 1934:1-1999:3. Ces données sont fortement hétérogènes et c'est certainement la raison pour laquelle la plupart des auteurs commencent leur échantillon au début des années soixante. En conséquence notre échantillon débute en 1965 :1. Les taux d'intérêt sont avant cette date très bas et peu volatils. Le graphique 1.a montre des données ainsi qu'une évaluation non paramétrique de la volatilité calculée sur une fenêtre de 13 observations pondérées par un noyau d'Epanechnikov. Ceci prouve que la majeure partie de la volatilité se situe dans les années 1979-1982. Pendant cette période, la Réserve Fédérale s'est écartée de sa politique habituelle visant à contrôler les taux d'intérêt nominaux et a choisi les réserves non empruntées comme nouvel instrument pour sa politique monétaire. Le même type de changement d'objectif a été opéré dans la plupart des pays européens pendant cette période ce qui a également conduit à des niveaux et à des volatilités des taux d'intérêts jusque-là inconnus.

Le graphique 1.b montre l'évaluation non paramétrique de la densité marginale de r obtenue par un noyau gaussien et une fenêtre déterminée au moyen de la règle de Silverman. La complexité des données apparaît dans ce graphique. Tout d'abord, cette densité est très asymétrique. Cette asymétrie ne peut être reproduite qu'avec un modèle hétéroscédastique où la volatilité est par exemple une fonction des niveaux passés. En second lieu, cette densité est multi-modale, indiquant le mélange d'au moins deux distributions. Enfin, la longue queue de droite correspond à des taux d'intérêt au-dessus de 10 pour cent provenant essentiellement des années 1979-1982.

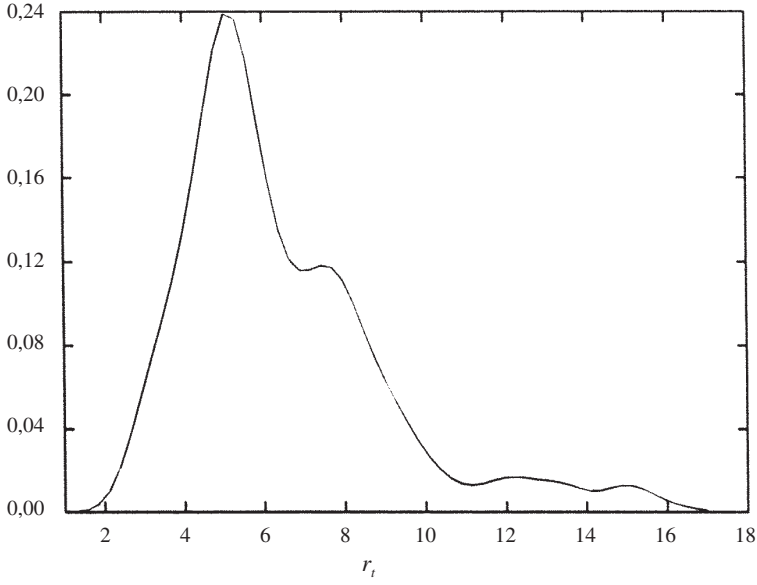
GRAPHIQUE 1

OBSERVATIONS, VOLATILITÉ ET DENSITÉ DU TAUX D'INTÉRÊT DE COURT TERME

a : données



b : densité



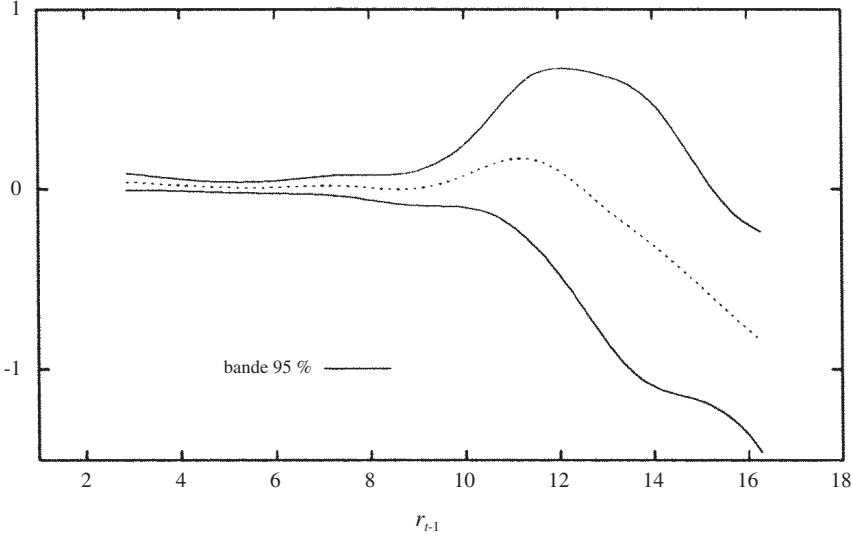
Nous allons maintenant évaluer la fonction de dérive $m(r)$ du modèle discrétisé (3) en employant une régression non paramétrique de Δr_t sur r_{t-1} . Le paramètre de lissage h est choisi égal à 2 après examen visuel du graphique de $\hat{m}(r)$. La fonction de dérive estimée, tracée dans le graphique 2.a en même temps qu'une bande de confiance uniforme à 95 %, semble être pratiquement constante et égale à zéro pour des taux jusqu'à 10 %. Ceci milite en faveur d'une promenade aléatoire. Pour des taux plus élevés, la fonction de dérive estimée adopte une pente linéaire négative. Cependant, un regard plus approfondi indique un retour à la moyenne très léger pour des taux plus bas que 6 %. Ait-Sahalia (1996b), pour le taux de dépôt des eurodollars à sept jours, considère la fonction de dérive $\mu(r) = \alpha_0 + \alpha_1 r + \alpha_2 r^2 + \alpha_3 / r$ dans son modèle en temps continu. Dans notre cas, la dérive semble être linéaire en r_{t-1} avec une rupture après 10 %. Ceci suggère un modèle à deux régimes pour la dérive comme dans par exemple Pfann, Schotman et Tschernig (1996). Notez cependant que l'effet de retour à la moyenne pour des valeurs de r supérieures à 10 % est très mal estimé. Il y a comparativement très peu d'observations dans ce régime (85 observations après 8 %, 34 après 10 % et 22 après 12 % sur 411 observations au total).

L'estimation non paramétrique de la fonction de volatilité $s(r)$ du modèle (3) au moyen de la formule (5) confirme, comme le montre le graphique 2.b, la relation apparente entre volatilité élevée et taux d'intérêt élevés. Cependant, la forme de cette courbe remet en cause fondamentalement la validité du modèle à élasticité constante $\sigma(r) = \sigma r^\gamma$. La volatilité semble être constante pour des taux bas et augmenter pour des taux plus élevés. Cependant, cette volatilité diminue après un certain niveau, niveau qui semble identique à celui à partir duquel la fonction de dérive devient négative. Pour conserver une certaine compatibilité avec les modèles existants d'évaluation, nous pouvons être tentés par une formulation où la volatilité suit deux régimes, mais reste constante à l'intérieur de chaque régime. La formulation paramétrique suggérée par Ait-Sahalia (1996b) pour le taux de dépôt d'eurodollar de sept jours $\sigma^2(r) = \sigma_0 + \sigma_1 r + \sigma^2 r^{2\gamma}$ peut constituer une alternative à ce modèle de deux régimes. Cependant, si nous voulons vraiment tenir compte de la diminution de la volatilité, la forme paramétrique $s^2(r) = \sigma / (1 + \gamma (r - c) 2)$ offre une solution plutôt parcimonieuse car avec deux paramètres, γ et c , elle arrive bien à reproduire la forme en cloche du graphique 2.b.

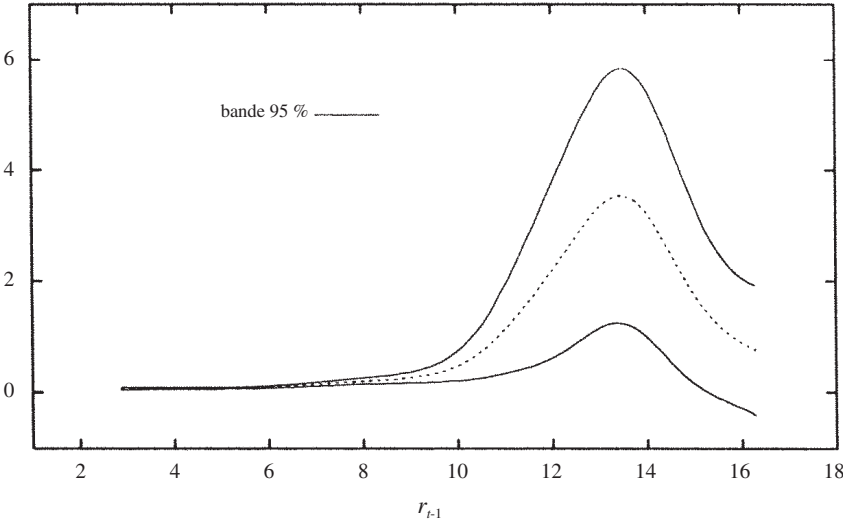
GRAPHIQUE 2

ESTIMATION NON PARAMÉTRIQUE DE LA DIFFUSION DISCRÉTISÉE

a : dérive



b : volatilité



5. UN AUTRE REGARD BAYÉSIEN SUR LE MODÈLE CEV

Nous passerons en revue et comparerons diverses versions usuelles du modèle à élasticité constante de la volatilité (CEV) mis en avant par CKLS (1992), en considérant une dérive linéaire et une dérive non linéaire. Nous voulons en particulier vérifier l'affirmation d'Ait-Sahalia (1996b) sur la source principale d'erreur de spécification de la linéarité de la dérive. Mais nous voulons également vérifier le rôle joué par la paramétrisation de la fonction de volatilité. Nous partons du modèle discrétisé suivant :

$$\Delta r_t = \alpha_0 + \alpha_1 r_{t-1} + \alpha_2 r_{t-1}^2 + \alpha_3 / r_{t-1} + \sigma s(r_{t-1}) \varepsilon_t \tag{35}$$

où $\varepsilon_t \sim N(0, 1)$. Nous considérerons les fonctions de volatilité suivantes :

- La formulation simple de Vasicek (1977) avec $s^2(r_{t-1}) = 1$.
- Le modèle CEV avec $s^2(r_{t-1}) = r_{t-1}^{2\gamma}$.
- Finalement, nous introduisons la fonction de volatilité Cauchy (CAV)

$$s^2(r_{t-1}) = \frac{1 + \gamma}{1 + \gamma(r_{t-1} - c_r)^2} \tag{36}$$

où la volatilité est une fonction croissante de r , atteint un maximum pour $r_{t-1} = c_r$ et diminue ensuite. Le paramètre γ mesure la vitesse de transition entre les périodes de taux faibles et de taux élevés⁶. Le paramètre c_r est d'une certaine manière un paramètre de seuil car il donne une indication sur le niveau de r après lequel la volatilité diminue.

Nous voulons comparer trois modèles (Vasicek, CEV, CAV) ayant soit une dérive linéaire soit une dérive non linéaire. Ceci fait six variantes différentes. Examinons tout d'abord dans le tableau 1 la probabilité *a posteriori* que la distance de Hellinger soit inférieure à un niveau donné.

La première conclusion que nous pouvons tirer de ce tableau est qu'il faut tenir compte de l'hétéroscédasticité ($s(r) \neq 1$). Le modèle de Vasicek est complètement dominé et pire il n'arrive même pas à générer des taux d'intérêt positifs avec une dérive linéaire. La deuxième conclusion est qu'une fonction de dérive non linéaire semble meilleure dans tous les cas. Cependant, une fois que l'on a pris en compte l'hétéroscédasticité, il est difficile de distinguer entre les différents modèles car CEV et CAV semblent équivalents.

6. Le paramètre γ , mesurant une vitesse, n'est pas estimable sans *a priori* informatif. La probabilité d'une vitesse infinie n'est pas nulle, ce qui rend l'*a posteriori* non intégrable sans information supplémentaire. Nous avons utilisé un *a priori* exponentiel avec $q(\gamma) \propto \exp(-\gamma)$.

TABLEAU 1

P-VALEURS PRÉDICTIVES *A POSTERIORI* POUR LES MODÈLES VASICEK, CEV ET CAV

Modèle	$s^2(r) = 1$			$s^2(r) = r^{2\gamma}$			$s^2(r) = \frac{1+\gamma}{1+\gamma(r-c_r)^2}$		
	0,01	0,05	0,10	0,01	0,05	0,10	0,01	0,05	0,10
a	0,01	0,05	0,10	0,01	0,05	0,10	0,01	0,05	0,10
N.L.	0,00	0,00	0,00	0,00	0,62	1,00	0,00	0,50	0,96
L.D.	–	–	–	0,00	0,56	0,82	0,00	0,28	0,76

NOTE : Les calculs sont basés sur 250 simulations. La fonction de volatilité est détaillée sur l'axe horizontal. La fonction de dérive change avec l'axe vertical. L.D. signifie dérive linéaire et N.L. dérive non linéaire. – indique la présence de tirages négatifs. Les nombres correspondent à $Pr(D_H^2 < a)$.

Donnons maintenant les résultats d'inférence pour les modèles CEV et CAV avec une dérive non linéaire. Du tableau 2, nous voyons que les paramètres de la fonction de dérive sont estimés avec plus de précision dans le modèle CAV. Sa fonction de volatilité parvient à capturer exactement la crête de la volatilité révélée autour de $r = 13$ par l'analyse non paramétrique. Pour le modèle CEV, l'espérance *a posteriori* de γ est plus grande que un, conduisant en théorie à un modèle non stationnaire quand la dérive est linéaire. Notez cependant que l'espérance *a posteriori* de α_2 est négative, rétablissant de ce fait la stationnarité dans le cas d'une dérive non linéaire.

TABLEAU 2

RÉSULTATS D'INFÉRENCE POUR LES MODÈLES CEV ET CAV
AVEC UNE DÉRIVE NON LINÉAIRE

Modèle	α_0	α_1	α_2	α_3	γ	c_r	σ^2
CEV	-0,32 [0,29]	0,14 [0,13]	-0,020 [0,017]	0,24 [0,20]	1,37 [0,08]	– –	0,013
CAV	-0,55 [0,28]	0,25 [0,12]	-0,035 [0,015]	0,40 [0,21]	4,42 [1,34]	13,08 [0,30]	0,60

NOTE : Les écarts-type *a posteriori* sont indiqués entre crochets.

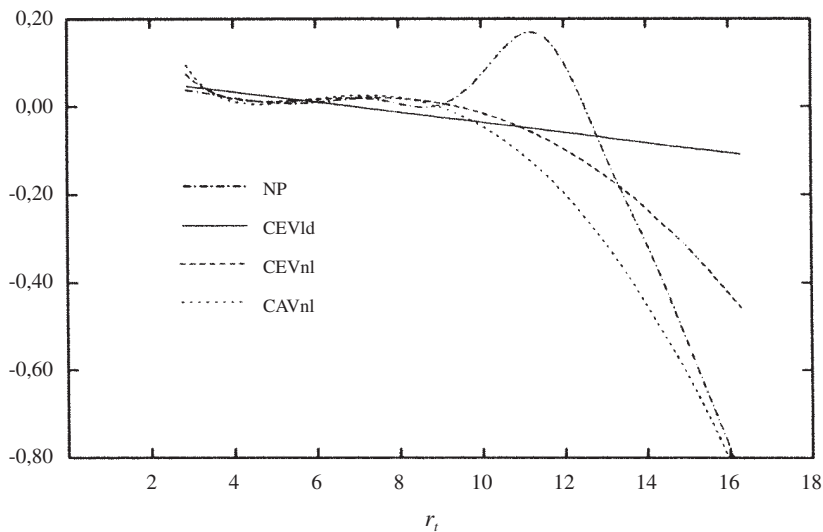
Les graphiques des fonctions implicites de dérive et de volatilité, calculées à l'espérance *a posteriori* des paramètres et tracées dans le graphique 3, donnent plus de détails au sujet de la capacité des différentes spécifications à reproduire certaines des caractéristiques des données. S'il n'y a pas beaucoup de différences entre les différentes fonctions non linéaires de dérive, les différences sont plus

marquées pour les fonctions de volatilité. Le modèle de CAV a une capacité bien supérieure à celle du modèle CEV pour reproduire la volatilité contenue dans les données.

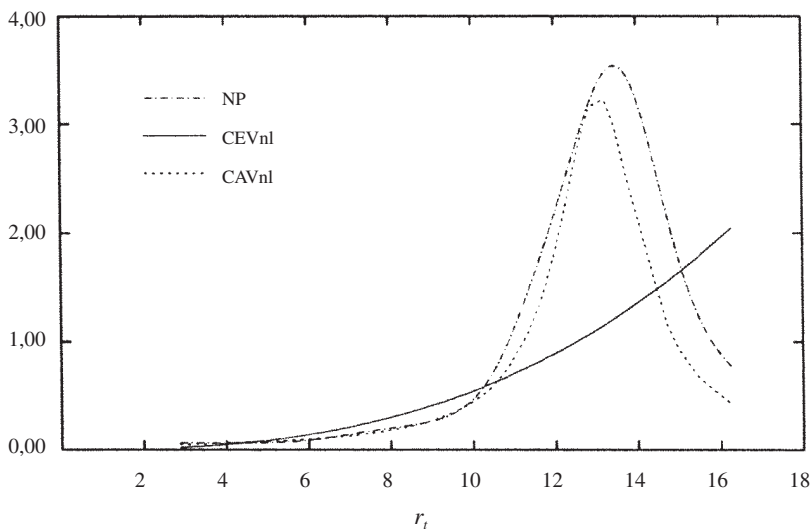
GRAPHIQUE 3

FONCTIONS DE DÉRIVE ET DE VOLATILITÉ IMPLICITES

a : dérive



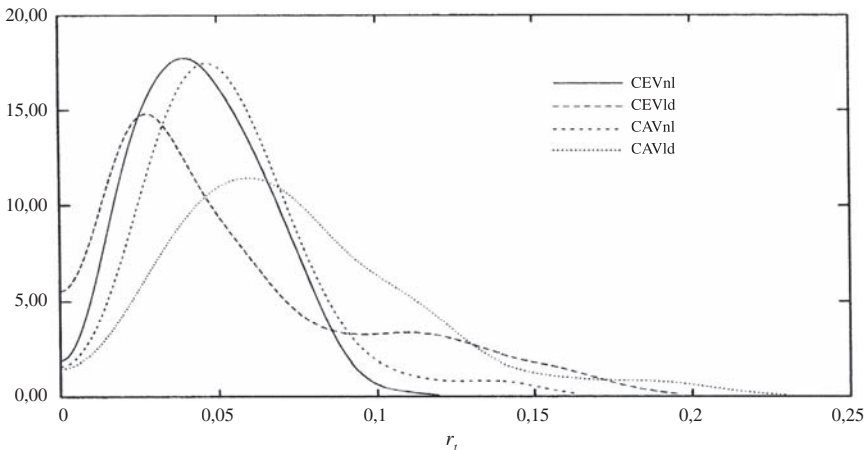
b : volatilité



Dans le graphique 4, nous montrons la distribution d'échantillonnage de la distance de Hellinger pour les quatre processus intéressants. Le modèle de Vasicek n'est pas considéré car il était clairement rejeté par les données dans le tableau 1. Il est évident à première vue que, en moyenne, la distance de Hellinger est proche de zéro pour tous les modèles. Mais il y a une grande différence dans le comportement de queue pour les modèles avec une dérive linéaire. En particulier, la grande queue droite produite par le modèle CEV avec dérive linéaire peut être due à un problème de non stationnarité, mais le modèle CAV avec une dérive linéaire reproduit le même comportement. Nous pouvons confirmer les conclusions émises par Ait-Sahalia (1996b) au sujet des dérives linéaires, mais nos conclusions sont moins fortes que les siennes, obtenues dans un cadre classique, et d'ailleurs avec un modèle différent.

GRAPHIQUE 4

DISTRIBUTION D'ÉCHANTILLONNAGE DE LA DISTANCE DE HELLINGER



NOTE : CEV et CAV indiquent le type de fonction de volatilité tandis que les lettres minuscules ld et nl indiquent respectivement une dérive linéaire et non linéaire.

6. L'IMPACT D'UNE MODÉLISATION À DEUX RÉGIMES

Beaucoup d'auteurs ont proposé de considérer des modèles à changement de régime comme une alternative possible au modèle CEV. Pfann, Schotman et Tschernig (1996), Gray (1996), Pai et Pedersen (1998) considèrent un modèle à deux régimes avec une dérive linéaire, mais avec différents coefficients de retour à la moyenne dans chaque régime. Gray (1996), Pai et Pedersen (1998) ou encore Bansal et Zhou (2002) considèrent un modèle avec chaîne de Markov cachée du type de celui proposé par Hamilton (1989) ce qui rend difficile une modélisation de la dépendance de la volatilité. Pfann, Schotman et Tschernig (1996) considèrent un modèle TAR (autorégressif à seuil) à la Tong. Ces modèles sont plus

robustes que les modèles à chaîne de Markov cachée comme l'a montré Boldin (1996). Leur modèle comporte une dérive linéaire différente dans chaque régime, mais impose une fonction commune de volatilité.

6.1 Spécification du modèle

Considérons maintenant une fonction de transition F à valeur dans $[0, 1]$ et définie par

$$F(c_1, r_{t-1}) = \begin{cases} 1 & \text{si } r_{t-1} > c_1 \\ 0 & \text{autrement.} \end{cases} \quad (37)$$

Notre modèle de départ sera

$$\Delta r_t = [1 - F(c_1, r_{t-1})] m_1(r_{t-1}) + F(c_1, r_{t-1}) m_2(r_{t-1}) + \sigma s_t \varepsilon_t. \quad (38)$$

Il est intéressant d'étudier diverses fonctions de dérive. La première question est de savoir si le mécanisme de commutation parvient à épuiser toute la non-linéarité contenue dans les données. Étant donné que le principal problème auquel nous devons faire face lors de l'étape de simulation en temps discret concerne les taux d'intérêt négatifs, nous partirons de la formulation générale suivante :

$$\begin{aligned} m_1(r_{t-1}) &= \alpha_{10} + \alpha_{11} r_{t-1} + \alpha_{12} / r_{t-1}, \\ m_2(r_{t-1}) &= \alpha_{20} + \alpha_{21} r_{t-1}. \end{aligned} \quad (39)$$

Dans le premier régime le terme $1 / r_{t-1}$ pénalise des taux d'intérêt trop petits. Il faudra tester si ce terme est nécessaire.

Maintenant indiquons trois options pour la fonction scédastique s_t^2 . Nous considérons tout d'abord une généralisation du modèle de Vasicek qui autorise des variances différentes entre chaque régime, mais constantes à l'intérieur d'un même régime

$$\begin{aligned} \sigma^2 s_t^2 &= [1 - F(c_2, r_{t-1})] \sigma_1^2 + F(c_2, r_{t-1}) \sigma_2^2, \\ &= \sigma_1^2 [1 - F(c_2, r_{t-1}) + F(c_2, r_{t-1}) \phi] \end{aligned} \quad (40)$$

où $\phi = \sigma_2^2 / \sigma_1^2$. Nous considérons deux paramètres de seuils c_1 et c_2 pour permettre aux fonctions de dérive et de volatilité de réagir différemment aux niveaux du taux d'intérêt. Nous considérons ensuite les mêmes options qu'auparavant : les fonctions de volatilité CEV et CAV.

Remarque : il convient de signaler que ce type de modèle peut difficilement être relié à un modèle en temps continu car le changement de régime viole l'hypothèse selon laquelle les fonctions de dérive et de volatilité sont indépendantes de t . C'est la raison pour laquelle tous les auteurs cités en début de section utilisent un modèle

en temps discret. Ait-Sahalia (1996b), pour rester dans un cadre de temps continu, favorise une modélisation non linéaire plutôt qu'un modèle à changement de régime. On peut toutefois se poser la question de savoir si cette objection tient toujours si F définie en (37) devient une fonction de transition lisse comme par exemple une fonction logistique.

6.2 Inférence bayésienne

L'inférence bayésienne peut être conduite sur ces huit variantes en utilisant le même cadre que celui présenté dans la première section, pour peu que nous considérons la transformation suivante des données

$$\begin{aligned} y_t(\theta) &= y_t / s_t(\theta), \\ x_t(\theta) &= [(1 - F(c_1) x'_t, F(c_1) x'_t) / s_t(\theta)] \end{aligned} \quad (41)$$

afin de récrire le modèle sous la même forme que (22). c_1 est maintenant le premier élément de θ . Nous rapportons dans le tableau 3 les P -valeurs prédictives *a posteriori* de a attachées à chaque variante.

TABLEAU 3

P-VALEURS PRÉDICTIVES *A POSTERIORI* POUR LES DIFFÉRENTES VERSIONS
DU MODÈLE À DEUX RÉGIMES

Modèle	σ_1^2, σ_2^2			$r^{2\gamma}$			$\frac{1+\gamma}{1+\gamma(r-c_r)^2}$		
	0,01	0,05	0,10	0,01	0,05	0,10	0,01	0,05	0,10
a	0,01	0,05	0,10	0,01	0,05	0,10	0,01	0,05	0,10
N.L.	0,00	0,62	0,99	0,01	0,66	0,96	0,04	0,90	1,00
L.D.	–	–	–	0,01	0,60	0,89	0,00	0,51	0,92

NOTE : Les calculs sont basés sur 250 simulations. La fonction de volatilité est détaillée sur l'axe horizontal. La fonction de dérive change avec l'axe vertical. L.D. signifie une dérive linéaire dans chaque régime, alors que N.L. correspond à une spécification complète. – indique des tirages négatifs. Les nombres correspondent à $Pr(D_H^2 < a)$.

Examinons tout d'abord l'impact du choix des différentes modélisations retenues pour les fonctions de dérive sur les performances prédictives du modèle. Clairement, tous les modèles préfèrent avoir une dérive non linéaire dans le premier régime. Le mécanisme de changement de régime couplé à une dérive linéaire fournit une modélisation satisfaisante pour le régime de taux d'intérêt élevé, mais le modèle a besoin de l'aide d'un terme non linéaire additionnel pour le régime de taux d'intérêt bas. Si nous nous tournons maintenant vers l'évaluation de la fonction de volatilité, il s'avère que le modèle CAV est de loin le meilleur choix. Un

mécanisme de changement de régime à l'intérieur de la fonction de volatilité du modèle de Vasicek donne de bons résultats et fournit un modèle intéressant quand la fonction de dérive est non linéaire.

Donnons maintenant les résultats d'inférence dans le tableau 4 pour les trois fonctions différentes de volatilité et un changement de régime avec non-linéarité pour la fonction de dérive. Quand la fonction de volatilité est de type Cauchy, les données indiquent que les fonctions de dérive et de volatilité ont des inflexions pour à peu près le même niveau de taux d'intérêt. Pour le modèle de Vasicek à changement de régime, c_1 et c_2 sont au contraire relativement différents, car dans ce cas les deux fonctions scédastiques peuvent s'interpréter comme une approximation de la fonction CAV. Nous devons maintenant savoir si le fait d'envisager un modèle à changement de régime vaut le coût supplémentaire en temps de calcul. Dans les graphiques 5.a et 5.b nous comparons les meilleurs modèles en provenance des tableaux 1 et 3. Nous avons un groupe de trois modèles qui semblent équivalents : CEV, CAV et 2rCEV. L'un d'eux possède un changement de régime dans la fonction de dérive. Mais un quatrième modèle domine ce groupe : le modèle CAV avec un changement de régime dans la fonction de dérive. En conséquence, un mécanisme de changement de régime constitue clairement une amélioration majeure pour certains modèles. Ceci est particulièrement vrai pour le modèle de Vasicek qui ne figure pas dans le graphique 6.

TABLEAU 4

RÉSULTATS D'INFÉRENCE POUR LES MODÈLES VASICEK, CEV ET CAV
 À CHANGEMENT DE RÉGIME AVEC UNE DÉRIVE NON LINÉAIRE

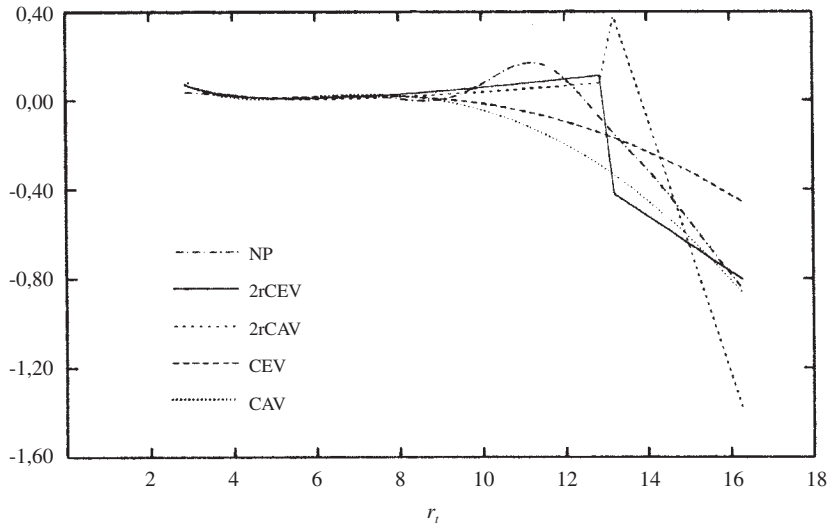
Modèle	α_{10}	α_{11}	α_{12}	α_{20}	α_{21}	c_1	ϕ ou γ	c_2 ou c_r
Vasicek	-0,07 [0,14]	0,01 [0,03]	0,09 [0,14]	1,36 [2,59]	-0,28 [0,45]	12,86 [0,41]	4,66 [0,67]	9,94 [0,37]
CEV	-0,10 [0,11]	0,02 [0,03]	0,12 [0,10]	0,47 [2,27]	-0,12 [0,40]	12,88 [0,36]	1,37 [0,08]	
CAV	-0,09 [0,12]	0,02 [0,03]	0,11 [0,11]	3,10 [2,80]	-0,57 [0,47]	12,97 [1,32]	4,33 [1,32]	13,06 [0,29]

NOTE : Les écarts-type *a posteriori* sont indiqués entre crochets.

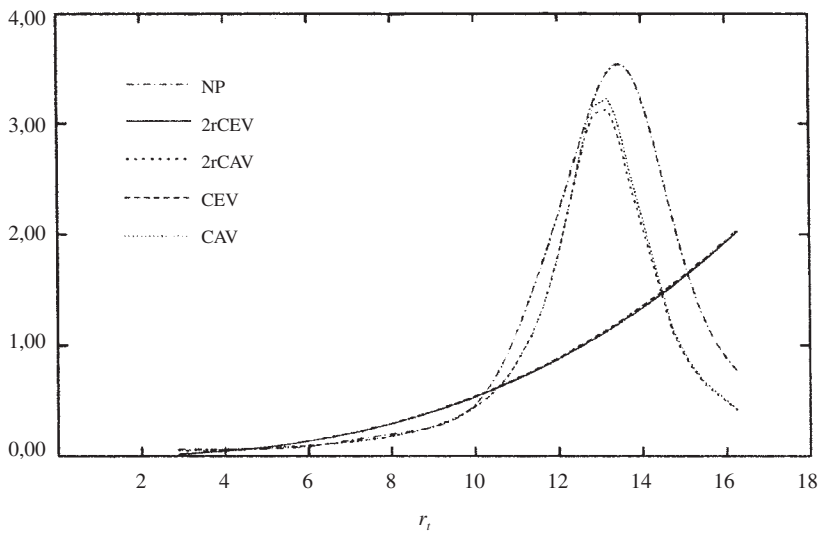
GRAPHIQUE 5

FONCTIONS DE DÉRIVE ET DE VOLATILITÉ IMPLICITES

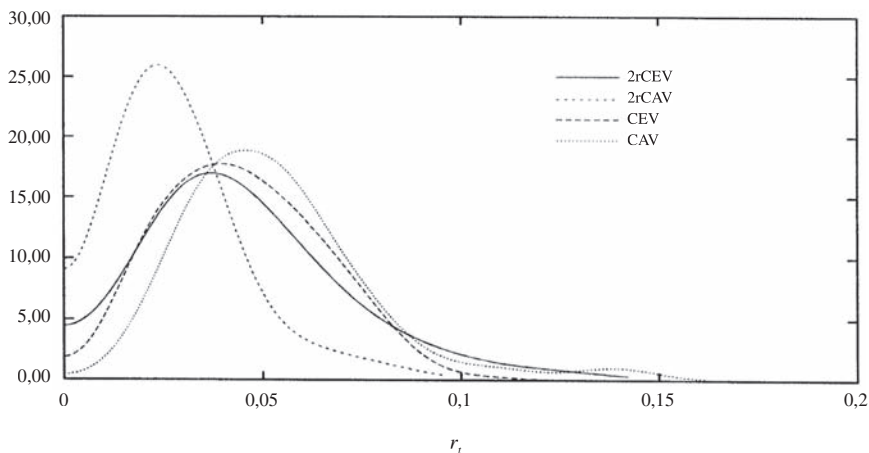
a : dérive



b : volatilité



GRAPHIQUE 6

COMPARAISONS DES MEILLEURS MODÈLES AU MOYEN DE D_H^2 

CONCLUSION

Dans cet article, nous avons combiné l'utilisation d'estimateurs non paramétriques classiques et de l'inférence bayésienne. L'exploration des propriétés de la fonction de vraisemblance est d'un intérêt commun pour les deux approches. L'approche non paramétrique donne des intuitions et des indications sur la forme potentielle des relations à modéliser. L'approche bayésienne détaille la densité *a posteriori* des paramètres d'intérêt et détecte le manque possible d'information fourni par l'échantillon dans certaines directions. D'ailleurs, l'approche bayésienne permet de simuler la densité prédictive *a posteriori* qui peut être confrontée à l'échantillon réel. Cette confrontation nécessite l'évaluation de densités, ce qui peut encore être fait au moyen d'une approche non paramétrique.

Nous avons développé une méthode pour évaluer et comparer des modèles. Cette méthode se concentre sur les propriétés dynamiques de la simulation du modèle estimé et est ainsi plutôt exigeante. Ait-Sahalia (1996b) a employé une approche semblable, c'est-à-dire qu'il compare la densité estimée de l'échantillon à la densité marginale produite par le processus paramétrique en temps continu qui est donné en (17). L'approche d'Ait-Sahalia (1996b) est différente car elle considère directement le processus en temps continu. Son processus d'évaluation a un taux de rejet bien supérieur au nôtre.

Quand un modèle est rejeté par l'approche prédictive, aucune indication n'est fournie quant à la direction où le modèle devrait être amélioré. Nous n'avons pas fourni ou utilisé de tests de mauvaise spécification. Cependant, les évaluations non paramétriques préliminaires nous fournissent tout d'abord un guide pour

proposer une formulation paramétrique particulière et ensuite un instrument d'évaluation *a posteriori* pour les fonctions implicites de dérive et de volatilité calculées en l'espérance *a posteriori*.

Un modèle peut présenter des propriétés prédictives très bonnes et en même temps ne pas être très fiable pour des applications en finance. Il y a un écart entre un modèle empirique du taux d'intérêt de court terme et une modélisation de la probabilité risque neutre. Certains de nos modèles ont pu très facilement reproduire la distribution empirique du taux d'intérêt sans risque. Mais il peut être très difficile d'introduire le prix du risque d'une manière raisonnable quand la fonction de dérive est non linéaire. Un modèle avec une dérive non linéaire pure, est très difficile à calibrer, alors qu'un mécanisme à changement de régime est beaucoup plus facile à manipuler.

BIBLIOGRAPHIE

- AIT-SAHALIA, Y. (1996a), « Non-parametric Pricing of Interest Rate Derivative Securities », *Econometrica*, 64(3) : 527-560.
- AIT-SAHALIA, Y. (1996b), « Testing Continuous-time Models of the Spot Interest Rate », *Review of Financial Studies*, 9(2) : 385-426.
- AIT-SAHALIA, Y. (2002), « Maximum-likelihood Estimation of Discretely-sampled Diffusions: A Closed-form Approximation Approach », *Econometrica*, 70(1) : 223-262.
- AUESTADT, B. et D. TJØSTHEIM (1990), « Identification of Non-linear Time Series: First Order Characterisation and Order Estimation », *Biometrika*, 77(3) : 669-687.
- BANSAL, R. et H. ZHOU (2002), « Term Structure of Interest Rates With Regime Shifts », *Journal of Finance*, 57(5) : 1 997-2 043.
- BAUWENS, L. et M. LUBRANO (1998), « Bayesian Inference in GARCH Models Using the Gibbs Sampler », *The Econometrics Journal*, 1 : C23-C46, <http://www.blackwellpublishers.co.uk/ectj>.
- BAUWENS, L., M. LUBRANO et J.-F. RICHARD (1999), *Bayesian Inference in Dynamic Econometric Models*. Oxford University Press, Oxford.
- BICKEL, P. et M. ROSENBLATT (1973), « One Some Global Measures of the Deviation of Density Function Estimates », *Annals of Statistics*, 1(6) : 1 071-1 091.
- BOLDIN, M. (1996), « A Check on the Robustness of Hamilton's Markov Switching Model Approach to the Economic Analysis of the Business Cycle », *Studies in Non-linear Dynamics and Econometrics*, 1(1) : 35-46.
- BOX, G. E. (1980), « Sampling and Bayes' Inference in Scientific Modelling and Robustness », *Journal of the Royal Statistical Society, A*, 143(4) : 383-430.
- BRENNER, R. J., R. H. HARJES et K. F. KRONER (1996), « Another Look at Models of the Short-term Interest Rate », *Journal of Financial and Quantitative Analysis*, 31(1) : 85-107.

- BROZE, L., O. SCAILLET et J.-M. ZAKOIAN (1995), « Testing for Continuous-time Models of the Short-term Interest Rate », *Journal of Empirical Finance*, 2(3) : 199-223.
- CHAN, K., G. KAROLYI, F. LONGSTAFF et A. SANDERS (1992), « An Empirical Comparison of Alternative Models of the Short-term Interest Rates », *Journal of Finance*, 47(3) : 1 209-1 227.
- CONLEY, T. G., L. P. HANSEN, E. G. LUTTMER et J. A. SCHEINKMAN (1997), « Short-term Interest Rates as Subordinated Diffusions », *The Review of Financial Studies*, 10(3) : 525-577.
- COX, J., J. INGERSOLL et S. ROSS (1985), « A Theory of the Term Structure of Interest Rates », *Econometrica*, 53(2) : 385-407.
- DOTHAN, L. U. (1978), « On the Term Structure of Interest Rates », *Journal of Financial Economics*, 6 : 59-69.
- DURHAM, G. et R. GALLANT (2002), « Numerical Techniques for Maximum Likelihood Estimation of Continuous-time Diffusion Processes », *The Journal of Business and Economic Statistics*, 20(3) : 297-316.
- ELERIAN, O. (1998), « A Note on the Existence of a Closed Form Conditional Transition Density for the Milstein Scheme », Discussion Paper 1998-W18, Nuffield College, Oxford.
- ELERIAN, O., S. CHIB et N. SHEPHARD (2001), « Likelihood Inference for Discretely Observed Non-linear Diffusions », *Econometrica*, 69(4) : 959-993.
- ERAKER, B. O. (2001), « MCMC Analysis of Diffusion Models with Application to Finance », *Journal of Business and Economic Statistics*, 19(2) : 177-191.
- FLORENS-ZMIROU, D. (1993), « On Estimating the Diffusion Coefficient from Discrete Observations », *Journal of Applied Probability*, 30 : 790-804.
- GOURIÉROUX, C. et A. MONFORT (1996), *Simulation-Based Econometric Methods*, Oxford University Press, Oxford.
- GRAY, S. F. (1996), « Modeling the Conditional Distribution of Interest Rates as a Regime Switching Process », *Journal of Financial Economics*, 42(1) : 27-62.
- HAMILTON, J. (1989), « A New Approach to the Economic Analysis of Non-stationary Time Series and the Business Cycle », *Econometrica*, 57(2) : 357-384.
- HÄRDLE, W. (1990), *Applied Non-parametric Regression*, Econometric Society Monographs, Cambridge University Press, Cambridge, New York et Melbourne.
- HÄRDLE, W., H. LÜTKEPOHL et R. CHEN (1997), « A Review of Non-parametric Time Series Analysis », *International Statistical Review*, 65(1) : 49-72.
- INGERSOLL, J. E. (1987), *Theory of Financial Decision Making*, Rowman and Littlefield Publishers, Savage, États-Unis.
- JIANG, G. et J. KNIGHT (1997), « A Non-parametric Approach to the Estimation of Diffusion Processes, with an Application to Short Term Interest Rate Model », *Econometric Theory*, 13(5) : 615-645.

- JIANG, G. J. (1998), « Non-parametric Modeling of U.S. Interest Rate Term Structure Dynamics and Implications on the Price of Derivative Securities », *Journal of Financial and Quantitative Analysis*, 33(4) : 465-497.
- JONES, C. (1998), « A Simple Bayesian Method for the Analysis of Diffusion Processes », Discussion paper, The Wharton School, University of Pennsylvania, Philadelphia.
- KESSLER, M. (1997), « Estimation of an Ergodic Diffusion from Discrete Observations », *Scandinavian Journal of Statistics*, 24 : 211-229.
- LI, Q. (1996), « Nonparametric Testing of Closeness between Two Unknown Distribution Functions », *Econometric Reviews*, 15(3) : 261-274.
- LUBRANO, M. (2000), « Bayesian Analysis of Non-linear Time Series Models with a Threshold », in W. A. BARNETT, D. F. HENDRY, S. HYLLEBERG, T. TERÄSVIRTA, D. TJØSTHEIM et A. WÜRTZ (éds), *Non-linear Econometric Modelling*, p. 79-118, Cambridge University Press, Cambridge.
- LUBRANO, M. (2001), « Smooth Transition GARCH Models: A Bayesian Perspective », *Recherches Économiques de Louvain*, 67(3) : 257-287.
- MENG, X.-L. (1994), « Posterior Predictive P-Values », *The Annals of Statistics*, 22(3) : 1 142-1 160.
- MILSTEIN, G. (1978), « A Method of Second Order Accuracy Integration of Stochastic Differential Equations », *Theory of Probability and its Applications*, 23 : 396-401.
- PAGAN, A. et A. ULLAH (1999), *Nonparametric Econometrics*, Themes in Modern Econometrics. Cambridge, New York et Melbourne.
- PAI, J. et H. PEDERSEN (1998), « The Bayesian Approach to Regime Switching Models of the Short Interest Rate: Theory, Calibration and Implementation », Mimeo.
- PEDERSEN, A. R. (1995), « A New Approach to Maximum Likelihood Estimation for Stochastic Differential Equations Based on Discrete Observations », *Scandinavian Journal of Statistics*, 22 : 55-71.
- PFANN, G. A., P. C. SCHOTMAN et R. TSCHERNIG (1996), « Non-linear Interest Rate Dynamics and Implications for the Term Structure », *Journal of Econometrics*, 74(1) : 149-176.
- RICE, J. (1984), « Bandwidth Choice for Non-parametric Regression », *Annals of Statistics*, 12(4) : 1 215-1 230.
- ROBERTS, G. et O. STRAMER (2001), « On Inference for Partially Observed Nonlinear Diffusion Models Using Metropolis-Hastings Algorithm », *Biometrika*, 88(3) : 603-621.
- RUBIN, D. B. (1984), « Bayesianly Justifiable and Relevant Frequency Calculations for the Applied Statistician », *The Annals of Statistics*, 12(4) : 1 151-1 172.
- SAHU, S. et G. ROBERTS (1999), « On Convergence of the EM Algorithm and the Gibbs Sampler », *Statistics and Computing*, 9 : 55-64.

- STANTON, R. (1997), « A Non-parametric Model of Term Structure Dynamics and the Market Price of Interest Rate Risk », *The Journal of Finance*, 52(5) : 1 973-2 002.
- TANNER, M. A. et W. H. WONG (1987), « The Calculation of Posterior Distributions by Data Augmentation », *Journal of the American Statistical Association*, 82(398) : 528-540.
- TJØSTHEIM, D. et B. AUESTADT (1994a), « Non-parametric Identification of Non-linear Time Series: Projection », *Journal of the American Statistical Association*, 89(428) : 1 398-1 409.
- TJØSTHEIM, D. et B. AUESTADT (1994b), « Non-parametric Identification of Non-linear Time Series: Selecting Significant Lags », *Journal of the American Statistical Association*, 89(428) : 1 410-1 419.
- VASICEK, O. (1977), « An Equilibrium Characterisation of the Term Structure », *Journal of Financial Economics*, 5 : 177-188.