

GREQAM

**Groupement de Recherche en Economie
Quantitative d'Aix-Marseille - UMR-CNRS 6579
Ecole des Hautes Etudes en Sciences Sociales
Universités d'Aix-Marseille II et III**

Document de Travail

n°03A15

Ludovic GIET

**Estimation par inférence indirecte des
équations de diffusion: l'impact du choix du
procédé de discrétisation**



Estimation par inférence indirecte des équations de diffusion: l'impact du choix du procédé de discrétisation

Ludovic GIET¹

12 juin 2003

Abstract

We propose an adaptation of the standard method of estimation by indirect inference of a stochastic differential equation, where the Euler scheme is replaced by the Milstein one. The latter, for a fixed discretization step, is more accurate than the first, and results in saving of time in the calibration step. This saving will be illustrated by a Monte Carlo experiment, and the method will be used in order to study american short-term interest rate.

Key words: Indirect inference, simulations, Euler and Milstein discretizations, short-term interest rate, stochastic differential equation.

JEL Classification: C15, C22, C32, G19

Résumé

Nous proposons une adaptation de la méthode standard d'estimation par inférence indirecte d'une équation différentielle stochastique, dans laquelle la discrétisation standard d'Euler sera remplacée par celle de Milstein. Cette dernière, pour un pas de discrétisation donnée, est plus précise que la première, et induit un gain de temps dans l'étape de calibration. Ce gain sera illustré par une expérience de Monte-Carlo, et la méthode sera utilisée pour la modélisation du taux d'intérêt de court terme américain.

Mots clés: Inférence indirecte, simulations, discrétisations d'Euler et de Milstein, taux d'intérêt de court-terme, équation différentielle stochastique.

1 Introduction

Les équations différentielles stochastiques (SDE) sont utilisées dans de nombreux domaines économiques et notamment en finance où elles offrent un cadre de

1. GREQAM-CNRS, Centre de la Vieille Charité, 2 rue de la Charité, F-13002 Marseille, France. Je remercie MM. Michel Lubrano et Russell Davidson pour leur aide et leurs précieux commentaires.

travail confortable pour la modélisation des taux d'intérêt ou encore pour le pricing d'option en temps continu. Mais toute la difficulté de l'exercice précédent provient moins de l'obtention des formules d'évaluation que de l'estimation du paramètre d'intérêt θ . En effet, alors qu'une SDE (ici scalaire) s'exprime en temps continu sous la forme d'une équation différentielle

$$dY(t) = \mu \{t, Y(t)\} dt + \sigma \{t, Y(t)\} dW(t) \quad (1.1)$$

où le drift μ et la volatilité σ dépendent implicitement de θ , les données disponibles restent quant à elles discrètes, ce qui rend l'estimation du paramètre d'intérêt très délicate. Ainsi, l'estimation par maximum de vraisemblance exact n'est possible que lorsque le processus admet une discrétisation exacte, puisqu'il est alors possible d'exprimer analytiquement les densités transitoires entre deux observations consécutives, et de former la vraisemblance. Mais, la plupart du temps, il est impossible de tirer de (1.1) l'expression exacte de ces densités transitoires si bien que l'estimateur par maximum de vraisemblance exact ne soit pas automatiquement implémentable.

Une première solution pour contourner ce problème consiste à estimer θ sur une version discrétisée de (1.1) par pseudo-maximum de vraisemblance (PML) ou par une méthode des moments, comme dans Chan, Karyoli, Longstaff & Sanders (1992); mais les estimateurs ainsi obtenus souffrent le plus souvent d'un biais de discrétisation, du fait de l'aspect approximatif de la discrétisation utilisée.

La principale alternative au MLE exact est la méthode des moments généralisés (GMM), qui dans le cadre de l'estimation des paramètres des équations de diffusion présente des problèmes d'identification levés par l'utilisation de coûteuses extensions (Gouriéroux & Monfort 1995).

Ces difficultés motivent l'utilisation des méthodes simulées, parmi lesquelles le maximum de vraisemblance simulé (Brandt & Santa-Clara 2001), la méthode des moments efficients (Gallant & Tauchen 1996) ou encore l'inférence indirecte (Gouriéroux, Monfort & Renault 1993). Le principe de base de ces méthodes est d'ajuster le paramètre θ par une procédure de calibration, afin d'obtenir des caractéristiques similaires pour les variables endogènes observées et les données simulées. Nous nous intéressons ici plus particulièrement à l'inférence indirecte qui consiste à minimiser, dans une métrique convenable, la distance entre l'estimateur d'un modèle auxiliaire sur les données observées et l'estimateur de ce même modèle auxiliaire sur des données simulées.

Il est donc primordial d'obtenir des données simulées à partir de (1.1), ce qui là encore n'est pas toujours possible. Pour contourner ce problème, il faut avoir recours à des discrétisations approximatives dont la plus populaire est la discrétisation d'Euler-Maruyama (EM dans la suite). Les discrétisations approximatives se basent toutes sur le principe de *data augmentation*, c'est-à-dire sur l'insertion de données fictives, correspondant à des dates d'observations supplémentaires. Alors, plus le nombre de dates insérées est important, ou de manière équivalente

plus la pas de discrétisation¹ δ est petit, meilleure sera la trajectoire approchée. Nous travaillerons ici avec la famille des discrétisations présentant un caractère de convergence forte, dont fait partie la discrétisation standard précédemment évoquée. Chaque membre de cette famille se caractérise par un certain ordre γ (ou ordre de convergence forte), indiquant à quel point il offre des simulations précises: plus l'ordre de convergence forte est élevé, plus l'approximation obtenue pour un pas δ fixé sera proche de la véritable trajectoire, sous un certain critère restant à définir. Or, la discrétisation EM se trouve être l'un des membres les moins précis de cette famille; Kloeden & Platen (1999) montrent comment il suffit d'ajouter à la discrétisation EM un terme correctif pour obtenir un membre plus précis de cette famille, ou discrétisation de Milstein (1978). Par construction, cette dernière fournit donc des données plus précises que la discrétisation d'Euler, et il semble donc numériquement intéressant de la mettre en œuvre pour estimer θ par inférence indirecte.

L'impact du choix du procédé de discrétisation est discuté dans ce papier qui s'organise comme suit. La section 2 présente plus en détail les procédés de discrétisation d'Euler et de Milstein et en étudie les propriétés. La section 3 propose une adaptation de l'estimation par inférence indirecte de θ . La section 4 fait l'objet d'une application empirique à la modélisation d'un taux court américain et la section 5 offre une conclusion.

2 Les procédés de discrétisation d'Euler et de Milstein

La question de la discrétisation des équations différentielles ordinaires (ODE) est un sujet bien connu, pour lequel de nombreuses méthodes ont été proposées. Pour autant, l'adaptation heuristique au cadre stochastique de ces méthodes n'est pas nécessairement une bonne stratégie, dans la mesure où le calcul différentiel déterministe n'est pas totalement compatible avec le calcul différentiel stochastique, ou calcul d'Ito. Ce dernier permet de voir (1.1) comme une formulation symbolique de

$$Y(t) = Y(t_0) + \int_{t_0}^t \mu \{s, Y(s)\} ds + \int_{t_0}^t \sigma \{s, Y(s)\} dW(s) \quad (2.1)$$

dans laquelle la dernière intégrale possède une espérance nulle et une variance donnée par l'isométrie d'Ito $\int_{t_0}^t E[\sigma \{s, Y(s)\}]^2 ds$.

Sachant la SDE (1.1), et ses observations y_1, \dots, y_T aux dates τ_1, \dots, τ_T , nous nous proposons de montrer ici comment sont construites les discrétisations d'Euler et de Milstein, puis d'exhiber leur précision respective. Comme précisé plus haut, un procédé de discrétisation se définit par un pas de discrétisation δ , et un point de départ coïncidant avec y_1 . Dans toute la suite, nous noterons $N = T/\delta$ le nombre de simulations à effectuer, et par $\tilde{y}_\delta, \tilde{y}_{2\delta}, \dots, \tilde{y}_{N\delta}$ les données simulées, correspondant aux dates $\tilde{\tau}_\delta, \tilde{\tau}_{2\delta}, \dots, \tilde{\tau}_{N\delta}$.

1. Le pas de discrétisation se définit comme le laps de temps séparant deux simulations consécutives

L'outil de base est l'équivalent stochastique de la *chain rule* déterministe, ou lemme d'Ito. Ce lemme permet d'exprimer la SDE suivie par $Z(t) = g\{t, Y(t)\}$, si $Y(t)$ vérifie (1.1) et si la fonction $g\{\cdot, \cdot\}$ est suffisamment régulière:

$$dZ(t) = L_0 g\{t, Y(t)\} dt + L_1 g\{t, Y(t)\} dW(t) \quad (2.2)$$

où les opérateurs L_0 et L_1 sont définis par

$$L_0 g(t, Y(t)) = g'_t + g'_y \mu + \frac{\sigma^2}{2} g''_{yy} \quad (2.3)$$

et par

$$L_1 g\{t, Y(t)\} = \sigma g'_y \quad (2.4)$$

Notons qu'ici, les fonctions g , σ , μ et leurs dérivées sont évaluées en $\{t, Y(t)\}$. Une expression équivalente de (2.2) est obtenue quand on considère son expression intégrale

$$g\{t, Y(t)\} = g\{t_0, Y(t_0)\} + \int_{t_0}^t L_0 g\{s, Y(s)\} ds + \int_{t_0}^t L_1 g\{s, Y(s)\} dW(s) \quad (2.5)$$

2.1 La discrétisation d'Euler-Mayurama

La version discrète de (1.1) la plus simple est le procédé d'Euler-Maryuama, se donnant par

$$\tilde{y}_{(k+1)\delta} = \tilde{y}_{k\delta} + \delta \mu\{k\delta, \tilde{y}_{k\delta}\} + \sqrt{\delta} \sigma\{k\delta, \tilde{y}_{k\delta}\} \varepsilon_{(k+1)\delta} \quad (2.6)$$

où $\sqrt{\delta} \varepsilon_{(k+1)\delta}$, distribué selon une normale centrée et de variance δ , représente l'incrément du mouvement brownien entre deux dates consécutives $\tilde{\tau}_{k\delta}$ et $\tilde{\tau}_{(k+1)\delta}$.

Cette relation récursive s'obtient en appliquant le lemme d'Ito sous forme intégrale aux deux fonctions $\sigma\{s, Y(s)\}$ et $\mu\{s, Y(s)\}$ présentes dans (2.1). On obtient alors un développement d'Ito-Taylor au premier ordre:

$$Y(t) = Y(t_0) + \mu\{t_0, Y(t_0)\} \int_{t_0}^t ds + \sigma\{t_0, Y(t_0)\} \int_{t_0}^t dW(s) + R_1(t_0, t) \quad (2.7)$$

où le reste $R_1(t_0, t)$ se donne par

$$\begin{aligned} R_1(t_0, t) &= \int_{t_0}^t \int_{t_0}^s L_0 \mu\{u, Y(u)\} du ds + \int_{t_0}^t \int_{t_0}^s L_1 \mu\{u, Y(u)\} dW(u) ds \\ &+ \int_{t_0}^t \int_{t_0}^s L_0 \sigma\{u, Y(u)\} du dW(s) \\ &+ \int_{t_0}^t \int_{t_0}^s L_1 \sigma\{u, Y(u)\} dW(u) dW(s) \end{aligned} \quad (2.8)$$

En négligeant le reste, et en posant $t_0 = k\delta$, $t = (k+1)\delta$, on obtient effectivement (2.6), si l'on remarque:

$$\int_{k\delta}^{(k+1)\delta} dt_s = \tilde{\tau}_{(k+1)\delta} - \tilde{\tau}_{k\delta} \equiv \delta \quad (2.9)$$

$$\int_{k\delta}^{(k+1)\delta} dW(s) = W((k+1)\delta) - W(k\delta) \equiv \sqrt{\delta}\varepsilon_{(k+1)\delta} \quad (2.10)$$

L'expression de la densité transitoire induite par la discrétisation d'EM se déduit immédiatement de (2.6), en tenant compte de l'aspect gaussien de $\sqrt{\delta}\varepsilon_{(k+1)\delta}$.

Théorème 2.1. *Etant donné la discrétisation EM de l'équation (1.1), la distribution conditionnelle introduite est normale*

$$\tilde{y}_{(k+1)\delta} | \tilde{y}_{k\delta} \sim N(\tilde{y}_{k\delta} + \delta\mu; \delta\sigma^2) \quad (2.11)$$

où les fonctions μ et σ sont évaluées en $\{k\delta, \tilde{y}_{k\delta}\}$.

La précision d'un procédé numérique peut être définie de nombreuses façons, selon son utilisation. Par exemple, pour approcher des moments, on considérera un *critère de convergence faible*; pour approcher les trajectoires elles-mêmes, comme c'est ici le cas, on utilise plutôt le *critère de convergence forte*. Un procédé de discrétisation présente une convergence forte si l'erreur commise en approchant la dernière donnée observée y_T par la dernière donnée simulée, notée précédemment $\tilde{y}_{N\delta}$, est en moyenne négligeable. Cette famille de discrétisation, ou *strong order schemes* selon Kloeden & Platen (1999), comprend notamment les discrétisations d'Euler et de Milstein. Mathématiquement, la définition de la convergence forte introduit un ordre γ , permettant de distinguer les différents membres:

$$\exists K, \exists \delta_0, \forall \delta \in [0, \delta_0], E |y_T - \tilde{y}_{\delta N}| \leq K\delta^\gamma \quad (2.12)$$

Ainsi, pour un pas de discrétisation δ fixé, plus le coefficient γ sera élevé, plus précise sera l'approximation. Kloeden & Platen (1999) prouvent que sous certaines conditions le procédé d'EM possède un ordre de convergence forte de 0.5, ce qui se trouve être l'ordre le plus faible de la famille considérée.

2.2 La discrétisation de Milstein

Il existe bon nombre de procédés plus précis que celui d'Euler et appartenant à la même famille; le plus immédiat est celui de Milstein (1978), qui s'obtient comme une version corrigée du précédent. Cette correction s'obtient en effectuant un développement d'Ito-Taylor au second ordre, ce qui revient à développer au premier ordre l'un des termes du reste $R_1(t_0, t)$ dont l'expression est donnée par (2.8). Notons qu'il existe plusieurs développements au second ordre, puisque plusieurs termes sont présents. Cela signifie que, contrairement au développement de Taylor déterministe, le développement d'Ito-Taylor n'admet pas d'expression simple d'utilisation. En effet, son expression, que l'on pourra trouver dans le chapitre 5

de Kloeden & Platen (1999), se base sur des notions mathématiques² qu'il serait trop coûteux d'exposer ici. C'est pourquoi, plutôt que de donner une expression peu intuitive, nous nous contenterons dans la suite d'une construction «manuelle» somme toute assez abordable, comme dans Kloeden (2001).

En développant la dernière intégrale du reste, nous obtenons effectivement une version du développement d'Ito-Taylor au second ordre:

$$\begin{aligned} Y(t) = Y(t_0) &+ \mu \{t_0, Y(t_0)\} \int_{t_0}^t ds + \sigma \{t_0, Y(t_0)\} \int_{t_0}^t dW(s) \\ &+ L_1 \sigma \{t_0, Y(t_0)\} \int_{t_0}^t \int_{t_0}^s dW(u) dW(s) + R_2(t_0, t) \end{aligned} \quad (2.13)$$

où le reste d'ordre 2 est:

$$\begin{aligned} R_2(t_0, t) &= \int_{t_0}^t \int_{t_0}^s L_0 \mu \{u, Y(u)\} du ds + \int_{t_0}^t \int_{t_0}^s L_1 \mu \{u, Y(u)\} dW(u) ds \\ &+ \int_{t_0}^t \int_{t_0}^s L_0 \sigma \{u, Y(u)\} du dW(s) \\ &+ \int_{t_0}^t \int_{t_0}^s \int_{t_0}^u L_0 L_1 \sigma \{v, Y(v)\} dv dW(u) dW(s) \\ &+ \int_{t_0}^t \int_{t_0}^s \int_{t_0}^u L_1 L_1 \sigma \{v, Y(v)\} dW(v) dW(u) dW(s) \end{aligned} \quad (2.14)$$

En négligeant le reste, et en posant $t_0 = k\delta$, $t = (k+1)\delta$, on obtient alors:

$$\begin{aligned} \tilde{y}_{(k+1)\delta} &= \tilde{y}_{k\delta} + \mu \{k\delta, \tilde{y}_{k\delta}\} \int_{k\delta}^{(k+1)\delta} dt_s + \sigma \{k\delta, \tilde{y}_{k\delta}\} \int_{k\delta}^{(k+1)\delta} dW(s) \\ &+ \frac{\sigma \{k\delta, \tilde{y}_{k\delta}\} \sigma'_x \{k\delta, \tilde{y}_{k\delta}\}}{2} \left(\left(\int_{k\delta}^{(k+1)\delta} dW(s) \right)^2 - \delta \right) \end{aligned} \quad (2.15)$$

où les relations (2.9) et (2.10) restent valables. En remarquant que

$$\left(\int_{k\delta}^{(k+1)\delta} dW(s) \right)^2 - \delta = [W((k+1)\delta) - W(k\delta)]^2 - \delta \equiv \delta(\varepsilon^2 - 1) \quad (2.16)$$

où $\varepsilon \sim N(0,1)$, nous obtenons l'expression de la discrétisation de Milstein:

$$\begin{aligned} \tilde{y}_{(k+1)\delta} &= \tilde{y}_{k\delta} + \delta \mu \{k\delta, \tilde{y}_{k\delta}\} + \sqrt{\delta} \sigma \{k\delta, \tilde{y}_{k\delta}\} \varepsilon_{(k+1)\delta} \\ &+ \frac{\sigma \{k\delta, \tilde{y}_{k\delta}\} \sigma'_y \{k\delta, \tilde{y}_{k\delta}\}}{2} \delta (\varepsilon_{(k+1)\delta}^2 - 1) \end{aligned} \quad (2.17)$$

La discrétisation de Milstein se déduit finalement de celle d'Euler par l'ajout d'un terme correctif, qui prend la forme

$$\frac{\sigma \{k\delta, \tilde{y}_{k\delta}\} \sigma'_y \{k\delta, \tilde{y}_{k\delta}\}}{2} \delta (\varepsilon_{(k+1)\delta}^2 - 1) \quad (2.18)$$

2. Citons par exemple les multi-indices, les coefficients d'Ito ou encore les intégrales multiples récursives.

Ce terme correctif provient du fait que la discrétisation de Milstein est définie à partir d'un développement d'Ito-Taylor à l'ordre 2 ce qui permet ainsi de mieux tenir compte de la volatilité. Une première conséquence est l'augmentation de l'ordre de convergence forte, qui atteint ici la valeur $\gamma = 1$, contre $\gamma = 0.5$ pour la discrétisation de base.

La présence du terme correctif se traduit aussi par une densité transitoire plus complexe, comme le prouve Elerian (1998).

Théorème 2.2. *Etant donné le changement de variable:*

$$z_{(k+1)\delta} = \frac{\tilde{y}_{(k+1)\delta} - B}{A} \quad (2.19)$$

avec

$$\lambda = \frac{1}{\delta(\sigma')^2} \quad (2.20)$$

$$A = \frac{\sigma\sigma'\delta}{2} \quad (2.21)$$

$$B = -\frac{\sigma}{2\sigma'} + \tilde{y}_{k\delta} + \mu\delta - A \quad (2.22)$$

la distribution conditionnelle de $z_{(k+1)\delta}$ sachant $z_{k\delta}$ est un khi-deux non centré, à un degré de liberté et admettant λ comme paramètre de non-centralité. La distribution conditionnelle de $\tilde{y}_{(k+1)\delta}$ sachant $\tilde{y}_{k\delta}$ admet alors pour expression

$$g\left(\tilde{y}_{(k+1)\delta} | \tilde{y}_{k\delta}\right) = \frac{\exp(-\lambda/2)}{|A| \sqrt{2\pi}} z_{(k+1)\delta}^{-1/2} \exp\left(-\frac{z_{(k+1)\delta}}{2}\right) \cosh\left(\sqrt{\lambda z_{(k+1)\delta}}\right)$$

où les fonctions σ , σ' et μ sont évaluées en $(k\delta, \tilde{y}_{k\delta})$. La fonction $\cosh(x)$ définit le cosinus hyperbolique, avec $\cosh(x) = (e^x + e^{-x})/2$.

Cette densité conditionnelle, contrairement à celle impliquée par le procédé d'EM, n'est pas symétrique. Par ailleurs, le terme correctif (2.18) modifie seulement la variance conditionnelle, par rapport à la discrétisation d'Euler:

$$\text{Var}\left(\tilde{y}_{(k+1)\delta} | \tilde{y}_{k\delta}\right) = \delta\sigma^2 + \frac{(\sigma\sigma'\delta)^2}{2} \quad (2.23)$$

Les autres membres de la famille de discrétisation «forte» s'obtiennent en appliquant le même principe; finalement, procéder à des développements d'Ito-Taylor à des ordres de plus en plus élevés permet d'obtenir des procédés qui se déduisent du procédé d'EM par l'ajout de termes correctifs, de plus en plus nombreux, mettant en œuvre les dérivées du drift et de la volatilité, et présentant des ordres de convergence forte de plus en plus importants. Kloeden & Platen (1999) obtiennent ainsi des ordres γ atteignant 1.5, voire 2.0. Voici par exemple l'expression du procédé atteignant $\gamma = 1.5$, obtenu comme une version corrigée de la discrétisation

de Milstein:

$$\begin{aligned}
\tilde{y}_{(k+1)\delta} &= \tilde{y}_{k\delta} + \delta\mu + \sigma\Delta W + \frac{\delta\sigma\sigma'}{2} [(\Delta W)^2 - 1] + \sigma\mu'\Delta Z \\
&+ \frac{1}{2} \left\{ \mu\mu' + \frac{\sigma^2\mu''}{2} \right\} \delta^2 + \left\{ \mu\sigma' + \frac{\sigma^2\sigma''}{2} \right\} \{\delta\Delta W - \Delta Z\} \quad (2.24) \\
&+ \frac{\sigma}{2} \left\{ \sigma\sigma'' + (\sigma')^2 \right\} \left\{ \frac{(\Delta W)^2}{3} - \delta \right\} \Delta W
\end{aligned}$$

où toutes les fonctions sont évaluées en $(k\delta, \tilde{y}_{k\delta})$, ΔW et ΔZ deux termes gaussiens vérifiant

$$\Delta W \sim N(0, \delta) \quad (2.25)$$

$$\Delta Z \sim N(0, \delta^3/3) \quad (2.26)$$

$$\text{Cov}(\Delta W, \Delta Z) = \delta^2/2 \quad (2.27)$$

On peut vérifier que (2.24) est bien une version corrigée de la discrétisation de Milstein, que les termes correctifs sont plus nombreux que dans (2.17), et qu'ils utilisent des dérivées d'ordre supérieur. Là encore, l'ajout des termes correctifs permet d'augmenter l'ordre de convergence forte de 1/2. Finalement, passer de la discrétisation d'Euler à celle de Milstein permet de gagner un demi point de convergence, en ajoutant un simple terme correctif; par contre passer de la discrétisation de Milstein à la discrétisation donnée par (2.24) présente ici un coût calculatoire beaucoup trop prohibitif. C'est pourquoi, dans la suite, nous nous limiterons à l'utilisation des discrétisations d'Euler et de Milstein.

3 Estimation par inférence indirecte d'une équation de diffusion

3.1 Le principe général

Gouriéroux et al. (1993) [GMR dans la suite] présentent le principe de l'estimation par inférence indirecte dans un cadre général. On considère un modèle initial exprimant la variable endogène y en fonction de son passé, d'une variable exogène x et d'un bruit ε , dont la distribution est connue:

$$y_t = r(y_{t-1}, x_t, \varepsilon_t; \theta), \quad t = 1, \dots, T \quad (3.1)$$

Dans le cas général, la vraisemblance peut s'avérer suffisamment complexe pour que l'estimateur par maximum de vraisemblance exact ne soit pas implémentable. Malgré cela, on suppose être capable de simuler des données issues de (3.1) pour toute valeur de θ . Pour contourner le problème de l'estimation, on peut introduire un modèle auxiliaire

$$y_t = r_a(y_{t-1}, x_t, \varepsilon_t; \kappa) \quad (3.2)$$

qui est un modèle plus simple que le modèle initial (sans être toutefois trop éloigné) et dont l'estimation est relativement aisée. Si le modèle auxiliaire est estimé sur les données observées, on obtiendra un estimateur inconsistant, du fait de la mauvaise spécification du modèle (3.2). L'inférence indirecte permet de réduire ce biais par l'intermédiaire de simulations issues du modèle initial et le mécanisme de réduction du biais mis en œuvre se base sur la comparaison de l'estimation du modèle auxiliaire sur données simulées et observées. Plus explicitement, l'estimateur par inférence indirecte s'obtiendra comme la valeur des paramètres $\hat{\theta}$ minimisant la distance entre l'estimation du modèle auxiliaire sur les données observées et l'estimation de ce même modèle auxiliaire sur des données simulées à partir de (3.1) obtenues pour $\theta = \hat{\theta}$.

En posant $y_1^T = \{y_1, \dots, y_T\}$ et $x_1^T = \{x_1, \dots, x_T\}$, considérons d'abord l'estimation du modèle auxiliaire sur les données observées

$$\hat{\kappa}_0 = \arg \max_{\kappa} Q_a(y_1^T, x_1^T; \kappa) \quad (3.3)$$

où $Q_a(y_1^T, x_1^T; \kappa)$ est un critère d'estimation du modèle auxiliaire, comme par exemple sa vraisemblance.

Pour une certaine valeur de θ , on simule ensuite des données issues du modèle initial, notées $\tilde{y}(\theta)_1^T$. L'estimation du modèle auxiliaire sur données simulées se donne alors par

$$\hat{\kappa}_T(\theta) = \arg \max_{\kappa} Q_a(\tilde{y}(\theta)_1^T, x_1^T; \kappa) \quad (3.4)$$

Enfin, l'estimateur de θ par inférence indirecte, $\hat{\theta}_T(\Omega)$, s'obtient par calibration de telle sorte que la distance entre $\hat{\kappa}_T(\theta)$ et $\hat{\kappa}_0$ soit minimale:

$$\hat{\theta}_T(\Omega) = \arg \min_{\theta} (\hat{\kappa}_T(\theta) - \hat{\kappa}_0)^\top \Omega (\hat{\kappa}_T(\theta) - \hat{\kappa}_0) \quad (3.5)$$

Notons enfin que pour être applicable, cette méthode impose que le modèle auxiliaire soit décrit par au moins autant de paramètres que le modèle initial, c'est-à-dire que $\dim \kappa \geq \dim \theta$.

La matrice Ω , symétrique et non-négative, permet de définir la métrique. GMR établissent que l'estimateur est consistant, asymptotiquement normal et donnent l'expression de la métrique optimale Ω^* , pour laquelle la matrice de variance-covariance de l'estimateur est minimale. Par ailleurs, lorsque κ et θ ont même dimension, la métrique optimale peut être choisie comme étant la métrique euclidienne conventionnelle, c'est-à-dire $\Omega^* = \mathbf{Id}$.

Afin d'améliorer la performance de la méthode, GMR considèrent plusieurs simulations du modèle initial pour une même valeur de θ . Si S représente le nombre d'échantillons, on sera amené à simuler la matrice $\tilde{y}(\theta)_1^{T(S)}$, où pour tout $s \in \{1, \dots, S\}$, le s ème échantillon $\tilde{y}(\theta)_1^{T(s)}$ s'obtient de la même manière que précédemment. L'estimateur (3.4) devient alors

$$\hat{\kappa}_T(\theta)^S = \arg \max_{\kappa} \sum_{s=1}^S Q_a(\tilde{y}(\theta)_1^{T(s)}, x_1^T; \kappa) \quad (3.6)$$

et l'estimateur par inférence indirecte s'obtient toujours par calibration, comme exposé dans (3.5), mais $\hat{\kappa}_T(\theta)$ sera remplacé par $\hat{\kappa}_T(\theta)^S$. On s'attend à augmenter la précision de l'estimation en augmentant le nombre de simulations S ; l'expérimentateur devra donc fixer la valeur de S en arbitrant entre temps de calcul et précision.

Enfin, le choix du modèle auxiliaire influence très fortement la précision de l'estimateur. Par exemple, lorsque GMR estime par inférence indirecte un modèle MA(1) en prenant pour modèle auxiliaire un modèle AR(p), et compare ces estimateurs à celui obtenu par maximum de vraisemblance, il apparaît que pour $p = 1$, la perte d'efficacité est assez importante. Au contraire, le choix $p = 2$, et $p = 3$, permet à l'estimateur par inférence indirecte d'atteindre quasiment l'efficacité de l'estimateur par maximum de vraisemblance. Par conséquent, il conviendra de choisir très soigneusement un bon modèle auxiliaire.

3.2 Application à l'estimation des SDE

Broze, Scaillet & Zakoian (1998) ont été les premiers à estimer les équations de diffusion par inférence indirecte et ont souligné que plusieurs aménagements étaient nécessaires.

Le premier aménagement est consécutif au fait qu'il n'est généralement pas possible de simuler à partir du modèle continu (1.1). Pour contourner cette difficulté, on peut toujours avoir recours à un procédé de discrétisation (Broze et al. utilisent par exemple une discrétisation d'Euler) mais, pour fournir des simulations convenables, ce procédé nécessite un pas de discrétisation assez fin et par conséquent un nombre important de calculs.

L'autre problème posé par l'estimation indirecte des SDE est le choix du modèle auxiliaire. Comme explicité plus haut, ce dernier peut être vu comme une version approximative du modèle initial, c'est pourquoi la version discrétisée de (1.1) au pas $\delta = 1$ est conventionnellement choisie comme modèle auxiliaire. En conservant le cadre de travail proposé par Broze et al. (1998), la discrétisation d'Euler conduit ainsi au modèle auxiliaire

$$y_{t+1} = y_t + \mu \{t, y_t; \kappa\} + \sigma \{t, y_t; \kappa\} \varepsilon_{t+1} \quad (3.7)$$

avec $\varepsilon \sim N(0; 1)$. L'estimation des SDE par inférence indirecte utilisant la discrétisation d'Euler se caractérise donc par une double utilisation de cette discrétisation, et se déroule en plusieurs étapes:

Étape 1: L'estimateur naïf $\hat{\kappa}_0$, ou estimateur de première étape, s'obtient comme l'estimateur du modèle auxiliaire sur les données observées. Le critère Q_a retenu pour mener l'optimisation présentée en (3.3) est généralement la log vraisemblance conditionnelle aux données initiales du modèle auxiliaire. Le critère s'exprime ainsi:

$$Q_a(y_1^T; \kappa) = \sum_{t=1}^{T-1} \log g(y_{t+1} | y_t; \kappa) \quad (3.8)$$

où $g(\cdot; \kappa)$ est la densité transitoire induite par le modèle auxiliaire. Ici, le théorème 2.1 indique que $y_{t+1}|y_t$ est normalement distribué, d'espérance $y_t + \mu \{t, y_t; \kappa\}$ et de variance $\sigma^2 \{t, y_t; \kappa\}$.

Étape 2: A un θ donné, les simulations $\tilde{y}(\theta)_1^{T(S)}$ sont tirées d'une discrétisation du modèle initial à un pas $\delta < 1$. Plus exactement, le s ème échantillon s'obtient par la relation récursive:

$$\begin{aligned} \tilde{y}(\theta)_{(k+1)\delta, (s)} = \tilde{y}(\theta)_{k\delta, (s)} &+ \delta\mu \left\{ k\delta, \tilde{y}(\theta)_{k\delta, (s)} \right\} \\ &+ \sqrt{\delta}\sigma \left\{ k\delta, \tilde{y}(\theta)_{k\delta, (s)} \right\} \varepsilon_{(k+1)\delta, (s)} \end{aligned} \quad (3.9)$$

avec $k = 1, \dots, T-1$ et avec $\varepsilon_{(k+1)\delta, (s)} \sim N(0,1)$. Puis $\hat{\kappa}_T(\theta)^S$ est estimé sur ces données simulées, en utilisant de nouveau le critère explicité dans l'étape 1.

Étape 3: On répète l'étape 2 jusqu'à obtenir un vecteur θ associé à un estimateur $\hat{\kappa}_T(\theta)^S$ aussi proche que possible de l'estimateur naïf $\hat{\kappa}_0$; ce vecteur est l'estimateur par inférence indirecte de θ . Puisque les modèles initial et auxiliaire présentent le même nombre de paramètres, nous prendrons comme métrique optimale la métrique euclidienne. L'étape de calibration de θ deviendra donc:

$$\hat{\theta}_T^S(\Omega) = \arg \min_{\theta} \left(\hat{\kappa}_T(\theta)^S - \hat{\kappa}_0 \right)^\top \left(\hat{\kappa}_T(\theta)^S - \hat{\kappa}_0 \right) \quad (3.10)$$

La minimisation de la forme quadratique précédente revient finalement à résoudre en θ l'équation suivante:

$$\hat{\kappa}_T(\theta)^S = \hat{\kappa}_0$$

Il convient de préciser que la matrice $[\varepsilon_{(1)}, \dots, \varepsilon_{(S)}]$ des aléas normaux utilisés lors de l'étape 2 reste la même lors de l'étape de calibration, puisque la distance considérée est minimisée sur θ .

Du fait de l'approximation du modèle initial (1.1) par un modèle discrétisé à un pas δ , Broze, et al. montrent que l'estimateur par inférence indirecte présente ici un biais asymptotique dépendant de δ . Ce dernier est appelé à décroître lorsque δ diminue, puisque les trajectoires approximatives deviennent dès lors très proches de la trajectoire réelle.

3.3 Estimation par inférence indirecte d'une SDE par le procédé de Milstein

L'estimation indirecte des SDE utilise le procédé standard de discrétisation d'Euler afin d'obtenir des données simulées issues du modèle initial (1.1). Pour produire des données suffisamment précises, la discrétisation standard nécessite un pas de discrétisation δ assez petit, ce qui augmente considérablement le temps

de calcul. Or, la discrétisation de Milstein présente un ordre de convergence forte plus élevé que celle d'Euler, ce qui signifie que pour un δ fixé, la discrétisation de Milstein fournit des données simulées plus précises que celles obtenues avec le procédé d'Euler standard. Autrement dit, à δ fixé, il existe un gain d'efficacité résultant de l'utilisation de la discrétisation de Milstein. La propriété précédente peut se reformuler d'une autre façon: pour atteindre un certain niveau de précision, la discrétisation de Milstein nécessite un pas δ moins faible (et par conséquent un temps de calcul moindre) que celui qu'il faudrait utiliser avec le procédé d'Euler.

Puisqu'il est numériquement plus intéressant d'utiliser la discrétisation de Milstein à la place de celle d'Euler, nous nous proposons d'adapter en conséquence la méthode standard d'inférence indirecte de Broze et al. (1998).

Plusieurs modifications sont donc à apporter à l'algorithme présenté précédemment. La première concerne le procédé de discrétisation utilisé lors de l'étape de simulation. La formule récursive (3.9) devra être corrigée par l'ajout d'un terme supplémentaire, devenant ainsi

$$\begin{aligned}\tilde{y}(\boldsymbol{\theta})_{(k+1)\delta,(s)} &= \tilde{y}(\boldsymbol{\theta})_{k\delta,(s)} + \delta\mu \left\{ k\delta, \tilde{y}(\boldsymbol{\theta})_{k\delta,(s)} \right\} \\ &+ \sqrt{\delta}\sigma \left\{ k\delta, \tilde{y}(\boldsymbol{\theta})_{k\delta,(s)} \right\} \varepsilon_{(k+1)\delta,(s)} \\ &+ \frac{\delta}{2}\sigma \left\{ k\delta, \tilde{y}(\boldsymbol{\theta})_{k\delta,(s)} \right\} \sigma' \left\{ k\delta, \tilde{y}(\boldsymbol{\theta})_{k\delta,(s)} \right\} \left[\varepsilon_{(k+1)\delta,(s)}^2 - 1 \right]\end{aligned}$$

Compte tenu de cette correction, le modèle auxiliaire (3.7) s'écrit désormais:

$$\begin{aligned}y_{t+1} = y_t &+ \mu \{t, y_t; \boldsymbol{\kappa}\} + \sigma \{t, y_t; \boldsymbol{\kappa}\} \varepsilon_{t+1} \\ &+ \frac{1}{2}\sigma \{t, y_t; \boldsymbol{\kappa}\} \sigma' \{t, y_t; \boldsymbol{\kappa}\} \left[\varepsilon_{t+1}^2 - 1 \right]\end{aligned}\quad (3.11)$$

où, comme d'habitude, $\varepsilon \sim N(0; 1)$. Mais, l'estimation de (3.11) par maximum de vraisemblance s'avère assez délicate, car, d'après le théorème 2.2, ce modèle auxiliaire induit une densité transitoire $g(y_{t+1}|y_t; \boldsymbol{\kappa})$ distribuée selon un khi-deux non-centré. Pour y remédier, on peut estimer (3.11) par pseudo-maximum de vraisemblance, du moment où la pseudo-densité de ce modèle auxiliaire appartient à la famille exponentielle. Par simplicité, nous supposons que

$$y_{t+1}|y_t, \boldsymbol{\kappa} \sim N\left(\tilde{\mu}(y_t, \boldsymbol{\kappa}), \tilde{\sigma}^2(y_t, \boldsymbol{\kappa})\right)$$

où $\tilde{\mu}(y_t, \boldsymbol{\kappa})$ et $\tilde{\sigma}^2(y_t, \boldsymbol{\kappa})$ sont les deux premiers moments de $y_{t+1}|y_t$ induits par (3.11):

$$\tilde{\mu}(y_t, \boldsymbol{\kappa}) = y_t + \mu \{t, y_t; \boldsymbol{\kappa}\} \quad (3.12)$$

$$\tilde{\sigma}^2(y_t, \boldsymbol{\kappa}) = \sigma^2 \{t, y_t; \boldsymbol{\kappa}\} + \frac{1}{2} \left[\sigma \{t, y_t; \boldsymbol{\kappa}\} \sigma' \{t, y_t; \boldsymbol{\kappa}\} \right]^2 \quad (3.13)$$

Finalement, les estimateurs naïfs des étapes 1 et 2 se donneront par la maximisation du critère:

$$Q_a(y_1^T; \boldsymbol{\kappa}) = \sum_{t=1}^{T-1} \log \phi \left(y_t; \tilde{\mu}(y_t, \boldsymbol{\kappa}); \tilde{\sigma}^2(y_t, \boldsymbol{\kappa}) \right) \quad (3.14)$$

où $\phi(\mu; \sigma^2)$ représente la pdf d'une normale d'espérance μ et de variance σ^2 . Remarquons que (3.14) possède la même structure gaussienne que (3.8), mais que seule l'expression de la variance change d'une expression à l'autre du fait de l'utilisation de la discrétisation de Milstein.

Afin d'illustrer la mise en œuvre de l'estimation indirecte d'une SDE se basant sur une discrétisation de Milstein, considérons une série d'expériences de Monte-Carlo. Afin de pouvoir comparer les efficacités respectives des deux types d'estimateurs indirects, il faut utiliser une SDE admettant une discrétisation exacte, ce qui nous permettra d'obtenir l'estimateur par maximum de vraisemblance en guise de référence. Du fait de l'utilisation de la discrétisation de Milstein, il convient de choisir en outre un procédé dont la volatilité ne soit pas constante. Nous avons choisi de travailler avec le mouvement brownien géométrique:

$$dy_t = \mu y_t dt + \sigma y_t dW_t$$

qui admet pour discrétisation exacte:

$$y_t = y_0 \exp \left[\left(\mu - \frac{\sigma^2}{2} \right) t + \sigma W_t \right]$$

Considérons alors 500 estimateurs de $\theta_0 = (\mu_0, \sigma_0) = (.125; .5)$ obtenus par maximum de vraisemblance et par inférence indirecte (avec $\delta = 1/10$ et $S = 3$) à partir d'échantillons constitués de $T = 500$ observations.

	Moyenne	Biais	Ecart-type	RMSE
<i>MLE</i>	0.50020884	0.00020884	0.01453333	0.01452029
<i>Naïf Euler</i>	0.60452382	0.10452382	0.03752444	0.11104277
<i>Naïf Milstein</i>	0.56202406	0.06202406	0.03058242	0.06914042

TAB. 1 – *Estimateurs naïfs*

Le tableau 1 présente quelques statistiques descriptives concernant les estimateurs naïfs du paramètre σ . Ces derniers sont fortement biaisés, et évidemment bien moins efficaces que l'estimateur par maximum de vraisemblance exact. En comparant les estimateurs indirects entre eux, on constate que la discrétisation de Milstein permet non seulement de réduire le biais mais aussi d'augmenter la précision des résultats (donnés par le RMSE), en comparaison de ceux obtenus à partir de la méthode standard.

	Moyenne	Biais	Ecart-type	RMSE
<i>MLE</i>	0.50020884	0.00020884	0.01453333	0.01452029
<i>Naïf Euler</i>	0.60452382	0.10452382	0.03752444	0.11104277
<i>Indirect Euler</i>	0.51078106	0.01078106	0.02579044	0.02792934
<i>Naïf Milstein</i>	0.56202406	0.06202406	0.03058242	0.06914042
<i>Indirect Milstein</i>	0.50815012	0.00815012	0.02628465	0.02749410

TAB. 2 – *Estimations indirectes*

Considérons maintenant les estimateurs indirects de θ_0 . Comme le montre le tableau 2, l'inférence indirecte réduit spectaculairement le biais de l'estimateur naïf et améliore par ailleurs la précision des résultats, quelle que soit la discrétisation utilisée; la figure 1 en donne une illustration plus parlante.

La comparaison des estimateurs indirects entre eux montre que pour un pas de discrétisation assez grand ($\delta = 1/10$), la discrétisation de Milstein réduit là encore le biais et augmente marginalement la précision des résultats par rapport à la méthode standard (cf. figure 2).

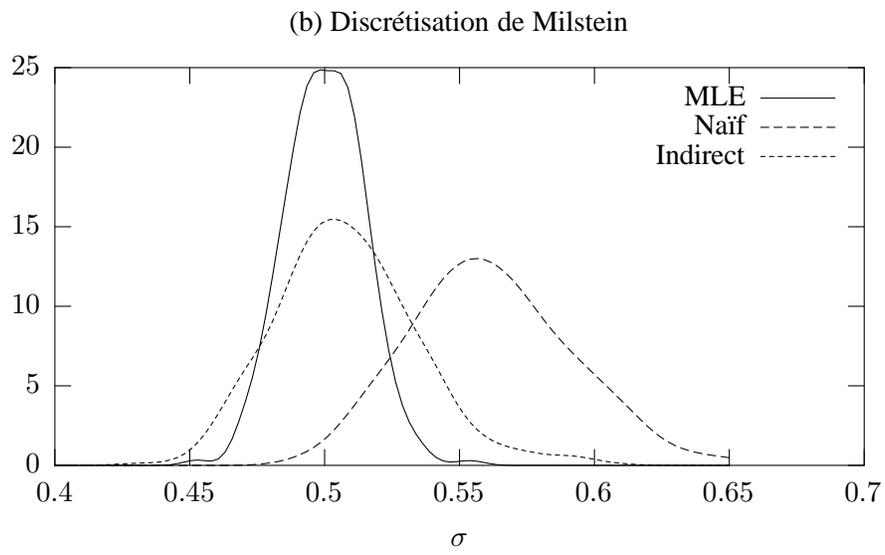
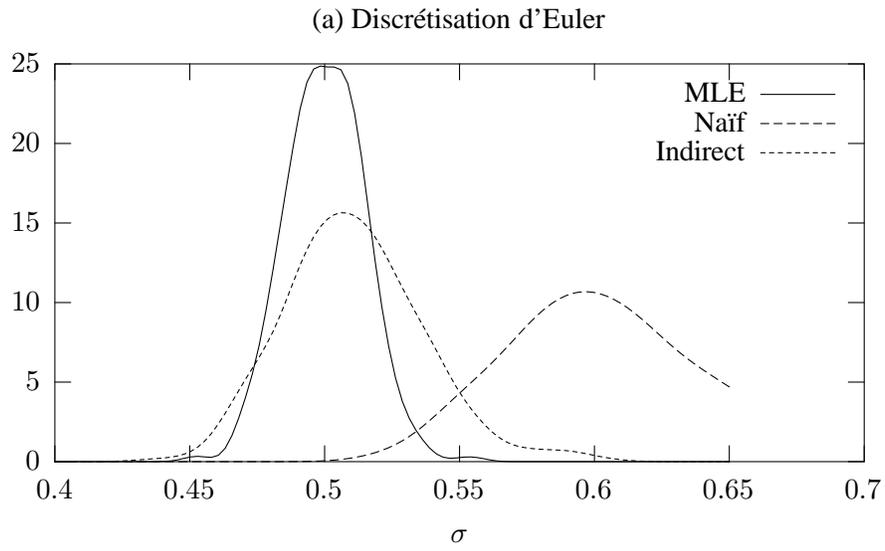


FIG. 1 – Comparaison des distributions des estimateurs de σ .

4 Modélisation empirique

La section précédente a montré comment l'utilisation du procédé de discrétisation de Milstein conduisait à des estimateurs indirects plus précis que ceux basés sur la discrétisation conventionnelle d'Euler. Ces résultats théoriques seront mis à contribution afin de modéliser la structure du taux sans risque du marché obligataire américain, vérifiant maintenant une SDE de la forme:

$$dr(t) = \mu(t, r(t), \theta)dt + \sigma(t, r(t), \theta)dW(t) \quad (4.1)$$

On considère à cet effet comme proxy de l'actif sans risque les données utilisées précédemment dans Lubrano (2000), se présentant sous la forme de 411 observations mensuelles du bon du Trésor américain à 3 mois, s'étendant de janvier 1965 à mars 1999, et obtenues sur la page web de la Banque de réserve fédérale de Saint-Louis³.

Sur la figure 3, on remarque deux principales périodes pour l'actif sans risque: une période d'exception assez courte, se situant au début des années quatre-vingt (1979-1982), marquée par des valeurs très élevées et volatiles, et qui correspond à un changement de politique monétaire de la réserve fédérale. La période de défaut se caractérise quant à elle par une durée plus longue, et surtout par des valeurs bien moins élevées. La période d'exception a donc été rapidement résorbée, ce qui milite en faveur d'une spécification *mean-reverting* du drift. Cette dernière spécification induit une force de rappel vers la moyenne de long terme, s'exerçant aussi bien pour des faibles que pour des fortes valeurs du taux.

Même si une étude préalable des données peut fournir de précieux renseignements sur les caractéristiques des fonctions de drift et de volatilité, l'étape de spécification paramétrique reste toujours délicate. La difficulté provient moins de la paramétrisation elle-même que de la capacité de ces spécifications de reproduire les caractéristiques des données. Par exemple, nous avons vu précédemment combien une spécification *mean-reverting* du drift était adaptée aux données utilisées; si le retour à la moyenne induite par le drift est modélisée de manière linéaire, alors on peut considérer une famille de spécifications paramétriques de (1.1) proposée par Chan et al. (1992), où la volatilité présente une élasticité constante⁴:

$$dr(t) = \beta(\alpha - r(t))dt + \sigma r(t)^\gamma dW(t) \quad (4.2)$$

$\alpha > 0$ représente alors la moyenne de long terme, et $\beta > 0$ l'intensité de la force de rappel vers cette moyenne de long terme. Mais la spécification *CEV*, qui permet de retrouver les modélisations classiques présentant un retour à la moyenne linéaire⁵ se trouve être assez peu pertinente, à en croire le test de spécification développé par Ait-Sahalia (1996).

3. <http://research.stlouisfed.org/fred2/series/tb3m>

4. Spécification *constant elasticity volatility* ou encore *CEV*.

5. En posant $\alpha = 0$ et $\gamma = 1$, on obtient le GBM, en posant $\gamma = 0$ on retrouve le modèle OU et le modèle CIR se donne par $\gamma = 1/2$.

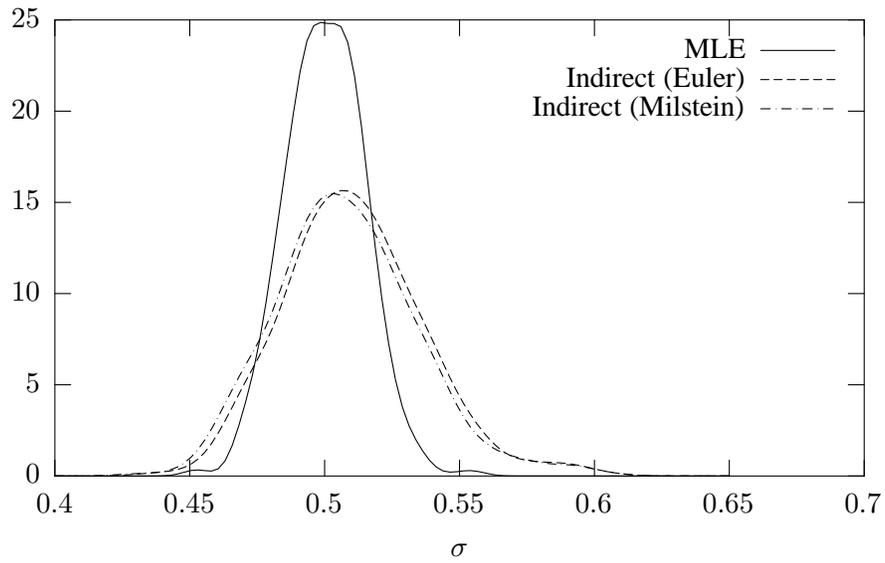


FIG. 2 – Comparaison des distributions des estimateurs indirects de σ .

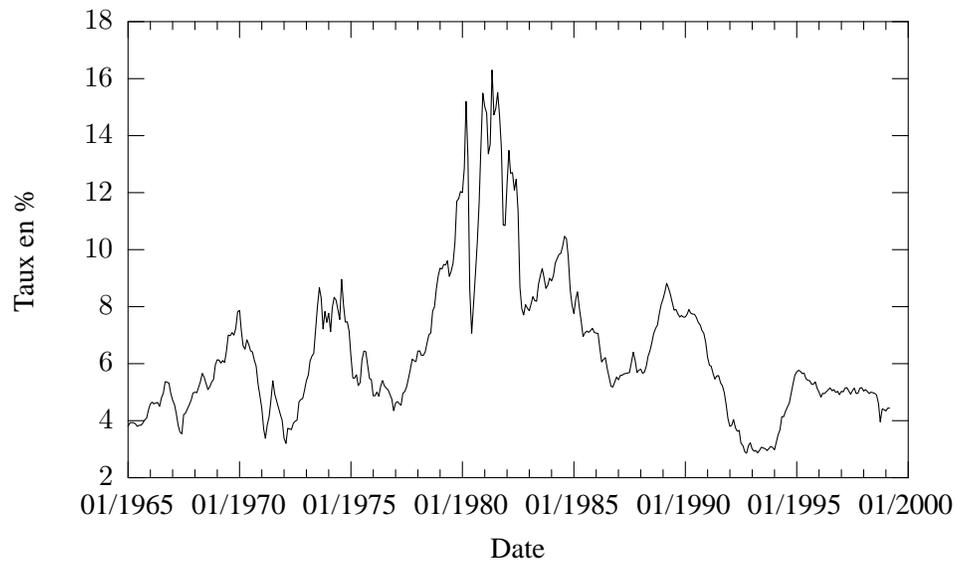


FIG. 3 – T-Bill à trois mois sur le marché secondaire (observations mensuelles).

Pour contourner le problème de sélection *a priori* d'une modélisation paramétrique posé notamment par l'exemple précédent, nous reproduisons la démarche utilisée dans Lubrano (2000). Cette dernière consiste à esquisser, à l'aide de l'outil non-paramétrique, les formes du drift et de la volatilité puis à rechercher ensuite une spécification paramétrique de ces deux fonctions aussi concordante que possible. En tant que méthode indicatrice, l'étape préliminaire n'a pas besoin d'être extrêmement précise, et peut se baser par conséquent sur une grossière approximation du processus continu. Cette approximation est obtenue à partir d'une discrétisation d'Euler à un pas unitaire de (4.1), que l'on écrira:

$$\Delta r_t = \mu(r_{t-1}) + \sigma(r_{t-1})\varepsilon_t \quad (4.3)$$

En tant qu'espérance de Δr_t conditionnellement à ε_t , la fonction $\mu(\cdot)$ peut être estimée par l'estimateur de Nadaraya-Watson

$$\hat{\mu}_h(r) = \frac{\sum_t K[(r_{t-1} - r)/h] \Delta r_t}{\sum_t K[(r_{t-1} - r)/h]} \quad (4.4)$$

où K est un noyau et h le paramètre de lissage convenablement choisi. La diffusion $\sigma^2(\cdot)$ s'estime par la même méthode:

$$\hat{\sigma}_h^2(r) = \frac{\sum_t K[(r_{t-1} - r)/h] (\Delta r_t)^2}{\sum_t K[(r_{t-1} - r)/h]} - \hat{\mu}_h^2(r) \quad (4.5)$$

Le choix de K est connu pour être sans grande importance pour les résultats (nous prendrons par défaut le noyau gaussien) mais la détermination de h est une étape cruciale et délicate. Il existe un grand nombre de méthodes plus ou moins sophistiquées et gourmandes en temps de calcul pour déterminer h (minimisation d'un critère d'erreur quadratique moyenne, emploi d'une fonction de pénalité, *cross validation*, etc.), mais la plus simple reste la détermination graphique, qui sera utilisée ici. Enfin, il est bon de noter que l'estimateur non-paramétrique approximatif (4.5) se trouve être numériquement proche de celui proposé par Florens-Zmirou (1993):

$$\hat{S}_h^2(r) = \frac{1}{\delta} \frac{\sum_t K[(r_{t-1} - r)/h] (\Delta r_t)^2}{\sum_t K[(r_{t-1} - r)/h]} \quad (4.6)$$

Les estimateurs non-paramétriques approximatif et exact se distinguent l'un de l'autre par les termes $-\hat{\mu}_h^2(r)$ et $1/\delta$. Sous certaines conditions de régularité, (4.6) est un estimateur consistant de la volatilité et dont la distribution asymptotique est normale. Par construction, ces deux estimateurs présentent des résultats numériquement voisins, comme l'illustre la figure 4. Les différences s'accroissent pour les valeurs élevées du taux d'intérêt, puisque le drift intervenant dans l'expression de l'estimateur approximatif (4.5) prend alors de fortes valeurs; ainsi, jusqu'à 15%, l'écart quadratique intégré entre les deux estimateurs s'élève à 0.0374, contre environ dix fois plus si on augmente la borne supérieure à 16%.

Les résultats de l'estimation non-paramétrique sont représentés sur la figure 5 et sont conformes aux investigations empiriques menées par Ait-Sahalia (1996):

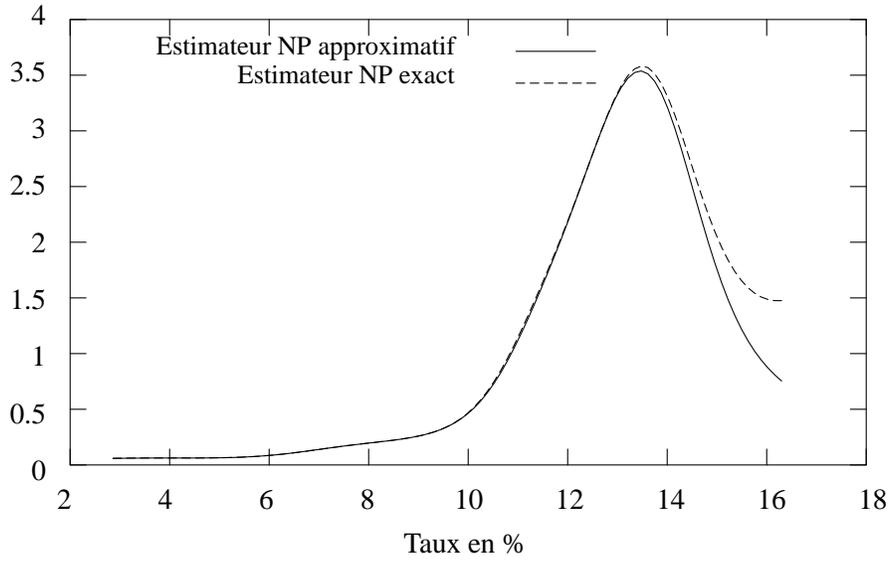


FIG. 4 – Estimateurs non-paramétriques exact et approximatif de la fonction scédastique.

la dérive est en effet constante et presque nulle sur un large intervalle de valeurs, disons jusqu'à 13%. Le drift présente aussi une propriété de force de rappel vers la moyenne, plus intense pour les fortes que pour les faibles valeurs du taux. Ces éléments ne correspondent pas à l'aspect uniformément linéaire de la dérive du modèle CEV et motivent l'utilisation d'un drift non-linéaire du type $\mu(r) = \alpha_0 + \alpha_1/r + \alpha_2 r + \alpha_3 r^2$. L'existence d'une force de rappel pour les valeurs extrêmes de r permet de signer certains coefficients: $\alpha_2 r + \alpha_3 r^2$ doit devenir suffisamment négatif pour des taux devenus trop grands, $\alpha_1 > 0$ assure le retour vers la moyenne pour des taux trop proches de zéro.

La volatilité $\sigma^2(\cdot)$ est une fonction tout d'abord croissante puis décroissante; ce comportement est typiquement incompatible avec la spécification (4.2) dans laquelle la fonction scédastique CEV ne peut être que monotone. Comme le maximum de la volatilité correspond assez bien au seuil au delà duquel le drift devient négatif, on peut concevoir une modélisation à deux régimes, où chaque régime se caractérisait par une formulation donnée des fonctions d'intérêt; ainsi, le premier régime se donnerait par une dérive quasi-nulle et une volatilité strictement croissante et le second régime par une dérive linéaire et une volatilité toutes les deux décroissantes. Nous préférons à cette formulation à changement de régime une spécification paramétrique unique, où pour tenir compte de la forme en cloche de la volatilité, cette dernière est modélisée par une *Cauchy volatility* (CAV):

$$\sigma^2(r) = \frac{\sigma^2}{1 + \gamma(r - c)^2}$$

Cette spécification permet de retrouver une formulation à changement de régime:

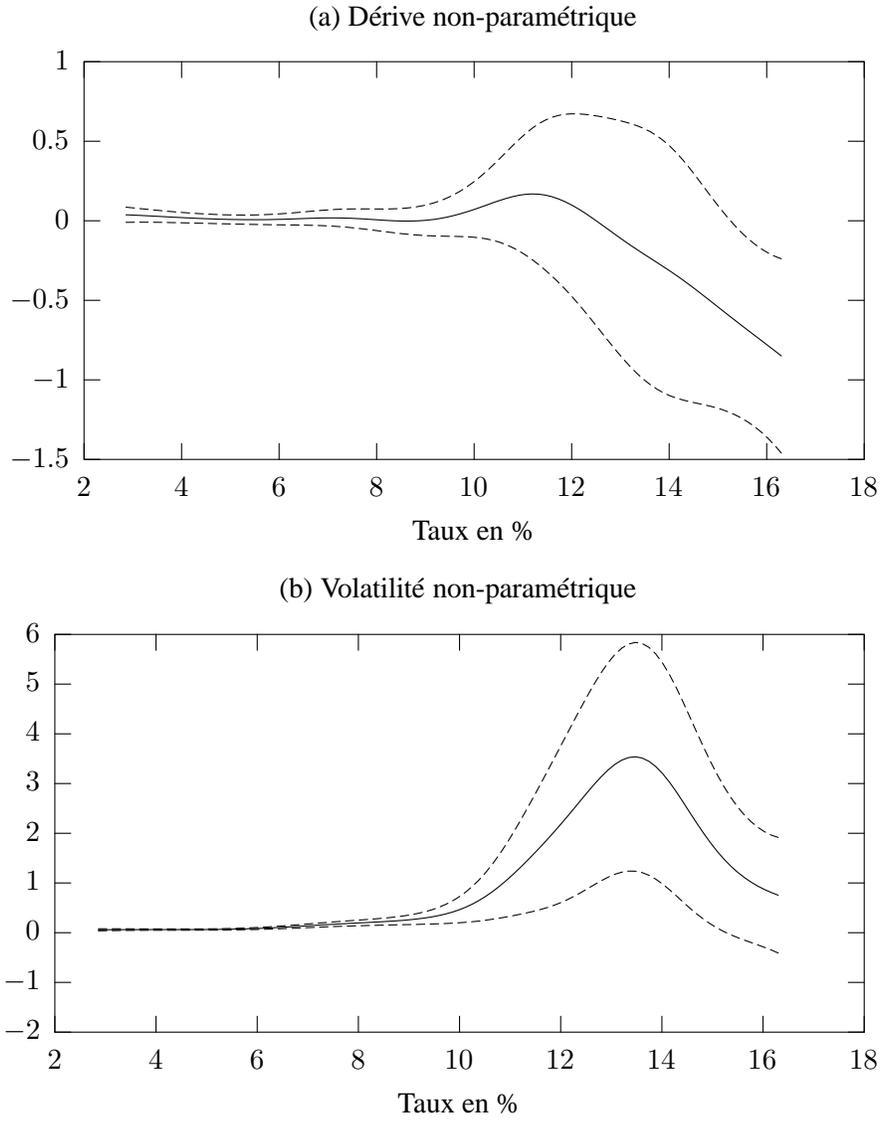


FIG. 5 – Fonctions drift et volatilité estimées non-paramétriquement (en trait plein) et les bornes de leurs intervalles de confiance (en pointillé).

la valeur c pour laquelle la volatilité devient maximale peut être vu comme le seuil, γ comme un paramètre mesurant la vitesse de transition entre les régimes de hautes et de faibles valeurs de taux.

Nous considérons finalement la spécification paramétrique suivante

$$dr(t) = \left[\alpha_0 + \frac{\alpha_1}{r} + \alpha_2 r + \alpha_3 r^2 \right] dt + \left[\frac{\sigma^2}{1 + \gamma (r - c)^2} \right]^{1/2} dW(t)$$

que nous nous proposons d'estimer par inférence indirecte, avec $\delta = 1/100$ et $S = 5$ réplifications.

Les résultats sont exposés dans le tableau 3, où on trouvera aussi les estimateurs naïfs basés sur les discrétisation de Milstein et d'Euler.

Les coefficients du drift sont en adéquation avec les intuitions tirées de l'analyse non-paramétrique; α_1 est bien positif, et $\alpha_2 r + \alpha_3 r^2$ est bien négatif pour de fortes valeurs. L'utilisation du procédé de Milstein ne modifie quasiment pas les estimations, comme on peut aussi le constater sur la figure 6.

Paramètres	Discrétisation d'Euler		Discrétisation de Milstein	
	Naïf	Indirect	Naïf	Indirect
α_0	-1,438719	-1,435348	-1,458275	-1,438492
α_1	2,624409	2,591287	2,650431	2,658226
α_2	0,256072	0,258967	0,260649	0,261900
α_3	-0,014374	-0,014357	-0,014709	-0,014538
σ^2	5,375238	5,840059	4,205558	4,346665
γ	1,195303	1,457712	0,977395	0,987439
c	12,945031	12,828262	12,788651	12,999055

TAB. 3 – Estimateurs naïfs et indirects ($S = 5$ réplifications et $\delta = 1/100$).

Les résultats sont beaucoup plus contrastés concernant la fonction scédastique puisque la discrétisation de Milstein corrige sensiblement les coefficients. Sur la figure 3, la correction due à la discrétisation de Milstein est matérialisée par un déplacement la courbe de la fonction scédastique estimée vers le sud-est; ceci correspond à une augmentation de c et à une diminution de σ^2 par rapport à l'estimation se basant sur la discrétisation d'Euler.

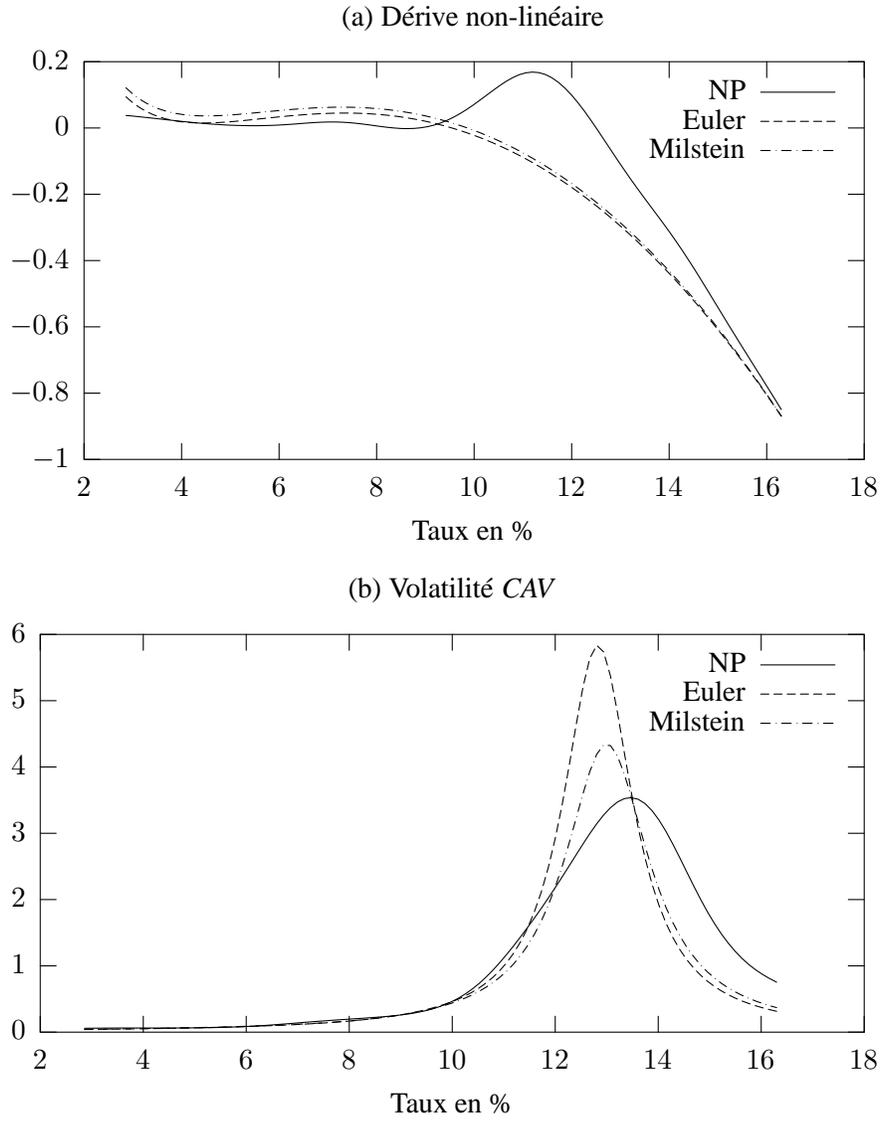


FIG. 6 – *Drift et diffusion estimés paramétriquement par inférence indirecte comparés aux estimateurs non-paramétriques.*

5 Conclusion

Dans ce papier nous avons proposé une adaptation de la méthode d'estimation par inférence indirecte d'une SDE, où la discrétisation de Milstein est utilisée en lieu et place de la discrétisation standard d'Euler.

Cette adaptation est motivée par le fait que la discrétisation de Milstein présente un ordre de convergence forte plus élevé que la discrétisation d'Euler, ce qui signifie qu'il existe un gain de temps de calcul à utiliser la première discrétisation lors des étapes de simulation. D'un point de vue pratique, la discrétisation de Milstein est une version corrigée de la discrétisation standard, s'obtenant par l'ajout d'un unique terme utilisant la dérivée première de la volatilité $\sigma(\cdot, \cdot)$. En ce qui concerne l'estimateur naïf, le modèle auxiliaire s'estime soit par maximum de vraisemblance exact, ce qui implique la manipulation d'un khi-deux non centré, soit par pseudo-maximum de vraisemblance. Dans ce dernier cas, la densité auxiliaire peut être spécifiée comme normale, et paramétrée par les deux premiers moments induits par le modèle auxiliaire.

Une extension potentielle consisterait à remplacer la discrétisation de Milstein par une discrétisation présentant un ordre de convergence forte encore plus élevé; Kloeden & Platen (1999) présentent par exemple les deux successeurs de la discrétisation de Milstein dans la famille des discrétisations "fortes", présentant des ordres respectifs de $3/2$ et 2 . Selon le principe présenté dans la section 2, ces procédés se déduisent de la discrétisation d'Euler par l'ajout de termes correctifs, de plus en plus nombreux, qui prennent en compte les dérivées successives de $\mu(\cdot, \cdot)$ et de $\sigma(\cdot, \cdot)$. Il ne sera alors plus possible de calculer la densité transitoire exacte induite par le modèle auxiliaire, et ce dernier devra par conséquent être estimé par pseudo-maximum de vraisemblance. En reproduisant notre approche, le PMLE pourrait se baser sur une pseudo-densité normale, paramétrée par les moments $E(y_{t+1}|y_t, \kappa)$ et $\text{Var}(y_{t+1}|y_t, \kappa)$.

Références

- Ait-Sahalia, Y. (1996), "Testing continuous-time models of the spot interest rate", *The review of financial studies* **9**(2), 385–426.
- Brandt, M. W. & Santa-Clara, P. (2001), "Simulated likelihood estimation of diffusions with an application to exchange rate dynamics in incomplete markets", Technical Report 274, National bureau for economic research (NBER).
- Broze, L., Scaillet, O. & Zakoian, J.-M. (1998), "Quasi-indirect inference for diffusion processes", Technical Report 9505, Center for operations research and econometrics - Université catholique de Louvain.
- Chan, K. C., Karyoli, A. G., Longstaff, F. A. & Sanders, A. B. (1992), "An empirical comparison of alternative models for the short-term interest rate", *The journal of finance* **XLVII**(3), 1209–1227.

- Elerian, O. (1998), “A note on the existence of a closed form conditional transition density for the milstein scheme”, Economics discussion paper, Nuffield College, Oxford.
- Florens-Zmirou, D. (1993), “On estimating the diffusion coefficient from discrete observations”, *Journal of applied probability* **30**, 790–804.
- Gallant, A. R. & Tauchen, G. (1996), “Which moments to match?”, *Econometric Theory* **12**, 657–681.
- Gouriéroux, C. & Monfort, A. (1995), *Simulation-based econometric methods*, CORE lecture series, Oxford.
- Gouriéroux, C., Monfort, A. & Renault, E. (1993), “Indirect inference”, *Journal of applied Econometrics* **8**, 85–118.
- Kloeden, P. E. (2001), “A brief overview of numerical methods for stochastic differential equations”. Fachbereich Mathematik, Johann Wolfgang Goethe Universität (tiré d’une conférence tenue à Naples le 25 mai 2001).
- Kloeden, P. & Platen, E. (1999), *Numerical solution of stochastic differential equations*, Applications of mathematics; Stochastic modelling and applied probability, troisième édition, Springer-Verlag.
- Lubrano, M. (2000), “Modélisation bayésienne non-linéaire du taux d’intérêt de court terme américain: l’aide des outils non-paramétriques”. CORE Discussion paper, 2000-38.
- Milstein, G. (1978), “A method of second order accuracy integration of stochastic differential equations”, *Theory of probability and its applications* **23**, 396–401.

Working Papers / Documents de Travail

Le GREQAM. diffuse une série de documents de travail qui prend la suite de celle diffusée depuis 1982 par le GREQE. Cette série comporte 4 sous séries - respectivement : "théories", "applications", "méthodologie", "Tirés à part"

Certains papiers sont disponibles sous la forme de fichiers PDF. Adobe Acrobat vous permet de les lire, de les transférer et de les imprimer.

<http://durandal.cnrs-mrs.fr/GREQAM/dt/dt.htm>

Adresses du GREQAM

GREQAM

Centre de la Vieille Charité
2 Rue de La Charité
13002 MARSEILLE
tél. 04.91.14.07.70
fax. 04.91.90.02.27
E-mail : greqam@ehess.cnrs-mrs.fr

GREQAM/ LEQAM

Château La Farge
Route des Milles
13290 LES MILLES
tél. 04.42.93.59.80
fax. 04.42.93.09.68
E-mail : leqam@romarin.univ-aix.fr

GREQAM/ C.R.I.D.E.S.O.P.E.

Faculté d'Économie Appliquée
Bât. Austerlitz
15-19 Allée Claude Forbin
13627 Aix-en-Provence Cedex 1
Tel : 04 42 96 12 31
Fax : 04 42 96 80 00