



**HAL**  
open science

# Schéma de discrétisation anticipatif et estimation du paramètre de dérive d'une diffusion

Sandie Souchet

► **To cite this version:**

Sandie Souchet. Schéma de discrétisation anticipatif et estimation du paramètre de dérive d'une diffusion. ESAIM: Probability and Statistics, 2000, 4, pp.233-258. hal-00276939

**HAL Id: hal-00276939**

**<https://hal.science/hal-00276939>**

Submitted on 2 May 2008

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Schémas de discrétisation anticipatifs et estimation du paramètre de dérive d'une diffusion

Sandie SOUCHET  
SAMOS, Université Paris 1\*

## Résumé

Soit  $Y^T = (Y_t)_{t \in [0, T]}$  une diffusion réelle ergodique, dont la dérive dépend d'un paramètre inconnu  $\theta_0 \in \mathbb{R}^p$ . Notre but est d'estimer  $\theta_0$  à partir d'une observation discrétisée de  $Y^T$ ,  $(Y_{k\delta})_{k=0, n}$ , lorsque  $T = n\delta$  tend vers l'infini et que  $\delta$  est petit mais fixé. Nous adaptons pour cela la méthode d'estimation des Moments Généralisés (*cf.* Hansen) au schéma d'approximation discret et anticipatif du trapèze, puis au schéma de Simpson. Sous certaines conditions générales, le schéma du trapèze (resp. de Simpson) fournit une estimation de  $\theta_0$  avec un biais explicite d'ordre  $\delta^2$  (resp.  $\delta^4$ ). L'estimateur obtenu est de plus asymptotiquement normal. Ces résultats généralisent ceux obtenus par Bergstrom [1], pour une diffusion gaussienne, à dérive linéaire en  $\theta$ .

## Abstract

Let  $Y^T = (Y_t)_{t \in [0, T]}$  be a real ergodic diffusion process which drift depends on an unknown parameter  $\theta_0 \in \mathbb{R}^p$ . Our aim is to estimate  $\theta_0$  from a discrete observation of the process  $Y^T$ ,  $(Y_{k\delta})_{k=0, n}$ , for a fixed and small  $\delta$ , as  $T = n\delta$  goes to infinity. For that purpose, we adapt the Generalized Method of Moments (*see* Hansen) to the anticipative and approximate discrete-time trapezoidal scheme, and then to Simpson's. Under some general assumptions, the trapezoidal scheme (respectively Simpson's scheme) provides an estimation of  $\theta_0$  with a bias of order  $\delta^2$  (resp.  $\delta^4$ ). Moreover, this estimator is asymptotically normal. These results generalize Bergstrom's [1], which were obtained for a Gaussian diffusion process, which drift is linear in  $\theta$ .

## Mots clefs

Schéma du trapèze; schéma de Simpson; schéma anticipatif; diffusion ergodique; estimation par variables instrumentales; méthode des moments généralisés; contraste; biais d'estimation; efficacité asymptotique en variance.

## Classification A M S

62 M 05 - 62 F 12.

## 1 Introduction

Nous abordons dans ce travail le problème de l'estimation du paramètre de dérive d'une diffusion ergodique observée à des instants équidistants de pas  $\delta > 0$ , constant et petit.

---

\*Centre Pierre Mendès France, 90, rue de Tolbiac 75013 PARIS

Le problème de l'estimation paramétrique pour les diffusions a fait l'objet de nombreux travaux, que l'on peut classer en fonction du type d'observations disponibles.

Si l'on note  $[0, T]$  l'intervalle d'étude et si l'on dispose d'une observation continue de la trajectoire du processus, l'estimateur du maximum de vraisemblance est consistant, asymptotiquement normal et efficace avec la vitesse  $\sqrt{T}$ , lorsque  $T$  tend vers l'infini ([15], [14]).

Cependant, le recours à une observation discrète paraît plus réaliste. Là encore, on distingue le cas d'un pas d'observation constant  $\delta > 0$  de celui où les temps d'observation sont asymptotiquement denses dans  $\mathbb{R}^+$ .

Dans un cadre asymptotique du type  $T = n\delta_n \rightarrow \infty$ ,  $\delta_n \rightarrow 0$ ,  $n\delta_n^p \rightarrow 0$ , Florens-Zmirou [5], Prakasa-Rao [17] et Yoshida [22] proposent des procédures d'estimation optimales issues de la discrétisation de la vraisemblance continue; Kessler [10] donne une approximation gaussienne de la vraisemblance. Pedersen propose une méthode d'évaluation numérique des densités de transition tout en contrôlant les propriétés théoriques de l'estimation ([18], [19]).

Dans le cas où le pas d'observation  $\delta$  est constant, Dacunha-Castelle et Florens-Zmirou [4] ont prouvé que l'estimateur du maximum de vraisemblance est consistant et asymptotiquement efficace à la vitesse  $\sqrt{n\delta}$ , quand  $n \rightarrow \infty$ . Mais on ne dispose pas, en général, de la forme analytique des densités de transition du processus; un autre point de vue est de considérer des estimateurs issus de fonctions d'estimation (Bibby et Sorensen [2]; Kessler [11] [12]; [21]).

Dans [5], Florens-Zmirou construit un contraste à partir du schéma d'Euler approchant le schéma discret exact. Nous adopterons une démarche identique mais en utilisant des schémas anticipatifs et la méthode d'estimation des moments généralisés (G.M.M).

Soit  $(\Omega, \mathcal{A}, P)$  un espace de probabilité et  $W$  un mouvement brownien unidimensionnel standard issu de 0 défini sur cet espace. On munit l'espace précédent de la filtration naturelle  $\mathcal{F} = \{\mathcal{F}_t\}_{t \geq 0}$  associée à  $W$ . On considère alors l'équation différentielle stochastique :

$$dY_t = f(\theta_0, Y_t) dt + \sigma(Y_t) dW_t, \quad 0 \leq t \leq T, \quad Y_0 = x \quad (1)$$

où  $f : \mathbb{R}^p \times \mathbb{R} \rightarrow \mathbb{R}$  et  $\sigma : \mathbb{R} \rightarrow \mathbb{R}$  sont des fonctions connues et  $\theta_0 \in \mathbb{R}^p$  est le paramètre à estimer.

Supposons que (1) admette une unique solution forte de condition initiale  $x$ , notée  $Y^T = (Y_t)_{t \in [0, T]}$  (Karatzas et Shreve [9]). La relation entre deux observations successives est :

$$Y_{(k+1)\delta} = Y_{k\delta} + \int_{k\delta}^{(k+1)\delta} f(\theta_0, Y_t) dt + \int_{k\delta}^{(k+1)\delta} \sigma(Y_t) dW_t$$

Le schéma d'approximation d'Euler est obtenu en approchant l'intégrale  $\int_{k\delta}^{(k+1)\delta} f(\theta_0, Y_t) dt$  par  $\delta f(\theta_0, Y_{k\delta})$ . L'erreur commise est alors en  $O_P(\delta^2)$ . La méthode d'estimation des moindres carrés associée au schéma d'approximation d'Euler à pas  $\delta$  donne ainsi une estimation de  $\theta_0$  à  $\delta$  près (*cf.* [5]). Dans le cas du schéma d'Euler, l'ordre du biais étant directement lié à celui de la variable d'erreur, la réduction de ce biais d'estimation passe par une réduction de l'erreur commise en approximant  $\int_{k\delta}^{(k+1)\delta} f(\theta_0, Y_t) dt$ . Or,

des résultats d'analyse classique montrent qu'il existe des méthodes plus fines permettant de mieux approcher l'intégrale d'une fonction déterministe. On peut citer la méthode du trapèze (méthode à deux points) et la méthode de Simpson (méthode à trois points). Elles donnent respectivement, sous de bonnes conditions de régularité pour  $f$  (cf. [16]) :

$$\int_a^{a+\delta} f(x)dx = \frac{\delta}{2} (f(a) + f(a + \delta)) + O(\delta^3)$$

$$\int_a^{a+2\delta} f(x)dx = \frac{\delta}{3} (f(a) + 4f(a + \delta) + f(a + 2\delta)) + O(\delta^5)$$

L'idée est de transposer ces approximations à notre contexte : les nouveaux schémas d'approximation ainsi construits, couplés à une méthode d'estimation adaptée, vont permettre d'obtenir un biais d'estimation d'un ordre supérieur à  $\delta$  et ceci, sans perte d'efficacité en variance.

La paternité de cette idée revient aux économètres (cf. Bergstrom [1], Sargan [20]): utilisant le schéma du trapèze couplé avec la méthode d'estimation par variable instrumentale (V.I.), ils obtiennent un contrôle du biais en  $\delta^2$  pour une diffusion vectorielle gaussienne ergodique, de dérive linéaire en  $\theta$  :  $f(\theta, Y_t) = \theta Y_t$ ,  $\theta$  étant une matrice de dimensions  $r \times r$  et  $Y_t$  un vecteur de  $\mathbb{R}^r$ .

Après avoir présenté les schémas anticipatifs du trapèze et de Simpson (§ 2), nous montrons dans la section 3 que le schéma du trapèze associée à la méthode d'estimation par variable instrumentale permet d'estimer  $\theta_0$  à un biais en  $O(\delta^2)$  près, pour une diffusion unidimensionnelle de dérive linéaire en  $\theta_0$ ,  $f(\theta_0, y) = \sum_{i=1}^p f_i(y) \theta_0^i$ . Dans le paragraphe 4, nous étendons ces résultats au cas d'une dérive  $f(\theta_0, y)$  non nécessairement linéaire en  $\theta_0$  en couplant les schémas anticipatifs à la méthode d'estimation des moments généralisés, méthode qui généralise celle par variable instrumentale au cas non linéaire (Hansen [7],[8]; Gourieroux et Monfort [6]). Le biais d'estimation obtenu en utilisant le schéma de Simpson est de l'ordre de  $\delta^4$ . La section 5 est une étude expérimentale qui illustre les résultats précédents pour trois modèles de diffusion particuliers.

## 2 Présentation des schémas d'approximation

Commençons par introduire quelques notations. Dans la suite,  $f(\theta, x)$  sera noté  $f_\theta(x)$ . Pour tout  $p \in \mathbb{N}^*$ ,  $C^p(\mathbb{R})$  représente l'espace des fonctions  $p$  fois continument dérivables sur  $\mathbb{R}$ . Pour toute fonction  $g$ ,  $g'$ ,  $g''$  et  $g^{(k)}$  sont respectivement les dérivées à l'ordre 1, 2 et  $k$  de  $g$  en  $x$ . Notons  $A_{\theta_0}$  le générateur infinitésimal associé à (1) défini par :  $\forall h \in C^2(\mathbb{R})$ ,  $A_{\theta_0}h = f_{\theta_0}h' + \frac{1}{2}\sigma^2h''$ ;  $A_{\theta_0}^l$  représente le  $l^{\text{ième}}$  itéré de  $A_{\theta_0}$ . La notation abrégée,  $l = m, r$ , se lit  $l \in \{m, \dots, r\}$ .

### Propriété 1 : Schéma du trapèze

Supposons que  $f_{\theta_0} \in C^4(\mathbb{R})$  et  $\sigma \in C^2(\mathbb{R})$ . Alors :

$$\forall t \geq 0, Y_{t+\delta} = Y_t + \frac{\delta}{2} (f_{\theta_0}(Y_t) + f_{\theta_0}(Y_{t+\delta})) + \eta_t + \xi_t \quad (2)$$

avec :

$$\begin{aligned} \eta_t &= \frac{1}{2} \int_t^{t+\delta} (t + \delta - v)(t - v) A_{\theta_0}^2 f_{\theta_0}(Y_v) dv \\ \xi_t &= \int_t^{t+\delta} \sigma(Y_v) \left[ 1 + \frac{1}{2} (t + \delta - v)(t - v) (A_{\theta_0} f_{\theta_0})'(Y_v) + \left( \frac{\delta}{2} + t - v \right) f'_{\theta_0}(Y_v) \right] dW_v \end{aligned}$$

Si de plus, pour  $l = 0, 1$ ,  $E_P \left( \int_0^t \left[ (A_{\theta_0}^l f_{\theta_0})' \sigma \right]^2(Y_v) dv \right) < \infty$  et  $E_P \left( \int_0^t \sigma^2(Y_v) dv \right) < \infty$ ,

alors  $(\xi_t)_{t \geq 0}$  est une suite de variables de carré intégrable telles que  $E_P(\xi_t | \mathcal{F}_t) \stackrel{p.s.}{=} 0$ .

On constate que l'erreur  $\eta_t$  est en  $O_P(\delta^3)$ , ce qui est conforme à l'approximation d'une intégrale ordinaire par la méthode du trapèze.

### Démonstration de la propriété 1 :

La démonstration repose sur le résultat suivant d'interversion de l'ordre d'intégration entre intégrale de Riemann et intégrale stochastique : soit  $\left( \int_0^t J_v dW_v \right)_{t \geq 0}$  une martingale locale continue et  $g$  une fonction continue sur  $\mathbb{R}$ ; par application de la formule de Ito pour les martingales locales [9], on obtient alors :

$$\int_t^{t+\delta} \left( \int_v^{t+\delta} g(u) du \right) J_v dW_v = \int_t^{t+\delta} \left( \int_t^u J_v dW_v \right) g(u) du \quad (3)$$

$$\int_t^{t+\delta} \left( \int_t^v g(u) du \right) J_v dW_v = \int_t^{t+\delta} \left( \int_u^{t+\delta} J_v dW_v \right) g(u) du \quad (4)$$

Posons :

$$\begin{aligned} R_t &= \int_t^{t+\delta} f_{\theta_0}(Y_v) dv - \frac{\delta}{2} (f_{\theta_0}(Y_t) + f_{\theta_0}(Y_{t+\delta})) \\ &= \frac{1}{2} \int_t^{t+\delta} [f_{\theta_0}(Y_v) - f_{\theta_0}(Y_t)] dv - \frac{1}{2} \int_t^{t+\delta} [f_{\theta_0}(Y_{t+\delta}) - f_{\theta_0}(Y_v)] dv \end{aligned}$$

Par application de la formule de Ito, on obtient :

$$\begin{aligned} R_t &= \frac{1}{2} \int_t^{t+\delta} \left[ \int_t^v A_{\theta_0} f_{\theta_0}(Y_u) du \right] dv + \frac{1}{2} \int_t^{t+\delta} \left[ \int_t^v (f'_{\theta_0} \sigma)(Y_u) dW_u \right] dv \\ &\quad - \frac{1}{2} \int_t^{t+\delta} \left[ \int_v^{t+\delta} A_{\theta_0} f_{\theta_0}(Y_u) du \right] dv - \frac{1}{2} \int_t^{t+\delta} \left[ \int_v^{t+\delta} (f'_{\theta_0} \sigma)(Y_u) dW_u \right] dv \end{aligned}$$

Par (3) et (4), on a :

$$R_t = \int_t^{t+\delta} \left(t + \frac{\delta}{2} - u\right) A_{\theta_0} f_{\theta_0}(Y_u) du + \int_t^{t+\delta} \left(t + \frac{\delta}{2} - u\right) (f'_{\theta_0} \sigma)(Y_u) dW_u$$

Or :  $\int_t^{t+\delta} \left(t + \frac{\delta}{2} - u\right) du = 0$ . Donc :

$$\begin{aligned} \int_t^{t+\delta} \left(t + \frac{\delta}{2} - u\right) A_{\theta_0} f_{\theta_0}(Y_u) du &= \frac{1}{2} \int_t^{t+\delta} \left(t + \frac{\delta}{2} - u\right) [A_{\theta_0} f_{\theta_0}(Y_u) - A_{\theta_0} f_{\theta_0}(Y_t)] du \\ &\quad - \frac{1}{2} \int_t^{t+\delta} \left(t + \frac{\delta}{2} - u\right) [A_{\theta_0} f_{\theta_0}(Y_{t+\delta}) - A_{\theta_0} f_{\theta_0}(Y_u)] du \end{aligned}$$

De manière similaire, on obtient :

$$\begin{aligned} \int_t^{t+\delta} \left(t + \frac{\delta}{2} - u\right) A_{\theta_0} f_{\theta_0}(Y_u) du &= \frac{1}{2} \int_t^{t+\delta} (t + \delta - v)(t - v) A_{\theta_0}^2 f_{\theta_0}(Y_v) dv \\ &\quad + \frac{1}{2} \int_t^{t+\delta} (t + \delta - v)(t - v) \left((A_{\theta_0} f_{\theta_0})' \sigma\right)(Y_v) dW_v \end{aligned}$$

Ceci permet d'écrire :

$$Y_{t+\delta} = Y_t + \frac{\delta}{2} (f_{\theta_0}(Y_t) + f_{\theta_0}(Y_{t+\delta})) + R_t + \int_t^{t+\delta} \sigma(Y_v) dW_v$$

ce qui n'est autre que le schéma du trapèze (2).  $\square$

### Propriété 2 : Schéma de Simpson

Supposons que  $f_{\theta_0} \in C^8(\mathbb{R})$  et  $\sigma \in C^6(\mathbb{R})$ . Alors :

$$\forall t \geq 0, Y_{t+2\delta} = Y_t + \frac{\delta}{3} (f_{\theta_0}(Y_t) + 4f_{\theta_0}(Y_{t+\delta}) + f_{\theta_0}(Y_{t+2\delta})) + \eta_t + \xi_t \quad (5)$$

avec :

$$\eta_t = \int_t^{t+\delta} \frac{(v-t)^3 (v-t-\frac{4\delta}{3})}{4!} A_{\theta_0}^4 f_{\theta_0}(Y_v) dv + \int_{t+\delta}^{t+2\delta} \frac{(t+2\delta-v)^3 (t+\frac{2\delta}{3}-v)}{4!} A_{\theta_0}^4 f_{\theta_0}(Y_v) dv$$

$$\xi_t = \xi_t^1 + \xi_t^2$$

$$\xi_t^1 = \int_t^{t+\delta} \sigma(Y_v) \left[ 1 + \sum_{k=0}^3 \frac{(-1)^k (v-t)^k}{(k+1)!} \left( t + \frac{(k+1)\delta}{3} - v \right) (A_{\theta_0}^k f_{\theta_0})'(Y_v) \right] dW_v$$

$$\xi_t^2 = \int_{t+\delta}^{t+2\delta} \sigma(Y_v) \left[ 1 + \sum_{k=0}^3 \frac{(t+2\delta-v)^k}{(k+1)!} \left( t + 2\delta - \frac{(k+1)\delta}{3} - v \right) (A_{\theta_0}^k f_{\theta_0})'(Y_v) \right] dW_v$$

Si de plus, pour  $l = 0, 3$ ,  $E_P \left( \int_0^t \left[ (A_{\theta_0}^l f_{\theta_0})' \sigma \right]^2(Y_v) dv \right) < \infty$  et  $E_P \left( \int_0^t \sigma^2(Y_v) dv \right) < \infty$ ,

alors  $(\xi_t^1)_{t \geq 0}$  et  $(\xi_t^2)_{t \geq 0}$  sont des suites de variables de carré intégrable qui vérifient

$$E_P(\xi_t^1 | \mathcal{F}_t) \stackrel{p.s.}{=} 0 \text{ et } E_P(\xi_t^2 | \mathcal{F}_{t+\delta}) \stackrel{p.s.}{=} 0$$

De manière similaire, la transposition de la méthode de Simpson à notre problème donne un résultat conforme à celui attendu, où l'erreur  $\eta_t$  est un  $O_P(\delta^5)$ .

### Démonstration de la propriété 2 :

Si on note  $R_t = \int_t^{t+\delta} f_{\theta_0}(Y_v) dv - \frac{\delta}{3} [f_{\theta_0}(Y_t) + 4f_{\theta_0}(Y_{t+\delta}) + f_{\theta_0}(Y_{t+2\delta})]$ , il suffit de décomposer ce reste sous la forme :

$$\begin{aligned} R_t &= \frac{1}{3} \int_t^{t+\delta} [f_{\theta_0}(Y_v) - f_{\theta_0}(Y_t)] dv - \frac{2}{3} \int_t^{t+\delta} [f_{\theta_0}(Y_{t+\delta}) - f_{\theta_0}(Y_v)] dv \\ &+ \frac{2}{3} \int_{t+\delta}^{t+2\delta} [f_{\theta_0}(Y_v) - f_{\theta_0}(Y_{t+\delta})] dv - \frac{1}{3} \int_{t+\delta}^{t+2\delta} [f_{\theta_0}(Y_{t+2\delta}) - f_{\theta_0}(Y_v)] dv \end{aligned}$$

On utilise alors les mêmes arguments que ceux développés dans la démonstration de la propriété 1.  $\square$

## 3 Le cas d'une dérive linéaire en $\theta$ :

### Schéma du trapèze et Estimation par variable instrumentale

Pour l'estimation de  $\theta_0$  basée sur le schéma d'approximation d'Euler [5], le contraste employé est celui des moindres carrés (M.C.). Si nous utilisons le schéma d'approximation du trapèze, la corrélation existant entre la variable explicative et le résidu biaise la procédure d'estimation par moindres carrés. Il faut donc utiliser une autre méthode.

Si le modèle est linéaire en  $\theta$ , une alternative classique utilisée en économétrie est l'estimation par *variable instrumentale* ([1], [6]) : pour un bon choix de *l'instrument*, instrument que l'on substitue alors à la variable explicative, cette méthode fournit une estimation convergente là où les moindres carrés conduisent à un biais systématique.

Nous examinerons ici le cas d'un modèle de diffusion pour laquelle la dérive s'écrit  $\theta \cdot f(y)$ , sans hypothèses de linéarité en  $y$ . Il est intéressant de mettre l'accent sur cet exemple à plusieurs titres :

- Il permet, d'une part, de comprendre le choix du contraste qui sera retenu dans le cas d'une dérive générale  $f(\theta, y)$ , sans condition de linéarité ni en  $\theta$  ni en  $y$ . Comme nous le verrons au paragraphe 4, ce choix (estimation des moments généralisés, G.M.M.) coïncidera avec celui de l'estimation par V.I. dans le cas d'un modèle linéaire en  $\theta$ .
- D'autre part, dans le cas linéaire, les hypothèses (notées H 1) assurant de bonnes propriétés asymptotiques sont simplifiées par rapport au cas général.
- Enfin, la linéarité permet d'obtenir pour l'estimateur par V.I., une expression explicite et linéaire dans les observations.

Dans la suite, pour toute matrice  $B = (B_k^i)_{k,i}$ ,  $k$  représente l'indice des lignes et  $i$  celui des colonnes. On note de plus  ${}^t B$  la matrice transposée de  $B$ .

### 3.1 La méthode d'estimation par variables instrumentales

Considérons l'e.d.s. définie par :

$$dY_t = \theta_0 \cdot f(Y_t) dt + \sigma(Y_t) dW_t, \quad 0 \leq t \leq T, \quad Y_0 = x \quad (6)$$

avec  $\theta_0 \cdot f(Y_t) = \sum_{i=1}^p f_i(Y_t) \theta_0^i$ . On suppose que  $f_1, \dots, f_p$  et  $\sigma$  sont des fonctions connues, définies de  $\mathbb{R}$  dans  $\mathbb{R}$ ,  $\sigma^2$  étant de plus strictement positive sur  $\mathbb{R}$ .

Pour  $i = 1, p$  et  $k = 1, n$ , on note :

$$X_k^i = \frac{f_i(Y_{k\delta}) + f_i(Y_{(k-1)\delta})}{2\sigma(Y_{(k-1)\delta})}, \quad X_k = (X_k^1, \dots, X_k^p), \quad X = (X_k^i)_{k=1, n, i=1, p}$$

$$Y_k = \frac{Y_{k\delta} - Y_{(k-1)\delta}}{\delta\sigma(Y_{(k-1)\delta})}, \quad Y = {}^t(Y_1, \dots, Y_n)$$

Dans le schéma d'approximation du trapèze (2), le fait que la variable explicative  $X_k$  et le bruit blanc  $\xi_{(k-1)\delta}$  soient corrélés conduit à un biais systématique dans l'estimation par moindres carrés : l'estimateur des M.C.,  $\theta_n = ({}^t X X)^{-1} ({}^t X Y)$ , doit être adapté.

La méthode d'estimation par V.I. consiste à remplacer la variable explicative  $X_k$  par un *instrument*  $\tilde{X}_k$  décorrélé de  $\xi_{(k-1)\delta}$ , mais bien lié à  $X_k$ . Dans le cas présent, il est naturel de choisir  $\tilde{X}_k = (\tilde{X}_k^1, \dots, \tilde{X}_k^p)$  avec  $\tilde{X}_k^i = \frac{f_i(Y_{(k-1)\delta})}{\sigma(Y_{(k-1)\delta})}$ . L'estimateur par V.I. est défini par (cf. [1], [6]) :

$$\theta_n^I = ({}^t \tilde{X} X)^{-1} ({}^t \tilde{X} Y), \quad \tilde{X} = (\tilde{X}_k^i)_{k=1, n, i=1, p}$$

Comme on le verra, le choix de cet instrument est bon car il permet de préserver l'efficacité asymptotique en variance à un facteur  $(I_p + O(\delta))$  près : la variance limite de  $\theta_n^I$  est, à un terme négligeable près, l'inverse de l'information de Fisher asymptotique.

### 3.2 Résultats asymptotiques

On note  $s(y, \theta_0) = \exp\left(-2 \int_{x_0}^y \frac{\theta_0 \cdot f(z)}{\sigma^2(z)} dz\right)$  la dérivée de la fonction d'échelle associée à l'équation (6).

**Hypothèses H 1 : Schéma du trapèze, dérive linéaire en  $\theta$**

**H 1.1** Pour tout  $i = 1, p$ ,  $f_i$  est de classe  $C^4$  sur  $\mathbb{R}$  et  $\sigma$  est de classe  $C^2$  sur  $\mathbb{R}$  et vérifie :

$$\forall x \in \mathbb{R}, \quad \sigma^2(x) > 0$$

Il existe de plus  $K > 0$  telle que :  $\forall i = 1, p, \forall x \in \mathbb{R}, |f_i(x)| + |\sigma(x)| \leq K(1 + |x|)$ .  
Sous ces conditions, (6) admet une unique solution forte  $Y^T = (Y_t)_{t \in [0, T]}$  non explosive de condition initiale  $Y_0 = x$  [9].

**H 1.2**

$$\int_0^\infty s(x, \theta_0) dx = \int_{-\infty}^0 s(x, \theta_0) dx = \infty \text{ et } \int_{-\infty}^\infty [s(x, \theta_0) \sigma^2(x)]^{-1} dx = C(\theta_0) < \infty$$

Sous (H 1.1) et (H 1.2),  $Y^T$  est récurrente positive sur  $\mathbb{R}$ , de loi invariante :  $\mu_{\theta_0}(dy) = [C(\theta_0) s(y, \theta_0) \sigma^2(y)]^{-1} dy$ . Notons  $P_{\theta_0}^T$  la loi du processus stationnaire solution de (6) et  $E_{\theta_0}$  l'espérance sous  $P_{\theta_0}^T$  :  $E_{\theta_0}[g(Y^T)] = \int_{y \in C([0, T])} g(y) dP_{\theta_0}^T(y)$ ,  $C([0, T])$  étant l'espace des fonctions continues sur  $[0, T]$ .

**H 1.3** La fonction  $c_{\theta_0}(x) = \frac{(\theta_0 \cdot f)^2(x)}{2\sigma^2(x)} + \frac{(\theta_0 \cdot f)'(x)}{2} - \frac{\sigma'(x)(\theta_0 \cdot f)(x)}{\sigma(x)} + \frac{(\sigma')^2(x)}{8}$  satisfait :

$$\min \left\{ \lim_{x \rightarrow +\infty} c_{\theta_0}(x), \lim_{x \rightarrow -\infty} c_{\theta_0}(x) \right\} = c_{\theta_0} > 0$$

**H 1.4** La loi invariante  $\mu_{\theta_0}$  admet des moments de tous ordres et vérifie :

$$\forall r \in \mathbb{N}, \int \frac{|x|^r}{\sigma^8(x)} d\mu_{\theta_0}(x) < \infty$$

**H 1.5** Pour tout  $i = 1, p$ , les dérivées successives de  $f_i$  jusqu'à l'ordre 4 sont à croissance polynomiale, ainsi que  $\sigma'$  et  $\sigma''$ .

**H 1.6**  $H_{\theta_0} = \left( E_{\theta_0} \left[ \frac{f_i(Y_0) f_j(Y_0)}{\sigma^2(Y_0)} \right] \right)_{1 \leq i, j \leq p}$  est inversible.

*Cette hypothèse implique l'identifiabilité paramétrique du modèle.*

Des hypothèses précédentes découlent deux résultats importants :

- Sous (H 1.1), (H 1.2) et (H 1.3),  $Y^T$  satisfait une propriété de  $\rho$ -mélange exponentiel. Si  $\pi_{\theta_0}^t$  est le semi-groupe associé à (6) et  $L_0^2(\mu_{\theta_0}) = \{f \in L^2(\mu_{\theta_0}), \mu_{\theta_0}(f) = 0\}$ , il existe  $\lambda > 0$  telle que [11]:

$$\forall f \in L_0^2(\mu_{\theta_0}), \|\pi_{\theta_0}^t(f)\|_{L^2(\mu_{\theta_0})} \leq \exp(-\lambda t) \|f\|_{L^2(\mu_{\theta_0})} \quad (7)$$

Cette propriété de mélange permet d'obtenir les résultats d'ergodicité et de normalité asymptotiques suivants (cf. [2], lemme 3.2; [5]). Notons  $Q_{\theta_0}^\delta = \mu_{\theta_0} \otimes \pi_{\theta_0}^\delta$  :

**Lemme 1** Si  $Q_{\theta_0}^\delta(g^2) < \infty$ , alors :  $\frac{1}{n} \sum_{k=1}^n g(Y_{k\delta}, Y_{(k-1)\delta}) \xrightarrow{L^2(P)} Q_{\theta_0}^\delta(g)$ .

Si de plus,  $Q_{\theta_0}^\delta(g) = 0$ , on a :  $\frac{1}{\sqrt{n}} \sum_{k=1}^n g(Y_{k\delta}, Y_{(k-1)\delta}) \xrightarrow{\mathcal{D}(P)} \mathcal{N}(0, V_{\theta_0}^\delta)$ ,

où  $V_{\theta_0}^\delta = \int [g^2(x, y) + 2g(x, y)[U_\delta(g)](y)] Q_{\theta_0}^\delta(dx, dy)$  avec  $U_\delta(g)(x) = \sum_{k=1}^{+\infty} \pi_{\theta_0}^{k\delta}(g)(x)$ .

La convergence et la normalité asymptotique de  $\theta_n^I$  s'appuient directement sur ce résultat.

- Sous (H 1.1), (H 1.2) et (H 1.4), il existe une constante  $C > 0$  telle que :

$$\forall t, 0 \leq t \leq T, E_{\theta_0} \left[ \max_{0 \leq s \leq t} |Y_s|^{2m} \right] \leq C (1 + E_{\theta_0} [|Y_0|^{2m}]) \exp(Ct) \quad (8)$$

Cette majoration associée à la formule de Ito permet d'établir les régularités précisées dans le lemme ci-après pour les fonctions suivantes :

$$\forall i, j = 1, p, M_{i,j}(v) = E_{\theta_0} \left[ \frac{f_i(Y_0) f_j(Y_v)}{\sigma^2(Y_0)} \right], P_{i,j}(v) = E_{\theta_0} \left[ (f_i f_j)(Y_0) \frac{\sigma^2(Y_v)}{\sigma^4(Y_0)} \right]$$

$$N_i(v) = E_{\theta_0} \left[ f_i(Y_0) \frac{[A_{\theta_0}^2(\theta_0 \cdot f)](Y_v)}{\sigma^2(Y_0)} \right]$$

**Lemme 2** *Sous (H 1),  $N_i$  est continue sur  $\mathbb{R}^+$ ;  $M_{i,j}$  et  $P_{i,j}$  sont de classe  $C^1$  sur  $\mathbb{R}_*^+$  et dérivables à droite en 0 avec :*

$$\forall v \in \mathbb{R}^+, M'_{i,j}(v) = E_{\theta_0} \left[ f_i(Y_0) \frac{(A_{\theta_0} f_j)(Y_v)}{\sigma^2(Y_0)} \right], P'_{i,j}(v) = E_{\theta_0} \left[ (f_i f_j)(Y_0) \frac{(A_{\theta_0} \sigma^2)(Y_v)}{\sigma^4(Y_0)} \right]$$

et  $\lim_{h \rightarrow 0} M'_{i,j}(h) = M'_{i,j}(0^+), \lim_{h \rightarrow 0} P'_{i,j}(h) = P'_{i,j}(0^+)$

Ces fonctions interviennent dans l'expression de la limite et de la variance asymptotiques de  $\theta_n^I$ . Les caractéristiques asymptotiques de  $\theta_n^I$  vont ainsi pouvoir être développées en fonction de  $\delta$ . Dans ce but, on définit :

$$M_{\theta_0} = (M'_{i,j}(0^+))_{i,j} = \left( E_{\theta_0} \left[ f_i(Y_0) \frac{(A_{\theta_0} f_j)(Y_0)}{\sigma^2(Y_0)} \right] \right)_{i,j} \quad (9)$$

$$P_{\theta_0} = (P'_{i,j}(0^+))_{i,j} = \left( E_{\theta_0} \left[ (f_i f_j)(Y_0) \frac{(A_{\theta_0} \sigma^2)(Y_0)}{\sigma^4(Y_0)} \right] \right)_{i,j} \quad (10)$$

$${}^t N_{\theta_0} = (N_i(0))_i = \left( E_{\theta_0} \left[ f_i(Y_0) \frac{[A_{\theta_0}^2(\theta_0 \cdot f)](Y_0)}{\sigma^2(Y_0)} \right] \right)_i \quad (11)$$

Notons :

$$[{}^t \tilde{X} Y]^\delta = \left( E_{\theta_0} \left[ f_i(Y_0) \left( \frac{Y_\delta - Y_0}{\delta \sigma^2(Y_0)} \right) \right] \right)_i \text{ et } [{}^t \tilde{X} X]^\delta = \left( E_{\theta_0} \left[ \frac{f_i(Y_0) (f_j(Y_\delta) + f_j(Y_0))}{2\sigma^2(Y_0)} \right] \right)_{i,j}$$

**Théorème 1** *Sous (H 1), pour  $H_{\theta_0}$  définie en (H 1.6),  $P_{\theta_0}$  en (10) et  $N_{\theta_0}$  en (11), il existe  $\delta_0 > 0$  tel que, pour tout  $\delta, 0 < \delta \leq \delta_0$ , on a, lorsque  $n \rightarrow +\infty$  :*

$$\theta_n^I = \left( {}^t \tilde{X} X \right)^{-1} \left( {}^t \tilde{X} Y \right) \xrightarrow{P} \theta_\delta = \left( [{}^t \tilde{X} X]^\delta \right)^{-1} [{}^t \tilde{X} Y]^\delta \text{ et } \sqrt{n\delta} (\theta_n^I - \theta_\delta) \xrightarrow{\mathcal{D}(P)} \mathcal{N}_p(0, V_\delta(\theta_0))$$

avec :

$$\theta_\delta - \theta_0 = -\frac{\delta^2}{12} H_{\theta_0}^{-1} N_{\theta_0} + o(\delta^2) \text{ et } V_\delta(\theta_0) = H_{\theta_0}^{-1} \left( I_p + \frac{\delta}{2} P_{\theta_0} H_{\theta_0}^{-1} + o(\delta^2) \right)$$

Comme nous l'avons souligné,  $H_{\theta_0}$  est l'information de Fisher asymptotique associée au modèle. L'efficacité asymptotique en variance est donc préservée à un facteur  $(I_p + O(\delta))$  près.

### Démonstration du théorème 1 :

- D'après le lemme 2, pour  $\delta$  dans un voisinage de 0, on a :  $M_{i,j}(\delta) = M_{i,j}(0) + \delta M'_{i,j}(0) + o(\delta)$ . Or :  $\left[ {}^t\tilde{X}X \right]^\delta = \frac{1}{2} (M_{i,j}(0) + M_{i,j}(\delta))_{i,j}$ . D'où :

$$\left[ {}^t\tilde{X}X \right]^\delta = H_{\theta_0} + \frac{\delta}{2} M_{\theta_0} + o(\delta) \quad (12)$$

$H_{\theta_0}$  étant inversible, il existe  $\delta_0 > 0$  tel que pour  $\delta \in ]0, \delta_0]$ ,  $\left[ {}^t\tilde{X}X \right]^\delta$  est inversible, ce qui implique l'existence de  $\theta_\delta$ .

- Par application du lemme 1, lorsque  $n \rightarrow \infty$ ,  $\frac{1}{n} {}^t\tilde{X}X \xrightarrow{L^2(P)} \left[ {}^t\tilde{X}X \right]^\delta$ . D'après ce qui précède, pour  $\delta \in ]0, \delta_0]$  et  $n$  assez grand,  ${}^t\tilde{X}X$  est inversible et donc  $\theta_n^I = \left( {}^t\tilde{X}X \right)^{-1} \left( {}^t\tilde{X}Y \right)$  existe.

Pour montrer la convergence de  $\theta_n^I$  vers  $\theta_\delta$ , il suffit d'appliquer à nouveau le lemme 1 :  $\frac{1}{n} {}^t\tilde{X}Y \xrightarrow{L^2(P)} \left[ {}^t\tilde{X}Y \right]^\delta$ .

- L'écart entre  $\theta_\delta$  et  $\theta_0$  se calcule en utilisant la décomposition du trapèze (2) :

$$\theta_\delta = \theta_0 + \left( \left[ {}^t\tilde{X}X \right]^\delta \right)^{-1} B_\delta, \quad {}^tB_\delta = \left( E_{\theta_0} \left[ f_i(Y_0) \frac{\eta_0}{\delta \sigma^2(Y_0)} \right] \right)_i$$

Par application du lemme 2, on montre que  $B_\delta = -\frac{\delta^2}{12} N_{\theta_0} + o(\delta^2)$ . Avec (12), on obtient le biais annoncé.

- On utilise à nouveau le lemme 1 pour prouver la normalité asymptotique. On a :

$$\sqrt{\frac{\delta}{n}} \left( {}^t\tilde{X}Y - {}^t\tilde{X}X\theta_\delta \right) \xrightarrow{\mathcal{D}(P)} \mathcal{N}_p(0, \Gamma_\delta(\theta_\delta)) \quad \text{et} \quad \frac{1}{n} {}^t\tilde{X}X \xrightarrow{P} \left[ {}^t\tilde{X}X \right]^\delta$$

Or :  $\sqrt{n\delta} (\theta_n^I - \theta_\delta) = \left( \frac{1}{n} {}^t\tilde{X}X \right)^{-1} \sqrt{\frac{\delta}{n}} \left( {}^t\tilde{X}Y - {}^t\tilde{X}X\theta_\delta \right)$ . Donc :

$$\sqrt{n\delta} (\theta_n^I - \theta_\delta) \xrightarrow{\mathcal{D}(P)} \mathcal{N}_p(0, V_\delta(\theta_0)), \quad V_\delta(\theta_0) = \left( \left[ {}^t\tilde{X}X \right]^\delta \right)^{-1} \Gamma_\delta(\theta_\delta) \left( \left[ {}^t\tilde{X}X \right]^\delta \right)^{-1}$$

Avec :

**Lemme 3** Sous (H 1), pour tout  $\delta$ ,  $0 < \delta \leq \delta_0$ , on a :  $\Gamma_\delta(\theta_\delta) = H_{\theta_0} + \frac{\delta}{2} P_{\theta_0} + o(\delta)$

La démonstration de ce résultat est donnée en annexe B. Utilisant (12), on obtient ainsi le développement annoncé pour  $V_\delta(\theta_0)$ .  $\square$

## 4 Le cas d'une dérive générale :

### Schéma du trapèze et estimation des moments généralisés

Ce paragraphe est consacré à l'étude du cas général, sans nécessité de linéarité en  $\theta$  pour la fonction de dérive. On utilise pour cela la méthode d'estimation des moments généralisés (G.M.M.) ([8],[7]; [6]). Cette méthode généralise au cas non linéaire la méthode des V.I.

Notons :

$$\mathcal{V}_\delta = {}^t(\mathcal{V}_\delta^1, \dots, \mathcal{V}_\delta^p), \quad \mathcal{V}_\delta^i(y, x, \theta) = \frac{\partial}{\partial \theta^i} f_\theta(x) \left( \frac{y - x - \frac{\delta}{2}(f_\theta(x) + f_\theta(y))}{\delta \sigma^2(x)} \right) \quad (13)$$

$$V_n(\theta) = \frac{1}{n} \sum_{k=1}^n \mathcal{V}_\delta(Y_{k\delta}, Y_{(k-1)\delta}, \theta), \quad V_n^i(\theta) = \frac{1}{n} \mathcal{V}_\delta^i(Y_{k\delta}, Y_{(k-1)\delta}, \theta) \quad (14)$$

L'estimateur des moments généralisés,  $\hat{\theta}_n$ , associé à  $\mathcal{V}_\delta$  et à la matrice de poids l'identité de dimension  $p$ , est défini par :

$$\hat{\theta}_n = \inf_{\theta \in \Theta} U_n(\theta), \quad U_n(\theta) = {}^t V_n(\theta) I_p V_n(\theta) = \|V_n(\theta)\|_2^2$$

On vérifie aisément que  $\theta_n^I$  et  $\hat{\theta}_n$  coïncident dans le cas d'un modèle linéaire car alors  $V_n(\theta) = {}^t \tilde{X} Y - {}^t \tilde{X} X \theta$  et donc  $U_n(\theta_n^I) = 0$ . Nous allons montrer que  $\hat{\theta}_n$  converge vers une limite  $\theta_\delta$  que l'on peut caractériser, et est asymptotiquement normal et efficace en variance à un facteur  $(I_p + O(\delta))$  près. Dans le cas des modèles linéaires, pour  $\delta$  petit,  $\theta_\delta$  est l'unique solution explicite du système linéaire :  $[{}^t \tilde{X} Y]^\delta - [{}^t \tilde{X} X]^\delta \theta_\delta = 0$ . Dans le cas général, nous allons montrer que,  $\theta_\delta$  est l'unique solution de l'équation  $E_{\theta_0}[\mathcal{V}_\delta(Y_\delta, Y_0, \theta)] = 0$ .

On note pour tout  $i = 1, p$  :

$$\begin{aligned} V_i(\theta, \delta) &= E_{\theta_0}[\mathcal{V}_\delta^i(Y_\delta, Y_0, \theta)] \text{ pour } \delta > 0, \text{ et} \\ V_i(\theta, 0) &= E_{\theta_0} \left[ \frac{\partial}{\partial \theta^i} f_\theta(Y_0) \left( \frac{f_{\theta_0}(Y_0) - f_\theta(Y_0)}{\sigma^2(Y_0)} \right) \right] \end{aligned}$$

#### 4.1 Propriétés asymptotiques

Soit  $\Theta$  un sous ensemble **compact** de  $\mathbb{R}^p$  tel que  $\theta_0 \in \overset{\circ}{\Theta}$ . Soit  $\mathbb{U}(\Theta)$  un ouvert de  $\mathbb{R}^p$  tel que  $\Theta \subset \mathbb{U}(\Theta)$ .

#### Hypothèses H 2 : Schéma du trapèze et G.M.M.

(H 1.2), (H 1.3), (H 1.4) et (H 1.6) sont maintenues en remplaçant le cas échéant  $\theta_0 \cdot f$  par  $f_{\theta_0}$ .

**H 2.1**  $f_{\theta_0}$  est de classe  $C^4$  sur  $\mathbb{R}$ ,  $\sigma$  est de classe  $C^2$  sur  $\mathbb{R}$  et vérifie :

$$\forall x \in \mathbb{R}, \sigma^2(x) > 0$$

De plus, il existe  $K > 0$  telle que :  $\forall x \in \mathbb{R}, |f_{\theta_0}(x)| + |\sigma(x)| \leq K(1 + |x|)$ .

- H 2.5** 1. Pour tout  $x \in \mathbb{R}$ ,  $\theta \rightarrow f_\theta(x)$  est de classe  $C^4$  sur  $\mathbb{U}(\Theta)$ . Cette fonction et ses dérivées partielles successives jusqu'à l'ordre 4 sont à croissance polynomiale en  $x$ , uniformément en  $\theta$  sur  $\mathbb{U}(\Theta)$ .  
De plus,  $f_\theta(x)$ ,  $\frac{\partial}{\partial \theta^i} f_\theta(x)$  et  $\frac{\partial^2}{\partial \theta^j \partial \theta^i} f_\theta(x)$  sont continues sur  $\mathbb{U}(\Theta) \times \mathbb{R}$ .
2. Pour  $\theta \in \mathbb{U}(\Theta)$ ,  $x \rightarrow f_\theta(x)$  est deux fois dérivable sur  $\mathbb{R}$ . La fonction  $(\theta, x) \rightarrow A_{\theta_0} f_\theta(x)$  est continue sur  $\mathbb{U}(\Theta) \times \mathbb{R}$  et est à croissance polynomiale en  $x$ , uniformément en  $\theta$  sur  $\mathbb{U}(\Theta)$ .  
De plus, pour tout  $x \in \mathbb{R}$ ,  $\theta \rightarrow A_{\theta_0} f_\theta(x)$  est de classe  $C^1$  sur  $\mathbb{U}(\Theta)$ .
3. Les fonctions  $f'_{\theta_0}$ ,  $(A_{\theta_0} f_{\theta_0})'$ ,  $A_{\theta_0}^2 f_{\theta_0}$  et  $A_{\theta_0}(\sigma^2 f'_{\theta_0})$  sont à croissance polynomiale en  $x$ , ainsi que  $\sigma'$  et  $\sigma''$ .
4. Pour tout  $i = 1, p$ ,  $x \rightarrow \frac{\partial}{\partial \theta^i} f_{\theta_0}(x)$  est de classe  $C^2$  sur  $\mathbb{R}$ . De plus,  $x \rightarrow [A_{\theta_0}(\frac{\partial}{\partial \theta^i} f_{\theta_0})](x)$  est à croissance polynomiale en  $x$ .

**H 2.7** Si, pour tout  $i = 1, p$ ,  $V_i(\theta, 0) = 0$  alors  $\theta = \theta_0$ .

*Dans le cas linéaire, cette hypothèse d'identifiabilité paramétrique est satisfaite sous (H 1.6)*

Pour tout  $i, j = 1, p$ , on redéfinit les fonctions  $M_{i,j}$ ,  $P_{i,j}$  et  $N_i$  introduites lors du paragraphe précédent en remplaçant  $\theta_0 \cdot f$  par  $f_{\theta_0}$  et  $f_i$  par  $\frac{\partial}{\partial \theta^i} f_{\theta_0}$ . Sous (H 2) et dans le cas général, les régularités décrites dans le lemme 2 restent vraies.

Nous commencerons par énoncer un lemme technique qui établit la régularité de la fonction  $(\theta, \delta) \rightarrow V_i(\theta, \delta)$ . Ce résultat nous permet en particulier de montrer que la méthode d'estimation proposée s'inscrit dans le cadre de l'estimation par minimum de contraste, ce qui simplifie l'étude des propriétés asymptotiques de  $\hat{\theta}_n$ . De plus, les deux développements mis en évidence dans ce lemme seront utilisés ultérieurement pour donner un développement des quantités asymptotiques associées à  $\hat{\theta}_n$ .

**Lemme 4** *Sous (H 2), pour tout  $i = 1, p$  et tout  $\delta \in \mathbb{R}^+$ ,  $\theta \rightarrow V_i(\theta, \delta)$  est de classe  $C^2$  sur  $\mathbb{U}(\Theta)$ . Ses dérivées partielles sont continues sur  $\mathbb{U}(\Theta) \times \mathbb{R}^+$  et on a :*

$$V_i(\theta_0, \delta) = -\frac{\delta^2}{12} N_i(0) + o(\delta^2) \quad (15)$$

$$\frac{\partial}{\partial \theta^j} V_i(\theta_0, \delta) = -\frac{1}{2} (M_{i,j}(\delta) + M_{i,j}(0)) + O(\delta^2) \quad (16)$$

*Pour tout  $\theta \in \mathbb{U}(\Theta)$ ,  $\delta \rightarrow V_i(\theta, \delta)$  est dérivable sur  $\mathbb{R}_*^+$  et dérivable à droite en 0. De plus,  $(\theta, \delta) \rightarrow V_i'(\theta, \delta)$  est continue sur  $\mathbb{U}(\Theta) \times \mathbb{R}^+$  et  $\theta \rightarrow V_i'(\theta, 0^+)$  est de classe  $C^1$  sur  $\mathbb{U}(\Theta)$ .*

La démonstration de ce lemme est donnée en annexe A. Montrons alors que  $U_n$  est un contraste au sens de la définition 3.2.7 de [3].

**Lemme 5** *Sous (H 2), il existe  $\delta_0 > 0$  tel que pour tout  $\delta$ ,  $0 < \delta \leq \delta_0$ ,  $U(\theta, \delta) = \sum_{i=1}^p (V_i(\theta, \delta))^2$  admet un unique minimum  $\theta_\delta = \phi(\delta) \in \overset{\circ}{\Theta}$  qui vérifie  $U(\theta_\delta, \delta) = 0$  et  $\lim_{\delta \rightarrow 0} \phi(\delta) = \theta_0$ .  $U(\theta, \delta)$  est la fonction de contraste associée au processus de contraste  $(U_n(\theta))_n$ .*

Cette caractérisation de  $\theta_\delta$  correspond à celle donnée dans le cas linéaire. Il suffit en effet de remarquer que pour une dérive linéaire en  $\theta$ , on a :

$$V(\theta, \delta) = {}^t(V_1(\theta, \delta), \dots, V_p(\theta, \delta)) = [{}^t\tilde{X}Y]^\delta - [{}^t\tilde{X}X]^\delta \theta$$

**Démonstration du lemme 5 :**

Pour tout  $\theta \in \mathbb{U}(\Theta)$ ,  $U_n(\theta)$  est  $\mathcal{F}_{n\delta}$ -mesurable et  $U(\theta, \delta)$  est positive. Les hypothèses (H 2.1), (H 2.4) et (H 2.5) impliquent que  $\mathcal{V}_\delta^i(y, x, \theta) \in L^2(Q_{\theta_0}^\delta)$ . Par application du lemme 1, on obtient pour tout  $i = 1, p$  :

$$\forall \theta \in \mathbb{U}(\Theta), V_n^i(\theta) \xrightarrow{L^2(P)} V_i(\theta, \delta) \text{ d'où } \forall \theta \in \Theta, U_n(\theta) \xrightarrow{L^1(P)} U(\theta, \delta)$$

Il nous reste à montrer que pour  $\delta$  petit,  $U(\cdot, \delta)$  admet un unique minimum  $\theta_\delta$  dans  $\overset{\circ}{\Theta}$ . Soit  $\bar{V}_i$  la fonction définie sur  $\mathbb{U}(\Theta) \times \mathbb{R}$  par :

$$\begin{aligned} \text{si } \delta > 0, \bar{V}_i(\theta, \delta) &= V_i(\theta, \delta) \\ \text{si } \delta \leq 0, \bar{V}_i(\theta, \delta) &= V_i(\theta, 0) + \delta V_i'(\theta, 0^+) \end{aligned}$$

En utilisant le lemme 4, on montre que pour tout  $i = 1, p$ ,  $\bar{V}_i$  est de classe  $C^1$  sur  $\mathbb{U}(\Theta) \times \mathbb{R}$ . Notons  $\bar{V}$  la fonction définie de  $\mathbb{U}(\Theta) \times \mathbb{R}$  dans  $\mathbb{R}^p$  par  $\bar{V}(\theta, \delta) = {}^t(\bar{V}_1(\theta, \delta), \dots, \bar{V}_p(\theta, \delta))$  :  $\bar{V}$  est de classe  $C^1$  sur  $\overset{\circ}{\Theta} \times \mathbb{R}$  et, d'après (H 2.7),  $\theta_0$  est l'unique point de  $\overset{\circ}{\Theta}$  tel que  $\bar{V}(\theta_0, 0) = V(\theta_0, 0) = 0$ . De plus, le gradient de  $\theta \rightarrow \bar{V}(\theta, 0)$  en  $\theta_0$  est  $-H_{\theta_0}$  et  $H_{\theta_0}$  est inversible. Par application du *théorème des fonctions implicites*, il existe  $\varepsilon > 0$  et  $\delta_0 > 0$  tels que pour tout  $\delta$ ,  $|\delta| < \delta_0$ , il existe un unique  $\theta_\delta$  dans la boule ouverte,  $B(\theta_0, \varepsilon) \subset \overset{\circ}{\Theta}$ , qui vérifie  $\bar{V}(\theta_\delta, \delta) = 0$  et  $\theta_\delta = \phi(\delta)$ . De plus,  $\phi$  est dérivable en 0. En particulier, pour tout  $\delta$ ,  $0 < \delta < \delta_0$ , il existe un unique  $\theta_\delta \in \overset{\circ}{\Theta}$  tel que  $U(\theta_\delta, \delta) = 0$  et  $\theta_\delta = \phi(\delta)$  avec  $\lim_{\delta \rightarrow 0} \phi(\delta) = \theta_0$ .  $\square$

#### 4.1.1 Etude de la convergence de $\hat{\theta}_n$

**Théorème 2** Sous (H 2), pour tout  $\delta$ ,  $0 < \delta \leq \delta_0$ ,  $\hat{\theta}_n \xrightarrow{P} \theta_\delta$

**Démonstration :**

La démonstration repose sur la vérification des hypothèses du théorème 3.2.8. de [3]. Sous (H 2), à  $\delta$  fixé,  $0 < \delta \leq \delta_0$ ,  $\theta \rightarrow V_i(\theta, \delta)$  et  $\theta \rightarrow V_n^i(\theta)$  sont de classe  $C^2$  sur  $\mathbb{U}(\Theta)$  (lemme 4); il en est donc de même pour  $U_n(\theta)$  et  $U(\theta, \delta)$ .

Une condition suffisante pour obtenir le point 2 du théorème 3.2.8. est qu'il existe une suite de variables aléatoires  $(Z_n)_n$ , qui converge en probabilité sous  $P$  vers une constante positive  $Z$ , telle que :

$$\forall \eta > 0, \forall (\theta, \beta) \in \Theta^2 : \|\theta - \beta\|_2 \leq \eta, |U_n(\theta) - U_n(\beta)| \leq Z_n \eta$$

Pour tout  $(\theta, \beta)$  tels que  $\|\theta - \beta\|_2 \leq \eta$ , on a :

$$|U_n(\theta) - U_n(\beta)| \leq \sum_{i=1}^p \left| (V_n^i(\theta))^2 - (V_n^i(\beta))^2 \right| \leq 2p \sum_{i=1}^p \left[ \left( \sum_{j=1}^p D_n^{i,j} \right) C_n^i \right] \eta$$

avec, par application du lemme 1 :

$$\begin{aligned} D_n^{i,j} &= \frac{1}{n} \sum_{k=1}^n \sup_{\theta \in \Theta} \left| \frac{\partial}{\partial \theta^j} \mathcal{V}_\delta^i(Y_{k\delta}, Y_{(k-1)\delta}, \theta) \right| \xrightarrow{L^2(P)} D^{i,j} = E_{\theta_0} \left[ \sup_{\theta \in \Theta} \left| \frac{\partial}{\partial \theta^j} \mathcal{V}_\delta^i(Y_\delta, Y_0, \theta) \right| \right] \\ C_n^i &= \frac{1}{n} \sum_{k=1}^n \sup_{\theta \in \Theta} \left| \mathcal{V}_\delta^i(Y_{k\delta}, Y_{(k-1)\delta}, \theta) \right| \xrightarrow{L^2(P)} C^i = E_{\theta_0} \left[ \sup_{\theta \in \Theta} \left| \mathcal{V}_\delta^i(Y_\delta, Y_0, \theta) \right| \right] \end{aligned}$$

où  $C^i$  et  $D^{i,j}$  sont des constantes indépendantes de  $\theta$ . Pour  $Z_n$  et  $Z$  définies par :

$$Z_n = 2p \sum_{i=1}^p \left[ \left( \sum_{j=1}^p D_n^{i,j} \right) C_n^i \right] \text{ et } Z = 2p \sum_{i=1}^p \left[ \left( \sum_{j=1}^p D^{i,j} \right) C^i \right]$$

On obtient donc le résultat voulu avec  $Z_n \xrightarrow{L^1(P)} Z$ .  $\square$

#### 4.1.2 Evaluation de l'écart $\theta_\delta - \theta_0$

**Théorème 3** *Sous (H 2), pour  $H_{\theta_0}$  définie en (H 2.6) et  $N_{\theta_0}$  en (11) et pour  $\delta$ ,  $0 < \delta \leq \delta_0$ , on a :*

$$\theta_\delta = \theta_0 + B_{\theta_0} \delta^2 + o(\delta^2) \text{ avec } B_{\theta_0} = -\frac{1}{12} H_{\theta_0}^{-1} N_{\theta_0}$$

#### Démonstration :

Sous (H 2), à  $\delta$  fixé,  $0 < \delta \leq \delta_0$ ,  $\theta \rightarrow V_i(\theta, \delta)$  est de classe  $C^2$  sur  $\mathbb{U}(\Theta)$ . De plus  $V_i(\theta_\delta, \delta) = 0$ . Par application du théorème des accroissements finis, il existe  $t_i \in ]0, 1[$  tel que  $\theta_{\delta,i} = t_i \theta_0 + (1 - t_i) \theta_\delta$  vérifie :

$$V_i(\theta_\delta, \delta) = 0 = V_i(\theta_0, \delta) + \sum_{j=1}^p \frac{\partial}{\partial \theta^j} V_i(\theta_{\delta,i}, \delta) (\theta_\delta^j - \theta_0^j) \quad (17)$$

D'après le lemme 5, lorsque  $\delta \rightarrow 0$ ,  $\theta_\delta \rightarrow \theta_0$  et donc  $\theta_{\delta,i} \rightarrow \theta_0$ . En utilisant le lemme 4, lorsque  $\delta \rightarrow 0$ , on obtient :

$$\frac{\partial}{\partial \theta^j} V_i(\theta_{\delta,i}, \delta) \rightarrow \frac{\partial}{\partial \theta^j} V_i(\theta_0, 0) = -M_{i,j}(0) \text{ d'où } \frac{\partial}{\partial \theta^j} V_i(\theta_{\delta,i}, \delta) = -M_{i,j}(0) + o(1)$$

D'après (15), (17) s'écrit alors :

$$0 = -\frac{\delta^2}{12} N_{\theta_0} - (H_{\theta_0} + o(1)) (\theta_\delta - \theta_0) + o(\delta^2)$$

L'hypothèse (H 2.6) conduit alors au résultat annoncé.  $\square$

### 4.1.3 Normalité asymptotique de $\widehat{\theta}_n$

On note pour tout  $\delta$ ,  $0 < \delta \leq \delta_0$  :

$$H_{\theta_0}^\delta = \left( \frac{\partial}{\partial \theta^j} V_i(\theta_\delta, \delta) \right)_{i,j} \quad (18)$$

Dans le cas des modèles linéaires en  $\theta$ ,  $H_{\theta_0}^\delta$  et  $-\left[ {}^t \widetilde{X} X \right]^\delta$  coïncident. Par ailleurs, le développement mis en évidence pour  $\left[ {}^t \widetilde{X} X \right]^\delta$  (cf. (12)) reste vrai pour  $H_{\theta_0}^\delta$  dans le cas général :

**Lemme 6** *Sous (H 2), pour  $H_{\theta_0}$  définie en (H 2.6) et  $M_{\theta_0}$  définie en (9), on a :*

$$\forall \delta, 0 < \delta \leq \delta_0, H_{\theta_0}^\delta = -H_{\theta_0} - \frac{\delta}{2} M_{\theta_0} + o(\delta)$$

*De plus, il existe  $\delta_1$ ,  $0 < \delta_1 \leq \delta_0$ , tel que pour tout  $\delta$ ,  $0 < \delta \leq \delta_1$ ,  $H_{\theta_0}^\delta$  est inversible.*

La démonstration de ce lemme est donnée en annexe C. Comme dans le cas linéaire, ce résultat permet d'établir un développement en  $\delta$  de la variance asymptotique de  $\widehat{\theta}_n$ .

**Théorème 4** *Sous (H 2), pour  $H_{\theta_0}$  définie en (H 2.6) et  $P_{\theta_0}$  en (10), pour tout  $\delta$ ,  $0 < \delta \leq \delta_1$ , on a lorsque  $n \rightarrow \infty$  :*

$$\sqrt{n\delta} \left( \widehat{\theta}_n - \theta_\delta \right) \xrightarrow{\mathcal{D}(P)} \mathcal{N}_p(0, V_\delta(\theta_0)), \quad V_\delta(\theta_0) = H_{\theta_0}^{-1} \left( I_p + \frac{\delta}{2} P_{\theta_0} H_{\theta_0}^{-1} + o(\delta) \right)$$

**Démonstration :**

Sous (H 2), pour  $\delta$  fixé,  $0 < \delta \leq \delta_0$ , on sait que  $\theta_\delta \in \overset{\circ}{\Theta}$  et que  $U_n(\theta)$  est de classe  $C^3$  sur  $\mathbb{U}(\Theta)$ . Notons pour tout  $\theta \in \Theta$  :

$$\begin{aligned} DU_n(\theta) &= 2 {}^t (DV_n(\theta)) V_n(\theta), \quad DV_n(\theta) = \left( \frac{\partial}{\partial \theta^j} V_n^i(\theta) \right)_{i,j} \\ D^2U_n(\theta) &= \left( 2 \sum_{i=1}^p \left( \frac{\partial^2}{\partial \theta^k \partial \theta^j} V_n^i(\theta) \right) V_n^i(\theta) + 2 \sum_{i=1}^p \left( \frac{\partial}{\partial \theta^j} V_n^i(\theta) \right) \left( \frac{\partial}{\partial \theta^k} V_n^i(\theta) \right) \right)_{j,k} \\ R_n &= \int_0^1 \left( D^2U_n(\theta_\delta + s(\widehat{\theta}_n - \theta_\delta)) - D^2U_n(\theta_\delta) \right) ds \end{aligned}$$

L'application de la formule de Taylor avec reste intégral donne :

$$\sqrt{n\delta} DU_n(\widehat{\theta}_n) = 0 = \sqrt{n\delta} DU_n(\theta_\delta) + (D^2U_n(\theta_\delta) + R_n) \sqrt{n\delta} (\widehat{\theta}_n - \theta_\delta)$$

Pour  $H_{\theta_0}^\delta$  définie en (18), la normalité asymptotique de  $\widehat{\theta}_n$  résulte des trois convergences suivantes :

$$\sqrt{n\delta} DU_n(\theta_\delta) \xrightarrow{\mathcal{D}(P)} \mathcal{N}_p(0, K(\theta_\delta)), \quad K(\theta_\delta) = 4 {}^t H_{\theta_0}^\delta \Gamma_\delta(\theta_\delta) H_{\theta_0}^\delta \quad (19)$$

$$D^2U_n(\theta_\delta) \xrightarrow{P} 2 {}^t H_{\theta_0}^\delta H_{\theta_0}^\delta, \quad H_{\theta_0}^\delta \text{ étant inversible} \quad (20)$$

$$R_n \xrightarrow{P} 0 \quad (21)$$

Pour (19), il suffit de vérifier par application du lemme 1 que :

$$\sqrt{n\delta}V_n(\theta_\delta) \xrightarrow{\mathcal{D}(P)} \mathcal{N}_p(0, \Gamma_\delta(\theta_\delta)) \text{ et } {}^tDV_n(\theta_\delta) \xrightarrow{L^2(P)} {}^tH_{\theta_0}^\delta$$

De même, la convergence de  $D^2U_n(\theta_\delta)$  est une conséquence directe du lemme 1.

Pour montrer que le reste  $R_n$  tend vers 0, on applique la formule de Rolle. Pour tout  $i = 1, p$  et tout  $j = 1, p$ , on a :

$$\left| \int_0^1 \left[ \frac{\partial^2}{\partial \theta^i \partial \theta^j} U_n(\theta_\delta + s(\hat{\theta}_n - \theta_\delta)) - \frac{\partial^2}{\partial \theta^i \partial \theta^j} U_n(\theta_\delta) \right] ds \right| \leq \sum_{k=1}^p D_n^{i,j,k} |\hat{\theta}_n^k - \theta_\delta^k|$$

avec, par une nouvelle application du lemme 1 :

$$D_n^{i,j,k} = \sup_{\theta \in \Theta} \left| \frac{\partial^3}{\partial \theta^k \partial \theta^j \partial \theta^i} U_n(\theta) \right| \xrightarrow{L^1(P)} D^{i,j,k}$$

$D^{i,j,k}$  étant une constante finie, la convergence en probabilité de  $\hat{\theta}_n$  vers  $\theta_\delta$  entraîne celle de  $R_n$  vers 0.

(19), (20), (21) et le lemme 6 impliquent que pour tout  $\delta$ ,  $0 < \delta \leq \delta_1$  :

$$\sqrt{n\delta}(\hat{\theta}_n - \theta_\delta) \xrightarrow{\mathcal{D}(P_{\theta_0})} \mathcal{N}_p(0, V_\delta(\theta_0)), \quad V_\delta(\theta_0) = (H_{\theta_0}^\delta)^{-1} \Gamma_\delta(\theta_\delta) ({}^tH_{\theta_0}^\delta)^{-1}$$

Pour obtenir le développement de  $V_\delta(\theta_0)$  annoncé, il suffit d'utiliser le lemme 3 qui reste valable, sous (H 2), dans le cas d'une dérive non nécessairement linéaire en  $\theta$  (cf. Annexe B). On obtient ainsi :

$$V_\delta(\theta_0) = (-H_{\theta_0})^{-1} \left[ H_{\theta_0} + \frac{\delta}{2} P_{\theta_0} \right] (-H_{\theta_0})^{-1} + A_\delta(\theta_0) \left[ H_{\theta_0} + \frac{\delta}{2} P_{\theta_0} \right] {}^tA_\delta(\theta_0) + o(\delta),$$

où  $A_\delta(\theta_0) = \left( [I_p + (H_{\theta_0})^{-1} H_{\theta_0}^\delta] (H_{\theta_0}^\delta)^{-1} \right)$

avec, par application du lemme 6 :  $\lim_{\delta \rightarrow 0} \frac{1}{\delta} A_\delta(\theta_0) = \frac{1}{2} H_{\theta_0}^{-1} M_{\theta_0} H_{\theta_0}^{-1} \square$

## 4.2 Adaptation de la méthode au schéma de Simpson

La méthode d'estimation des moments généralisés peut être adaptée au schéma d'approximation de Simpson (5). On définit pour tout  $i = 1, p$  :

$$\begin{aligned} \tilde{V}_n^i(\theta) &= \frac{2}{n} \sum_{k=1}^{\frac{n}{2}} \tilde{\mathcal{V}}_\delta^i(Y_{2k\delta}, Y_{2k\delta-1}, Y_{2(k-1)\delta}, \theta) \\ \tilde{\mathcal{V}}_\delta^i(z, y, x, \theta) &= \frac{\partial}{\partial \theta^i} f_\theta(x) \left( \frac{z - x - \frac{\delta}{3} (f_\theta(x) + 4f_\theta(y) + f_\theta(z))}{2\delta\sigma^2(x)} \right) \end{aligned}$$

Soit  $\tilde{U}_n(\theta) = \sum_{i=1}^p \left( \tilde{V}_n^i(\theta) \right)^2$  et  $\tilde{\theta}_n = \inf_{\theta \in \Theta} \tilde{U}_n(\theta)$  l'estimateur des moments généralisés. Moyennant quelques adaptations dans les hypothèses, on peut montrer que cet estimateur converge vers  $\theta_0$ , à un biais en  $\delta^4$  près, et qu'il est asymptotiquement normal et efficace en variance à un facteur  $(I_p + O(\delta))$  près.

### Hypothèses H 3 : Schéma de Simpson et G.M.M.

Toutes les hypothèses de (H 2) sont maintenues à l'identique à l'exception de (H 2.1) et (H 2.5.3) respectivement remplacées par :

**H 3.1** (H 2.1) mais en supposant  $f_{\theta_0}$  et  $\sigma$  respectivement de classe  $C^8$  et  $C^6$  sur  $\mathbb{R}$ .

**H 3.5.3** Pour tout  $l = 0, 3$ ,  $(A_{\theta_0}^l f_{\theta_0})'$ ,  $A_{\theta_0}^4 f_{\theta_0}$  et  $A_{\theta_0}(\sigma^2 f'_{\theta_0})$  sont à croissance polynomiale en  $x$ , ainsi que celles de  $\sigma$  jusqu'à l'ordre 2.

Notons, pour tout  $i$  :

$$\tilde{N}_i(v) = E_{\theta_0} \left[ \left( \frac{\partial}{\partial \theta^i} f_{\theta_0}(Y_0) \right) \frac{(A_{\theta_0}^4 f_{\theta_0})(Y_v)}{\sigma^2(Y_0)} \right] \quad \text{et} \quad {}^t \tilde{N}_{\theta_0} = \left( \tilde{N}_i(0) \right)_i \quad (22)$$

On définit de plus :

$$\begin{aligned} \tilde{V}_i(\theta, \delta) &= E_{\theta_0} \left[ \tilde{\mathcal{V}}_\delta^i(Y_{2\delta}, Y_\delta, Y_0, \theta) \right] \quad \text{pour } \delta > 0, \text{ et} \\ \tilde{V}_i(\theta, 0) &= E_{\theta_0} \left[ \frac{\partial}{\partial \theta^i} f_\theta(Y_0) \left( \frac{f_{\theta_0}(Y_0) - f_\theta(Y_0)}{\sigma^2(Y_0)} \right) \right] \end{aligned}$$

**Théorème 5** Sous (H 3), pour  $H_{\theta_0}$  définie en (H 3.6),  $P_{\theta_0}$  en (10),  $\tilde{N}_{\theta_0}$  en (22), il existe  $\delta_1 > 0$  tel que pour tout  $\delta$ ,  $0 < \delta \leq \delta_1$ , il existe un unique  $\tilde{\theta}_\delta \in \overset{\circ}{\Theta}$  qui vérifie, pour tout  $i = 1, p$ ,  $\tilde{V}_i(\tilde{\theta}_\delta, \delta) = 0$  et

$$\tilde{\theta}_\delta - \theta_0 = -\frac{\delta^4}{180} H_{\theta_0}^{-1} \tilde{N}_{\theta_0} + o(\delta^4)$$

D'autre part, pour  $\tilde{V}_\delta(\theta_0) = H_{\theta_0}^{-1} (I_p + \delta P_{\theta_0} H_{\theta_0}^{-1} + o(\delta))$ , on a lorsque  $n \rightarrow \infty$  :

$$\tilde{\theta}_n \xrightarrow{P} \tilde{\theta}_\delta \quad \text{et} \quad \sqrt{n\delta} \left( \tilde{\theta}_n - \tilde{\theta}_\delta \right) \xrightarrow{\mathcal{D}(P)} \mathcal{N}_p \left( 0, \tilde{V}_\delta(\theta_0) \right)$$

## 5 Etude expérimentale

Ce paragraphe est consacré à la mise en oeuvre des méthodes précédentes pour trois modèles de diffusion. Pour les deux premiers (Ornstein-Uhlenbeck, (O.U.) et Cox-Ingersoll-Ross, (C.I.R.)) la dérive est linéaire dans le paramètre et affine en  $x$ . Pour le troisième (L.N.), la fonction de dérive  $f_\theta(x) = \theta x \ln(x) + \frac{1}{2}x$  n'est plus linéaire en  $x$ . Nous comparerons les résultats d'estimation associés aux schémas anticipatifs à ceux obtenus par la méthode des moindres carrés et le schéma d'Euler (*cf.* [5]).

Nous indiquerons par  $e$ ,  $t$  et  $s$ , les quantités se rapportant respectivement au schéma d'Euler, du trapèze et de Simpson. On va illustrer expérimentalement quelques unes des propriétés annoncées pour  $\hat{\theta}_n$  : convergence vers  $\theta_\delta$ , écart entre  $\hat{\theta}_n$  et  $\theta_0$ , biais théorique et empirique, variance asymptotique. Pour cela, on simule la diffusion sur un intervalle  $[0, T]$  en utilisant un schéma d'Euler sur une grille fine de pas 0.005. Puis, on estime le paramètre par la méthode des moments généralisés pour les schémas  $t$  et  $s$  et par les moindres carrés pour  $e$ , ceci pour un choix  $(\delta, n)$ ,  $n\delta = T$  avec :  $T = 40$  et  $\delta \in \{0.01, 0.1\}$ ,  $T = 400$  et  $\delta \in \{0.1, 1\}$ ,  $T = 4000$  et  $\delta = 1$ . On répète 100 fois cette expérience afin de calculer la moyenne et l'écart quadratique moyen empiriques notées respectivement  $m(\hat{\theta}_n) = \frac{1}{n} \sum_{i=1}^{100} \hat{\theta}_n^i$  et  $eqm(\hat{\theta}_n) = \frac{1}{100} \sum_{i=1}^{100} (\hat{\theta}_n^i - \theta_0)^2$ .

### 5.1 Processus de Ornstein-Uhlenbeck

On considère l'e.d.s. suivante :

$$dY_t = -\theta_0 Y_t dt + dW_t, Y_0 = x \quad (23)$$

(23) admet une unique solution forte, à trajectoires continues :

$$Y_t = \exp(-\theta_0 t) \left( x + \int_0^t \exp(\theta_0 s) dW_s \right)$$

Si  $\theta_0 > 0$ ,  $(Y_t)_{t \geq 0}$  est récurrente positive sur  $\mathbb{R}$ , de loi invariante  $\mu_{\theta_0} = \mathcal{N}\left(0, \frac{1}{2\theta_0}\right)$ . On vérifie facilement (H 1) et (H 3). On montre de plus que :  $A'_{\theta_0} f_{\theta_0}(x) = (-\theta_0)^{l+1} x$  et

$$\begin{aligned} M(v) &= \frac{\exp(-\theta_0 v)}{2\theta_0}, P(v) = \frac{1}{2\theta_0} \\ N(v) &= \frac{\theta_0^2}{2} \exp(-\theta_0 v), \tilde{N}(v) = \frac{\theta_0^4}{2} \exp(-\theta_0 v) \end{aligned}$$

La variance asymptotique théorique de  $\sqrt{T}(\hat{\theta}_n - \theta_\delta)$  est encore notée  $V_\delta(\theta_0)$ . Pour le schéma d'Euler, on a :

$$\begin{aligned} \hat{\theta}_n^e &= -\frac{1}{\delta} \frac{\sum_{k=1}^n Y_{(k-1)\delta} (Y_{k\delta} - Y_{(k-1)\delta})}{\sum_{k=0}^{n-1} (Y_{k\delta})^2} \\ \theta_\delta^e &= -\frac{1}{\delta} (\exp(-\theta_0 \delta) - 1) = \theta_0 - \frac{\theta_0^2}{2} \delta + o(\delta) \\ V_\delta^e(\theta_0) &= 2\theta_0 (1 - \theta_0 \delta + (\delta)) \end{aligned}$$

Pour le schéma du trapèze, on a :

$$\begin{aligned} \hat{\theta}_n^t &= -\frac{2}{\delta} \frac{\sum_{k=1}^n Y_{(k-1)\delta} (Y_{k\delta} - Y_{(k-1)\delta})}{\sum_{k=1}^n Y_{(k-1)\delta} (Y_{k\delta} + Y_{(k-1)\delta})} \\ \theta_\delta^t &= -\frac{2}{\delta} \left( \frac{\exp(-\theta_0 \delta) - 1}{1 + \exp(-\theta_0 \delta)} \right) = \theta_0 - \frac{\theta_0^3}{12} \delta^2 + o(\delta^2) \\ V_\delta^t(\theta_0) &= 2\theta_0 (1 + o(\delta)) \end{aligned}$$

Pour le schéma de Simpson :

$$\begin{aligned}\widehat{\theta}_n^s &= -\frac{3}{\delta} \frac{\sum_{k=1}^{\frac{n}{2}} Y_{2(k-1)\delta} (Y_{2k\delta} - Y_{2(k-1)\delta})}{\sum_{k=1}^{\frac{n}{2}} Y_{2(k-1)\delta} (Y_{2(k-1)\delta} + 4Y_{2k\delta-1} + Y_{2k\delta})} \\ \theta_\delta^s &= -\frac{3}{\delta} \left( \frac{\exp(-2\delta\theta_0) - 1}{\exp(-2\delta\theta_0) + 4\exp(-\delta\theta_0) + 1} \right) = \theta_0 - \frac{\theta_0^5}{180} \delta^4 + o(\delta^4) \\ V_\delta^s(\theta_0) &= 2\theta_0(1 + o(\delta))\end{aligned}$$

On choisit :  $Y_0 = 0$ ,  $\theta_0 = 1$ . On obtient les trois tableaux suivants :

$\delta$	$\theta_\delta^e$	$\theta_\delta^t$	$\theta_\delta^s$
0.01	0.995017	0.999992	1
0.1	0.951626	0.999167	0.999999
1	0.632121	0.92434	0.995067

Table 1: Modèle de O.U. : comparaison des valeurs  $\theta_\delta$  pour les trois schémas.

	$\delta$	$n$	$m(\widehat{\theta}_n^e)$	$m(\widehat{\theta}_n^t)$	$m(\widehat{\theta}_n^s)$
$T = 40$	0.1	400	0.846248	0.884295	0.928478
	0.01	4000	1.012309	1.017524	1.019922
$T = 400$	1	400	0.638412	0.940662	1.007697
	0.1	4000	0.956130	1.004378	1.004885
$T = 4000$	1	4000	0.634981	0.930612	1.004823

Table 2: Modèle de O.U. : comparaison des moyennes empiriques pour 100 réalisations.

	$\delta$	$n$	$eqm(\widehat{\theta}_n^e)$	$eqm(\widehat{\theta}_n^t)$	$eqm(\widehat{\theta}_n^s)$
$T = 40$	0.1	400	0.03527	0.02707	0.01996
	0.01	4000	0.01295	0.01336	0.01591
$T = 400$	1	400	0.1325	0.01223	0.03111
	0.1	4000	0.00613	0.00515	0.00552
$T = 4000$	1	4000	0.13339	0.005546	0.00276

Table 3: Modèle de O.U. : comparaison des écarts quadratiques moyens pour 100 réalisations.

On constate que les estimateurs associés aux schémas anticipatifs diminuent le biais et l'écart quadratique en comparaison du schéma d'Euler pour l'ensemble, sauf lorsque  $n = 4000$  et  $\delta = 0.01$ .

Lorsque  $\delta = 0.1$  et  $n = 400$ , c'est la valeur relativement faible de  $T$ ,  $T = 40$ , qui explique l'importance du biais empirique. Dans les cas où  $\delta = 1$  ( $T = 400$  ou  $T = 4000$ ), le biais provient principalement de la différence entre  $\theta_\delta$  et  $\theta_0$  : pour les schémas d'Euler et du trapèze, on constate que l'écart entre l'estimation et  $\theta_0$  coïncide avec  $\theta_\delta - \theta_0$ .

L'influence de  $\delta$  et du type de schéma de discrétisation apparait bien dès que  $T$  est suffisamment grand.

## 5.2 Processus de Cox-Ingersoll-Ross

On considère l'e.d.s. suivante :

$$dY_t = b_0 - a_0 Y_t dt + \sqrt{Y_t} dW_t, \quad Y_0 = x$$

$$\theta = (b, a), \quad \sigma(x) = \sqrt{x} \text{ et } f = {}^t(f_1, f_2) \text{ avec } f_1(x) = -x \text{ et } f_2(x) = 1.$$

$f$  est  $C^\infty$  sur  $\mathbb{R}$  et lipschitzienne,  $\sigma$  est  $C^\infty$  sur  $\mathbb{R}_+^*$  et hölderienne de rapport  $\frac{1}{2}$ . L'e.d.s. admet donc une solution forte unique, à trajectoires continues. Si  $b_0 \geq \frac{1}{2}$  et  $a_0 > 0$ , la diffusion est récurrente positive sur  $\mathbb{R}_+^*$ , de loi invariante  $\mu_{\theta_0} = \Gamma(2b_0, 2a_0)$ , admettant des moments de tous ordres. On vérifie sans difficulté que, dès que  $b_0 > 2$ , (H 1) et (H 3) sont satisfaites et que :

$$\begin{aligned} M_{1,1}(v) &= \frac{b_0}{a_0}, \quad M_{1,2}(v) = -1, \quad M_{2,1}(v) = \frac{\exp(-a_0 v) - 2b_0}{2b_0 - 1}, \quad M_{2,2}(v) = \frac{2a_0}{2b_0 - 1} \\ N_1(v) &= 0, \quad N_2(v) = \frac{a_0^3}{2b_0 - 1} \exp(-a_0 v) \\ P_{1,1}(v) &= \frac{b_0}{a_0}, \quad P_{1,2}(v) = P_{2,1}(v) = \frac{\exp(-a_0 v) - 2b_0}{2b_0 - 1}, \quad P_{2,2}(v) = 2 \frac{a_0 b_0 (b_0 - \exp(-a_0 v))}{(2b_0 - 1)(b_0 - 1)} \end{aligned}$$

Pour le schéma d'Euler, on a :

$$\begin{aligned} a_\delta^e &= -\frac{1}{\delta} (\exp(-a_0 \delta) - 1) = a_0 - \frac{a_0^2}{2} \delta + o(\delta) \\ b_\delta^e &= \frac{b_0}{a_0} a_\delta = b_0 - \frac{b_0 a_0}{2} \delta + o(\delta) \end{aligned}$$

Pour le schéma du trapèze :

$$a_\delta^t = -\frac{\delta \exp(-a_0 \delta) - 1}{2 \exp(-a_0 \delta) + 1} = a_0 - \frac{a_0^3}{12} \delta^2 + o(\delta^2), \quad b_\delta^t = \frac{b_0}{a_0} a_\delta^t = b_0 - \frac{b_0 a_0^2}{12} \delta^2 + o(\delta^2)$$

Enfin, pour le schéma de Simpson, on obtient :

$$\begin{aligned} a_\delta^s &= -\frac{3}{\delta} \left( \frac{\exp(-2\delta a_0) - 1}{\exp(-2\delta a_0) + 4 \exp(-\delta a_0) + 1} \right) = a_0 - \frac{a_0^5}{180} \delta^4 + o(\delta^4) \\ b_\delta^s &= \frac{b_0}{a_0} a_\delta^s = b_0 - \frac{b_0 a_0^4}{180} \delta^4 + o(\delta^4) \end{aligned}$$

Bien qu'il soit possible de calculer le développement à l'ordre 1 en  $\delta$  des différentes variances asymptotiques, nous nous contenterons de donner leur terme principal commun :

$$H_{\theta_0}^{-1} = (2b_0 - 1) \begin{pmatrix} \frac{2a_0}{2b_0 - 1} & 1 \\ 1 & \frac{b_0}{a_0} \end{pmatrix}.$$

$\delta$	$a_\delta (a_0 = 1)$			$b_\delta (b_0 = 3)$		
	$a_\delta^e$	$a_\delta^t$	$a_\delta^s$	$b_\delta^e$	$b_\delta^t$	$b_\delta^s$
0.01	0.995	0.999	1	2.985	2.999	3
0.1	0.951	0.999	0.999	2.854	2.997	2.999
1	0.632	0.924	0.995	1.896	2.772	2.985

Table 4: Modèle de C.I.R. : comparaison des valeurs  $(a_\delta, b_\delta)$  pour les trois schémas.

	$\delta$	$n$	$a_n (a_0 = 1)$			$b_n (b_0 = 3)$		
			$m(a_n^e)$	$m(a_n^t)$	$m(a_n^s)$	$m(b_n^e)$	$m(b_n^t)$	$m(b_n^s)$
$T = 40$	0.1	400	1.097	1.1659	1.171	3.235	3.429	3.4447
	0.01	4000	1.0759	1.102	1.099	3.208	3.279	3.268
$T = 400$	1	400	0.644	0.957	1.052	1.932	2.867	3.151
	0.1	4000	0.967	1.016	1.016	2.903	3.050	3.049
$T = 4000$	1	4000	0.633	0.926	1.006	1.897	2.776	3.015

Table 5: Modèle de C.I.R. : comparaison des moyennes empiriques pour 100 réalisations.

	$\delta$	$n$	$a_n (a_0 = 1)$			$b_n (b_0 = 3)$		
			$eqm(a_n^e)$	$eqm(a_n^t)$	$eqm(a_n^s)$	$eqm(b_n^e)$	$eqm(b_n^t)$	$eqm(b_n^s)$
$T = 40$	0.1	400	0.0734	0.0938	0.1	0.581	0.7	0.78
	0.01	4000	0.0869	0.0846	0.086	0.1125	0.72	0.73
$T = 400$	1	400	0.129	0.0147	0.045	1.16	0.134	0.412
	0.1	4000	0.0065	0.0062	0.0082	0.053	0.054	0.065
$T = 4000$	1	4000	0.135	0.0064	0.003	1.219	0.061	0.032

Table 6: Modèle de C.I.R. : comparaison des écarts quadratiques moyens pour 100 réalisations.

On choisit :  $Y_0 = 3$ ,  $(a_0, b_0) = (1, 3)$ . On obtient les trois tableaux suivants :

L'amélioration apportée par le schéma du trapèze par rapport au schéma d'Euler est particulièrement nette pour  $T = 400$  avec  $\delta = 1$  et  $T = 4000$ . Lorsque  $T = 40$ , l'imprécision des résultats obtenus s'expliquent par la faible valeur de  $T$ . L'influence de  $\delta$  sur la variance asymptotique est d'autant plus forte que la discrétisation associée au schéma anticipatif est importante et que la valeur de  $T$  est faible. Ceci explique en particulier que l'écart quadratique moyen pour le schéma de Simpson soit supérieur à celui obtenu pour le schéma du trapèze lorsque  $\delta = 1$  et  $T = 400$ .

### 5.3 Cas d'une dérive non linéaire en $x$ : diffusion Log-Normale

Soit l'e.d.s. :

$$dY_t = \left( -\theta_0 \ln(Y_t) Y_t + \frac{1}{2} Y_t \right) dt + Y_t dW_t, \quad Y_0 = x$$

Cette équation admet une unique solution forte à trajectoires continues (cf. [13], p.125) :

$$Y_t = \exp \left( \exp(-\theta_0 t) \ln(x) + \exp(-\theta_0 t) \int_0^t \exp(\theta_0 s) dW_s \right)$$

On montre que  $\pi_{\theta_0}^t(x, \cdot)$  est une loi log-normale de paramètre  $\left( \ln(x) \exp(-\theta_0 t), \frac{1 - \exp(-2\theta_0 t)}{2\theta_0} \right)$ . Si  $\theta_0 > 0$ , la diffusion est récurrente positive de loi invariante  $\mu_{\theta_0}$  log-normale de paramètres  $\left(0, \frac{1}{2\theta_0}\right)$ . Dans ce cas,  $f_{\theta}(x) = -\theta_0 \ln(x) x + \frac{1}{2}x$  ne vérifie pas l'intégralité de (H 2) et (H 3) : en particulier, l'hypothèse de croissance polynomiale faite sur  $f_{\theta_0}$  n'est pas satisfaite. Cependant, puisque la transition  $\pi_{\theta_0}^t$  est explicite, on peut prouver l'ensemble des régularités requises pour l'obtention des résultats asymptotiques. On a ainsi, en notant  $\alpha(v) = \frac{1 - \exp(-\theta_0 v)}{2\theta_0}$  :

$$\begin{aligned} M(v) &= \left( \frac{\exp(-\theta_0 v)}{2\theta_0} - \alpha^2(v) \right) \exp(\alpha(v)) \\ P(v) &= \left( \frac{1}{2\theta_0} + 4\alpha^2(v) \right) \exp(4\alpha(v)) \\ R(v) &= \alpha(v) \exp(\alpha(v)) \end{aligned}$$

Les estimateurs associés aux trois schémas d'approximation vont converger respectivement vers :

$$\begin{aligned} \theta_{\delta}^e &= \frac{1}{\delta} \frac{R(\delta)}{M(0)} = \theta_0 + \frac{1}{2} (\theta_0 - \theta_0^2) \delta + o(\delta) \\ \theta_{\delta}^t &= \frac{2}{\delta} \left( 1 - \frac{\delta}{4} \right) \frac{R(\delta)}{M(0) + M(\delta)} = \theta_0 + \left( -\frac{\theta_0^3}{12} + \frac{\theta_0^2}{4} - \frac{\theta_0}{16} \right) \delta^2 + o(\delta^2) \\ \theta_{\delta}^s &= \frac{\left(\frac{3}{\delta} - \frac{1}{2}\right) R(2\delta) - 2R(\delta)}{M(0) + 4M(\delta) + M(2\delta)} = \theta_0 + \frac{1}{12} \left( -\frac{\theta_0^5}{15} + \theta_0^4 - \frac{65\theta_0^3}{4} + \frac{13\theta_0^2}{3} - \frac{\theta_0}{48} \right) \delta^4 + o(\delta^4) \end{aligned}$$

Si, comme pour le modèle de C.I.R. les schémas anticipatifs n'améliorent pas sensiblement les résultats obtenus par le schéma d'Euler pour des valeurs importantes de  $\delta$  et faibles de  $T$ , le schéma du trapèze semble un bon choix en situation de discrétisation assez générale. On choisit :  $Y_0 = \exp\left(\frac{1}{4}\right)$  et  $\theta_0 = 1$ . On obtient les trois tableaux suivants :

$\delta$	$\theta_{\delta}^e$	$\theta_{\delta}^t$	$\theta_{\delta}^s$
0.01	0.999979	1	1
0.1	0.998	1.001	1
1	0.867	1.0569	1.0268

Table 7: Diffusion Log-Normale : comparaison des valeurs  $\theta_{\delta}$  pour les trois schémas.

	$\delta$	$n$	$m(\hat{\theta}_n^e)$	$m(\hat{\theta}_n^t)$	$m(\hat{\theta}_n^s)$
$T = 40$	0.1	400	1.100	1.107	1.117
	0.01	4000	1.094	1.094	1.102
$T = 400$	1	400	0.863	1.068	0.682
	0.1	4000	1.005	1.008	1.007
$T = 4000$	1	4000	0.872	1.067	1.104

Table 8: Diffusion Log-Normale : comparaison des moyennes empiriques pour 100 réalisations.

	$\delta$	$n$	$eqm(\hat{\theta}_n^e)$	$eqm(\hat{\theta}_n^t)$	$eqm(\hat{\theta}_n^s)$
$T = 40$	0.1	400	0.048	0.054	0.048
	0.01	4000	0.0638	0.064	0.0735
$T = 400$	1	400	0.03	0.0765	4.85
	0.1	4000	0.0043	0.0048	0.00635
$T = 4000$	1	4000	0.0173	0.011	0.08

Table 9: Diffusion Log-Normale : comparaison des écarts quadratiques moyens pour 100 réalisations.

**Remerciements :** Je tiens à remercier Xavier Guyon pour son encadrement et ses conseils précieux.

## References

- [1] Bergstrom, A.R. (1976) *Statistical inference in Continuous Time Series*.- In *Statistical inference in Continuous Time Economic Models*., Bergstrom, Ed., North Holland, Amsterdam.
- [2] Bibby, B.M. & Sorensen, M. (1995) *Martingale Estimation Functions for Discretely Observed Diffusion Processes*. - Bernoulli 1, 17-39.
- [3] Dacunha-Castelle, D. & Duflo, M. (1993) *Probabilité et Statistiques. Tome 2- 2<sup>eme</sup> Ed.* Masson.
- [4] Dacunha-Castelle, D. & Florens-Zmirou, D. (1986) *Estimation of the coefficient of a diffusion from discrete observations*.- Stochastics 19, 263-284.
- [5] Florens-Zmirou, D. (1989) *Approximate discrete schemes for statistics of diffusion processes*.- Statistics 20, 547-557.
- [6] Gourieroux, C. & Monfort, A. *Statistique et Modèles Econométriques. Tome 1- Economica*.
- [7] Hansen, L (1982) *Large Sample Properties of Generalized Method of Moments Estimators*.- Econometrica, 50 (4), 1029-1054.

- [8] Hansen, L & Singleton, K (1982) *Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models*.- *Econometrica*, 50 (5), 1269-1286.
- [9] Karatzas, I. & Shreve S.E. (1996) *Brownian Motion and Stochastic Calculus*.- 2<sup>nd</sup> Ed. Springer.
- [10] Kessler, M. (1997) *Estimation of an ergodic diffusion from discrete observations*.- *Scand. J. Stat.* 24 (2), 211-229.
- [11] Kessler, M. (1996) *Simple and Explicit Estimating Functions for a Discretely Observed Diffusion Process*.- Research Reports 336, Department of theoretical statistics, University of Aarhus.
- [12] Kessler, M. & Sorensen, M. (1995) *Estimating Equations Based on Eigenfunctions for a Discretely Observed Diffusion Process*.- Research Reports 332, Department of theoretical statistics, University of Aarhus.
- [13] Kloeden, P.E. & Platen, E. (1995) *Numerical Solution of Stochastic Differential Equations*.- Springer.
- [14] Kutoyants, Yu. A. (1984) *Parameter estimation for stochastic processes*.- Research and Exposition in math. 6, Heldermann Verlag, Berlin.
- [15] Liptser, R.S. & Shiryaev, A.N. (1977) *Statistics of random processes. Tome 1,2*- Springer-Verlag.
- [16] Press, W.H., Teukolsky, S.A., Vetterling, W.T. & Flannery, B.P. *Numerical Recipes in C*- 2<sup>nd</sup> Ed. Cambridge University Press, 132-133.
- [17] Prakasa-Rao, B. L. S. (1983) *Asymptotic theory for non linear least squares estimator for diffusion proceses*.- *Math. Operationsforsch. Statist Ser. Berlin* 14 (2), 195-209.
- [18] Pedersen, A. R. (1995) *Consistency and asymptotic normality of an approximate maximum likelihood estimator for discretely observed diffusion processes*.- *Bernoulli* 1, 257-279.
- [19] Pedersen, A. R. (1995) *A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations*.- *Scand. J. Statist.* 22, 55-71.
- [20] Sargan, J.D. (1976) *Some discrete approximations to continuous times stochastic models*- In *Statistical inference in Continuous Time Economic Models*., Bergstrom, Ed., North Holland, Amsterdam, 27-80.
- [21] Sorensen, M. (1996) *Estimating functions for discretely observed diffusions: A review*.- Research Reports 348, Department of theoretical statistics, University of Aarhus.
- [22] Yoshida, N. (1992) *Estimation for diffusion processes from discrete observations*.- *J. Multivariate Anal.* 41, 220-242.

## Annexe A : démonstration du lemme 4

La démonstration du lemme 4 repose sur les trois lemmes suivants : considérons  $(\theta, x) \rightarrow h(\theta, x)$  et  $(\theta, x) \rightarrow g(\theta, x)$  deux fonctions définies sur  $\mathbb{U}(\Theta) \times \mathbb{R}$  et notons  $m(\theta, \delta) = E_{\theta_0} [h(\theta, Y_0) g(\theta, Y_\delta)]$ .

**Lemme 7** *Supposons que  $h$  et  $g$  soient continues sur  $\mathbb{U}(\Theta) \times \mathbb{R}$  et à croissance polynomiale en  $x$  uniformément en  $\theta$  sur  $\mathbb{U}(\Theta)$ . Sous (H 2.1), (H 2.2) et si  $\mu_{\theta_0}$  admet des moments de tous ordres,  $m$  est continue sur  $\mathbb{U}(\Theta) \times \mathbb{R}^+$*

Ce résultat se déduit du théorème de Lebesgue qui s'applique du fait de la majoration (8).

**Lemme 8** *On suppose que, pour tout  $x \in \mathbb{R}$ ,  $h(\cdot, x)$  et  $g(\cdot, x)$  sont de classe  $C^m$  sur  $\mathbb{U}(\Theta)$ , ces fonctions ainsi que leur dérivées par rapport à  $\theta$  étant à croissance polynomiale en  $x$  uniformément en  $\theta$  sur  $\mathbb{U}(\Theta)$ .*

*Alors, sous (H 2.1), (H 2.2) et si  $\mu_{\theta_0}$  admet des moments de tous ordres, pour  $\delta \in \mathbb{R}^+$ ,  $\theta \rightarrow m(\theta, \delta)$  est de classe  $C^m$  sur  $\mathbb{U}(\Theta)$  et ses dérivées sont obtenues par dérivation sous le signe  $E_{\theta_0}$ .*

La démonstration de ce lemme repose également sur l'application du théorème de Lebesgue.

**Lemme 9** *Supposons que, pour tout  $\theta \in \mathbb{U}(\Theta)$ ,  $x \rightarrow g(\theta, x)$  soit de classe  $C^2$  sur  $\mathbb{R}$ ,  $h$ ,  $g$  et  $A_{\theta_0}g$  étant à croissance polynomiale en  $x$  uniformément en  $\theta$  sur  $\mathbb{U}(\Theta)$ .*

*Alors, sous (H 2.1), (H 2.2) et si  $\mu_{\theta_0}$  admet des moments de tous ordres,  $\delta \rightarrow m(\theta, \delta)$  est de classe  $C^1$  sur  $\mathbb{R}_*^+$  et dérivable à droite en 0 avec :  $m'(\theta, \delta) = E_{\theta_0} [h(\theta, Y_0) (A_{\theta_0}g)(\theta, Y_\delta)]$*

Pour montrer ce résultat, on utilise la formule de Ito appliquée à  $g(\theta, Y_u)$  et à nouveau le théorème de Lebesgue avec la majoration (8).

### • Etude de la différentiabilité en $\theta$ :

Sous (H 2.5.1),  $\mathcal{V}_\delta^i$  satisfait les conditions du lemme 8 pour  $m = 2$ . Ainsi, pour tout  $\delta \in \mathbb{R}^+$ ,  $\theta \rightarrow V_i(\theta, \delta)$  est de classe  $C^2$  sur  $\mathbb{U}(\Theta)$ .

La continuité de  $V_i$  et de ses dérivées en  $\theta$  sur  $\mathbb{U}(\Theta) \times \mathbb{R}_*^+$  est une conséquence directe de l'application du lemme 7 sous (H 2.5.1).

En revanche, la continuité de ces fonctions aux points de la forme  $(\theta, 0)$  doit être étudiée séparément. On utilise alors la décomposition du trapèze (2). Pour  $V_i$  par exemple, on a pour tout  $\theta \in \mathbb{U}(\Theta)$  et tout  $\delta > 0$  :

$$\begin{aligned} V_i(\theta, \delta) &= -\frac{\delta^2}{2} \int_0^1 (1-u) u E_{\theta_0} \left[ \frac{\partial}{\partial \theta^i} f_\theta(Y_0) \frac{(A_{\theta_0}^2 f_{\theta_0})(Y_{u\delta})}{\sigma^2(Y_0)} \right] du \\ &+ \frac{1}{2} E_{\theta_0} \left[ \frac{\partial}{\partial \theta^i} f_\theta(Y_0) \left( \frac{f_{\theta_0}(Y_0) - f_\theta(Y_0)}{\sigma^2(Y_0)} \right) \right] \\ &+ \frac{1}{2} E_{\theta_0} \left[ \frac{\partial}{\partial \theta^i} f_\theta(Y_0) \left( \frac{f_{\theta_0}(Y_\delta) - f_\theta(Y_\delta)}{\sigma^2(Y_0)} \right) \right] \end{aligned}$$

Les trois fonctions précédentes sont continues sur  $\mathbb{U}(\Theta) \times \mathbb{R}^+$  (lemme 7). Pour toute suite  $((\theta_n, \delta_n))_n \subset \mathbb{U}(\Theta) \times \mathbb{R}_*^+$  convergeant vers  $(\theta, 0)$ , on obtient :

$$\lim_{n \rightarrow \infty} V_i(\theta_n, \delta_n) = E_{\theta_0} \left[ \frac{\partial}{\partial \theta^i} f_\theta(Y_0) \left( \frac{f_{\theta_0}(Y_0) - f_\theta(Y_0)}{\sigma^2(Y_0)} \right) \right] = V_i(\theta, 0)$$

$V_i$  est ainsi globalement continue sur  $\mathbb{U}(\Theta) \times \mathbb{R}^+$ . On procède de même pour ses dérivées partielles.

• **Etude de la dérivabilité en  $\delta$  :**

Utilisant (H 2.5.1), (H 2.5.2), et le lemme 9, on montre que, pour tout  $\theta \in \mathbb{U}(\Theta)$ ,  $\delta \rightarrow V_i(\theta, \delta)$  est dérivable sur  $\mathbb{R}_*^+$  avec :

$$\begin{aligned} \forall \delta > 0, V_i'(\theta, \delta) &= -\frac{1}{\delta^2} E_{\theta_0} \left[ \frac{\partial}{\partial \theta^i} f_\theta(Y_0) \left( \frac{Y_\delta - Y_0}{\sigma^2(Y_0)} \right) \right] \\ &+ \frac{1}{\delta} E_{\theta_0} \left[ \frac{\partial}{\partial \theta^i} f_\theta(Y_0) \frac{f_{\theta_0}(Y_\delta)}{\sigma^2(Y_0)} \right] - \frac{1}{2} E_{\theta_0} \left[ \frac{\partial}{\partial \theta^i} f_\theta(Y_0) \frac{(A_{\theta_0} f_\theta)(Y_\delta)}{\sigma^2(Y_0)} \right] \end{aligned}$$

Sous ces mêmes hypothèses,  $V_i'$  est de plus continue sur  $\mathbb{U}(\Theta) \times \mathbb{R}_*^+$  (lemme 7).

La dérivabilité à droite en 0 de  $\delta \rightarrow V_i(\theta, \delta)$  repose sur (2) et sur la formule de Ito. On a :

$$\begin{aligned} \frac{1}{\delta} (V_i(\theta, \delta) - V_i(\theta, 0)) &= -\frac{\delta}{2} \int_0^1 (1-u) u E_{\theta_0} \left[ \frac{\partial}{\partial \theta^i} f_\theta(Y_0) \frac{(A_{\theta_0}^2 f_\theta)(Y_{u\delta})}{\sigma^2(Y_0)} \right] du \\ &+ \frac{1}{2} \int_0^1 E_{\theta_0} \left[ \frac{\partial}{\partial \theta^i} f_\theta(Y_0) \frac{A_{\theta_0}(f_{\theta_0} - f_\theta)(Y_{u\delta})}{\sigma^2(Y_0)} \right] du \end{aligned}$$

Sous (H 2.5.2) et (H 2.5.3), le lemme 7 implique que :

$$\lim_{h \rightarrow 0} \frac{1}{\delta} (V_i(\theta, \delta) - V_i(\theta, 0)) = \frac{1}{2} E_{\theta_0} \left[ \frac{\partial}{\partial \theta^i} f_\theta(Y_0) \frac{A_{\theta_0}(f_{\theta_0} - f_\theta)(Y_0)}{\sigma^2(Y_0)} \right]$$

$V_i'(\theta, 0^+) = \frac{1}{2} E_{\theta_0} \left[ \frac{\partial}{\partial \theta^i} f_\theta(Y_0) \frac{A_{\theta_0}(f_{\theta_0} - f_\theta)(Y_0)}{\sigma^2(Y_0)} \right]$  existe donc. De plus, sous (H 2.5.2) et par application du lemme 7, on établit que cette fonction est de classe  $C^1$  sur  $\mathbb{U}(\Theta)$ .

Comme dans le paragraphe précédent, la continuité de  $(\theta, \delta) \rightarrow V_i'(\theta, \delta)$  aux points de la forme  $(\theta, 0)$  repose sur (2). On montre que :  $\forall ((\theta_n, \delta_n))_n \subset \mathbb{U}(\Theta) \times \mathbb{R}_*^+$  telle que  $(\theta_n, \delta_n) \rightarrow (\theta, 0)$ ,  $\lim_{n \rightarrow \infty} V_i'(\theta_n, \delta_n) = V_i'(\theta, 0^+)$ .

• **Equations (15) et (16) :**

Utilisant la décomposition du trapèze (2), on obtient :

$$V_i(\theta_0, \delta) = -\frac{\delta^2}{2} \int_0^1 (1-u) u N_i(u\delta) du$$

Par continuité de  $v \rightarrow N_i(v) : V_i(\theta_0, \delta) = -\frac{\delta^2}{12}N_i(0) + o(\delta^2)$ .

Pour (16), on se base de nouveau sur (2). On a ainsi :

$$\begin{aligned} \forall \delta > 0, \frac{\partial}{\partial \theta^j} V_i(\theta_0, \delta) &= -\frac{\delta^2}{2} \int_0^1 (1-u) u E_{\theta_0} \left[ \frac{\partial^2}{\partial \theta^j \partial \theta^i} f_{\theta_0}(Y_0) \frac{(A_{\theta_0}^2 f_{\theta_0})(Y_{\delta u})}{\sigma^2(Y_0)} \right] du \\ &\quad - \frac{1}{2} E_{\theta_0} \left[ \frac{\partial}{\partial \theta^i} f_{\theta_0}(Y_0) \left( \frac{\frac{\partial}{\partial \theta^j} f_{\theta_0}(Y_0) + \frac{\partial}{\partial \theta^j} f_{\theta_0}(Y_\delta)}{\sigma^2(Y_0)} \right) \right] \end{aligned}$$

Or, d'après le lemme 7:

$$\lim_{\delta \rightarrow 0} \int_0^1 (1-u) u E_{\theta_0} \left[ \frac{\partial^2}{\partial \theta^j \partial \theta^i} f_{\theta_0}(Y_0) \frac{(A_{\theta_0}^2 f_{\theta_0})(Y_{\delta u})}{\sigma^2(Y_0)} \right] du = \frac{1}{6} E_{\theta_0} \left[ \frac{\partial^2}{\partial \theta^j \partial \theta^i} f_{\theta_0}(Y_0) \frac{(A_{\theta_0}^2 f_{\theta_0})(Y_0)}{\sigma^2(Y_0)} \right].$$

Donc :  $\forall \delta > 0, \frac{\partial}{\partial \theta^j} V_i(\theta_0, \delta) = -\frac{1}{2} (M_{i,j}(\delta) + M_{i,j}(0)) + O(\delta^2)$ .  $\square$

## Annexe B : démonstration du lemme 3

Que la dérive soit linéaire en  $\theta$  ou non, le lemme 1 implique que  $\sqrt{n\delta} \left( {}^t \tilde{X} Y - {}^t \tilde{X} X \theta_\delta \right)$  et plus généralement  $\sqrt{n\delta} V_n(\theta_\delta)$  converge sous  $P$  vers une loi  $\mathcal{N}_p(0, \Gamma_\delta(\theta_\delta))$ , avec  $\Gamma_\delta(\theta_\delta) = (\Gamma_\delta^{i,j}(\theta_\delta))_{i,j}$  où, pour  $U_\delta$  définie au lemme 1 :

$$\begin{aligned} \Gamma_\delta^{i,j}(\theta_\delta) &= T_\delta^{i,j}(\theta_\delta) + R_\delta^{i,j}(\theta_\delta) \\ T_\delta^{i,j}(\theta_\delta) &= E_{\theta_0} [\delta \mathcal{V}_\delta^i(Y_\delta, Y_0, \theta_\delta) \mathcal{V}_\delta^j(Y_\delta, Y_0, \theta_\delta)] \\ R_\delta^{i,j}(\theta_\delta) &= E_{\theta_0} [\delta \mathcal{V}_\delta^i(Y_\delta, Y_0, \theta_\delta) [U_\delta(\mathcal{V}_\delta^j)](Y_\delta)] \end{aligned}$$

**Etude de  $T_\delta^{i,j}(\theta_\delta)$  :** montrons que  $T_\delta^{i,j}(\theta_\delta) = M_{i,j}(0) + \frac{\delta}{2} P'_{i,j}(0) + o(\delta)$ .

Sous (H 2), le lemme 8 implique que  $\theta \rightarrow T_\delta^{i,j}(\theta)$  est de classe  $C^3$  sur  $\mathbb{U}(\Theta)$ , ses dérivées étant obtenues par dérivation sous l'espérance. Par application de la formule de Taylor-Young, on obtient ainsi :

$$T_\delta^{i,j}(\theta_\delta) = T_\delta^{i,j}(\theta_0) + \sum_{k=1}^p \left( \frac{\partial}{\partial \theta^k} T_\delta^{i,j}(\theta_0) \right) (\theta_\delta^k - \theta_0^k) + \varepsilon(\theta_\delta - \theta_0) \|\theta_\delta - \theta_0\|_2$$

où  $\varepsilon(h) \rightarrow 0$ , lorsque  $h \rightarrow 0$ .

Si on montre que  $T_\delta^{i,j}(\theta_0) = M_{i,j}(0) + \frac{\delta}{2} P'_{i,j}(0) + o(\delta)$  et que, pour  $k = 1, p, \frac{\partial}{\partial \theta^k} T_\delta^{i,j}(\theta_0)$  est bornée, nous aurons alors le développement annoncé car  $\theta_\delta - \theta_0 = O(\delta^2)$ .

• D'après la décomposition du trapèze (2), on a :

$$\begin{aligned} T_\delta^{i,j}(\theta_0) &= \frac{1}{\delta} E_{\theta_0} \left[ \left( \frac{\partial}{\partial \theta^i} f_{\theta_0}(Y_0) \right) \left( \frac{\partial}{\partial \theta^j} f_{\theta_0}(Y_0) \right) \left( \frac{\xi_0}{\sigma^2(Y_0)} \right)^2 \right] \\ &+ \frac{1}{\delta} E_{\theta_0} \left[ \left( \frac{\partial}{\partial \theta^i} f_{\theta_0}(Y_0) \right) \left( \frac{\partial}{\partial \theta^j} f_{\theta_0}(Y_0) \right) \left( \frac{\eta_0}{\sigma^2(Y_0)} \right)^2 \right] \\ &+ \frac{2}{\delta} E_{\theta_0} \left[ \left( \frac{\partial}{\partial \theta^i} f_{\theta_0}(Y_0) \right) \left( \frac{\partial}{\partial \theta^j} f_{\theta_0}(Y_0) \right) \frac{\eta_0 \xi_0}{\sigma^2(Y_0)} \right] \end{aligned}$$

Or, le lemme 2 entraîne que :  $\frac{1}{\delta} E_{\theta_0} \left[ \left( \frac{\partial}{\partial \theta^i} f_{\theta_0}(Y_0) \right) \left( \frac{\partial}{\partial \theta^j} f_{\theta_0}(Y_0) \right) \left( \frac{\xi_0}{\sigma^2(Y_0)} \right)^2 \right] = M_{i,j}(0) + \frac{\delta}{2} P'_{i,j}(0^+) + o(\delta)$ . De plus, par deux applications successives de l'inégalité de Schwarz, on a :  $\left| E_{\theta_0} \left[ \left( \frac{\partial}{\partial \theta^i} f_{\theta_0}(Y_0) \right) \left( \frac{\partial}{\partial \theta^j} f_{\theta_0}(Y_0) \right) \left( \frac{\eta_0}{\sigma^2(Y_0)} \right)^2 \right] \right| \leq O(\delta^6)$ . Enfin, on montre aisément grâce à l'inégalité de Schwarz et aux résultats précédents que le double produit restant est un  $o(\delta^3)$ . On a donc le résultat attendu pour  $T_\delta^{i,j}(\theta_0)$ .

• Utilisant à nouveau (2) et par des arguments similaires à ceux développés précédemment, on obtient :

$$\forall k = 1, p, \frac{\partial}{\partial \theta^k} T_\delta^{i,j}(\theta_0) = E_{\theta_0} \left[ \frac{\partial^2}{\partial \theta^k \partial \theta^i} f_{\theta_0}(Y_0) \frac{\frac{\partial}{\partial \theta^j} f_{\theta_0}(Y_0)}{\sigma^2(Y_0)} + \frac{\partial^2}{\partial \theta^k \partial \theta^j} f_{\theta_0}(Y_0) \frac{\frac{\partial}{\partial \theta^i} f_{\theta_0}(Y_0)}{\sigma^2(Y_0)} \right] + o(1).$$

**Etude de  $R_\delta^{i,j}(\theta_\delta)$  :** nous allons montrer que :  $R_\delta(\theta_\delta) = o(\delta)$ .

Notons  $v_\delta^j(x) = E[\mathcal{V}_\delta^j(Y_\delta, Y_0, \theta_\delta) \mid Y_0 = x]$ . D'après la propriété 1 du trapèze, on a :

$$\begin{aligned} \mathcal{V}_\delta^j(Y_\delta, Y_0, \theta_\delta) &= \frac{\partial}{\partial \theta^j} f_{\theta_\delta}(Y_0) \left[ \frac{f_{\theta_0}(Y_0) - f_{\theta_\delta}(Y_0)}{\sigma^2(Y_0)} + \frac{f_{\theta_0}(Y_\delta) - f_{\theta_\delta}(Y_\delta)}{\sigma^2(Y_0)} \right] \\ &+ \frac{\partial}{\partial \theta^j} f_{\theta_\delta}(Y_0) \left[ \frac{\eta_0}{\delta \sigma^2(Y_0)} + \frac{\xi_0}{\delta \sigma^2(Y_0)} \right] \end{aligned}$$

D'où :

$$\begin{aligned} v_\delta^j(x) &= \frac{\partial}{\partial \theta^j} f_{\theta_\delta}(x) \left[ \frac{f_{\theta_0}(x) - f_{\theta_\delta}(x)}{\sigma^2(x)} + \frac{\pi_{\theta_0}^\delta(f_{\theta_0} - f_{\theta_\delta})(x)}{\sigma^2(x)} \right] \\ &- \delta^2 \frac{\partial}{\partial \theta^j} f_{\theta_\delta}(x) \frac{\int_0^1 (1-u) u \pi_{\theta_0}^{u\delta}(A_{\theta_0}^2 f_{\theta_0})(x) du}{2\sigma^2(x)} \end{aligned}$$

Compte tenu des hypothèses de croissance polynomiale uniforme en  $\theta$  faites sur les fonctions et de l'écart en  $\delta^2$  calculé entre  $\theta_\delta$  et  $\theta_0$ , il existe une constante  $K$  finie telle que :  $\|v_\delta^j(\theta_\delta)\|_{L^2(\mu_{\theta_0})} \leq K\delta^2$ .

De plus, par définition de  $\theta_\delta$ , on a :  $\mu_{\theta_0}(v_\delta^j(\theta_\delta)) = 0$ . D'après la propriété de mélange (7), on obtient alors :

$$\|U_\delta(\mathcal{V}_\delta^j)\|_{L^2(\mu_{\theta_0})} \leq \sum_{k=0}^{+\infty} \|\pi_{\theta_0}^{k\delta}(v_\delta^j(\theta_\delta))\|_{L^2(\mu_{\theta_0})} \leq K \frac{\delta^2}{1 - \exp(-\lambda\delta)} \leq O(\delta)$$

On a donc :

$$R_\delta^{i,j}(\theta_\delta) \leq \sqrt{\delta} \sqrt{T_\delta^{i,i}(\theta_\delta)} \|U_\delta(\mathcal{V}_\delta^j)\|_{L^2(\mu_{\theta_0})} \leq O\left(\delta^{\frac{3}{2}}\right)$$

Puisque  $T_\delta^{i,j}(\theta_\delta) = M_{i,j}(0) + \frac{\delta}{2}P'_{i,j}(0) + o(\delta)$  et  $R_\delta^{i,j}(\theta_\delta) = o(\delta)$ , on obtient ainsi :

$$\Gamma_\delta(\theta_\delta) = H_{\theta_0} + \frac{\delta}{2}P_{\theta_0} + o(\delta). \quad \square$$

## Annexe C : démonstration du lemme 6

Pour tout  $\delta > 0$  fixé et tout couple  $(i, j) \in \{1, \dots, p\}^2$ ,  $\theta \rightarrow \frac{\partial}{\partial \theta^j} V_i(\theta, \delta)$  est de classe  $C^1$  sur  $\mathbb{U}(\Theta)$  (lemme 4). Par application de la formule de Taylor-Young, on a alors :

$$\frac{\partial}{\partial \theta^j} V_i(\theta_\delta, \delta) = \frac{\partial}{\partial \theta^j} V_i(\theta_0, \delta) + \sum_{k=1}^p \frac{\partial^2}{\partial \theta^k \partial \theta^j} V_i(\theta_0, \delta) (\theta_\delta^k - \theta_0^k) + \varepsilon(\theta_\delta - \theta_0) \|\theta_\delta - \theta_0\|_2 \quad (24)$$

avec  $\lim_{h \rightarrow 0} \varepsilon(h) = 0$ .

Or, d'après (16) et le lemme 2 :  $\frac{\partial}{\partial \theta^j} V_i(\theta_0, \delta) = -M_{i,j}(0) - \frac{\delta}{2}M'_{i,j}(0^+) + o(\delta)$ . De plus, par application du lemme 4, on obtient :  $\lim_{\delta \rightarrow 0} \frac{\partial^2}{\partial \theta^k \partial \theta^j} V_i(\theta_0, \delta) = \frac{\partial^2}{\partial \theta^k \partial \theta^j} V_i(\theta_0, 0)$ . Enfin, on a :  $\theta_\delta - \theta_0 = O(\delta^2)$ . On en déduit ainsi que :

$$\lim_{\delta \rightarrow 0} \frac{1}{\delta} \left[ \frac{\partial}{\partial \theta^j} V_i(\theta_\delta, \delta) + M_{i,j}(0) \right] = -\frac{1}{2}M'_{i,j}(0^+) \quad \text{donc } H_{\theta_0}^\delta = -H_{\theta_0} - \frac{\delta}{2}M_{\theta_0} + o(\delta)$$

D'après (H 2.6),  $H_{\theta_0}$  est inversible. Donc, le développement précédent implique qu'il existe  $\delta_1$ ,  $0 < \delta_1 \leq \delta_0$ , tel que pour tout  $\delta$ ,  $0 < \delta \leq \delta_1$ ,  $H_{\theta_0}^\delta$  est inversible.  $\square$