THÈSE DE DOCTORAT DE MATHÉMATIQUES

UNIVERSITÉ DE MARNE–LA–VALLÉE

Arnaud GLOTER

Estimation des paramètres d'une diffusion cachée:

intégrales de processus de diffusion et modèles à volatilité stochastique.

Thèse soutenue le 14 Janvier 2000 devant le jury composé de:

Mme Valentine GENON–CATALOT	(Directeur)
Mme Dominique PICARD	
Mme Monique PONTIER	(Rapporteur)
M. Jean JACOD	
M. Damien LAMBERTON	
M. Gilles PAGES	
M. Michael SØRENSEN	(Rapporteur)

Je tiens tout d'abord à remercier Valentine Genon–Catalot pour la disponibilité et la grande gentillesse dont elle a fait preuve durant l'élaboration de cette thèse. Par la pertinence de ses observations, elle a su parfaitement orienter mes recherches. Sa relecture minutieuse de ce manuscript a aussi été une aide précieuse.

J'exprime ma gratitude à Monique Pontier et Michael Sørensen pour avoir consacré du temps à rapporter cette thèse.

Je suis très honoré que Dominique Picard, Jean Jacod, Damien Lamberton et Gilles Pagès aient accepté d'être membres du jury. Je les en remercie vivement.

Je remercie parmi l'équipe de Marne–la–Vallée, tout ceux qui par leur bonne humeur ont contribué à rendre l'ambiance agréable.

Je remercie, enfin, ma famille pour son soutien, indirect, à la rédaction de cette thèse.

Table des matières

Introduction

Considérons un processus de diffusion unidimensionnel (X_t) défini par une équation differentielle stochastique de la forme:

$$dX_t = b(X_t, \mu, \sigma)dt + a(X_t, \sigma)dB_t, \quad X_0 = \eta \tag{1}$$

où B est un mouvement brownien standard, et les fonctions b et a dépendent de paramètres inconnus μ, σ. Le problème étudié dans cette thèse est celui de l'estimation de ces paramètres, à partir de deux types d'observations pour lesquelles la trajectoire X n'est pas observée directement, mais se trouve cachée.

D'une part (Partie I de la thèse), nous considérons une observation discrétisée de pas Δ du processus

$$I_t = \int_0^t X_s ds. \tag{2}$$

D'autre part (Partie II), nous étudions le cas de modèles définis par:

$$dY_t = \rho(X_t, t)dt + \sigma_t dW_t, \quad Y_0 = \eta' \tag{3}$$

où $X_t = \sigma_t^2$ est une diffusion positive vérifiant une équation differentielle stochastique de la forme (1), W est un autre mouvement brownien pouvant être corrélé avec B et ρ une fonction connue ou inconnue mais dont la connaissance présente peu d'intérêt. Nous supposons disposer d'une observation discrétisée de la trajectoire (Y_t) uniquement.

Les deux cas correspondent à des modèles de diffusions bidimensionnels pour lesquels on n'observe que l'une des coordonnées, l'autre demeurant cachée et inaccessible à l'observation.

- en première partie, l'observation est (I_t) avec:

$$\begin{cases} dI_t = X_t dt \\ dX_t = b(X_t, \mu, \sigma)dt + a(X_t, \sigma)dB_t \end{cases}$$

- en seconde partie, l'observation est (Y_t) avec:

$$\begin{cases} dY_t = \rho(X_t, t)dt + \sigma_t dW_t \\ \sigma_t^2 = X_t, \quad dX_t = b(X_t, \mu, \sigma)dt + a(X_t, \sigma)dB_t. \end{cases}$$

L'origine et la motivation de cette étude proviennent des modèles stochastiques utilisés en finance et des problèmes qu'ils soulèvent tant en probabilité qu'en statistique.

En effet, en 1973, F. Black et M. Scholes suggèrent de modéliser le prix S_t d'un actif financier par un mouvement brownien géométrique:

$$\frac{dS_t}{S_t} = \mu dt + \sigma dB_t$$

où μ et σ sont des constantes et σ est la volatilité de l'actif. (Par consequent, $Y_t = \log S_t$ vérifie $dY_t = (\mu - \frac{\sigma^2}{2})dt + \sigma dB_t$).

De ce modèle, il déduisent la très célèbre "formule de Black et Scholes" d'évaluation des options européennes. Depuis lors, des constatations empiriques portant sur les prix observés d'options, ou sur les queues de distributions des prix d'actifs ont conduit à rejeter le modèle à volatilité constante et d'autres modèles ont été proposés par Hull et White (1987) où σ^2 n'est plus une constante, mais est une autre diffusion $X_t = \sigma_t^2$, positive, régie par un autre mouvement brownien. Ces modèles sont connus sous le nom de "modèles à volatilité stochastique". Notons cependant que, dans la littérature, cette dénomination recouvre souvent aussi les modèles à temps discret du type ARCH introduits par Engle (1982) toujours pour améliorer le modèle de Black et Scholes en tenant compte des propriétés observées des queues de distributions d'actifs. Les deux approches temps continu–temps discret ne sont pas sans rapport puisque les modèles à volatilité stochastique continus peuvent être obtenus comme approximations diffusions de modèles ARCH (Nelson (1990)).

Cette dénomination (volatilité stochastique) est aussi employée lorsque $\sigma = \sigma(S_t)$ ne dépend que de l'actif (il n'y a qu'un seul mouvement brownien) ou encore lorsque $\sigma = \sigma_t$ est un processus stochastique autre qu'une diffusion (voir par exemple Barndorff–Nielsen et Shephard (1998), Drost et Werker (1996), Frey et Runggaldier (1999)).

Dans un modèle à volatilité stochastique de type Hull et White (1987), les prix d'options dépendent des lois des intégrales $I_t = \int_0^t X_s ds$. Sur le plan statistique, les données disponibles sont:

- les prix d'options observés sur les marchés financiers
- les prix des actifs eux–mêmes.

A partir des prix d'options, il est possible de reconstituer des données pour les intégrales $\int_0^t X_s ds$ à differents instants t (cf Patorello *et al.* (1994)). Ceci fournit une première justification à notre partie I. On peut y ajouter la modélisation de phénomènes d'accumulation (fatigue de machine, niveau de cours d'eau) (cf Lefebvre (1997)): X est la vitesse d'évolution du phénomène, et I son niveau qui est observable.

Enfin, dans notre cas, la partie I a été le soubassement de la partie II. En effet, les deux types de modèles sont liés puisque la variation quadratique de Y_t dans le modèle (3) est justement $< Y >_t = I_t$.

Avant d'entrer dans les détails des chapitres, précisons que ceux–ci ont été écrits pour être lus de façon autonome. Ainsi, chaque chapitre contient les rappels des chapitres précédents nécessaires à sa compréhension ce qui implique des redites pour l'ensemble du texte de la thèse.

Partie I. Statistique basée sur l'observation discrétisée de l'intégrale d'une diffusion.

Le problème statistique d'estimation des paramètres d'une diffusion multidimensionnelle a été étudié par de nombreux auteurs. Parmi eux, citons, quoique la liste soit loin d'être exhaustive, Bibby et Sørensen (1995), Dacunha–Castelle et Florens–Zmirou (1986), Genon–Catalot et Jacod (1993), Kessler (1997), Kessler et Sørensen (1999), Kutoyants (1984). Cependant aucun de leur résultats ne s'appliquent à la diffusion (I_t, X_t). Nous n'avons trouvé aucune référence statistiques pour le couple (I_t, X_t).

Le modèle répond aux équations (1)–(2). L'échantillon observé $(I_{i\Delta}, i = 1, \ldots, n)$ n'est en fait plus markovien. C'est un modèle de Markov caché: la chaîne cachée est ici $(I_{i\Delta}, X_{i\Delta})$. De nombreux auteurs se sont intéressés à de tels modèles (cf, Leroux (1992), Ryden (1994), Bickel et Ritov (1996), par exemple). Toutefois les références statistiques sur le modèles de Markov cachés supposent, pour la plupart, que la chaîne cachée est à espace d'état fini.

Pour simplifier le problème et préparer d'éventuelles généralisations, nous avons étudié, en premier lieu, le cas particulier où X est un processus d'Ornstein–Uhlenbeck.

Chapitre I.1. Estimation de paramètres pour une discrétisation de l'intégrale d'un processus d'Ornstein–Uhlenbeck.

L'intérêt de cet exemple particulier est que l'on peut mener tous les calculs de façon explicite. Considérons (X_t) le processus solution stationnaire de $dX_t = \mu X_t dt + \sigma dB_t$ et introduisons pour Δ réel strictement positif et i entier, $J_i = \int_{i\Delta}^{(i+1)\Delta} X_s ds$.

Nous vérifions que (J_i) est un processus gaussien stationnaire ARMA(1,1) de coefficient de α-mélange exponentiellement décroissant. Cette structure ARMA(1,1) est mise en évidence par la relation suivante entre deux observations successives.

$$J_{i+1} - e^{\mu\Delta} J_i = \frac{\sigma}{\mu} \int_{i\Delta}^{(i+1)\Delta} (e^{\mu\Delta} - e^{\mu((i+1)\Delta - s)}) dB_s + \frac{\sigma}{\mu} \int_{(i+1)\Delta}^{(i+2)\Delta} (e^{\mu((i+2)\Delta - s)} - 1) dB_s \qquad (4)$$

La densité spectrale de (J_i) vérifie les conditions qui assurent que l'estimateur de Whittle, noté $\hat{\theta}_n = (\hat{\mu}_n, \hat{\sigma}_n^2)$ est consistant et asymptotiquement gaussien. Il est alors équivalent au maximum de vraisemblance exacte (voir par exemple Dzhaparidze et Yaglom (1983)). Nous obtenons une expression explicite pour le contraste de Whittle. Si μ est connu, l'estimateur $\hat{\sigma}_n^2$ est alors lui aussi explicite.

Nous considérons ensuite un estimateur moins spécifique au caractère gaussien des observations et donc a priori plus adapté à une généralisation pour d'autres diffusions. Pour cela nous utilisons la méthode que Ryden (1994) a introduite dans le cadre des chaînes de Markov cachées. Nous nous donnons un entier m superieur ou égal à 1 et découpons notre observation $(J_0, \ldots, J_{nm-1})$ en paquets de taille m, $(J_{km}, \ldots, J_{k(m+1)-1})$, pour $k = 0, \ldots, n-1$. Nous utilisons, alors, comme contraste la log–vraisemblance qu'aurait l'observation si ces paquets étaient indépendants. Le contraste est donc la somme des log–vraisemblances de chacun de ces paquets. On note $\tilde{\theta}_n^{(m)} = (\tilde{\mu}_n^{(m)}, \tilde{\sigma^2}_n^{(m)})$ l'estimateur résultant de ce contraste, appelé en anglais "Maximum likelihood split data estimator" (MLSDE). En utilisant la propriété de α–mélange des J_i, nous démontrons la consistance et normalité asymptotique pour $\tilde{\theta}_n^{(m)}$ et donnons une expression pour la variance asymptotique.

En évaluant numériquement cette variance asymptotique théorique, il apparait que, plus m est grand, plus l'estimateur de Ryden se rapproche de l'efficacité (ce qui n'est pas étonnant). Il est, au contraire, plus inattendu que, quand Δ tend vers 0, la variance asymptotique de $\tilde{\sigma^2}_n^{(m)}$ augmente fortement, même si la vraie valeur de μ est connue. L'estimation de σ^2 par cette méthode est donc bien moins bonne que par $\hat{\sigma^2}_n$ dont la variance asymptotique est indépendante de Δ.

Nous éclairons ce résultat numérique par un résultat théorique, en calculant l'équivalent suivant pour la variance asymptotique de $\tilde{\sigma^2}_n^{(m)}$ quand $m = 2$ et μ est connu:

$$var\ \tilde{\sigma^2}_n^{(m)} \sim_{\Delta\to 0} \frac{\sigma^4}{4\,|\mu|\,\Delta}.$$

Bien que restreinte au processus d'Ornstein–Uhlenbeck, l'étude du Chapitre 1 est de portée moins limitée qu'il n'y paraît au premier abord. En effet, le contraste de Whittle peut se calculer lorsque X est un diffusion à dérive linéaire de la même manière que lorsque X est un processus d'Ornstein–Uhlenbeck.

De plus, ce chapitre nous a éclairé sur les comportements différents pour l'estimation des paramètres de dérive μ et de variance σ. La variance d'estimation pour $\hat{\mu}_n$ est d'ordre $(n\Delta)^{-1}$, celle pour $\hat{\sigma^2}_n$ d'ordre n^{-1}. Au contraire, l'estimateur de Ryden ne fait pas la difference entre paramètres de dérive et variance, tout deux sont estimés avec une variance en $(n\Delta)^{-1}$.

Ce comportement à double vitesse a aussi été mis en évidence par Kessler (1997), lorsque l'on observe la discrétisation d'une diffusion (quelconque) avec un pas $\Delta = \Delta_n$ tendant vers 0. Dans les prochains chapitres, nous nous placerons aussi sous l'hypothèse d'un pas de temps tendant vers 0.

Chapitre I.2. Discrétisation d'une intégrale de diffusion et comparaison avec le schéma d'Euler.

Quand X n'est plus un processus d'Ornstein–Uhlenbeck, la vraisemblance exacte de $(J_0, \dots, J_{n-1})$ n'est en général plus explicite. En effet, la densité de ce vecteur est liée à la densité de transition de la diffusion (I_t, X_t) qui n'est explicite que dans très peu de modèles (voir Donati–Martin et Yor (1997), Leblanc (1997), pour X processus de Cox–Ingersoll–Ross). Nous avons, par contre, démontré dans l'appendice de cette thèse, que sous de conditions faibles sur X, cette vraisemblance existe (mais est incalculable).

Comme la vraisemblance exacte n'est pas explicite, afin de simplifier l'obtention de contrastes lorsque X est une diffusion quelconque solution de (1), nous nous plaçons sous l'hypothèse classique d'un pas de temps $\Delta = \Delta_n$ tendant vers 0. Dans le cas d'une observation de X, cette hypothèse permet de trouver des contrastes explicites, conduisant à une estimation équivalente au maximum de vraisemblance (Kessler (1997)).

Notre observation, $J_{i,n} = \int_{i\Delta_n}^{(i+1)\Delta_n} X_s ds$ pour $i = 0, \dots, n-1$, dépend donc maintenant de n.

L'objet de ce chapitre est d'obtenir des développements de $\Delta_n^{-1} J_{i,n}$ comme approximation de $X_{i\Delta_n}$, afin de voir s'il est possible de déduire des résultats classiques de statistique sur $(X_{i\Delta_n})$ des résultats sur $(\Delta_n^{-1} J_{i,n})$.

Une difficulté provient du fait que nous considérons une diffusion X à valeurs dans (l, r) avec $-\infty \le l < r \le \infty$. Il nous faut formuler une hypothèse qui assure que la diffusion n'approche pas trop les bords l, r qui sont en général des points d'irrégularité pour les fonctions que nous considérerons dans la suite. Pour cela nous introduisons deux fonctions $\mathcal{B}_l$ et $\mathcal{B}_r$ sur (l, r) tendant respectivement en l et r vers l'infini (on prend par exemple $\mathcal{B}_l(x) = 1 + (x - l)^{-1}$ et $\mathcal{B}_r(x) = 1 + (r - x)^{-1}$ quand les bornes sont finies et $\mathcal{B}_l(x) = 1 + |x|^k$ ou $\mathcal{B}_r(x) = 1 + |x|^k$ quand $l = -\infty$

ou $r = \infty$). Nous introduisons l'hypothèse (R) suivante:

$$\forall k \geq 0,\ \exists c > 0,\ \forall t \geq 0, \quad E\left(\sup_{s\in[t,t+1]} \mathcal{B}_l(X_s)^k \mid \mathcal{G}_t\right) \leq c\mathcal{B}_l^k(X_t) \tag{5}$$

$$\forall k \geq 0,\ \exists c > 0,\ \forall t \geq 0, \quad E\left(\sup_{s\in[t,t+1]} \mathcal{B}_r(X_s)^k \mid \mathcal{G}_t\right) \leq c\mathcal{B}_r^k(X_t) \tag{6}$$

où $\mathcal{G}_t = \sigma(X_s, s \leq t)$.

Nous démontrons que, pour une diffusion sur $(-\infty, \infty)$, cette hypothèse est vérifiée si les coefficients a et b vérifient la condition usuelle de croissance sous–linéaire.

Vérifier (R) est plus difficile quand l'un des bords est fini. Nous avons étudié notamment les diffusions suivantes sur $(0, \infty)$,

$$dX_t = \mu(X_t - m)dt + \sigma X_t^{\psi} dB_t, \quad \mu < 0,\ m, \sigma > 0 \text{ et } \psi \in [\frac{1}{2}, 1].$$

Pour $\psi = 1$ (diffusion bilinéaire), on peut ramener le problème de la borne 0 à une borne infinie en considérant X^{-1}.

Le cas $\psi = \frac{1}{2}$ (processus en racine carrée) est plus difficile à obtenir. Nous montrons que si $c_0 = \frac{2|\mu|m}{\sigma^2} > 1$, la condition (R) est verifiée avec la restriction que (5) n'est vraie que pour $k < c_0 - 1$. Remarquons que, si $c_0 < 1$, la borne 0 est atteignable et donc (R) ne peut être vraie.

Nous déduisons ensuite le cas $\psi \in (\frac{1}{2}, 1)$ du cas précédent.

Les résultats principaux de ce chapitre sont les développements suivants obtenus sous (R) et sous des conditions de régularité des fonctions a et b. On a:

$$\Delta_n^{-1} J_{i,n} - X_{i\Delta_n} = a(X_{i\Delta_n})\Delta_n^{\frac{1}{2}} \xi'_{i,n} + e_{i,n} \tag{7}$$

$$\Delta_n^{-1} J_{i+1,n} - \Delta_n^{-1} J_{i,n} - b(\Delta_n^{-1} J_{i,n})\Delta_n = a(X_{i\Delta_n})\Delta_n^{\frac{1}{2}}(\xi_{i,n} + \xi'_{i+1,n}) + \varepsilon_{i,n} \tag{8}$$

où

$$\xi_{i,n} = \Delta_n^{-\frac{3}{2}} \int_{i\Delta_n}^{(i+1)\Delta_n} (s - i\Delta_n) dB_s \quad \text{et} \quad \xi'_{i+1,n} = \Delta_n^{-\frac{3}{2}} \int_{(i+1)\Delta_n}^{(i+2)\Delta_n} ((i+2)\Delta_n - s) dB_s,$$

sont deux variables gaussiennes indépendantes de variance $\frac{1}{3}$. De plus, les termes $e_{i,n}$ et $\varepsilon_{i,n}$ sont des termes de reste, d'ordre Δ_n; $\varepsilon_{i,n}$ vérifie la condition de centrage suivante: $E(|E(\varepsilon_{i,n} \mid \mathcal{G}_{i\Delta_n})|) \leq c\Delta_n^2$.

Le développement (7) montre que la différence entre $\Delta_n^{-1} J_{i,n}$ et $X_{i\Delta_n}$ est d'ordre $\Delta_n^{\frac{1}{2}}$ et calcule le terme de premier ordre de cette différence.

Le développement (8) est une formule de type schéma d'Euler pour $(J_{i,n})$, qui est l'approximation pour une diffusion X quelconque de la formule exacte (4) obtenue quand X est un processus d'Ornstein–Uhlenbeck. Dans le cas particulier où a est constante, comme $(\xi_{i,n} + \xi'_{i+1,n})$ a la structure de covariance d'un processus MA(1), on voit que le processus $(\Delta_n^{-1} J_{i,n})$ est proche d'un processus gaussien ARMA(1,1) quand $\Delta_n \to 0$. Son comportement est donc très différent de celui de $(X_{i\Delta_n})$ qui est proche d'un AR(1) gaussien.

On déduit des développements précédents, le comportement de la variation quadratique des moyennes de X. Si $\Delta_n = \frac{T}{n}$ (T est fixé), on montre, en effet, que

$$\sum_{i=0}^{n-1}(\Delta_n^{-1}J_{i+1,n} - \Delta_n^{-1}J_{i,n})^2 \xrightarrow[\mathbf{P}]{n\to\infty} \frac{2}{3}\int_0^T a^2(X_s)ds. \tag{9}$$

Ce résultat est à rapprocher de ceux de Dellatre et Jacod (1997), qui étudie la variation quadratique du processus X observé avec erreurs d'arrondies.

On voit ainsi que l'on ne pourra pas remplacer $X_{i\Delta_n}$ par $\Delta_n^{-1}J_{i,n}$ dans les théorèmes usuels.

Chapitre I.3. Estimation de paramètres pour une observation discrète de l'intégrale d'une diffusion.

Nous reprenons le cadre de travail du chapitre précédent, la diffusion X est de plus récurrente positive sur (l,r), avec une mesure invariante ν_0. Le temps total d'observation $n\Delta_n$ tend vers l'infini quand $n \to \infty$.

Les premiers résultats de ce chapitre sont des théorèmes limites pour des fonctionnelles de $\overline{X}_{i,n} = \Delta_n^{-1}J_{i,n}$. Nous établissons, en utilisant les développements du chapitre 2, les convergences suivantes pour $f : (l,r) \to \mathbb{R}$ suffisamment régulière.

$$n^{-1}\sum_{i=0}^{n-1} f(\overline{X}_{i,n}) \xrightarrow[\mathbf{P}]{n\to\infty} \nu_0(f) \tag{10}$$

$$(n\Delta_n)^{-1}\sum_{i=0}^{n-1} f(\overline{X}_{i,n})(\overline{X}_{i+1,n} - \overline{X}_{i,n} - \Delta_n b(\overline{X}_{i,n})) \xrightarrow[\mathbf{P}]{n\to\infty} \frac{1}{6}\nu_0(f') \tag{11}$$

$$(n\Delta_n)^{-1}\sum_{i=0}^{n-1} f(\overline{X}_{i,n})(\overline{X}_{i+1,n} - \overline{X}_{i,n})^2 \xrightarrow[\mathbf{P}]{n\to\infty} \frac{2}{3}\nu_0(f) \tag{12}$$

Les résultats (11)–(12) sont à première vue suprenants et soulignent les différences entre $\overline{X}_{i,n}$ et $X_{i\Delta_n}$. En effet, si l'on remplace dans le terme de gauche de (11), $\overline{X}_{i,n}$ par $X_{i\Delta_n}$ et $\overline{X}_{i+1,n}$ par $X_{(i+1)\Delta_n}$, alors la limite est 0. Dans (12), par la même transformation la limite est $\nu_0(f)$.

Cependant, ces résultats sont compréhensibles grâces aux développements (7)–(8). Le facteur $\frac{2}{3}$ dans (11) est la variance de $(\xi_{i,n} + \xi'_{i,n})$. La limite non nulle dans (12) provient de la corrélation entre $\overline{X}_{i,n}$ et $\overline{X}_{i+1,n} - \overline{X}_{i,n} - \Delta_n b(\overline{X}_{i,n})$ due à la corrélation $cor(\xi'_{i,n}, (\xi_{i,n} + \xi'_{i+1,n})) = \frac{1}{6}$.

Nous prouvons, si $n\Delta_n^2 \to 0$, des théorèmes de limite centrale associés à (11)–(12),

$$(n\Delta_n)^{-\frac{1}{2}}\sum_{i=0}^{n-1}\left\{f(\overline{X}_{i,n})(\overline{X}_{i+1,n} - \overline{X}_{i,n} - \Delta_n b(\overline{X}_{i,n})) - f'(\overline{X}_{i,n})(\overline{X}_{i+1,n} - \overline{X}_{i,n})^2\right\} \xrightarrow[\mathcal{D}]{n\to\infty} \mathcal{N}\left(0, \nu_0(a^2f^2)\right)$$

$$n^{-\frac{1}{2}}\sum_{i=0}^{n-1} g(\overline{X}_{i,n})\left\{\frac{(\overline{X}_{i+1,n} - \overline{X}_{i,n})^2}{\Delta_n} - a^2(\overline{X}_{i,n})\right\} \xrightarrow[\mathcal{D}]{n\to\infty} \mathcal{N}\left(0, \nu_0(a^4g^2)\right).$$

Les deux expressions ci-dessus sont, de plus, asymptotiquements indépendantes.

Supposons, maintenant, que les coefficients de diffusion et de dérive, $a(x,\sigma)$ et $b(x,\mu)$ dépendent d'un paramètre $\theta = (\mu,\sigma) \in \Theta$, où $\Theta = \Theta_1 \times \Theta_2$ et Θ_1, Θ_2 sont deux compacts de $\mathbb{R}$. Pour estimer ce paramètre, nous intoduisons le contraste suivant

$$\mathcal{L}_n(\theta) = \sum_{i=0}^{n-1} \frac{\left(\overline{X}_{i+1,n} - \overline{X}_{i,n} - b(\overline{X}_{i,n},\mu)\Delta_n\right)^2}{\frac{2}{3}\Delta_n a^2(\overline{X}_{i,n},\sigma)} + \sum_{i=0}^{n-1} \log a^2(\overline{X}_{i,n},\sigma)$$
$$+ \frac{\Delta_n}{2} \sum_{i=0}^{n-1} \frac{\partial}{\partial x}\left(\frac{b(\overline{X}_{i,n},\mu)}{a^2(\overline{X}_{i,n},\sigma)}\right)\left(\overline{X}_{i+1,n} - \overline{X}_{i,n}\right)^2 \frac{3}{2\Delta_n} \qquad (13)$$

Ce contraste est une modification du contraste usuel basé sur $(X_{i\Delta_n})$ (voir Kessler (1997) par exemple) qui tient compte des différences entre fonctionnelles de $X_{i\Delta_n}$ et de $\overline{X}_{i,n}$ mises en évidence par (10)–(12).

Grâce aux théorèmes limites précédemments établis nous montrons la consistance de l'estimateur associé, $(\hat{\mu}_n, \hat{\sigma}_n) = \text{arginf}_{(\mu,\sigma)\in\Theta_1\times\theta_2} \mathcal{L}_n(\mu,\sigma)$.

Nous établissons, ensuite, la convergence en loi de $\left((n\Delta_n)^{\frac{1}{2}}(\hat{\mu}_n - \mu_0), n^{\frac{1}{2}}(\hat{\sigma}_n - \sigma_0)\right)$ vers une loi gaussienne

$$\mathcal{N}\left(0, \left\{\nu_0\left(\frac{(\partial_\mu b)^2(.,\mu_0)}{a^2(.,\sigma_0)}\right)\right\}^{-1}\right) \otimes \mathcal{N}\left(0, \frac{9}{16}\left\{\nu_0\left(\frac{(\partial_\sigma a)^2(.,\sigma_0)}{a^2(.,\sigma_0)}\right)\right\}^{-1}\right).$$

Le paramètre de diffusion est estimé à la vitesse $n^{\frac{1}{2}}$, celui de dérive à la vitesse $(n\Delta_n)^{\frac{1}{2}}$. L'estimation des paramètres de dérive est efficace. Notons que la variance asymptotique de $\hat{\sigma}_n$ est légèrement supérieure à celle obtenue par Kessler (1997), pour estimer σ, lorsque la diffusion X elle-même est observée.

Des exemples (incluant les modèles classiques de diffusions utilisées en finance) et des résultats numériques sur données simulées sont présentés. L'estimation s'avere bonne sur des échantillons finis.

Partie II. Modèles à volatité stochastique.

De nombreux auteurs (Gourieroux *et al.* (1993); Ruiz (1994); Renault et Touzi (1996); Genon-Catalot, Jeantheau et Laredo (1999); Sørensen (1999)) se sont intéressés au problème de l'estimation des paramètres de (1) lorsque seule Y solution de (3) est observée.

Par exemple, une méthode explicite basée sur l'échantillon $(Y_{i\Delta_n})_{0\leq i\leq n-1}$ est proposé par Genon-Catalot *et al.* (1999). Pour une diffusion ergodique, les auteurs estiment les paramètres inconnus de la distribution stationnaire de la diffusion cachée X. Notre but est, soit, estimer le coefficient de diffusion sur un intervalle de temps fini, soit, estimer tous les paramètres de X lorsqu'elle est ergodique.

Ceci est l'objet des deux prochains chapitres. Nous introduisons un nouveau type d'observation de Y. La trajectoire de Y est observée selon une double discrétisation, aux instants $(i + \frac{j}{m})\Delta$ avec $i = 0, \ldots, n-1$, $j = 0, \ldots, m$. Nous avons donc $N = nm + 1$ observations de pas $m^{-1}\Delta_n$.

Chapitre II.1. Discrétisation du modèle à volatilité stochastique sur un intervalle de temps fini, approximation de la volatilité intégrée et applications statistiques.

Nous reprenons les notations du chapitre 2 de la Partie I, mais ici X est à valeurs dans $(0,\infty)$ et Y est solution de (3). Pour $m \geq 1$, nous introduisons l'approximation, basée sur une discrétisation de pas $\frac{\Delta_n}{m}$ de Y, de la variable désormais non observée $J_{i,n}$:

$$\hat{J}^m_{i,n} = \sum_{j=0}^{m-1} (Y_{(i+\frac{j+1}{m})\Delta_n} - Y_{(i+\frac{j}{m})\Delta_n})^2.$$

Dans un premier temps nous montrons que l'erreur d'approximation $E_{i,n,m} = \Delta_n^{-1}\hat{J}^m_{i,n} - \Delta_n^{-1}J_{i,n}$ est d'ordre $m^{-\frac{1}{2}}$. Nous calculons, lorsque B et W sont indépendants, un développement au premier ordre de la variance conditionnelle à X de cette erreur. Cette étude de $E_{i,n,m}$ nous permet dans le cas où $\Delta_n = n^{-1}$ d'établir le comportement de l'approximation de la variation quadratique de X sur $[0,1]$: on montre que si m_n est une suite d'entiers tels que $n^{\frac{1}{2}}m_n^{-1} \to 0$, alors

$$\sum_{i=0}^{n-2}(\Delta_n \hat{J}^{m_n}_{i+1,n} - \Delta_n \hat{J}^{m_n}_{i,n})^2 = \frac{2}{3}\int_0^1 a^2(X_s)ds + 4\frac{n}{m_n}\int_0^1 X_s^2 ds + o_{\mathbf{P}}(1). \tag{14}$$

En comparant avec (9), on voit qu'un nouveau terme de biais, dû à la variance des erreurs $E_{i,n,m}$ est apparu.

Si le modèle sur $(X_t)_{t\in[0,1]}$ a la forme multiplicative

$$dX_t = \theta a(X_t)dB_t + b(X_t)dt,$$

l'équation (14) nous suggère d'introduire l'estimateur suivant de θ basé sur $(Y_{(i+\frac{j}{m})\frac{1}{n}})_{\substack{i=0,\dots,n-1\\ j=0,\dots,m}}$ (observation sur $[0,1]$).

$$\hat{\theta^2}_{n,m} = \frac{3}{2}\sum_{i=0}^{n-2}\left\{\frac{(n\hat{J}^m_{i+1,n} - \hat{J}^m_{i,n})^2}{a^2(n\hat{J}^m_{i,n})} - \frac{4}{m}\frac{(n\hat{J}^m_{i,n})^2}{a^2(n\hat{J}^m_{i,n})}\right\}.$$

Comme a^{-1} peut être borné en 0 ou ∞, pour étudier $\hat{\theta^2}_{n,m}$, nous avons besoin de l'existence et de majorations pour les moments de $\hat{J}^m_{i,n}$ et de $(\hat{J}^m_{i,n})^{-1}$. Les bornes pour les moments de $\hat{J}^m_{i,n}$ sont plus aisés à obtenir que celles des moments de $(\hat{J}^m_{i,n})^{-1}$. Pour ce faire, nous démontrons d'abord une inégalité sur les petites déviations autour de 0 de la somme de variables aléatoires positives. Nous déduisons alors le résultat suivant.

$$\forall k \geq 0, \exists c \geq 0, \forall i,n \geq 0, \forall m \geq 2k+3, \quad E\left(\mathcal{B}_0^k(\Delta_n^{-1}\hat{J}^m_{i,n}) \mid \mathcal{G}_{i\Delta_n}\right) \leq c\mathcal{B}_0^k(X_{i\Delta_n}), \quad \text{pour } \mathcal{B}_0(x) = 1 + x^{-1},$$

$$\forall k \geq 0, \exists c \geq 0, \forall i,n,m \geq 0, \quad E\left(\mathcal{B}_\infty^k(\Delta_n^{-1}\hat{J}^m_{i,n}) \mid \mathcal{G}_{i\Delta_n}\right) \leq c\mathcal{B}_\infty^{2k}(X_{i\Delta_n}), \quad \text{pour } \mathcal{B}_\infty(x) = 1 + x.$$

Grâce à ces inégalités, nous pouvons montrer la consistance suivante:

$$\text{Si } n^{\frac{2}{3}}m_n^{-1} \to 0, \qquad \hat{\theta^2}_{n,m_n} \xrightarrow[\mathbf{P}]{n\to\infty} \theta^2.$$

De plus,

- si $m_n = n$ alors $n^{\frac{1}{2}}(\hat{\theta^2}_{n,m_n} - \theta^2)$ est tendue,

- si $nm_n^{-1} \to 0$ alors $n^{\frac{1}{2}}(\hat{\theta^2}_{n,m_n} - \theta^2)$ converge en loi vers une gaussienne centrée de variance $\frac{9}{4}\theta^4$.

Remarquant que $\hat{\theta^2}_{n,n}$ utilise n^2 observations de Y et estime à vitesse $n^{\frac{1}{2}}$, nous avons construit un estimateur à vitesse $N^{\frac{1}{4}}$, où N désigne le nombre total d'observations.

Cette vitesse lente, $N^{\frac{1}{4}}$, se fait ressentir lors de l'étude numérique de $\hat{\theta}_{n,m}$ sur données simulées (pour $X_t = e^{\theta B_t}$). Il faut en effet beaucoup de données pour obtenir une estimation précise. On s'aperçoit en particulier que l'estimation des petites valeurs de θ est très difficile. Nous expliquons ce comportement par une étude théorique de la loi de l'estimateur sous l'hypothèse $\theta = 0$.

Chapitre II.2. Estimation de paramètres pour une diffusion cachée

Nous reprenons le cadre de travail du chapitre précédent, avec X récurrente positive sur $(0, \infty)$, de mesure stationnaire ν_0 et $n\Delta_n \to \infty$. Grâce aux majorations du Chapitre II.1. sur l'erreur $E_{i,n,m}$, nous trouvons les conditions sur la vitesse à laquelle m_n doit tendre vers l'infini qui assurent que dans les énoncés (10)–(12), on peut remplacer $\overline{X}_{i,n} = \Delta_n^{-1} J_{i,n}$ par son approximation $\hat{X}_{i,n}^m = \Delta_n^{-1}\hat{J}_{i,n}^m$. Nos résultats sont les suivants.

$$\text{Si } m_n \to 0, \quad n^{-1}\sum_{i=0}^{n-2} f(\hat{X}_{i,n}^{m_n}) \xrightarrow{n\to\infty} \nu_0(f)$$

$$\text{Si } m_n^{-1} = o(\Delta_n), \quad (n\Delta_n)^{-1}\sum_{i=0}^{n-2} f(\hat{X}_{i,n}^{m_n})(\hat{X}_{i+1,n}^{m_n} - \hat{X}_{i,n}^{m_n} - \Delta_n b(\hat{X}_{i,n}^{m_n})) \xrightarrow{n\to\infty} \frac{1}{6}\nu_0(f')$$

$$\text{Si } m_n^{-1} = o(\Delta_n), \quad (n\Delta_n)^{-1}\sum_{i=0}^{n-2} f(\hat{X}_{i,n}^{m_n})(\hat{X}_{i+1,n}^{m_n} - \hat{X}_{i,n}^{m_n})^2 \xrightarrow{n\to\infty} \frac{2}{3}\nu_0(f)$$

Sous les conditions plus réstrictives $n\Delta_n^2 \to 0$ et $m_n^{-1} = o(\frac{\Delta_n}{n})$, nous prouvons des théorèmes de limites centrales, qui sont analogues à ceux obtenus dans le chapitre I.3. (avec $\hat{X}_{i,n}$ remplaçant $\overline{X}_{i,n}$).

Nous supposons maintenant que $a(x, \vartheta)$ dépend d'un paramètre $\vartheta \in \Theta_2$ et $b(x, \mu, \vartheta)$ d'un paramètre supplémentaire $\mu \in \Theta_1$, où Θ_1 et Θ_2 sont des compacts de $\mathbb{R}$. (Il est pratique pour les modèles de diffusions positives de supposer que b depend du couple (μ, ϑ). Ceci est par exemple, le cas si X est l'exponentielle d'un processus d'Ornstein–Uhlenbeck.)

Nous reprenons alors le contraste (13), en remplacant les $\overline{X}_{i,n}$ par $\hat{X}_{i,n}^m$ et $b(x, \mu)$ par $b(x, \mu, \vartheta)$.

En minimisant ce contraste nous obtenons un estimateur $\hat{\theta}_{n,m} = (\hat{\mu}_{n,m}, \hat{\vartheta}_{n,m})$. Grâce aux théorèmes limites précédemment établies, nous montrons alors que si $m_n^{-1} = o(\Delta_n)$, l'estimateur $\hat{\theta^2}_{n,m_n}$ est consistent.

De plus si $n\Delta_n^2 \to 0$ et $m_n = o(n^{-1}\Delta_n)$, il est asymptotiquement normal:

$$\left((n\Delta_n)^{\frac{1}{2}}(\hat{\mu}_{n,m_n} - \mu_0), n^{\frac{1}{2}}(\hat{\vartheta}_{n,m_n} - \vartheta_0)\right) \xrightarrow[\mathcal{D}]{n\to\infty}$$

$$\mathcal{N}\left(0, \left\{\nu_0\left(\frac{(\partial_\mu b)^2(., \mu_0, \vartheta_0)}{a^2(., \vartheta_0)}\right)\right\}^{-1}\right) \otimes \mathcal{N}\left(0, \frac{9}{16}\left\{\nu_0\left(\frac{(\partial_\vartheta a)^2(., \vartheta_0)}{a^2(., \vartheta_0)}\right)\right\}^{-1}\right).$$

On s'aperçoit que notre estimateur $\hat{\mu}_{n,m_n}$ a le même comportement asymptotique que les estimateurs efficaces basés sur l'observation directe de X (c'est à dire une vitesse $(n\Delta_n)^{\frac{1}{2}}$ et une variance asymptotique $\nu_0\left(\frac{(\partial_\mu b)^2(.,\mu_0,\vartheta_0)}{a^2(.,\vartheta_0)}\right)$).

La vitesse d'estimation de ϑ, $n^{\frac{1}{2}}$, est lente devant le nombre d'observations disponibles nm_n. Des exemples, avec simulations numériques sont présentés.

Appendice. Un problème d'existence de densité.

Dans cet appendice, nous établissons l'existence de densité, par rapport à la mesure de Lebesgue, pour le vecteur $(J_0, \ldots, J_q) = (\int_0^\Delta X_s ds, \ldots, \int_{q\Delta}^{(q+1)\Delta} X_s ds)$. En effet, nous montrons que sous les conditions assurant l'existence de densité de transition pour (1), le vecteur $(J_0, \ldots, J_q)$ admet, pour tout q, une densité.

Comme pour les observations du q-uple $(X_{k\Delta})_{k\leq q-1}$, cette densité est difficilement exploitable sur le plan statistique.

References

- O.E. Barndorff-Nielsen and N. Shephard (1998). Aggregation and model construction for volatility models. Working paper series no.**10**, Center for Analytical Finance, University of Aarhus.
- B.M. Bibby and M. Sørensen M. (1995). Martingale estimations functions for discretely observed diffusion processes. *Bernoulli*, **1**, 17–39.
- P. Bickel and Y. Ritov (1996). Inference in hidden Markov models I: local asymptotic normality in the stationary case. *Bernoulli*, **2**, 199–228.
- F. Black and M. Scholes (1973). The valuation of option and corporate liabilities. *Journal of political economy*, **81**, 637–654.
- D. Dacunha-Castelle and D. Florens-Zmirou (1986). Estimation of the coefficient of a diffusion from discrete observations. *Stochastics*, **19**, 263–284.
- S. Delattre and J. Jacod (1997). A central limit theorem for normalized functions of the increments of a diffusion process, in the presence of round–off errors. *Bernoulli,* **3***(1)*, 1–28.
- C. Donati-Matin and M. Yor (1997). Some brownian functionals and their laws. *Annals of probability*, **25**(3), 1011–1058.

- F. Drost and B. Werker. (1996). Closing the GARCH gap: Continuous time GARCH modeling. *J. Econom.*, **74**(1),31–57.
- K.O. Dzhaparidze and A.M. Yaglom (1983). Spectrum parameter estimation in time series analysis. In *Developments in Statistics* (P.R. Krishnaiah, ed.) **4** p. 1–181 Academic, New York.
- R.F. Engle (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of U.K. inflation. *Econometrica*, **50**, 987–1008.
- R. Frey and W.J. Runggaldier (1999). A nonlinear filtering approach to volatility estimation with a view towards high frequency data. *Preprint.*
- V. Genon-Catalot and J. Jacod (1993). On the estimation of the diffusion coefficient for multidimensional diffusion processes. *Annales de l'I.H.P., probabilité–Statistiques*, **29**, 119–151.
- V. Genon-Catalot, T. Jeantheau and C. Laredo (1999). Parameter estimation for discretely observed stochastic volatility models. *To appear in Bernoulli* **5**(5).
- C. Gourieroux, A. Montfort, and E. Renault (1993). Indirect inference. *Journal of applied econometrics*, **8**, 85–118.
- J. Hull and A. White (1987). The pricing of options on assets with stochastic volatilities. *Journal of finance*, **42**, 281–300.
- M. Kessler (1997). Estimation of an ergodic diffusion from discrete observations. *Scandinavian journal of statistics*, **24**(2), 211–229.
- M. Kessler and M. Sørensen (1999). Estimating equations based on eigenfunctions for a discretely observed diffusion process. *Research Report No.* **332**, *Dept. Theor. Statist. Univ. of Aarhus. To appear in Bernoulli.*
- Y. Kutoyants (1984). Parameter estimation for stochastic processes *Heldermann*, Berlin.
- B.Leblanc (1997). *Modélisation de la volatilité d'un actif financier et applications.* PhD thesis, Université Paris 7.
- M. Lefebvre (1997). On the inverse of the first hitting time problem for bidimensional processes. *Journal of applied probability*, **34**(3), 610–622.
- B. Leroux (1992). Maximum–likelihood estimation for hidden Markov models. *Stochastic processes appl.*, **40**(1), 127–143.
- D. Nelson (1990). ARCH models as diffusion approximation *J. Econ.*, **45**(1/2), 7–38.
- E. Renault and N. Touzi (1996). Option hedging and implied volatilities in a stochastic volatility model. *Mathematical finance*, **3**, 279–302.

- E. Ruiz (1994). Quasi–maximum likelihood estimation of stochastic volatility models. *Journal of econometrics*, **63**, 289–306.
- T. Ryden (1994). Consistent and asymptotically normal parameter estimates for hidden Markov models. *The annals of statistics*, **22**(4), 1884–1895.
- M. Sørensen (1999). Prediction–based estimating functions. *Preprint.*

Partie I. Statistique basée sur l'observation discrétisée de l'intégrale d'une diffusion

Chapitre I.1. Estimation de paramètres pour une discrétisation de l'intégrale d'un processus d'Ornstein–Uhlenbeck

Parameter estimation for a discrete sampling of an integrated Ornstein–Uhlenbeck Process.

Abstract

We study the estimation of parameters $\theta = (\mu, \sigma^2)$ for a diffusion $dX_t = a(X_t, \sigma^2)dB_t + b(X_t, \mu)dt$, when we observe a discretization with step Δ of the integral $I_t = \int_0^t X_s ds$. To keep computations tractable we focus on the case of an Ornstein–Uhlenbeck process, but our results provide information on how to deal with other processes. We study an efficient estimator $\hat{\theta}_n$ based on the Gaussian property of the process $(\int_{i\Delta}^{(i+1)\Delta} X_s ds)_{i \geq 0}$, and we give an estimator $\bar{\theta}_n$ based on Ryden's idea of maximum likelihood split data. We compare these different estimators: first we give some numerical results, then we give a theoretical explanation for these results.

AMS 1991 *subject classification.* 62F12, 62M09, 62M10.
key words: diffusion processes, discrete time observation, parametric inference, minimum contrast, Whittle approximation.

(To appear in Statistics.)

1 Introduction

Parameter estimation for a diffusion process $(X_t)_{t \geq 0}$, which solves $dX_t = a(X_t, \sigma^2)dB_t + b(X_t, \mu)dt$ is now classical. It has been treated under many different assumptions for the observation of the sample path: (X_t) may be continuously observed throughout a time interval $[0, T]$ (see Kutoyants (1981)); or only a discretization may be observed, the sampling interval Δ being fixed or tending to zero as the number of observations tends to infinity (see Bibby and Sørensen (1995), Dacunha–Castelle and Florens-Zmirou (1986), Genon–Catalot and Jacod (1993), Kessler (1997)).

In this paper we suppose that we don't observe the process (X_t) itself but a discrete sampling of the integrated process $I_t = \int_0^t X_s ds$.

Integrals of diffusion processes have recently been considered in the field of finance in relation with the stochastic volatility models (see e.g. Ghysels *et al.* (1996) for a survey of these models). Data may be obtained from option prices and their associated implied volatilities (see e.g., Pastorello *et al* (1994), Taylor and Xu (1994, 1995)).

The process $(I_t)_{t \geq 0}$ is discretely observed with a regular sampling interval Δ. For a general diffusion X, the exact distribution of a n–sample $(I_{i\Delta},\ i \leq n)$ is not explicit. Therefore, we consider one of the few models for which computations are possible in order to try and compare different inference methods in view of further generalizations. In this paper, we study the case where (X_t) is a strictly stationary Ornstein–Uhlenbeck process:

$$dX_t = \mu X_t dt + \sigma dB_t.$$

The unknown parameter $\theta = (\mu, \sigma^2)$ is to be estimated from the observation of $(I_{i\Delta}, i \leq n)$ which is equivalent to the observation of the increments $(J_i = I_{(i+1)\Delta} - I_{i\Delta},\ i \leq n-1)$.

We first investigate the probabilistic properties of the process $(J_i, i \in \mathbb{N})$ (Section 2). It is a Gaussian ARMA(1,1) process with exponentially decaying α–mixing coefficient.

In Section 3, we study efficient estimators of θ. Although the exact distribution of $(J_i, i \leq n)$ is explicit, the likelihood function is hardly tractable. So we study the Whittle estimator of θ which is known to be asymptotically equivalent to the maximum likelihood estimator (see Dzhaparidze and Yaglom (1983)). It turns out that the Whittle contrast is explicit (Theorem 3.1). In particular, when μ is known, the Whittle estimator $\hat{\sigma^2}_n$ is explicit and its asymptotic variance is equal to $2\sigma^4$.

In Section 4, noting that $(J_i, i \in \mathbb{N})$ is a special case of Hidden Markov Chain, we use Ryden's approach to build other estimators namely the maximum likelihood split data estimator (MLSDE) (see Ryden (1994)). For m integer, the MLSDE $\tilde{\theta}_n^{(m)}$ is based on the maximization of the likelihood of a n-sample distributed as $(J_0, J_1, \ldots, J_{m-1})$. Thus $\tilde{\theta}_n^{(m)}$ uses nm datas. We prove that it is consistent, asymptotically Gaussian and give an expression for its asymptotic covariance matrix.

Section 5 deals with the comparison of asymptotic covariances of the previous estimators. We give numerical values and a theoretical comparison. The conclusions are the following. For large values of m, the MLSDE and the Whittle estimator have a close asymptotic behaviour. For small Δ, the results are more surprising. For the drift parameter, the asymptotic variances of estimators are similar. For the diffusion coefficient parameter, they are quite different. The asymptotic variance of $\hat{\sigma^2}_n$ does not depend on Δ, whereas the asymptotic variance of $\tilde{\sigma^2}_n^{(m)}$ is of order $\frac{1}{\Delta}$ (Theorem 5.1).

In Section 6, we discuss possibilities of extensions of the two methods to more general diffusion processes.

2 Model and assumptions

In this section, we describe the probabilistic properties of the observed process.

Let $(C = C(\mathbb{R}_+, \mathbb{R}), \mathcal{C}, (\mathcal{C}_t)_{t\geq 0}, (X_t, t \geq 0), \mathbb{P}_\theta)$ be the canonical probability space associated with the observation of a strictly stationary Ornstein-Uhlenbeck process with parameter $\theta = (\mu, \sigma^2)$: for $t \geq 0$, X_t is defined on C by $X_t(w) = w_t$, $\mathcal{C}_t = \sigma(X_s,\ 0 \leq s \leq t)$, $\mathcal{C} = \sigma(X_t,\ t \geq 0)$; $\theta = (\mu, \sigma^2)$ is a two-dimensional parameter: $\theta_1 = \mu$ is negative, and $\theta_2 = \sigma^2$ positive; $\mathbb{P}_\theta$ is the probability on $(C, \mathcal{C})$, such that, under $\mathbb{P}_\theta$, there exists a standard Brownian motion $(B_t^\theta, t \geq 0)$, adapted to $(\mathcal{C}_t)_{t\geq 0}$ and such that the canonical process $(X_t)_{t\geq 0}$ is solution of:

$$dX_t = \mu X_t dt + \sigma dB_t^\theta \tag{1}$$

with X_0 centered, Gaussian, with variance $\frac{\sigma^2}{2|\mu|}$ and independent of B^θ.

Under $\mathbb{P}_\theta$, the process $(X_t, t \geq 0)$ is a stationary Ornstein-Uhlenbeck process. Solving (1), we obtain, for all $t \geq 0$ and $h \geq 0$:

$$X_{t+h} = e^{\mu h} X_t + e^{\mu(t+h)} \sigma \int_t^{t+h} e^{-\mu s} dB_s^\theta \tag{2}$$

For Δ a positive real and $i \in \mathbb{N}$, let

$$J_i = \int_{i\Delta}^{(i+1)\Delta} X_s ds \quad \text{for } i \in \mathbb{N} \tag{3}$$

The process $(J_i)_{i\geq 0}$ is not Markov, but we can link J_i and J_{i+1} by a relation of ARMA(1,1) type.

Proposition 2.1. *Under $\mathbb{P}_\theta$, for all $i \geq 0$,*

$$J_{i+1} - e^{\mu\Delta} J_i = \frac{\sigma}{\mu} \int_{i\Delta}^{(i+1)\Delta} (e^{\mu\Delta} - e^{\mu((i+1)\Delta - s)}) dB_s^\theta + \frac{\sigma}{\mu} \int_{(i+1)\Delta}^{(i+2)\Delta} (e^{\mu((i+2)\Delta - s)} - 1) dB_s^\theta$$

Hence for all $i \geq 1$, $J_{i+1} - e^{\mu\Delta} J_i$ is independent of $(J_0, \ldots, J_{i-1})$

Proof. See the appendix. □

We define the following expressions

$$r(0,\mu) = \frac{1}{\mu^2}\left(\Delta + \frac{1-e^{\mu\Delta}}{\mu}\right) \tag{4}$$

$$r(k,\mu) = -\frac{1}{2\mu^3}e^{\mu|k|\Delta}e^{-\mu\Delta}\left(e^{\mu\Delta}-1\right)^2 \quad \text{for} \quad k \neq 0 \tag{5}$$

$$A_0(\mu) = \frac{1}{\mu^2}\left(\Delta + \frac{1-e^{2\mu\Delta}}{\mu} + \Delta e^{2\mu\Delta}\right) \tag{6}$$

$$A_1(\mu) = \frac{1}{2\mu^2}\left(\frac{e^{2\mu\Delta}-1}{\mu} - 2e^{\mu\Delta}\Delta\right) \tag{7}$$

$$B_0(\mu) = 1 + e^{2\mu\Delta}, \quad B_1(\mu) = -e^{\mu\Delta} \tag{8}$$

Proposition 2.2. *The process $(J_i)_{i\in\mathbb{N}}$ is strictly stationary and Gaussian with for $i,j \geq 0$,*

$$E(J_i) = 0, \quad Var(J_i) = r(0,\mu)\sigma^2, \quad Cov(J_i,J_j) = r(i-j,\mu)\sigma^2 \quad \text{for} \quad i \neq j.$$

Its spectral density $f(\lambda,\theta)$ has the explicit form:

$$f(\lambda,\theta) = \sigma^2\frac{A_0(\mu) + 2A_1(\mu)\cos\lambda}{B_0(\mu) + 2B_1(\mu)\cos\lambda} \tag{9}$$

with

$$B_0(\mu) - 2B_1(\mu) > 0\ ,\ A_1(\mu) > 0\ ,\ A_0(\mu) - 2A_1(\mu) > 0. \tag{10}$$

Proof. See the appendix □

An immediate consequence of (10) is that $\inf_{\lambda\in\mathbb{R}} f(\lambda,\theta) > 0$. Due to the form of its spectral density, the process $(J_i)_{i\in\mathbb{N}}$ has an ARMA(1,1) representation (see e.g. Brockwell and Davis (1991)). Moreover, we can study its α–mixing coefficient.

Proposition 2.3. *Let $\alpha_J(k)$ denote the α–mixing coefficient of $(J_i)_{i\in\mathbb{N}}$ (see e.g. Doukhan (1994) chap.1). We have*

$$\alpha_J(k) \leq e^{\mu(k+1)\Delta}.$$

Hence, the process is ergodic.

Proof. Let Q_θ denote the stationary distribution of $(X_t)_{t\geq 0}$, i.e. $Q_\theta = \mathcal{N}\left(0, \frac{\sigma^2}{2|\mu|}\right)$. The infinitesimal generator of the Ornstein-Uhlenbeck process considered as an operator on $L^2(Q_\theta)$ is self-adjoint and has discrete spectrum equal to $\{n\mu, n \in \mathbb{N}\}$ (see Karlin and Taylor (1981) p.332). Thus, using Proposition 1 p.112 in Doukhan (1994), we know that the α-mixing coefficient $\alpha_X(t)$ of $(X_t)_{t\geq 0}$ satisfies:

$$\alpha_X(t) = \alpha\left(\sigma(X_0), \sigma(X_t)\right) \leq e^{\mu t}$$

(The first equality above is valid because $(X_t)_{t\geq 0}$ is a strictly stationary Markov process). Since J_i is $\sigma(X_s, i\Delta \leq s \leq (i+1)\Delta)$ measurable, $\alpha_J(k-1) \leq e^{\mu k\Delta}$. Now, $\mu < 0$ implies $\lim_{k\to\infty}\alpha_J(k) = 0$, which gives the ergodicity. □

3 An efficient estimator

The likelihood function of the $(J_0, \dots, J_{n-1})$ is explicitly known but its exact formula is difficult to compute. Therefore, instead of the exact likelihood, we shall use its Whittle approximation which provides efficient estimators (see e.g., Dzhaparidze and Yaglom (1983)). This approximation is also studied in Dacunha-Castelle and Duflo (1986) (chapter 3) and called the Whittle contrast.

3.1 The Whittle contrast

Recall the definition of the periodogram $I_n(\lambda)$ for $n \in \mathbb{N}$ and $\lambda \in \mathbb{R}$:

$$I_n(\lambda) = \frac{1}{n}\left|\sum_{p=0}^{n-1} J_p e^{-ip\lambda}\right|^2 \tag{11}$$

The Whittle contrast is given by:

$$U_n(\theta) = \frac{1}{2\pi}\int_{-\pi}^{\pi}\left[\ln f(\lambda,\theta) + \frac{I_n(\lambda)}{f(\lambda,\theta)}\right] d\lambda \tag{12}$$

Let $\hat{\theta}_n = \operatorname{arginf}_{\theta\in\Theta} U_n(\theta)$. This estimator is called the Whittle estimator.

We can actually compute the Whittle contrast explicitely.

Theorem 3.1. *We have*

$$U_n(\theta) = \ln\left(\sigma^2\frac{A_1(\mu)}{-\xi(\mu)}\right) + \frac{1}{\sigma^2 n}\sum_{k,l=0}^{n-1} J_k J_l c(k-l,\mu) \tag{13}$$

where (see (6) - (8)*)*

$$c(0,\mu) = \left(\frac{B_1(\mu)\xi^2(\mu) + B_0(\mu)\xi(\mu) + B_1(\mu)}{\xi(\mu)\sqrt{A_0^2(\mu) - 4A_1^2(\mu)}} + \frac{B_1(\mu)}{A_1(\mu)}\right) \tag{14}$$

$$c(k,\mu) = \xi(\mu)^{|k|-1}\left(\frac{B_1(\mu)\xi^2(\mu) + B_0(\mu)\xi(\mu) + B_1(\mu)}{\sqrt{A_0^2(\mu) - 4A_1^2(\mu)}}\right) \quad \textit{if } |k| \neq 0 \tag{15}$$

$$\xi(\mu) = \frac{-A_0(\mu) + \sqrt{A_0^2(\mu) - 4A_1^2(\mu)}}{2A_1(\mu)} \tag{16}$$

(Note that, by (10)*, $\xi(\mu)$ is well defined and is negative.)*

Proof. For the sake of simplicity, we omit μ in all expressions depending on μ only. Using (11) and (12) we have to prove that

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} \ln f(\lambda,\theta) d\lambda = \ln(\sigma^2\frac{A_1}{-\xi}) \tag{17}$$

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} f(\lambda,\theta)^{-1} d\lambda = \left(\frac{B_1\xi^2 + B_0\xi + B_1}{\xi\sqrt{A_0^2 - 4A_1^2}} + \frac{B_1}{A_1}\right)\frac{1}{\sigma^2} \tag{18}$$

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} e^{ik\lambda} f(\lambda,\theta)^{-1} d\lambda = \xi^{|k|-1}\left(\frac{B_1\xi^2 + B_0\xi + B_1}{\sqrt{A_0^2 - 4A_1^2}}\right)\frac{1}{\sigma^2} \text{ if } |k| \neq 0 \tag{19}$$

First, we state a useful equality (see e.g. Theorem 15.18 p.299 in Rudin (1966)):

$$\text{For } |x| \leq 1, \quad \int_{-\pi}^{\pi} \ln(1 + x^2 - 2x\cos\lambda)d\lambda = 0 \tag{20}$$

Let us prove (17). Using (9),(8) we have

$$\ln f(\lambda,\theta) = \ln(\sigma^2 \frac{A_1}{-\xi}) + \ln(-\frac{A_0\xi}{A_1} - 2\xi\cos\lambda) - \ln(1 + e^{2\mu\Delta} - 2e^{\mu\Delta}\cos\lambda)$$

But using (16), we have $A_1\xi^2 + A_0\xi + A_1 = 0$, and so

$$\ln f(\lambda,\theta) = \ln(\sigma^2 \frac{A_1}{-\xi}) + \ln(1 + \xi^2 - 2\xi\cos\lambda) - \ln(1 + e^{2\mu\Delta} - 2e^{\mu\Delta}\cos\lambda)$$

Since, by (16), $|\xi| \leq 1$ and $e^{\mu\Delta} \leq 1$, we can apply (20), and this gives (17).

Let us prove (18). We compute:

$$\begin{aligned}\frac{1}{2\pi}\int_{-\pi}^{\pi} f(\lambda,\theta)^{-1}d\lambda &= \frac{1}{2\pi\sigma^2}\int_{-\pi}^{\pi}\left(\frac{B_0 + 2B_1\cos\lambda}{A_0 + 2A_1\cos\lambda}\right)d\lambda \\ &= \frac{1}{\sigma^2 2i\pi}\int_{\mathbb{U}} \frac{B_0 + B_1(z+z^{-1})}{(A_0 + A_1(z+z^{-1}))z}dz, \text{ where } \mathbb{U} \text{ is the unit circle} \\ &= \frac{1}{\sigma^2 2i\pi}\int_{\mathbb{U}} \frac{B_1z^2 + B_0z + B_1}{(A_1z^2 + A_0z + B_1)z}dz\end{aligned}$$

Using the residue Theorem we have (with $G(z) = \frac{B_1z^2+B_0z+B_1}{(A_1z^2+A_0z+B_1)z}$ for the sake of simplicity)

$$c(0) = \sum_{\alpha \text{ pole of } G,\ |\alpha|<1} res(G,\alpha)$$

The pole zero has residue $\frac{B_1}{A_1}$, and ξ is the only other pole with residue $\frac{B_1\xi^2+B_0\xi+B_1}{\xi\sqrt{A_0^2-4A_1^2}}$. This gives (18).

To get (19), we compute the expression of $\frac{1}{2\pi}\int_{-\pi}^{\pi} e^{ik\lambda}f(\lambda,\theta)^{-1}d\lambda$, in the same way as for (18). □

3.2 Properties of the Whittle estimator

To study the Whittle estimator, we assume that $\Theta = [\underline{\mu},\overline{\mu}] \times [\underline{\sigma}^2,\overline{\sigma^2}]$ with $\underline{\mu} < \overline{\mu} < 0$, $0 < \underline{\sigma}^2 < \overline{\sigma^2}$. The assumption of compacity for Θ is used in Dacunha-Castelle and Duflo (1986) in order to simplify the proof of consistency of minimum contrast estimators (see Dacunha-Castelle and Duflo (1986), Theorem 3.2.8). We denote by $\theta_0 = (\mu_0, \sigma_0^2)$ the true value of the parameter and assume that $\theta_0 \in \overset{\circ}{\Theta}$.

Let $I(\theta)$ be the 2×2 matrix defined for $\theta \in \overset{\circ}{\Theta}$, by

$$\text{for } i,j \in \{1,2\} \quad I(\theta)_{i,j} = \frac{1}{4\pi}\int_{-\pi}^{\pi} \frac{\partial}{\partial\theta_i}\ln f(\lambda,\theta)\frac{\partial}{\partial\theta_j}\ln f(\lambda,\theta)d\lambda \tag{21}$$

Proposition 3.2. 1) *For all $\theta \in \overset{\circ}{\Theta}$, the matrix $I(\theta)$ is non singular.*
2) $\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{n\to\infty} \mathcal{N}(0, I^{-1}(\theta_0))$ *in distribution under $\mathbb{P}_{\theta_0}$.*

Proof. Provided $I(\theta_0)$ is non singular, 2) is easily obtained by a classical proof (see Dacunha-Castelle and Duflo (1986), Dzhaparidze and Yaglom (1983)).

Let us prove 1). By noticing, using (9), that $f(\lambda, \theta) = \sigma^2 g(\lambda, \mu)$, we obtain:

$$\frac{\partial}{\partial \sigma^2} \ln(f(\lambda, \theta)) = \frac{1}{\sigma^2}$$

Hence the matrix $I(\theta)$ is:

$$I(\theta) = \begin{bmatrix} \frac{1}{4\pi} \int_{-\pi}^{\pi} \left(\frac{\partial}{\partial \mu} \ln g(\lambda, \mu)\right)^2 d\lambda & \frac{1}{4\pi\sigma^2} \int_{-\pi}^{\pi} \left(\frac{\partial}{\partial \mu} \ln g(\lambda, \mu)\right) d\lambda \\ \frac{1}{4\pi\sigma^2} \int_{-\pi}^{\pi} \left(\frac{\partial}{\partial \mu} \ln g(\lambda, \mu)\right) d\lambda & \frac{1}{2\sigma^4} \end{bmatrix} \tag{22}$$

Suppose that $I(\theta)$ is singular, then $\det I(\theta) = 0$. Because of (22) we have $\int_{-\pi}^{\pi} \left(\frac{\partial}{\partial \mu} \ln g(\lambda, \mu)\right)^2 d\lambda = \left(\int_{-\pi}^{\pi} \left(\frac{\partial}{\partial \mu} \ln g(\lambda, \mu)\right) d\lambda\right)^2$; but equality in the Cauchy-Schwarz inequality implies that $\frac{\partial}{\partial \mu} \ln g(\lambda, \mu)$ is independent of λ. We deduce that

$$\frac{\partial}{\partial \mu}\left(\frac{\partial}{\partial \lambda} \ln g(\lambda, \mu)\right) = \frac{\partial}{\partial \lambda}\left(\frac{\partial}{\partial \mu} \ln g(\lambda, \mu)\right) = 0$$

Derivating (9), we find

$$\left(\frac{\partial}{\partial \lambda} \ln g(\lambda, \mu)\right)_{|\lambda = \frac{\pi}{2}} = \frac{-2A_1}{A_0} - \frac{-2B_1}{B_0}$$

where this expression should not depend on μ. Replacing (6) and (7) above yields the fact that

$$-\frac{\frac{e^{2\mu\Delta}-1}{\mu} - 2e^{\mu\Delta}\Delta}{\Delta + \frac{1-e^{2\mu\Delta}}{\mu} + \Delta e^{2\mu\Delta}} - \frac{2e^{\mu\Delta}}{1+e^{2\mu\Delta}} \text{ should be independent of } \mu$$

By letting $\mu \to -\infty$ we find it is equal to zero. Hence, for all μ,

$$-\left(e^{2\mu\Delta} - 1 - 2\mu\Delta e^{\mu\Delta}\right)\left(1 + e^{2\mu\Delta}\right) - \left(\mu\Delta + 1 - e^{\mu\Delta} + \mu\Delta e^{2\mu\Delta}\right) 2e^{\mu\Delta} = 0$$

This is absurd. So $I^{-1}(\theta)$ exists. □

Remark 3.3. *If μ_0 is known, then the Whittle estimator $\hat{\sigma^2}_n$ of σ_0^2 is given by*

$$\hat{\sigma^2}_n = \frac{1}{n} \sum_{k,l=0}^{n-1} J_k J_l c(k-l, \mu_0)$$

and satisfies $\sqrt{n}(\hat{\sigma^2}_n - \sigma_0^2) \xrightarrow{n\to\infty} \mathcal{N}(0, 2\sigma_0^4)$. This is the same asymptotic distribution as the MLE of σ^2 based on the observation of $(X_{i\Delta})_{i\leq n}$.

4 Maximum likelihood split data estimator

4.1 Introduction and notations

The process $(J_i)_{i\in\mathbb{N}}$ is not Markov, but is a deterministic function of the 2-dimensional Markov chain $(J_i, X_{(i+1)\Delta})_{i\in\mathbb{N}}$. This is a special case of Hidden Markov Model, therefore we use Ryden's idea (Ryden (1994)). We split the observation into groups of fixed size, consider these groups as independent and then maximize the resulting likelihood. The resulting estimator is called the maximum likelihood split data estimator (MLSDE). In this section, we prove the consistency and asymptotic normality of the MLSDE.

Let m be an integer, $m \geq 1$. Define for $i = 0, 1, \ldots, n-1$,

$$K_i^{(m)} = (J_{im}, J_{im+1}, \ldots, J_{im+m-1})^* \tag{23}$$

If x is a vector or a matrix, we denote by x^* its transpose. Since m is fixed throughout this section, we shall set,

$$K_i^{(m)} = K_i \tag{24}$$

The process $(K_i, i \in \mathbb{N})$ is ergodic and its α-mixing coefficient satisfies $\alpha_K(k) \leq \alpha_J((k+1)m)$. By $\inf_{\lambda\in\mathbb{R}} f(\lambda,\theta) > 0$, the covariance matrix of K_0,

$$\left(\sigma^2 M_{i,j}^{(m)}(\mu)\right)_{0\leq i,j\leq m-1} = \sigma^2 \left(r(i-j,\mu)\right)_{0\leq i,j\leq m-1}, \tag{25}$$

is invertible.

Let $P_\theta^{(m)}$ be the distribution of K_0, $p^{(m)}(.,\theta)$ its density under $\mathbb{P}_\theta$, and:

$$\begin{aligned} l^{(m)}(.,\theta) &= \ln p^{(m)}(.,\theta) \\ U_n^{(m)}(\theta) &= \frac{1}{n}\sum_{i=0}^{n-1} l^{(m)}(K_i,\theta), \quad \tilde{\theta}_n^{(m)} = \operatorname*{argmax}_{\theta\in\Theta} U_n^{(m)}(\theta) \end{aligned} \tag{26}$$

For $i, j \in \{1, 2\}$,

$$I_{i,j}^{(m)}(\theta) = E_\theta\left[\frac{\partial}{\partial\theta_i} l^{(m)}(K_0,\theta) \frac{\partial}{\partial\theta_j} l^{(m)}(K_0,\theta)\right] \tag{27}$$

For $i, j \in \{1, 2\}$, $k \geq 1$,

$$\gamma_{i,j}^{(m)}(k,\theta) = E_\theta\left[\frac{\partial}{\partial\theta_i} l^{(m)}(K_0,\theta) \frac{\partial}{\partial\theta_j} l^{(m)}(K_k,\theta)\right] \tag{28}$$

4.2 Asymptotic behaviour of the maximum likelihood split data estimator

Before stating results for $\tilde{\theta}_n^{(m)}$ we need two preliminary propositions. The first one is the identifiability assumption.

Proposition 4.1. *If $m \geq 2$, then $P_\theta^{(m)} = P_{\theta'}^{(m)}$ if and only if $\theta = \theta'$.*

Proof. Assume that $\theta = (\mu, \sigma^2)$, $\theta' = (\mu', \sigma^{2'})$ and $P_\theta^{(m)} = P_{\theta'}^{(m)}$. Since $P_\theta^{(m)}$ is a m-dimensional Gaussian law and $m \geq 2$, $P_\theta^{(m)} = P_{\theta'}^{(m)}$ implies the equality between the variance of J_0 and the covariance of (J_0, J_1) under $\mathbb{P}_\theta$ and $\mathbb{P}_{\theta'}$. By Proposition 2.2 and (4), (5):

$$\frac{\sigma^2}{\mu^2}\left(\Delta + \frac{1-e^{\mu\Delta}}{\mu}\right) = \frac{\sigma^{2'}}{\mu'^2}\left(\Delta + \frac{1-e^{\mu'\Delta}}{\mu'}\right) \tag{29}$$

$$\frac{\sigma^2}{2\mu^3}\left(1-e^{\mu\Delta}\right)^2 = \frac{\sigma^{2'}}{2\mu'^3}\left(1-e^{\mu'\Delta}\right)^2$$

It follows by a simple calculation that $\mu = \mu'$ and $\sigma^2 = \sigma^{2'}$. □

Remark 4.2. *If $m = 1$, then only one parameter may be identified.*

Now, for the asymptotic normality, we need the following result.

Proposition 4.3. *For $m \geq 2$ and $\theta \in \overset{\circ}{\Theta}$, $I^{(m)}(\theta)$ is non singular.*

Proof. Assume that $I^{(m)}(\theta)$ is singular, then $\det I^{(m)}(\theta) = 0$. By (27),

$$E_\theta\left[\left(\frac{\partial}{\partial\mu} l^{(m)}(K_0,\theta)\right)^2\right] E_\theta\left[\left(\frac{\partial}{\partial\sigma^2} l^{(m)}(K_0,\theta)\right)^2\right] = \left(E_\theta\left[\frac{\partial}{\partial\mu} l^{(m)}(K_0,\theta)\frac{\partial}{\partial\sigma^2} l^{(m)}(K_0,\theta)\right]\right)^2$$

This equality in the Cauchy-Schwarz inequality implies that there exits a constant $c(\theta)$ such that (recall (26)):

$$\frac{\partial}{\partial\mu}\ln p^{(m)}(x,\theta) = c(\theta)\frac{\partial}{\partial\sigma^2}\ln p^{(m)}(x,\theta) \quad \forall x \in \mathbb{R}^m \tag{30}$$

Using the fact that the covariance matrix of K_0 is $\sigma^2 M^{(m)}(\mu)$, we get:

$$\begin{aligned}
\ln p^{(m)}(x,\theta) &= -\frac{1}{2}\left[m2\pi + m\ln\sigma^2 + \ln\det M^{(m)}(\mu) + \frac{1}{\sigma^2}x^*\left(M^{(m)}(\mu)\right)^{-1}x\right] \\
\frac{\partial}{\partial\sigma^2}\ln p^{(m)}(x,\theta) &= -\frac{1}{2}\left[\frac{m}{\sigma^2} - \frac{1}{\sigma^4}x^*(M^{(m)}(\mu))^{-1}x\right] \\
\frac{\partial}{\partial\mu}\ln p^{(m)}(x,\theta) &= -\frac{1}{2}\left[\frac{\partial}{\partial\mu}(\ln\det M^{(m)}(\mu)) + \frac{1}{\sigma^2}x^*\frac{\partial}{\partial\mu}\left((M^{(m)}(\mu))^{-1}\right)x\right]
\end{aligned} \tag{31}$$

So, by (30), we must have $\frac{\partial}{\partial\mu}\left((M^{(m)}(\mu))^{-1}\right) = -\frac{c(\theta)}{\sigma^2}(M^{(m)}(\mu))^{-1}$; since $M^{(m)}(\mu)$ does not depend on σ^2, the same is true for $-\frac{c(\theta)}{\sigma^2}$. Set $\tilde{c}(\mu) = -\frac{c(\theta)}{\sigma^2}$, then we have:

$$\frac{\partial}{\partial\mu}\left((M^{(m)}(\mu))^{-1}\right) = \tilde{c}(\mu)(M^{(m)}(\mu))^{-1}$$

We can solve this equation: $(M^{(m)}(\mu))^{-1} = (M^{(m)}(\mu_0))^{-1}\exp\left(\int_{\mu_0}^{\mu}\tilde{c}(s)ds\right)$. Hence,

$$M^{(m)}(\mu) = M^{(m)}(\mu_0)\exp\left(-\int_{\mu_0}^{\mu}\tilde{c}(s)ds\right)$$

But this implies that $M_{0,0}^{(m)}(\mu)$ and $M_{0,1}^{(m)}(\mu)$ have the same asymptotic behaviour as $\mu \to -\infty$ ($\sim$ constant $\exp\left(-\int_{\mu_0}^{\mu}\tilde{c}(s)ds\right)$). And this is absurd: $M_{0,0}^{(m)}(\mu)$ is of order μ^{-2}, and $M_{0,1}^{(m)}(\mu)$ is of order μ^{-3} (using (4) and (5)). □

We can now prove that $\tilde{\theta}_n^{(m)}$ is asymptotically normal.

Theorem 4.4. *Assume $m \geq 2$ then*

$$H_{i,j}^{(m)}(\theta) = \sum_{k=1}^{\infty} \gamma_{i,j}^{(m)}(k,\theta), \quad \Gamma_{i,j}^{(m)}(\theta) = I_{i,j}^{(m)}(\theta) + 2H_{i,j}^{(m)}(\theta) \tag{32}$$

are well defined (for $i,j \in \{1,2\}$ and $\theta \in \mathring{\Theta}$) and

$$\sqrt{n}(\tilde{\theta}_n^{(m)} - \theta_0) \xrightarrow{n\to\infty} \mathcal{N}\left(0, I^{(m)}(\theta_0)^{-1}\Gamma^{(m)}(\theta_0)I^{(m)}(\theta_0)^{-1}\right)$$
$$= \mathcal{N}\left(0, I^{(m)}(\theta_0)^{-1} + 2I^{(m)}(\theta_0)^{-1}H^{(m)}(\theta_0)I^{(m)}(\theta_0)^{-1}\right)$$

in distribution under $\mathbb{P}_{\theta_0}$.

Proof. Consistency is obtained by adapting Ryden's proof (Ryden (1994)) to this model. We only consider the asymptotic normality.

Denote $\nabla l^{(m)}(x,\theta) = \left(\frac{\partial}{\partial\mu} l^{(m)}(x,\theta), \frac{\partial}{\partial\sigma^2} l^{(m)}(x,\theta)\right)$ and $\nabla^2 l^{(m)}(x,\theta) = \left[\frac{\partial^2}{\partial\theta_i\theta_j} l^{(m)}(x,\theta)\right]_{i,j\in\{1,2\}}$. By using standard arguments it is enough to prove

1) $\frac{1}{n}\sum_{i=0}^{n-1} \nabla^2 l^{(m)}(K_i,\theta_0) \xrightarrow{n\to\infty} -I^{(m)}(\theta_0)$ $\mathbb{P}_{\theta_0} a.s.$

2) $\frac{1}{\sqrt{n}}\sum_{i=0}^{n-1} \nabla l^{(m)}(K_i,\theta_0) \xrightarrow{n\to\infty} \mathcal{N}(0,\Gamma^{(m)}(\theta_0))$, in law under $\mathbb{P}_{\theta_0}$ and $\Gamma^{(m)}(\theta_0)$ is well defined.

3) For i,j,k in $\{1,2\}$, $\sup_{n\in\mathbb{N},\theta\in\Theta} \frac{1}{n}\sum_{i=0}^{n-1}\left|\frac{\partial^3}{\partial\theta_i\theta_j\theta_k} l^{(m)}(K_i,\theta)\right|$ is bounded in $\mathbb{P}_{\theta_0}$ probability.

Point 1) is obtained by the ergodicity of $(K_i,\ i\geq 0)$, since $E_{\theta_0}(\nabla^2 l^{(m)}(K_0,\theta_0)) = -I^{(m)}(\theta_0)$.

To obtain 2) by Theorem 1 p.46 in Doukhan (1994), it is enough to show that $H^{(m)}(\theta)$ is well defined . (Because the other assumptions of the theorem are easy to check: $E[|K_0|^{2+\delta}] < \infty$ and $\sum_{k=0}^{\infty} \alpha_K(k)^{\frac{\delta}{2+\delta}} < \infty$, for some $\delta > 0$, since $\alpha_K(k-1) \leq \alpha_J(km)$.)

Applying the first covariance inequality given in Doukhan (1994), Theorem 3 p.9 we get, for all $\theta \in \mathring{\Theta}$, (see (28))

$$\gamma_{i,j}^{(m)}(k,\theta) \leq 8\alpha_J^{\frac{1}{2}}(mk)\left\{E_\theta\left[\frac{\partial}{\partial\theta_i} l^{(m)}(K_0,\theta)\right]^4 E_\theta\left[\frac{\partial}{\partial\theta_j} l^{(m)}(K_k,\theta)\right]^4\right\}^{\frac{1}{4}}$$
$$= 8\alpha_J^{\frac{1}{2}}(mk)\left\{E_\theta\left[\frac{\partial}{\partial\theta_i} l^{(m)}(K_0,\theta)\right]^4 E_\theta\left[\frac{\partial}{\partial\theta_j} l^{(m)}(K_0,\theta)\right]^4\right\}^{\frac{1}{4}} .$$

Using Proposition 2.3, we see that $\sum_{k=0}^{\infty} \alpha_J^{\frac{1}{2}}(mk) \leq \infty$. Hence, $H^{(m)}(\theta)$ is well defined.

Finally, for i,j,k in $\{1,2\}$ and $n > 0$,

$$\sup_{\theta\in\Theta} \frac{1}{n}\sum_{i=0}^{n-1}\left|\frac{\partial^3}{\partial\theta_i\theta_j\theta_k} l^{(m)}(K_i,\theta)\right| \leq \frac{1}{n}\sum_{i=0}^{n-1}\sup_{\theta\in\Theta}\left|\frac{\partial^3}{\partial\theta_i\theta_j\theta_k} l^{(m)}(K_i,\theta)\right|,$$

which converges when, $n\to\infty$, to $E_{\theta_0}\left[\sup_{\theta\in\Theta}\left|\frac{\partial^3}{\partial\theta_i\theta_j\theta_k} l^{(m)}(\theta)\right|\right]$.

So we have the result. □

Remark 4.5. *If μ_0 is known, using* (31) *we get:*

$$\tilde{\sigma^2}_n^{(m)} = \frac{1}{mn}\sum_{k=0}^{n-1} K_k^*(M^{(m)}(\mu_0))^{-1}K_k$$

5 Comparison of the theoretical asymptotic variances

In this section our aim is to compare the efficient estimator $\hat{\theta}_n$ with the MLSD estimators $\tilde{\theta}_n^{(m)}$ by means of their asymptotic theoretical variances for different values of Δ and m.

Clearly, when m increases, the MLSD estimator must behave better: indeed, the asymptotic covariance matrices of $\sqrt{nm}(\hat{\theta}_{mn} - \theta_0)$ and $\sqrt{nm}(\tilde{\theta}_n^{(m)} - \theta_0)$ tend to be similar as m becomes large (see Table 2).

When Δ varies, for fixed m, the results are more surprising. The asymptotic variances of $\hat{\mu}_n$ and $\tilde{\mu}_n^{(m)}$ are very similar: Table 1 shows that this variance is high for small Δ. This is consistent with the usual results of drift estimation for diffusions based on the discrete observation of the diffusion itself (Dacunha-Castelle and Florens-Zmirou (1986)), where the asymptotic variance is shown to be of order $O(\frac{1}{\Delta})$.

On the contrary, the asymptotic variances of $\hat{\sigma^2}_n$ and $\tilde{\sigma^2}_n^{(m)}$ behave differently as shown in Table 2. When Δ is small, the variance of the MLSD estimator is high. Indeed, the numerical results are confirmed by the theorical result of Theorem 5.1.

Table 1: We assume that $\sigma_0 = 1$ is known, and $\mu_0 = -1$, the figures take account of the fact that $\tilde{\mu}_n^{(2)}$ uses $2n$ datas.

Theoretical asymptotic variances of the estimator [1]				
	$\Delta = 2$	$\Delta = 1$	$\Delta = 0.1$	$\Delta = 0.01$
$Var\ \sqrt{n}(\hat{\mu}_n - \mu_0)$	1.1	2.0	20.0	200.0
$Var\ \sqrt{2n}(\tilde{\mu}_n^{(2)} - \mu_0)$	1.1	2.0	20.0	200.0

Table 2: We assume that $\mu_0 = -1$ is known, and $\sigma_0 = 1$, the figures take account of the fact that $\tilde{\sigma^2}_n^{(m)}$ uses mn datas.

Theoretical asymptotic variances of the estimator [1]				
	$\Delta = 2$	$\Delta = 1$	$\Delta = 0.1$	$\Delta = 0.01$
$Var\ \sqrt{n}(\hat{\sigma^2}_n - \sigma^2)$	2	2	2	2
$Var\ \sqrt{2n}(\tilde{\sigma^2}_n^{(2)} - \sigma_0^2)$	2.2	2.7	7.4	52.4
$Var\ \sqrt{4n}(\tilde{\sigma^2}_n^{(4)} - \sigma_0^2)$	2.0	2.3	3.6	14.8
$Var\ \sqrt{8n}(\tilde{\sigma^2}_n^{(8)} - \sigma_0^2)$	2.1	2.2	2.6	5.3

[1] We have calculated $var\hat{\mu}_n$ numerically with the formulae (9) and (21) the integral being classically approximated.

The rest of the section is now devoted to the theoretical study of matrices $I^{(m)}(\theta)$, $H^{(m)}(\theta)$, and $\Gamma^{(m)}(\theta)$. As $m \to \infty$, it is possible to prove that $I^{(m)}(\theta)^{-1}\Gamma^{(m)}(\theta)I^{(m)}(\theta)^{-1} \sim_{m\to\infty} m^{-1}I(\theta)$ (a detailed proof is available upon request). The latter property is consistent with the numerical results of Table 2.

For $m = 2$, as $\Delta \to 0$, the following theorem precises the difference between both types of estimator.

Theorem 5.1. *In the case $m = 2$, we may precise the following expressions for $I^{(2)}(\theta)$ and $\Gamma^{(2)}(\theta)$*

$$I^{(2)}_{2,2}(\theta) = \frac{1}{\sigma^4}, \quad \Gamma^{(2)}_{2,2}(\theta) \sim_{\Delta\to 0} \frac{1}{4\sigma^4\,|\mu|\,\Delta}. \tag{33}$$

Hence, if μ_0 is known, the asymptotic variance of $\tilde{\sigma^2}^{(2)}_n$ is equivalent when $\Delta \to 0$ to $\frac{\sigma_0^4}{4|\mu_0|\Delta}$.

Proof. $I^{(2)}_{2,2}(\theta)$ is the Fisher information for the parameter σ^2 of a Gaussian vector with covariance matrix $\sigma^2 M^{(2)}(\mu)$ of size 2×2. So we know that $I^{(2)}_{2,2}(\sigma^2) = \frac{1}{\sigma^4}$.

To prove the second part of (33), let us calculate the sum $H^{(2)}_{2,2}(\theta)$. We write the following diagonalization of $M^{(2)}(\mu)$ (recall that, by (25), $M^{(2)}(\mu) = \begin{bmatrix} r(0,\mu) & r(1,\mu) \\ r(1,\mu) & r(0,\mu)\end{bmatrix}$):

$$M^{(2)}(\mu) = V^{(2)*}D^{(2)}(\mu)V^{(2)}$$

$$V^{(2)} = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ -1 & 1\end{bmatrix}, \quad D^{(2)}(\mu) = \begin{bmatrix} r(0,\mu)+r(1,\mu) & 0 \\ 0 & r(0,\mu)-r(1,\mu)\end{bmatrix}$$

Denote $L^{(2)}_k = V^{(2)}K_k$, for $k = 0, \dots, n-1$ and $N^{(2)}_i$, $\quad i = 2k,\ 2k+1$ the components of $L^{(2)}_k$:

$$L^{(2)}_0 = \begin{bmatrix} N^{(2)}_0 \\ N^{(2)}_1 \end{bmatrix}, \; L^{(2)}_1 = \begin{bmatrix} N^{(2)}_2 \\ N^{(2)}_3 \end{bmatrix}, \; \dots, \; L^{(2)}_{n-1} = \begin{bmatrix} N^{(2)}_{2(n-1)} \\ N^{(2)}_{2n-1} \end{bmatrix}$$

Now, formula (31) writes:

$$\frac{\partial}{\partial\sigma^2}\ln p^{(2)}(K_k,\theta) = \frac{1}{2\sigma^4}\left[\left(\frac{N^{(2)2}_{2k}}{r(0,\mu)+r(1,\mu)} - \sigma^2\right) + \left(\frac{N^{(2)\;2}_{2k+1}}{r(0,\mu)-r(1,\mu)} - \sigma^2\right)\right]$$

To compute $\gamma^{(2)}_{2,2}(k,\theta)$ (see (28)), we use the fact that if (Z, Z') is a Gaussian vector with law $\mathcal{N}\left(0, \begin{bmatrix} a & b \\ b & a\end{bmatrix}\right)$, $E[(Z^2-a)(Z'^2-a)] = 2b^2$, and that $Cov_\theta(N^{(2)}_i, N^{(2)}_{2k+j}) = e^{2(k-1)\mu\Delta}Cov_\theta(N^{(2)}_i, N^{(2)}_j)$, we obtain:

$$\gamma^{(2)}_{2,2}(k,\theta) = \frac{e^{4(k-1)\mu\Delta}}{2\sigma^8}S(\Delta) \quad ,$$

We have calculated $var\tilde{\mu}^{(2)}_n$ and $var\tilde{\sigma^2}^{(m)}_n$ numerically with the formulae (27), (28), (32) the expectations beeing calculated by using the special Gaussian form of the considered r.v. We numerically diagonalize their covariance matrices to get computation on independent variables.

where

$$S(\Delta) = \frac{Cov_\theta(N_0^{(2)}, N_2^{(2)})^2}{(r(0,\mu)+r(1,\mu))^2} + \frac{Cov_\theta(N_0^{(2)}, N_3^{(2)})^2}{r(0,\mu)^2 - r(1,\mu)^2} + \frac{Cov_\theta(N_1^{(2)}, N_2^{(2)})^2}{r(0,\mu)^2 - r(1,\mu)^2} + \frac{Cov_\theta(N_1^{(2)}, N_3^{(2)})^2}{(r(0,\mu)-r(1,\mu))^2}$$

Hence,

$$\sum_{k=1}^{\infty} \gamma_{2,2}^{(2)}(k,\theta) = \frac{1}{2\sigma^8(1-e^{4\mu\Delta})} S(\Delta)$$

So, we have to give the limit of $S(\Delta)$ when $\Delta \to 0$. Using $N_0^{(2)} = \frac{1}{\sqrt{2}}(J_0 + J_1), N_1^{(2)} = \frac{1}{\sqrt{2}}(-J_0 + J_1), N_2^{(2)} = \frac{1}{\sqrt{2}}(J_2 + J_3), N_3^{(2)} = \frac{1}{\sqrt{2}}(J_2 - J_3)$, and Proposition 2.2, we obtain:

$$S(\Delta) = \frac{\sigma^4}{4}\frac{(2r(2,\mu)+r(1,\mu)+r(3,\mu))^2}{(r(0,\mu)+r(1,\mu)^2} + \frac{\sigma^4}{4}\frac{(r(1,\mu)-r(3,\mu))^2}{r(0,\mu)^2 - r(1,\mu)^2} + \frac{\sigma^4}{4}\frac{(r(3,\mu)-r(1,\mu))^2}{r(0,\mu)^2 - r(1,\mu)^2} + \frac{\sigma^4}{4}\frac{(2r(2,\mu)-r(1,\mu)-r(3,\mu))^2}{(r(0,\mu)+r(1,\mu))^2}$$

And we easily deduce that $S(\Delta) \xrightarrow{\Delta\to 0} \sigma^4$, by using the following straightforward equalities (by (4)–(5)):

$$r(0,\mu) = -\frac{\Delta^2}{2\mu} - \frac{\Delta^3}{6} + o(\Delta^3), \quad r(1,\mu) = -\frac{\Delta^2}{2\mu} - \frac{\Delta^3}{2} + o(\Delta^3)$$
$$r(2,\mu) = -\frac{\Delta^2}{2\mu} - \Delta^3 + o(\Delta^3), \quad r(3,\mu) = -\frac{\Delta^2}{2\mu} - \frac{3\Delta^3}{2} + o(\Delta^3).$$

□

6 Conclusions and possible extensions

Let us now draw some conclusions on the two methods in view of possible extensions. Suppose we want to estimate unknown parameters of an ergodic one-dimensional diffusion (X_t) from the observation of the sample $J_i = \int_{i\Delta}^{(i+1)\Delta} X_s ds$, $0 \le i \le n-1$. First note that the exact distribution of a m-tuple $(J_i, i \le m-1)$ is hardly tractable for large m (hence for the whole sample $m = n$). So, actually, we started with the idea of using Ryden's method for small values of m in the general case. In fact, our result enlight the fact that the method will not be appropriate at least for estimating the diffusion coefficient parameters.

Moreover, even for small values of m ($m = 1, 2$) Ryden's likelihood will not be easily computable.

On the contrary, the Whittle contrast seems more suitable for generalization since it relies only on the covariance structure of the J_i's and there are several ergodic diffusions for which these covariances are explicit and simple. Further work is in progess in this direction.

7 Appendix

7.1 Proof of the proposition 2.1

We integrate (2), (with $h = \Delta$), between $i\Delta$ and $(i+1)\Delta$:

$$J_{i+1} - e^{\mu\Delta} J_i = \sigma \int_{i\Delta}^{(i+1)\Delta} e^{\mu(t+\Delta)} \int_t^{t+\Delta} e^{-\mu s} dB_s^\theta dt$$

Using the Fubini Theorem, we get:

$$J_{i+1} - e^{\mu\Delta} J_i = \sigma \int_{i\Delta}^{(i+1)\Delta} dB_s^\theta \left(e^{-\mu s} \int_{i\Delta}^{s} e^{\mu(t+\Delta)} dt \right) + \sigma \int_{(i+1)\Delta}^{(i+2)\Delta} dB_s^\theta \left(e^{-\mu s} \int_{s-i\Delta}^{(i+1)\Delta} e^{\mu(t+\Delta)} dt \right)$$

This gives the results

□

7.2 Proof of proposition 2.2

Since $(X_s)_{s\geq 0}$ is a strictly stationary Gaussian process, so is the process $(J_i)_{i\in\mathbb{N}}$. Because the expectation of X_s is zero, by the Fubini Theorem $E[J_i] = 0$, for all i.

Let us calculate the covariance function of $(J_i)_{i\in\mathbb{N}}$. Elementary computations show that covariance function of $(X_s)_{s\geq 0}$ is given by: $Cov(X_s, X_{s'}) = \frac{\sigma^2}{-2\mu} e^{\mu|s'-s|}$.

So, with some computations, for $0 \leq i \leq j$,

$$E(J_i J_j) = \int_{i\Delta}^{(i+1)\Delta} \int_{j\Delta}^{(j+1)\Delta} E[X_s X_{s'}] ds ds' = \sigma^2 r(j-i, \mu)$$

The spectral density $f(\lambda, \theta)$ is given by:

$$f(\lambda, \theta) = \sigma^2 r(0, \mu) + \sum_{k=1}^{\infty} \sigma^2 (e^{i\lambda k} r(k, \mu) + e^{-i\lambda k} r(k, \mu)) = \sigma^2 \frac{A_0 + 2A_1 \cos\lambda}{B_0 + 2B_1 \cos\lambda}$$

The inequalities of (10) are obtained by elementary computations.

□

References

- Bibby B.M. and Sørensen S. (1995). Martingale estimations functions for discretely observed diffusion processes. *Bernoulli*, **1** p.17–39.
- Brockwell P.J. and Davis R.A. (1991). *Time Series: Theory and Methods.* Springer-Verlag.
- Dacunha-Castelle D. and Duflo M. (1986). *Probability and statistics Volume 2.* Springer-Verlag.
- Dacunha-Castelle D. and Florens-Zmirou D. (1986). Estimation of the coefficients of a diffusion from discrete obsrevations. *Stochastics*, **19**.

- Doukhan P. (1994). *Mixing. Properties and examples.* Springer-Verlag.

- Dzhaparidze K.O. and Yaglom A.M. (1983). Spectrum parameter estimation in time series analysis. In *Developments in Statistics* (P.R. Krishnaiah, ed.) **4** p. 1–181 Academic, New York.

- Genon-Catalot V. and Jacod J. (1993). On the estimation of the diffusion coefficient for multidimensional diffusion processes. *Annales de l'I.H.P.,Probabilité-Statistiques*, **29** p.119–151.

- Ghysels E., Harvey A. and Renault E. (1996), Stochastic volatility. *Handbook of statistics*, **14** p.119–192.

- Karlin S. and Taylor H.M. (1981). *A second course in stochastic processes.* Academic Press, New–York.

- Kessler M. (1997). Estimation of an ergodic diffusion from discrete observations. *Scandinavian Journal of Statistics No.* **24***(2)* p.211–229.

- Kutoyants Yu. A. (1981). *Parameter Estimation for Stochastic Process.* Heldermann, Berlin.

- Pastorello S., Renaud E. and Touzi N. (1994). Statistical inference for random variance option pricing. *Southern European Economics Discussion Series*, D.P.**136**.

- Rudin W. (1966). *Real and complex analysis.* Mc.Graw Hill.

- Ryden T. (1994). Consistent and asymptotically normal parameter estimates for hidden markov models. *The annals of statistics*, **22**(4) p.1884–1895.

- Taylor S.J. and Xu X. (1994). The magnitude of implied volatility smiles: theory and empirical evidence for exchange rates. *Rev. Futures Markets*, **13**, p.355–380.

- Taylor S.J. and Xu X. (1995). The incremental volatility information in one million foreign exchange quotations. *Preprint, Conference in high frequency data in finance, Zürich.*

Chapitre I.2. Discrétisation d'une intégrale de diffusion et comparaison avec le schéma d'Euler

Discrete sampling of an integrated diffusion process and a comparison with the Euler scheme.

Abstract

Let (X_t) be a diffusion on the interval (l, r) and Δ_n a sequence of positive numbers tending to zero. We define J_i as the integral between $i\Delta_n$ and $(i+1)\Delta_n$ of X_s. We give an approximation of the law of $(J_0, \ldots, J_{n-1})$ useful for numerical simulations and statistical applications. This approximation is based on a Euler scheme expansion for the Markov process $(J_i, X_{(i+1)\Delta_n})$. In some special cases, an approximation by an explicit Gaussian ARMA(1,1) process is obtained.

1 Introduction

Consider the stochastic process $I_t = \int_0^t X_s ds$ where X is a one-dimensional diffusion process given by

$$dX_t = a(X_t)dB_t + b(X_t)dt, \quad X_0 = \eta, \tag{1}$$

with B a standard Brownian motion and η a random variable independent of B. The process I_t appears naturally in many problems studied recently.

First, the two-dimensional process (I_t, X_t) solves the system:

$$\begin{cases} dI_t = X_t \, dt \\ dX_t = a(X_t)dB_t + b(X_t)dt \end{cases} \tag{2}$$

which is a special case of two-dimensional model without noise in the first equation. In [8], the component I_t is used for modelling a non Markovian processes.

Second, integrals of stochastic processes play an important role in finance. For instance, in the continuous stochastic volatility models, introduced by Hull and White [5], the logarithm Y_t of the stock price is modelled by:

$$\begin{cases} dY_t = \rho(X_t)dt + \sqrt{X_t}dW_t, \\ dX_t = a(X_t)dB_t + b(X_t)dt \end{cases} \tag{3}$$

where X_t is a positive diffusion process called the volatility of the stock price. The quadratic variation of Y_t is $I_t = \int_0^t X_s ds$. This integrated volatility plays a crucial role in finance. For instance, to derive option prices formulae, it is necessary to compute the distribution of I_t. (see e.g Leblanc [7], see also Genon–Catalot *et al.* [3], Barndorff–Nielsen and Shephard [1])

Now, the exact distribution of the integrated process (I_t) is generally not explicit except for very few models.

In this paper, our concern is the study of the distribution of a discrete sampling of (I_t). We have in view statistical application to the inference of unknown parameters in the drift and diffusion coefficient of model (1) when the observation is $(I_{i\Delta}, i \leq n)$ for a positive sampling interval Δ. This will be the subject of a forthcoming paper.

For $i \geq 0$, let us set $J_i = \int_{i\Delta}^{(i+1)\Delta} X_s ds = I_{(i+1)\Delta} - I_{i\Delta}$. We study here the joint distribution of (J_i).

The case of X a stationary Ornstein–Uhlenbeck process,

$$dX_t = \mu X_t dt + \sigma dB_t$$

has been investigated in a previous work (see Gloter [4]). Explicit computations, for this special model, yield that (J_i) is a Gaussian ARMA(1,1) process.

There are difficulties to deal with a general case, and our approach is to obtain approximations when the sampling interval $\Delta = \Delta_n$ depends on n and tends to 0 as $n \to \infty$.

Recall that, under the assumption $\Delta_n \to 0$, using the Euler scheme,

$$X_{(i+1)\Delta_n} \simeq X_{i\Delta_n} + b(X_{i\Delta_n})\Delta_n + a(X_{i\Delta_n})(B_{(i+1)\Delta_n} - B_{i\Delta_n}) \tag{4}$$

we can approach the distribution of $(X_{i\Delta_n}, i \leq n)$: conditionally on $(X_{j\Delta_n}, j < i)$, $X_{i\Delta_n}$ is almost Gaussian with mean $X_{i\Delta_n} + \Delta_n b(X_{i\Delta_n})$ and variance $\Delta_n a^2(X_{i\Delta_n})$. This approximation has been fruitfully used for statistical applications (see e.g. Kessler [6]).

Here, we obtain expansions for $J_i^n = \int_{i\Delta_n}^{(i+1)\Delta_n} X_s ds = I_{(i+1)\Delta_n} - I_{i\Delta_n}$. For simplicity, we omit the superscript n and simply write $J_i = J_i^n$. Noting that $\frac{J_i}{\Delta_n}$ is close to $X_{i\Delta_n}$, we should in particular answer the following question: can we approximate the law of $(J_0, \dots, J_{n-1})$, by the law of a Markov process? Actually, we prove that $(\frac{J_i}{\Delta_n})$ is different from $(X_{i\Delta_n})$, and this has consequences for the statistical inference.

The paper is organized as follows. Assumptions on the model are presented in Section 2.1. Then, in Section 2.2, we give our asymptotic expansions. First we compare $\frac{J_i}{\Delta_n}$ and $X_{(i+1)\Delta_n}$ (Proposition 2.2): the difference is of order $\Delta_n^{\frac{1}{2}}$.

The process (J_i) is not Markov, but function of the Markov chain $Z_i = \begin{bmatrix} \frac{J_{i-1}}{\Delta_n} \\ X_{i\Delta_n} \end{bmatrix}$. In Theorem 2.4, we obtain a discrete Euler scheme formula, for this process:

$$Z_{i+1} \simeq P Z_i + \Delta_n b(X_{i\Delta_n}) \begin{bmatrix} \frac{1}{2} \\ 1 \end{bmatrix} + a(X_{i\Delta_n}) \Delta_n^{\frac{1}{2}} V_i,$$

where $P = \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix}$, V_i is a centered Gaussian vector with covariance matrix $\begin{bmatrix} \frac{1}{3} & \frac{1}{2} \\ \frac{1}{2} & 1 \end{bmatrix}$, independent of Z_j for $j \leq i$. Let us stress the fact that this scheme can be used for the numerical simulation of the vector $(J_0, \dots, J_{n-1})$ instead of the naive method which consists in simulating a Euler scheme for X with grid $\frac{\Delta_n}{m}$ and then approximate J_i by $\sum_{j=0}^{m-1} \frac{\Delta_n}{m} X_{(i+\frac{j}{m})\Delta_n}$.

However, the above scheme is not enough for the statistical applications we have in mind. So, in Theorem 2.6, we give an expansion of $\frac{J_{i+1}}{\Delta_n} - \frac{J_i}{\Delta_n}$:

$$\frac{J_{i+1}}{\Delta_n} - \frac{J_i}{\Delta_n} - b(\frac{J_i}{\Delta_n})\Delta_n = a(X_{i\Delta_n})(\xi_{i,n} + \xi'_{i+1,n})\Delta_n^{\frac{1}{2}} + \epsilon_{i,n}$$

where $\epsilon_{i,n}$ is a remainder term and the vector $(\xi_{i,n} + \xi'_{i+1,n})_{0 \leq i \leq n-1}$ is Gaussian with the same covariance matrix as a MA(1) process. Let us notice that the analogous term in (4), $\left(\frac{B_{(i+1)\Delta_n} - B_{i\Delta_n}}{\sqrt{\Delta_n}}\right)_{0 \leq i \leq n-1}$ is Gaussian with i.i.d. components. In particular, when a is constant, our expansion provides an approximation by a Gaussian ARMA(1,1) process. Hence, $(\frac{J_i}{\Delta_n})$ is really different from $X_{i\Delta_n}$.

A surprising consequence of our results is that, when $\Delta_n = \frac{T}{n}$, $T > 0$ fixed, then $\sum_{i=0}^{n-1} \left(\frac{J_{i+1}}{\Delta_n} - \frac{J_i}{\Delta_n}\right)^2$ is not a consistent estimator of the quadratic variation of (X_t) but converges to $\frac{2}{3} \int_0^T a^2(X_s) ds$.

Section 3 is devoted to the extension of our results when we replace the uniform mean $\frac{1}{\Delta} \int_{i\Delta}^{(i+1)\Delta} X_s ds$ by the more general form $\int_{i\Delta}^{(i+1)\Delta} X_s \phi(\frac{s - i\Delta}{\Delta}) ds$ with $\phi \geq 0$ and $\int_0^1 \phi(s) ds = 1$.

In Section 4, we give examples of classical models satisfying our set of assumptions.

2 Asymptotic expansions for small sampling interval.

2.1 Assumptions on the diffusion model

We assume that (X_t) is the one dimensional diffusion process defined by:

$$dX_t = a(X_t)dB_t + b(X_t)dt, \quad X_0 = \eta \tag{5}$$

where $(B_t)_{t\geq 0}$ is a standard Brownian motion, η is a random variable independent of (B_t).

Let $-\infty \leq l < r \leq \infty$ and consider the following assumptions.

(A1) Equation (5) admits a unique strong solution taking value in (l, r); a and b are two real valued functions defined on (l, r) with continuous second derivatives on (l, r).

Let us consider two positive measurable functions, $\mathcal{B}_l$ and $\mathcal{B}_r$ defined on (l, r) satisfying the following property: for all five non negative real numbers $\alpha, \beta, \alpha', \beta', p$, there exists a constant c such that for all $x \in (l, r)$:

$$(\mathcal{B}_l^{\alpha}(x) + \mathcal{B}_r^{\beta}(x)) \times (\mathcal{B}_l^{\alpha'}(x) + \mathcal{B}_r^{\beta'}(x)) \leq c(\mathcal{B}_l^{\alpha+\alpha'}(x) + \mathcal{B}_r^{\beta+\beta'}(x))$$
$$(\mathcal{B}_l^{\alpha}(x) + \mathcal{B}_r^{\beta}(x))^p \leq c(\mathcal{B}_l^{p\alpha}(x) + \mathcal{B}_r^{p\beta}(x))$$

These functions are introduced to bound the growth of other functions near the boundaries l and r. For instance, if $-\infty < l < \infty$ (respectively $-\infty < r < \infty$) we may take $\mathcal{B}_l(x) = 1 + \frac{1}{x-l}$ (respectively $\mathcal{B}_r(x) = 1 + \frac{1}{r-x}$). And if $l = -\infty$ (resp. $r = \infty$), we may take $\mathcal{B}_l(x) = 1 + |x|$ (resp. $\mathcal{B}_r(x) = 1 + |x|$).

(A2) There exist non negative constants c, $\alpha_1, \alpha_2, \beta_1, \beta_2$ such that, for all $x \in (l, r)$,

$$|a(x)| + |b(x)| \leq c(1 + \mathcal{B}_r(x)),$$
$$\left|a'(x)\right| \leq c(\mathcal{B}_l^{\alpha_1}(x) + \mathcal{B}_r^{\alpha_1}(x)), \quad \left|a''(x)\right| \leq c(\mathcal{B}_l^{\alpha_2}(x) + \mathcal{B}_r^{\alpha_2}(x)),$$
$$\left|b'(x)\right| \leq c(\mathcal{B}_l^{\beta_1}(x) + \mathcal{B}_r^{\beta_1}(x)), \quad \left|b''(x)\right| \leq c(\mathcal{B}_l^{\beta_2}(x) + \mathcal{B}_r^{\beta_2}(x)),$$

Now, let Δ_n be a sequence of positive numbers with $\Delta_n \to 0$ as $n \to \infty$ and assume that $\Delta_n \leq 1$ for all n.

We set $\mathcal{G}_t = \sigma(B_s,\ s \leq t;\ \eta)$ and $\mathcal{G}_i^n = \mathcal{G}_{i\Delta_n}$. Below, the values of the constant c may change from a line to another but never depends on i or n.

(A3) There exists a positive constant K_l, such that:

$\forall\ k \in [0, K_l)$, $\exists c$, $\forall i, n$ $(i \leq n)$, $E\left(\sup_{s\in[i\Delta_n,(i+1)\Delta_n]} \mathcal{B}_l^k(X_s) \mid \mathcal{G}_i^n\right) \leq c\mathcal{B}_l^k(X_{i\Delta_n})$

$\forall\ k \in [0, \infty)$, $\exists c$, $\forall i, n$ $(i \leq n)$, $E\left(\sup_{s\in[i\Delta_n,(i+1)\Delta_n]} \mathcal{B}_r^k(X_s) \mid \mathcal{G}_i^n\right) \leq c\mathcal{B}_r^k(X_{i\Delta_n})$

Assumption (A3) means that the diffusion will not approach too abruptly the end points l and r.

In the Appendix, we prove several propositions useful for checking (A3). The reason why our condition in not symetric in l and r appears in Section 4. For all the models considered here, we can prove that $\forall\ k \geq 0$, $\exists c$, $\forall i, n$, $E\left(\sup_{s\in[i\Delta_n,(i+1)\Delta_n]} \mathcal{B}_r^k(X_s) \mid \mathcal{G}_i^n\right) \leq c\mathcal{B}_r^k(X_{i\Delta_n})$, but for one model (the C.I.R. model), we can not prove an analogous result near the left end point $l = 0$ for all k, but only for k lower than a constant K_l.

2.2 Main results

$$\text{Let } J_i = J_i^n = \int_{i\Delta_n}^{i\Delta_n+\Delta_n} X_s ds,$$

and consider the following random variables which will appear in our expansions,

$$\xi_{i,n} = \frac{1}{\Delta_n^{\frac{3}{2}}} \int_{i\Delta_n}^{(i+1)\Delta_n} (s - i\Delta_n) dB_s \quad \text{for } i,\ n \geq 0 \tag{6}$$

$$\xi'_{i+1,n} = \frac{1}{\Delta_n^{\frac{3}{2}}} \int_{(i+1)\Delta_n}^{(i+2)\Delta_n} (i\Delta_n + 2\Delta_n - s) dB_s \quad \text{for } i \geq -1,\ n \geq 0 \tag{7}$$

Lemma 2.1. *The r.v. $\xi_{i,n}$ and $\xi'_{i+1,n}$ are independent and Gaussian; $\xi_{i,n}$ is $\mathcal{G}_{i+1}^n$ measurable and independent of $\mathcal{G}_i^n$; $\xi'_{i+1,n}$ is $\mathcal{G}_{i+2}^n$ measurable and independent of $\mathcal{G}_{i+1}^n$. The following expectations are useful for the sequel:*

$$\begin{aligned}
E\left(\xi_{i,n} \mid \mathcal{G}_i^n\right) &= E\left(\xi'_{i+1,n} \mid \mathcal{G}_i^n\right) &&= 0 \\
E\left(\xi_{i,n}^2 \mid \mathcal{G}_i^n\right) &= E\left(\xi'^2_{i+1,n} \mid \mathcal{G}_i^n\right) &&= \tfrac{1}{3} \\
E\left((\xi_{i,n}^2 - \tfrac{1}{3})^2 \mid \mathcal{G}_i^n\right) &= E\left((\xi'^2_{i+1,n} - \tfrac{1}{3})^2 \mid \mathcal{G}_i^n\right) &&= \tfrac{2}{9} \\
E\left((\xi_{i,n}^2 - \tfrac{1}{3})\xi'_{i,n} \mid \mathcal{G}_i^n\right) &= E\left((\xi'^2_{i+1,n} - \tfrac{1}{3})\xi'_{i,n} \mid \mathcal{G}_i^n\right) &&= 0 \\
E\left(\xi_{i,n}\xi'_{i,n} \mid \mathcal{G}_i^n\right) &= \tfrac{1}{6}
\end{aligned}$$

Proof. Easy computations based on (6) and (7) give the result. For example

$$E\left(\xi_{i,n}\xi'_{i,n} \mid \mathcal{G}_i^n\right) = \frac{1}{\Delta_n^3} \int_{i\Delta_n}^{i\Delta_n+\Delta_n} (s - i\Delta_n)(i\Delta_n + \Delta_n - s) ds = \frac{1}{6}$$

□

Our first result is a first order comparison between $\frac{J_i}{\Delta_n}$ and $X_{i\Delta_n}$.

Proposition 2.2. *Assume that $2\alpha_1 < K_l$, (with α_1 and K_l given in (A3)) then:*

$$\frac{J_i}{\Delta_n} - X_{i\Delta_n} = a(X_{i\Delta_n})\Delta_n^{\frac{1}{2}}\xi'_{i,n} + e_{i,n} \tag{8}$$

where,

$$\left|E\left(e_{i,n} \mid \mathcal{G}_i^n\right)\right| \leq \Delta_n c(1 + \mathcal{B}_r(X_{i\Delta_n})) \tag{9}$$

$$E\left(e_{i,n}^2 \mid \mathcal{G}_i^n\right) \leq \Delta_n^2 c(\mathcal{B}_l^{2\alpha_1}(X_{i\Delta_n}) + \mathcal{B}_r^{2(1+\alpha_1)}(X_{i\Delta_n})) \tag{10}$$

Moreover, if k is a real number ≥ 1, then for all i, n ($i \leq n-1$):

$$E\left(\left|\frac{J_i}{\Delta_n} - X_{i\Delta_n}\right|^k \mid \mathcal{G}_i^n\right) \leq \Delta_n^{\frac{k}{2}} c(1 + \mathcal{B}_r^k(X_{i\Delta_n})) \tag{11}$$

Proof. We have:

$$\frac{J_i}{\Delta_n} - X_{i\Delta_n} = \frac{1}{\Delta_n}\int_{i\Delta_n}^{(i+1)\Delta_n} (X_v - X_{i\Delta_n})dv$$

and $X_v - X_{i\Delta_n} = \int_{i\Delta_n}^{v} b(X_s)ds + \int_{i\Delta_n}^{v} a(X_s)dB_s$.

So by the Fubini theorem, we get:

$$\frac{J_i}{\Delta_n} - X_{i\Delta_n} = \frac{1}{\Delta_n} a(X_{i\Delta_n})\int_{i\Delta_n}^{(i+1)\Delta_n} ((i+1)\Delta_n - v)dB_v + e_{i,n} = \Delta_n^{\frac{1}{2}} a(X_{i\Delta_n})\xi'_{i,n} + e_{i,n} \tag{12}$$

where $e_{i,n} = \alpha_{i,n} + \beta_{i,n}$, and

$$\alpha_{i,n} = \frac{1}{\Delta_n}\int_{i\Delta_n}^{i\Delta_n+\Delta_n} (a(X_v) - a(X_{i\Delta_n}))(i\Delta_n + \Delta_n - v)dB_v \tag{13}$$

$$\beta_{i,n} = \frac{1}{\Delta_n}\int_{i\Delta_n}^{i\Delta_n+\Delta_n}\int_{i\Delta_n}^{v} b(X_s)dsdv \tag{14}$$

Using assumption (A2), we get $|\beta_{i,n}| \le c\Delta_n(1 + \sup_{s\in[i\Delta_n,(i+1)\Delta_n]} \mathcal{B}_r(X_s))$.
Now by Assumption (A3), for all $k \ge 0$

$$E\left(|\beta_{i,n}|^k \mid \mathcal{G}_i^n\right) \le c\Delta_n^k(1 + \mathcal{B}_r^k(X_{i\Delta_n})) \tag{15}$$

Also $E(\alpha_{i,n} \mid \mathcal{G}_i^n) = 0$, so we get $|E(e_{i,n} \mid \mathcal{G}_i^n)| \le c\Delta_n(1 + \mathcal{B}_r(X_{i\Delta_n}))$.
Now, for $k \ge 2$, applying the Burkhölder-Davis-Gundy and the Jensen inequalities yields:

$$E\left(\left|\alpha_{i,n}^k\right| \mid \mathcal{G}_i^n\right) \le \frac{c}{\Delta_n^k} E\left(\left(\int_{i\Delta_n}^{(i+1)\Delta_n} ((i+1)\Delta_n - v)^2 (a(X_v) - a(X_{i\Delta_n})^2 dv\right)^{\frac{k}{2}} \mid \mathcal{G}_i^n\right)$$
$$\le c\int_{i\Delta_n}^{(i+1)\Delta_n} \phi_k(v)dv$$

with $\phi_k(v) = E\left(|a(X_v) - a(X_{i\Delta_n})|^k \mid \mathcal{G}_i^n\right)$. By Proposition B of the Appendix, there exists $c > 0$ such that, for all $v \in [i\Delta_n, (i+1)\Delta_n]$,

$$\phi_k(v) \le c\Delta_n^{\frac{k}{2}}(\mathcal{B}_l^{\alpha_1 k}(X_{i\Delta_n}) + \mathcal{B}_r^{(\alpha_1+1)k}(X_{i\Delta_n}))$$

Finally,

$$E\left(\left|\alpha_{i,n}^k\right| \mid \mathcal{G}_i^n\right) \le c\Delta_n^{\frac{k}{2}+1}(\mathcal{B}_l^{\alpha_1 k}(X_{i\Delta_n}) + \mathcal{B}_r^{(\alpha_1+1)k}(X_{i\Delta_n})) \tag{16}$$

So, by (15), (16), with $k = 2$,

$$E\left(e_{i,n}^2 \mid \mathcal{G}_i^n\right) \le c\Delta_n^2(\mathcal{B}_l^{2\alpha_1}(X_{i\Delta_n}) + \mathcal{B}_r^{2(\alpha_1+1)}(X_{i\Delta_n})).$$

Using that Proposition B (in the Appendix), we have

$$E\left(\sup_{s\in[i\Delta_n,(i+1)\Delta_n]} |X_s - X_{i\Delta_n}|^k \mid \mathcal{G}_i^n\right) \le \Delta_n^{\frac{k}{2}}(1 + \mathcal{B}_r^k(X_{i\Delta_n})).$$

This implies (11). □

The following corollary is useful for statistical applications.

Corollary 2.3. *Let $f \in \mathcal{C}^2(l,r)$ satisfy that there exists γ such that $|f'(x)| + |f''(x)| \leq c(\mathcal{B}_l^\gamma(x) + \mathcal{B}_r^\gamma(x))$, for all $x \in (l,r)$. Then,*

$$f(\frac{J_i}{\Delta_n}) - f(X_{i\Delta_n}) = \Delta_n^{\frac{1}{2}} f'(X_{i\Delta_n}) a(X_{i\Delta_n}) \xi'_{i,n} + \tilde{e}_{i,n}(f)$$

where, if $2\alpha_1 \vee 2\gamma < K_l$,

$$|E(\tilde{e}_{i,n}(f) \mid \mathcal{G}_i^n)| \leq \Delta_n c(\mathcal{B}_l^\gamma(X_{i\Delta_n}) + \mathcal{B}_r^{2+\gamma}(X_{i\Delta_n}))$$

if $2\alpha_1 \vee 4\gamma < K_l$,

$$E(\tilde{e}_{i,n}^2(f) \mid \mathcal{G}_i^n) \leq \Delta_n^2 c(\mathcal{B}_l^{2\gamma+2\alpha_1}(X_{i\Delta_n}) + \mathcal{B}_r^{4+2\gamma+2\alpha_1}(X_{i\Delta_n}))$$

Proof. We write

$$f(\frac{J_i}{\Delta_n}) - f(X_{i\Delta_n}) = f'(X_{i\Delta_n})(\frac{J_i}{\Delta_n} - X_{i\Delta_n}) + \frac{1}{2} f''(\hat{X})(\frac{J_i}{\Delta_n} - X_{i\Delta_n})^2$$

with $\hat{X} \in [X_{i\Delta_n}, \frac{J_i}{\Delta_n}]$. Using Proposition 2.2,

$$f(\frac{J_i}{\Delta_n}) - f(X_{i\Delta_n}) = \Delta_n^{\frac{1}{2}} f'(X_{i\Delta_n}) a(X_{i\Delta_n}) \xi'_{i,n} + \tilde{e}_{i,n}(f)$$

where $\tilde{e}_{i,n}(f) = \tilde{\alpha}_{i,n} + \tilde{\beta}_{i,n}$ and

$$\tilde{\alpha}_{i,n} = f'(X_{i\Delta_n}) e_{i,n} \tag{17}$$

$$\tilde{\beta}_{i,n} = \frac{1}{2} f''(\hat{X})(\frac{J_i}{\Delta_n} - X_{i\Delta_n})^2 \tag{18}$$

We have the bound $\left| f''(\hat{X}) \right| \leq c \sup_{s \in [i\Delta_n, (i+1)\Delta_n]} (\mathcal{B}_l^\gamma(X_s) + \mathcal{B}_r^\gamma(X_s))$.

Now, Cauchy–Schwarz's inequality, (17), (18), (9),(10), and (11) for $k = 4$ and $k = 8$ give the result. □

The following theorem gives the second order expansion for $\frac{J_i}{\Delta_n} - X_{i\Delta_n}$.

Theorem 2.4. *Assume that $2\alpha_1 \vee \beta_1 \vee \beta_2 < K_l$. We have a Euler scheme expansion for the Markov process* $\begin{bmatrix} \frac{J_{i-1}}{\Delta_n} \\ X_{i\Delta_n} \end{bmatrix}$.

$$\begin{bmatrix} \frac{J_i}{\Delta_n} \\ X_{(i+1)\Delta_n} \end{bmatrix} = P \begin{bmatrix} \frac{J_{i-1}}{\Delta_n} \\ X_{i\Delta_n} \end{bmatrix} + b(X_{i\Delta_n}) \begin{bmatrix} \frac{\Delta_n}{2} \\ \Delta_n \end{bmatrix} + a(X_{i\Delta_n}) \begin{bmatrix} \Delta_n^{\frac{1}{2}} \xi'_{i,n} \\ B_{(i+1)\Delta_n} - B_{i\Delta_n} \end{bmatrix} + \eta_{i,n}$$

where $P = \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix}$, *$\eta_{i,n}$ is $\mathcal{G}_{i+1}^n$ measurable, and*

$$|E(\eta_{i,n} \mid \mathcal{G}_i^n)| \leq \Delta_n^2 c(\mathcal{B}_l^{\beta_1 \vee \beta_2}(X_{i\Delta_n}) + \mathcal{B}_r^{1+\beta_1 \vee 2+\beta_2}(X_{i\Delta_n})) \tag{19}$$

$$E(\eta_{i,n}^2 \mid \mathcal{G}_i^n) \leq \Delta_n^2 c(\mathcal{B}_l^{2\alpha_1}(X_{i\Delta_n}) + \mathcal{B}_r^{2(1+\alpha_1)}(X_{i\Delta_n})) \tag{20}$$

Proof. We write the classic Euler scheme: $X_{(i+1)\Delta_n} = X_{i\Delta_n} + b(X_{i\Delta_n})\Delta_n + a(X_{i\Delta_n})(B_{(i+1)\Delta_n} - B_{i\Delta_n}) + \eta^{(2)}_{i,n}$, with

$$\left|E\left(\eta^{(2)}_{i,n} \mid \mathcal{G}^n_i\right)\right| \leq c\Delta_n^2(\mathcal{B}_l^{\beta_1\vee\beta_2}(X_{i\Delta_n}) + \mathcal{B}_r^{1+\beta_1\vee 2+\beta_2}(X_{i\Delta_n}))$$
$$E\left({\eta^{(2)}_{i,n}}^2 \mid \mathcal{G}^n_i\right) \leq \Delta_n^2(\mathcal{B}_l^{2\alpha_1}(X_{i\Delta_n}) + \mathcal{B}_r^{2+2\alpha_1}(X_{i\Delta_n}))$$

Hence, to prove the proposition it suffices to improve (8) by a term of order Δ_n:

$$\frac{J_i}{\Delta_n} - X_{i\Delta_n} = a(X_{i\Delta_n})\xi'_{i,n} + \frac{\Delta_n}{2}b(X_{i\Delta_n}) + \eta^{(1)}_{i,n},$$

where $\eta^{(1)}_{i,n}$ satisfies conditions (19) and (20).

This is done by writing, with the notation of Proposition 2.2, $e_{i,n} = \alpha_{i,n} + \beta_{i,n}$ and using (13) and (14). □

Remark 2.5.
- *The above scheme is Markovian, since the vector* $(\Delta_n^{\frac{1}{2}}\xi'_{i,n}, B_{(i+1)\Delta_n} - B_{i\Delta_n})$ *is independent of* $(\frac{J_{j-1}}{\Delta_n}, X_{j\Delta_n})$ *for* $j \leq i$. *Furthermore, by* (7), $(\Delta_n^{\frac{1}{2}}\xi'_{i,n}, B_{(i+1)\Delta_n} - B_{i\Delta_n})$ *is centered and Gaussian with covariance matrix* $\begin{bmatrix} \frac{\Delta_n}{3} & \frac{\Delta_n}{2} \\ \frac{\Delta_n}{2} & \Delta_n \end{bmatrix}$.
- *This scheme is simple and useful to simulate the vector* $(J_0, \ldots, J_{n-1})$.

The Euler scheme of Theorem 2.4 does not provide an approximation of the law of $(\frac{J_i}{\Delta_n})$. This is done by the following further result.

Theorem 2.6. *We have*

$$\frac{J_{i+1}}{\Delta_n} - \frac{J_i}{\Delta_n} - b(\frac{J_i}{\Delta_n})\Delta_n = a(X_{i\Delta_n})(\xi_{i,n} + \xi'_{i+1,n})\Delta_n^{\frac{1}{2}} + \varepsilon_{i,n} \tag{21}$$

where $\varepsilon_{i,n}$ *is* $\mathcal{G}^n_{i+2}$ *measurable, and there exists a constant* c *such that for all* i, n:

If $\beta_1 \vee 2\beta_2 \vee 4\alpha_1 < K_l$,

$$|E(\varepsilon_{i,n} \mid \mathcal{G}^n_i)| \leq \Delta_n^2 c(\mathcal{B}_l^{(2\alpha_1+\beta_2)\vee\beta_1}(X_{i\Delta_n}) + \mathcal{B}_r^{(2+2\alpha_2+\beta_2)\vee(1+\beta_1)}(X_{i\Delta_n})) \tag{22}$$

If $4\alpha_1 \vee 2\alpha_2 < K_l$,

$$E\left(\varepsilon^2_{i,n} \mid \mathcal{G}^n_i\right) \leq \Delta_n^2 c(\mathcal{B}_l^{2\alpha_1\vee\alpha_2}(X_{i\Delta_n}) + \mathcal{B}_r^{3+2\alpha_1+\alpha_2}(X_{i\Delta_n})) \tag{23}$$
$$E\left(\varepsilon^4_{i,n} \mid \mathcal{G}^n_i\right) \leq \Delta_n^4 c(\mathcal{B}_l^{4\alpha_1\vee 2\alpha_2}(X_{i\Delta_n}) + \mathcal{B}_r^{6+4\alpha_1+2\alpha_2}(X_{i\Delta_n})) \tag{24}$$

Proof. We integrate, between $i\Delta_n$ and $(i+1)\Delta_n$, the following equality:

$$X_{s+\Delta_n} - X_s = \int_s^{s+\Delta_n}(a(X_v)dB_v + b(X_v)dv)$$

Hence, $J_{i+1} - J_i = A_i + B_i$, with

$$A_i = \int_{i\Delta_n}^{(i+1)\Delta_n} ds \int_s^{s+\Delta_n} a(X_v)dB_v, \quad B_i = \int_{i\Delta_n}^{(i+1)\Delta_n} ds \int_s^{s+\Delta_n} b(X_v)dv.$$

Interchanging the order of integrations, we obtain

$$A_i = \int_{i\Delta_n}^{(i+1)\Delta_n} a(X_v)(v - i\Delta_n)dB_v + \int_{(i+1)\Delta_n}^{(i+2)\Delta_n} a(X_v)((i+2)\Delta_n - v)dB_v$$

Analogously,

$$B_i = \int_{i\Delta_n}^{(i+1)\Delta_n} b(X_v)(v - i\Delta_n)dv + \int_{(i+1)\Delta_n}^{(i+2)\Delta_n} a(X_v)((i+2)\Delta_n - v)dv$$

Introducing $a(X_{i\Delta_n})$ in A_i yields

$$\begin{aligned} A_i &= a(X_{i\Delta_n})\left(\int_{i\Delta_n}^{(i+1)\Delta_n} (v - i\Delta_n)dB_v + \int_{(i+1)\Delta_n}^{(i+2)\Delta_n} ((i+2)\Delta_n - v)dB_v\right) + a_{i,n} + a'_{i+1,n} \\ &= a(X_{i\Delta_n})\Delta_n^{\frac{3}{2}}(\xi'_{i,n} + \xi'_{i+1,n}) + a_{i,n} + a'_{i+1,n} \end{aligned}$$

with

$$a_{i,n} = \int_{i\Delta_n}^{(i+1)\Delta_n} (v - i\Delta_n)(a(X_v) - a(X_{i\Delta_n}))dB_v \tag{25}$$

$$a'_{i,n} = \int_{(i+1)\Delta_n}^{(i+2)\Delta_n} ((i+2)\Delta_n - v)(a(X_v) - a(X_{i\Delta_n}))dB_v \tag{26}$$

Analogously, introducing now $b(\frac{J_i}{\Delta_n})$ in B_i yields

$$\begin{aligned} B_i &= b(\frac{J_i}{\Delta_n})\left(\int_{i\Delta_n}^{(i+1)\Delta_n} (v - i\Delta_n)dv + \int_{(i+1)\Delta_n}^{(i+2)\Delta_n} ((i+2)\Delta_n - v)dv\right) + b_{i,n} + b'_{i+1,n} \\ &= b(\frac{J_i}{\Delta_n})\Delta_n^2 + b_{i,n} + b'_{i+1,n} \text{with} \end{aligned}$$

$$b_{i,n} = \int_{i\Delta_n}^{(i+1)\Delta_n} (v - i\Delta_n)(b(X_v) - b(\frac{J_i}{\Delta_n}))dv \tag{27}$$

$$b'_{i,n} = \int_{(i+1)\Delta_n}^{(i+2)\Delta_n} ((i+2)\Delta_n - v)(b(X_v) - b(\frac{J_i}{\Delta_n}))dv \tag{28}$$

Therefore, we get the expansion

$$\frac{J_{i+1}}{\Delta_n} - \frac{J_i}{\Delta_n} - b(\frac{J_i}{\Delta_n})\Delta_n = a(X_{i\Delta_n})(\xi_{i,n} + \xi'_{i+1,n})\Delta_n^{\frac{1}{2}} + \varepsilon_{i,n}$$

$$\text{with} \quad \varepsilon_{i,n} = \frac{a_{i,n}}{\Delta_n} + \frac{a'_{i+1,n}}{\Delta_n} + \frac{b_{i,n}}{\Delta_n} + \frac{b'_{i+1,n}}{\Delta_n} \tag{29}$$

• Let us prove (22).

$$E\left(\frac{b_{i,n}}{\Delta_n} \mid \mathcal{G}_i^n\right) = \frac{1}{\Delta_n}\int_{i\Delta_n}^{i\Delta_n+\Delta_n} (v - i\Delta_n) E\left(b(X_v) - b(J_i) \mid \mathcal{G}_i^n\right) dv$$

Since we know by Ito's Formula, Assumptions (A3) and (A2), that

$$\sup_{v\in[[i\Delta_n,\ (i+1)\Delta_n]} \left|E\left(b(X_v) - b(X_{i\Delta_n}) \mid \mathcal{G}_i^n\right)\right| \leq \Delta_n c(\mathcal{B}_l^{\beta_1\vee\beta_2}(X_{i\Delta_n}) + \mathcal{B}_r^{(1+\beta_1)\vee(2+\beta_2)}(X_{i\Delta_n}))$$

an application of Corollary 2.3, yields

$$\sup_{v\in[i\Delta_n,\ (i+1)\Delta_n]} \left|E\left(b(X_v) - b(J_i) \mid \mathcal{G}_i^n\right)\right| \leq \Delta_n c(\mathcal{B}_l^{2\alpha_1+\beta_2}(X_{i\Delta_n}) + \mathcal{B}_r^{(2+2\alpha_1+\beta_2)}(X_{i\Delta_n}))$$

Hence

$$\left|E\left(\frac{b_{i,n}}{\Delta_n} \mid \mathcal{G}_i^n\right)\right| \leq \Delta_n^2 c(\mathcal{B}_l^{(2\alpha_1+\beta_2)\vee\beta_1}(X_{i\Delta_n}) + \mathcal{B}_r^{(2+2\alpha_1+\beta_2)\vee(1+\beta_1)}(X_{i\Delta_n}))$$

We bound analogously $\left|E\left(\frac{b'_{i+1,n}}{\Delta_n} \mid \mathcal{G}_i^n\right)\right|$ (see (28)).

Now, $E\left(a_{i,n} \mid \mathcal{G}_i^n\right) = E\left(a'_{i+1,n} \mid \mathcal{G}_i^n\right) = 0$, so (22)is proved.

• We now prove (23) and (24).

Using the Cauchy-Schwarz inequality it is enough to show (24).

By (A2), we write (see (27), (28)):

$$\left|\frac{b_{i,n}}{\Delta_n}\right| \leq \Delta_n 2c \sup_{s\in[i\Delta_n, i\Delta_n+\Delta_n]} (1 + \mathcal{B}_r(X_s)), \quad \left|\frac{b'_{i+1,n}}{\Delta_n}\right| \leq \Delta_n 2c \sup_{s\in[i\Delta_n+\Delta_n, i\Delta_n+2\Delta_n]} (1 + \mathcal{B}_r(X_s)).$$

Now, using Assumption (A3), we get

$$E\left(\left|\frac{b_{i,n}}{\Delta_n}\right|^4 \mid \mathcal{G}_i^n\right) \leq \Delta_n^4 c(1 + \mathcal{B}_r^4(X_{i\Delta_n})), \quad E\left(\left|\frac{b'_{i+1,n}}{\Delta_n}\right|^4 \mid \mathcal{G}_i^n\right) \leq \Delta_n^4 c(1 + \mathcal{B}_r^4(X_{i\Delta_n})).$$

To end the proof, we have to bound $E\left(\left|\frac{a_{i,n}}{\Delta_n}\right|^4 \mid \mathcal{G}_i^n\right)$. Using the Burkhölder-Davis-Gundy inequality we obtain:

$$E\left(\left|\frac{a_{i,n}}{\Delta_n}\right|^4 \mid \mathcal{G}_i^n\right) = \frac{c}{\Delta_n^4} E\left(\left(\int_{i\Delta_n}^{i\Delta_n+\Delta_n} (a(X_v) - a(X_{i\Delta_n}))^2 (v - i\Delta)^2 dv\right)^2 \mid \mathcal{G}_i^n\right) \tag{30}$$

But using the Ito formula, and assumption (A2) we can write: $(a(X_v) - a(X_{i\Delta_n}))^2 = M_v + A_v$, where

$$M_v = 2\int_{i\Delta_n}^{v} \psi_{i,n}(s) dB_s, \quad \text{with } \psi_{i,n}(s) = (a(X_s) - a(X_{i\Delta_n}))a'(X_s)a(X_s) \tag{31}$$

and:

$$\sup_{v\in[i\Delta_n,\ (i+1)\Delta_n]} |A_v| \le c\Delta_n \sup_{s\in[i\Delta_n,\ (i+1)\Delta_n]} (\mathcal{B}_r^{2\alpha_1\vee\alpha_2}(X_s) + \mathcal{B}_l^{(2+2\alpha_1)\vee(3+\alpha_2)}(X_s)). \tag{32}$$

So, replacing in (30), after some easy computations, and applying (A3) to the right hand side of (32), we get

$$E\left(\left|\frac{a_{i,n}}{\Delta_n}\right|^4 \mid \mathcal{G}_i^n\right) \le \frac{2}{\Delta_n^4} E\left(\gamma_{i,n}^2 \mid \mathcal{G}_i^n\right) + \Delta_n^4 (B_l^{4\alpha_1\vee 2\alpha_2}(X_{i\Delta_n}) + B_r^{6+4\alpha_1+2\alpha_2} X_{i\Delta_n}))$$

with $\gamma_{i,n} = \int_{i\Delta_n}^{(i+1)\Delta_n} M_v(v-i\Delta_n)^2 dv$ (see (31)).

It remains to bound $E\left(\gamma_{i,n}^2 \mid \mathcal{G}_i^n\right)$. Using the Fubini Theorem, we have

$$\gamma_{i,n} = \int_{i\Delta_n}^{(i+1)\Delta_n} \psi_{i,n}(s)\left(\int_s^{i\Delta_n+\Delta_n} (v-i\Delta_n)^2 dv\right) dB_s$$

Hence,

$$E\left(\gamma_{i,n}^2 \mid \mathcal{G}_i^n\right) \le \int_{i\Delta_n}^{(i+1)\Delta_n} E\left(\psi_{i,n}^2(s) \mid \mathcal{G}_i^n\right) \Delta_n^6 ds \tag{33}$$

But by the Cauchy–Schwarz inequality,

$$E\left(\psi_{i,n}^2(s) \mid \mathcal{G}_i^n\right) \le E\left((a(X_s)-a(X_{i\Delta_n}))^4 \mid \mathcal{G}_i^n\right)^{\frac{1}{2}} E\left(a'^4(X_s)a^4(X_s) \mid \mathcal{G}_i^n\right)^{\frac{1}{2}}.$$

Then using (A2), (A3) and Proposition B of the Appendix, we obtain for $s \in [i\Delta_n, (i+1)\Delta_n]$:

$$E\left(\psi_{i,n}^2(s) \mid \mathcal{G}_i^n\right) \le \Delta_n c(\mathcal{B}_l^{4\alpha_1}(X_{i\Delta_n}) + \mathcal{B}_r^{4+2\alpha_1}(X_{i\Delta_n}))$$

Replacing the last inequation in (33), we get $E\left(\gamma_{i,n}^2 \mid \mathcal{G}_i^n\right) \le \Delta_n^8 c(\mathcal{B}_l^{4\alpha_1}(X_{i\Delta_n}) + \mathcal{B}_r^{4+2\alpha_1})$. We obtain a similar bound for $E\left(\left|\frac{a'_{i+1,n}}{\Delta_n}\right|^4 \mid \mathcal{G}_i^n\right)$, and hence the theorem is proved. □

Corollary 2.7. *Assume that,* $4\alpha_1$, α_2 *and* $4\beta_1 < K_l$. *Then, we have the following inequalities:*

$$|E\left(\varepsilon_{i,n}\xi_{i,n} \mid \mathcal{G}_i^n\right)| \le \Delta_n^{\frac{3}{2}} c(\mathcal{B}_l^{(\alpha_1+\beta_1)\vee\alpha_2}(X_{i\Delta_n}) + \mathcal{B}_r^{(1+\alpha_1+\beta_1)\vee(2+\alpha_2)}(X_{i\Delta_n})) \tag{34}$$

$$\left|E\left(\varepsilon_{i,n}\xi'_{i+1,n} \mid \mathcal{G}_i^n\right)\right| \le \Delta_n^{\frac{3}{2}} (\mathcal{B}_l^{(\alpha_1+\beta_1)\vee\alpha_2)}(X_{i\Delta_n}) + \mathcal{B}_r^{(1+\alpha_1+\beta_1)\vee(2+\alpha_2)}(X_{i\Delta_n})) \tag{35}$$

Proof. We only prove (34). With the notations of Theorem 2.6, we shall show:

$$|E\left(a_{i,n}\xi_{i,n} \mid \mathcal{G}_i^n\right)| \le \Delta_n^{\frac{5}{2}} c(\mathcal{B}_l^{\alpha_1\vee\alpha_2}(X_{i\Delta_n}) + \mathcal{B}_r^{(1+\alpha_1)\vee(2+\alpha_1)}(X_{i\Delta_n}))$$

$$\left|E\left(a'_{i+1,n}\xi_{i,n} \mid \mathcal{G}_i^n\right)\right| = 0$$

$$|E\left(b_{i,n}\xi_{i,n} \mid \mathcal{G}_i^n\right)| \le \Delta_n^{\frac{5}{2}} c(\mathcal{B}_l^{\alpha_1+\beta_1}(X_{i\Delta_n}) + \mathcal{B}_r^{1+\alpha_1+\beta_1}(X_{i\Delta_n}))$$

$$\left|E\left(b'_{i+1,n}\xi_{i,n} \mid \mathcal{G}_i^n\right)\right| \le \Delta_n^{\frac{5}{2}} c(\mathcal{B}_l^{\alpha_1+\beta_1}(X_{i\Delta_n}) + \mathcal{B}_r^{1+\alpha_1+\beta_1}(X_{i\Delta_n}))$$

These inequalities follow from the expressions of $a_{i,n}$, $a'_{i+1,n}$, $b_{i,n}$, $b'_{i+1,n}$ and $\xi_{i,n}$. □

Remark 2.8. *1. In the case where X is a stationary Ornstein–Uhlenbeck process $dX_t = \mu X_t dt + \sigma dB_t$ and for a fixed sampling interval Δ, we have an exact formula, analogous to our expansion* (21) *(see [4]).*

$$J_{i+1} - e^{\mu\Delta} J_i = \frac{\sigma}{\mu} \int_{i\Delta}^{(i+1)\Delta} (e^{\mu\Delta} - e^{\mu((i+1)\Delta - s)}) dB_s + \frac{\sigma}{\mu} \int_{(i+1)\Delta}^{(i+2)\Delta} (e^{\mu((i+2)\Delta - s)} - 1) dB_s$$

Furthermore, in this case, the covariance structure of (J_i) is the one of an ARMA(1,1) process.

2. $(U_i)_{i=0,\dots,n-1} = (\xi_{i,n} + \xi'_{i+1,n})_{i=0,\dots,n-1}$ is a Gaussian vector with covariance function : Var $U_i = \frac{2}{3}$, Cov $(U_i, U_{i+1}) = \frac{1}{6}$ and Cov $(U_i, U_{i+k}) = 0$ if $k \geq 2$.

This is the covariance function of an MA(1) vector. Therefore through the expansion (21) *we do not recover a Markovian property for $\frac{J_i}{\Delta_n}$. In the special case where a is constant, the expansion means that the process $(\frac{J_i}{\Delta_n})$ may be approximated by an ARMA(1,1) process.*

2.3 A statistical application

Let us assume that $\Delta_n = \frac{T}{n}$, for $T > 0$. We show that, we can not replace $X_{i\Delta_n}$ by $\frac{J_i}{\Delta_n}$ in the classical approximation of the quadratic variation of X:

$$\sum_{i=0}^{n-1} (X_{(i+1)\Delta_n} - X_{i\Delta_n})^2 \xrightarrow{n\to\infty} \int_0^T a^2(X_s) ds.$$

Proposition 2.9. *Let $(X_t)_{t\in[0,T]}$ be a diffusion satisfying (A1) - (A3) with $K_l = \infty$. Furthermore assume that:*

(A4) $\forall\, k \geq 0$, $\forall\, t \in [0,T]$ $E\left(\mathcal{B}_l^k(X_t)\right) < \infty$, $E\left(\mathcal{B}_r^k(X_t)\right) < \infty$.

Then,

$$\sum_{i=0}^{n-1} \left(\frac{J_{i+1}}{\Delta_n} - \frac{J_i}{\Delta_n} \right)^2 \xrightarrow{n\to\infty} \frac{2}{3} \int_0^T a^2(X_s) ds$$

Proof. First, we remark that by Assumptions (A3) and (A4), we have

$$\sup_{t\in[0,T]} E\left(\mathcal{B}_l^k(X_t)\right) < \infty \quad \text{and} \quad \sup_{t\in[0,T]} E\left(\mathcal{B}_r^k(X_t)\right) < \infty.$$

By the continuity of a and X, $\frac{2}{3}\sum_{i=0}^{n-1} a^2(X_{i\Delta_n})\Delta_n \xrightarrow[\mathbf{P}]{n\to\infty} \frac{2}{3}\int_0^T a^2(X_s)ds$. Hence it is enough to prove:

$$\sum_{i=0}^{n-1} \left\{ \left(\frac{J_{i+1}}{\Delta_n} - \frac{J_i}{\Delta_n} \right)^2 - \frac{2\Delta_n}{3} a^2(X_{i\Delta_n}) \right\} \xrightarrow[\mathbf{P}]{n\to\infty} 0.$$

For this, we use the expansion (21):

$$\sum_{i=0}^{n-1} \left\{ \left(\frac{J_{i+1}}{\Delta_n} - \frac{J_i}{\Delta_n} \right)^2 - \frac{2\Delta_n}{3} a^2(X_{i\Delta_n}) \right\} = D_n^{(1)} + D_n^{(2)} + D_n^{(3)}.$$

with:
$$D_n^{(1)} = \sum_{i=0}^{n-1} \Delta_n a^2(X_{i\Delta_n})\{(\xi_{i,n} + \xi'_{i+1,n})^2 - \frac{2}{3}\}$$
$$D_n^{(2)} = \sum_{i=0}^{n-1} (b(\frac{J_i}{\Delta_n})\Delta_n + \varepsilon_{i,n})^2$$
$$D_n^{(3)} = 2\sum_{i=0}^{n-1} \sqrt{\Delta_n} a(X_{i\Delta_n})(\xi_{i,n} + \xi'_{i+1,n})(b(\frac{J_i}{\Delta_n})\Delta_n + \varepsilon_{i,n})$$

Now, we bound, using (A2), (A3) and (23):
$$E\left(\left|D_n^{(2)}\right|\right) \leq c\Delta_n^2 \sup_{t\in[0,T]} E\left(\mathcal{B}_l^k(X_t) + \mathcal{B}_r^k(X_t)\right), \text{ with } k > 0.$$

We deduce $D_n^{(2)} \xrightarrow[\mathbf{P}]{n\to\infty} 0$.

Analogously: $E\left(\left|D_n^{(3)}\right|\right) \leq c\Delta_n^{\frac{3}{2}} \sup_{t\in[0,T]} E\left(\mathcal{B}_l^{k'}(X_t) + \mathcal{B}_r^{k'}(X_t)\right)$, with $k' > 0$. Hence $D_n^{(3)} \xrightarrow[\mathbf{P}]{n\to\infty}$ 0.

Using that, by Lemma 2.1, $E\left((\xi_{i,n} + \xi'_{i+1,n})^2 - \frac{2}{3} \mid \mathcal{G}_i^n\right) = 0$ and then Cauchy–Schwarz's inequality we compute:

$$E\left({D_n^{(1)}}^2\right) = \sum_{\substack{0\leq i,j\leq n-1\\ |i-j|\leq 1}} \Delta_n^2 E\left(a^2(X_{i\Delta_n})\{(\xi_{i,n} + \xi'_{i+1,n})^2 - \frac{2}{3}\}^2 a^2(X_{j\Delta_n})\{(\xi_{j,n} + \xi'_{j+1,n})^2 - \frac{2}{3}\}^2\right)$$
$$\leq cn\Delta_n^2 \sup_{t\in[0,T]} E(1 + \mathcal{B}_r^8(X_t))^{\frac{1}{2}}$$

Therefore, $D_n^{(1)}$ converges to zero in $\mathbf{L}^2$ sense. □

We deduce that if X solves $dX_t = b(X_t)dt + \sigma dB_t$ (and satisfies (A1)-(A4)) then,
$$\hat{\sigma^2}_n = \frac{3}{2}\sum_{i=0}^{n-1}(\frac{J_{i+1}}{\Delta_n} - \frac{J_i}{\Delta_n})^2 \tag{36}$$

is a consistent estimator of σ^2 based on (J_i). Further statistical applications will be investigated in a forthcoming paper. (see Delattre [2] for the close issue of parameters estimation for a diffusion which is observed with round off errors.)

3 Extension

In Section 2 we gave results for the process $(\frac{J_0}{\Delta_n}, \ldots, \frac{J_{n-1}}{\Delta_n})$, where $\frac{J_i}{\Delta_n}$ is the mean, in the interval $[i\Delta_n, (i+1)\Delta_n]$ of X. Here, we suppose that ϕ is a measurable, bounded, non negative fonction defined on $[0,1]$ and such that $\int_0^1 \phi(s)ds = 1$; we define
$$\hat{X}_i(\phi) = \int_{i\Delta_n}^{(i+1)\Delta_n} X_s\phi(\frac{s - i\Delta_n}{\Delta_n})ds.$$

In this Section, we extend results of Section 2.2 to $(\hat{X}_i(\phi))$. We do not give proofs for these results, since they are analogous to those of Section 2.2.

We define:

$$\xi_{i,n}(\phi) = \frac{1}{\Delta_n^{\frac{3}{2}}} \int_{i\Delta_n}^{(i+1)\Delta_n} \int_{i\Delta_n}^{s} \phi(\frac{v - i\Delta_n}{\Delta_n}) dv dB_s \quad \text{for } i,\ n \geq 0$$

$$\xi'_{i+1,n}(\phi) = \frac{1}{\Delta_n^{\frac{3}{2}}} \int_{(i+1)\Delta_n}^{(i+2)\Delta_n} \int_{s-\Delta_n}^{(i+1)\Delta_n} \phi(\frac{v - i\Delta_n}{\Delta_n}) dv dB_s \quad \text{for } i \geq -1,\ n \geq 0$$

Proposition 3.1. *Assume that $2\alpha_1 < K_l$, then:*

$$\hat{X}_i(\phi) - X_{i\Delta_n} = a(X_{i\Delta_n})\Delta_n^{\frac{1}{2}} \xi'_{i,n}(\phi) + e_{i,n}(\phi)$$

where,

$$|E(e_{i,n}(\phi) \mid \mathcal{G}_i^n)| \leq \Delta_n c(1 + \mathcal{B}_r(X_{i\Delta_n}))$$

$$E\left(e_{i,n}(\phi)^2 \mid \mathcal{G}_i^n\right) \leq \Delta_n^2 c(\mathcal{B}_l^{2\alpha_1}(X_{i\Delta_n}) + \mathcal{B}_r^{2(1+\alpha_1)}(X_{i\Delta_n}))$$

Moreover, if k is a real number ≥ 1, for all i, n:

$$E\left(\left|\hat{X}_i(\phi) - X_{i\Delta_n}\right|^k \mid \mathcal{G}_i^n\right) \leq \Delta_n^{\frac{k}{2}} c(1 + \mathcal{B}_r^k(X_{i\Delta_n}))$$

Proposition 3.2. *Assume that $2\alpha_1 \vee \beta_1 \vee \beta_2 < K_l$. We have a Euler scheme expansion for the Markov process $\begin{bmatrix} \hat{X}_{i-1}(\phi) \\ X_{i\Delta_n} \end{bmatrix}$.*

$$\begin{bmatrix} \hat{X}_i(\phi) \\ X_{(i+1)\Delta_n} \end{bmatrix} = P \begin{bmatrix} \hat{X}_{i-1}(\phi) \\ X_{i\Delta_n} \end{bmatrix} + b(X_{i\Delta_n}) \begin{bmatrix} \Delta_n \int_0^1 \int_0^s \phi(v) dv ds \\ \Delta_n \end{bmatrix}$$
$$+ a(X_{i\Delta_n}) \begin{bmatrix} \Delta_n^{\frac{1}{2}} \xi'_{i,n}(\phi) \\ B_{(i+1)\Delta_n} - B_{i\Delta_n} \end{bmatrix} + \eta_{i,n}(\phi)$$

where $P = \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix}$, $\eta_{i,n}(\phi)$ is $\mathcal{G}_{i+1}^n$ measurable, and

$$|E(\eta_{i,n}(\phi) \mid \mathcal{G}_i^n)| \leq \Delta_n^2 c(\mathcal{B}_l^{\beta_1 \vee \beta_2}(X_{i\Delta_n}) + \mathcal{B}_r^{1+\beta_1 \vee 2+\beta_2}(X_{i\Delta_n}))$$

$$E\left(\eta_{i,n}(\phi)^2 \mid \mathcal{G}_i^n\right) \leq \Delta_n^2 c(\mathcal{B}_l^{2\alpha_1}(X_{i\Delta_n}) + \mathcal{B}_r^{2(1+\alpha_1)}(X_{i\Delta_n}))$$

Theorem 3.3. *We have*

$$\hat{X}_{i+1}(\phi) - \hat{X}_i(\phi) - b(\hat{X}_i(\phi))\Delta_n = a(X_{i\Delta_n})(\xi_{i,n}(\phi) + \xi'_{i+1,n}(\phi))\Delta_n^{\frac{1}{2}} + \varepsilon_{i,n}(\phi)$$

where $\varepsilon_{i,n}(\phi)$ is $\mathcal{G}_{i+2}^n$ measurable, and there exists a constant c such that for all i, n:

If $\beta_1 \vee 2\beta_2 \vee 4\alpha_1 < K_l$,

$$|E(\varepsilon_{i,n}(\phi) \mid \mathcal{G}_i^n)| \leq \Delta_n^2 c(\mathcal{B}_l^{(2\alpha_1+\beta_2)\vee\beta_1}(X_{i\Delta_n}) + \mathcal{B}_r^{(2+2\alpha_2+\beta_2)\vee(1+\beta_1)}(X_{i\Delta_n}))$$

If $(4\alpha_1) \vee 2\alpha_2 < K_l$,

$$E\left(\varepsilon_{i,n}(\phi)^2 \mid \mathcal{G}_i^n\right) \leq \Delta_n^2 c(\mathcal{B}_l^{2\alpha_1 \vee \alpha_2}(X_{i\Delta_n}) + \mathcal{B}_r^{3+2\alpha_1+\alpha_2}(X_{i\Delta_n}))$$

$$E\left(\varepsilon_{i,n}(\phi)^4 \mid \mathcal{G}_i^n\right) \leq \Delta_n^4 c(\mathcal{B}_l^{4\alpha_1 \vee 2\alpha_2}(X_{i\Delta_n}) + \mathcal{B}_r^{6+4\alpha_1+2\alpha_2}(X_{i\Delta_n}))$$

4 Examples

4.1 Diffusion on $\mathbb{R}$

Here, $(l, r) = (-\infty, \infty)$ and we set $\mathcal{B}_{-\infty}(x) = 1$, $\mathcal{B}_{\infty}(x) = 1 + |x|$ (by this choice, we decide, for simplification, to use the same function, $\mathcal{B}_{\infty}$, to bound other functions near the two different bounds $-\infty$, ∞). Let X be the solution of $dX_t = a(X_t)dB_t + b(X_t)dt$, and assume that there exists c such that for all $x \in \mathbb{R}$, $|a(x)| + |b(x)| \leq c(1 + |x|)$; a and b are twice continuously diffentiable and their second derivatives have polynomial growth.

Then, it is immediate to check Assumptions (A1) and (A2). By Proposition A of the Appendix, Assumption (A3) is satisfied with $K_l = \infty$.

4.2 Positive diffusions

In these models $l = 0$ and $r = \infty$.

4.2.1 Exponantial of a diffusion on $\mathbb{R}$

Here $(l, r) = (0, \infty)$ and we set $\mathcal{B}_0(x) = 1 + \frac{1}{x}$ and $\mathcal{B}_{\infty}(x) = 1 + x^2$.

We assume that $X_t = \exp(Z_t)$, where Z_t is a diffusion on $\mathbb{R}$ defined as the solution of the equation: $dZ_t = \tilde{a}(Z_t)dB_t + \tilde{b}(Z_t)dt$ and $Z_0 = \eta$, is independent of B. Consider the following assumptions.

Functions $\tilde{a}$ and $\tilde{b}$ are defined on $\mathbb{R}$; $\tilde{a}$ is bounded; $\limsup_{z\to\infty} \tilde{b}(z) < \infty$; $\liminf_{z\to-\infty} \tilde{b}(z) > -\infty$; there exists c such that: for all $z \in \mathbb{R}$, $\left|\tilde{b}(z)\right| \leq c(1 + |z|)$; $\tilde{a}$ and $\tilde{b}$ are twice continuously differentiable; their second derivatives have polynomial growth and $\tilde{a}'$ is bounded.

Then diffusion X satisfies the stochastic differential equation:

$$dX_t = a(X_t)dB_t + b(X_t)dt, \tag{37}$$

with $a(x) = x\tilde{a}(\ln x)$ and $b(x) = x\tilde{b}(\ln x) + \frac{1}{2}x\tilde{a}^2(\ln x)$. By assumptions, a' is bounded, so a has linear growth.

Proposition 4.1. *The diffusion (X_t) satisfies Assumptions (A1)-(A3), with $K_l = \infty$.*

Proof. (A1) and (A2) are clear. For (A3), we first show the inequality with $\mathcal{B}_{\infty}$. Using the Markov property for X, it suffices to show that, for all $k \geq 0$, and Y solution of (37) starting with $y_0 > 0$, there exists c such that:

$$E\left(\sup_{t\in[0,1]} Y_t^k\right) \leq c(1 + y_0^k). \tag{38}$$

Since $\tilde{a}$ is bounded and $\limsup_{x\to\infty} \tilde{b}(x) < \infty$ there exists a constant M such that $\forall\ x > 0$, $b(x) \leq M(x + 1)$. We define $\hat{Y}$ as the strong solution, with the same Brownian motion B, of

$d\hat{Y}_t = a(\hat{Y}_t)dB_t + M(\hat{Y}_t + 1)dt$, $\hat{Y}_0 = y_0$. Using a comparison theorem (see Revuz–Yor p.375 [9]), we get:

$$\forall t \geq 0, \quad Y_t \leq \hat{Y}_t \text{ a.s..} \tag{39}$$

Now, $\hat{Y}$ solves a stochastic differential equation with coefficients having at most a linear growth, so by Proposition A of the Appendix:

$$E\left(\sup_{s\in[0,1]} \left|\hat{Y}_s\right|^k\right) \leq c(1 + y_0^k). \tag{40}$$

Then, (39) and (40) imply (38), hence we get the first inequality of (A3).

Noticing that $\frac{1}{X} = \exp(-Z)$, and $-Z$ has the same properties as Z, we obtain the second inequality (with $\mathcal{B}_0$) analogously.

□

Remark 4.2. *Hence, our results are valid for the exponantial of a Brownian motion or the exponantial an Ornstein–Uhlenbeck process.*

4.2.2 Cox–Ingersoll–Ross process

Again, here, $(l, r) = (0, \infty)$ and we set $\mathcal{B}_0(x) = 1 + \frac{1}{x}$ and $\mathcal{B}_\infty(x) = 1 + x$.

Let X be given by:

$$dX_t = (\alpha X_t + \beta)dt + \sigma\sqrt{X_t}dB_t, \quad X_0 = \eta, \tag{41}$$

with $\alpha < 0$, $\sigma, \beta > 0$ and η is a positive random variable independent of (B_t), and $\frac{2\beta}{\sigma^2} > 1$.

Assumptions (A1) and (A2) holds. Combining Propositions A and A' of the Appendix, we see that (X_t) satisfies Assumption (A3), for $K_l = \frac{2\beta}{\sigma^2} - 1$.

4.2.3 Bilinear diffusion

We set $\mathcal{B}_0(x) = 1 + \frac{1}{x}$ and $\mathcal{B}_\infty(x) = 1 + x$.

We suppose that:

$$dX_t = (\alpha X_t + \beta)dt + \sigma X_t dB_t, \quad X_0 = \eta, \tag{42}$$

with $\alpha < 0$, $\sigma, \beta > 0$ and η is a positive and independent of (B_t).

Using Propositions A and A" of the Appendix, we get that the diffusion (X_t) satisfies Assumptions (A1)–(A3) with $K_l = \infty$.

4.2.4 Elastic diffusion process

Again, here, $(l, r) = (0, \infty)$ and we set $\mathcal{B}_0(x) = 1 + \frac{1}{x}$ and $\mathcal{B}_\infty(x) = 1 + x$.

Let X be given by:

$$dX_t = (\alpha X_t + \beta)dt + \sigma X_t^\psi dB_t, \quad X_0 = \eta, \tag{43}$$

with $\alpha < 0$, $\sigma, \beta > 0$, $\psi \in (\frac{1}{2}, 1)$ and η is a positive random variable independent of (B_t).

Combining Propositions A and A''' of the Appendix, we see that (X_t) satisfies Assumptions (A1)–(A3) with $K_l = \infty$.

5 Appendix

The following proposition (also given in Kessler [6]) shows that Assumption (A3) holds for any general diffusion process on $\mathbb{R}$ under the usual standard assumptions on the drift and diffusion coefficient.

Proposition A. *Assume that X is the solution of* (5) *and that (A1) holds. Furthermore, a and b are supposed such that for all $x \in \mathbb{R}$, $|a(x)| + |b(x)| \leq c(1 + |x|)$. Then, for all integer $k \geq 1$, there exists a constant $c(k)$ depending only on c and k such that:*

$$E\left(\sup_{s\in[t,t+1]} |X_s|^k \mid \mathcal{G}_t\right) \leq c(k)(1 + |X_t|^k) \tag{44}$$

Proof. We can suppose (by the Hölder inequality) that $k \geq 2$. For $s \in [t, t+1]$ we write:

$$X_s = X_t + \int_t^s a(X_v)dB_v + \int_t^s b(X_v)dv \tag{45}$$

Using the Burkhölder inequality, we get ($c(k)$ may change from a line to another):

$$E\left(\sup_{u\in[t,s]} \left|X_u^k\right| \mid \mathcal{G}_t\right) \leq c(k)\left|X_t\right|^k + c(k)E\left(\left(\int_t^s a^2(X_v)dv\right)^{\frac{k}{2}} \mid \mathcal{G}_t\right)$$
$$+ c(k)E\left(\left(\int_t^s |b|(X_v)dv\right)^k \mid \mathcal{G}_t\right)$$

Using that $k \geq 2$:

$$E\left(\sup_{u\in[t,s]} \left|X_u^k\right| \mid \mathcal{G}_t\right) \leq c(k)\left|X_t\right|^k + c(k)E\left(\int_t^s |a|^k(X_v) + |b|^k(X_v)dv \mid \mathcal{G}_t\right)$$

Using the bound on $|a| + |b|$, and the Fubini Theorem:

$$E\left(\sup_{u\in[t,s]} \left|X_u^k\right| \mid \mathcal{G}_t\right) \leq c(k)\left|X_t\right|^k + c(k)\int_t^s E\left((1 + |X_v|^k) \mid \mathcal{G}_t\right)dv \tag{46}$$

Hence we have, if we set $\phi(s) = \sup_{u\in[t,s]} E\left(\left|X_u^k\right| \mid \mathcal{G}_t\right)$:

$$\phi(s) \leq c(k)(1 + |X_t|^k) + c(k)\int_t^s E\left(|X_v|^k) \mid \mathcal{G}_t\right)dv \leq c(k)(1 + |X_t|^k) + c(k)\int_t^s \phi(v)dv$$

By (A1), $\phi(s)$ is almost surely finite, and we may apply Gronwall's Lemma to obtain:

$$\phi(s) \leq c(1 + |X_t|^k)e^{c(k)(s-t)}.$$

Using that $s \in [t, t+1]$, we deduce:

$$\sup_{u\in[t,t+1]} E\left(\left|X_u^k\right| \mid \mathcal{G}_t\right) \leq c(k)(1 + |X_t|^k)$$

Reporting the last inequation in (46) gives (44). □

Proposition A'. *Assume that X is the solution of $dX_t = (\alpha X_t + \beta)dt + \sigma\sqrt{X_t}dB_t$, $X_0 = \eta > 0$ with $\alpha < 0$, $\beta > 0$, $\sigma > 0$ and $\frac{2\beta}{\sigma^2} > 1$. Let $k \in [0, \frac{2\beta}{\sigma^2} - 1)$, then $\exists c$ such that $\forall\ t \geq 0$:*

$$E\left(\sup_{s\in[t,t+1]}\left(\frac{1}{X_s}\right)^k \mid \mathcal{G}_t\right) \leq c\left(\frac{1}{X_t}\right)^k$$

Proof. Using the Markov property, we get for $x > 0$:

$$\mathbb{1}_{\{X_t=x\}}E\left(\sup_{s\in[t,t+1]}\left(\frac{1}{X_s}\right)^k \mid \mathcal{G}_t\right) = E\left(\sup_{s\in[0,1]}\left(\frac{1}{\tilde{X}_s}\right)^k\right),$$

where $\tilde{X}$ solves the same stochastic differential equation as X, with the initial condition $\tilde{X}_0 = x$.

Now, we use that the process $\tilde{X}$ can be represented as (see Leblanc [7]): $\tilde{X}_s = e^{\alpha s}R_{\tau(s)}$, where (R_s) is the square of a Bessel process of dimension δ, with $\delta = \frac{4\beta}{\sigma^2}$, starting from $\tilde{X}_0 = x$, and $\tau(s) = \frac{\sigma^2}{4}\frac{e^{|\alpha|s}-1}{|\alpha|}$ is a deterministic change of time.

We deduce that $\inf_{s\in[0,1]}\tilde{X}_s \geq e^{\alpha}\inf_{s\geq 0}R_s$, and hence:

$$E\left(\sup_{s\in[0,1]}\left(\frac{1}{\tilde{X}_s}\right)^k\right) \leq e^{|\alpha|}E\left(\frac{1}{\inf_{s\geq 0}R_s^k}\right)$$

But we know that, since $\delta > 2$, the law of $\inf_{s\geq 0}R_s$ is $xU^{\frac{2}{\delta-2}}$ where U is uniformly distributed on $[0,1]$ (See Revuz–Yor [9] p.430 Exercice 1.18). So,

$$E\left(\frac{1}{\inf_{s\geq 0}R_s^k}\right) = \frac{1}{x^k}E\left(U^{\frac{-2k}{\delta-2}}\right) = \frac{1}{x^k}\int_0^1 u^{\frac{-2k}{\delta-2}}du \leq c\frac{1}{x^k}$$

since the integral above is finite for $\frac{2k}{\delta-2} < 1$ i.e. for $k < \frac{\delta}{2} - 1 = \frac{2\beta}{\sigma^2} - 1$. □

Proposition A". *Assume that X solves the equation*

$$dX_t = (\alpha X_t + \beta)dt + \sigma X_t dB_t,\ X_0 > 0 \tag{47}$$

with $\alpha < 0$, β, $\sigma > 0$.

Then $\forall k \geq 0$, $\exists c$ such that for all t:

$$E\left(\sup_{s\in[t,t+1]}\frac{1}{X_s^k} \mid \mathcal{G}_t\right) \leq c(1 + \frac{1}{X_t^k})$$

Proof. Again by the Markov property of X, for $x > 0$,

$$\mathbb{1}_{\{X_t=x\}}E\left(\sup_{s\in[t,t+1]}\frac{1}{X_s^k} \mid \mathcal{G}_t\right) = E\left(\sup_{s\in[0,1]}\frac{1}{\tilde{X}_s^k}\right) \tag{48}$$

where $\tilde{X}$ is the solution of (47) starting with $\tilde{X}_0 = x$.

We set $Z_s = \frac{1}{\tilde{X}_s}$, then Z solves $dZ_s = -\sigma Z_s dB_s + \{(\sigma^2 - \alpha)Z_s - \beta Z_s^2\}ds$, with $Z_0 = \frac{1}{x}$. Now, we define Z' as the solution: $dZ'_s = -\sigma Z'_s dB_s + (\sigma^2 - \alpha)Z'_s ds$, with $Z'_0 = \frac{1}{x} = Z_0$.

Since $\beta < 0$, we can use a comparison theorem (see Revuz–Yor [9] p.375) to obtain:

$$\forall s, \quad Z_s \le Z'_s \quad a.s. \tag{49}$$

We apply Proposition A to Z':

$$E\left(\sup_{s\in[0,1]} {Z'_s}^k\right) \le c(1 + \left|{Z'_0}^k\right|) = c(1 + \frac{1}{x^k}) \tag{50}$$

Definition of Z, (48), (49) and (50) yield the result. □

Proposition A‴. *Assume X solves*

$$dX_t = (\alpha X_t + \beta)dt + \sigma X_t^{\psi} dB_t, \; X_0 > 0, \tag{51}$$

with $\alpha < 0$, $\sigma, \beta > 0$, $\psi \in (\frac{1}{2}, 1)$. Then, $\forall k \ge 0$, $\exists c$, $\forall t \ge 0$,

$$E\left(\sup_{s\in[t,t+1]} X_s^{-k} \mid \mathcal{G}_t\right) \le c(1 + X_t^{-k})$$

Proof. Set $k \ge 0$, by the Markov property of X, as in Proposition A, we can assume that $X_0 = x \in (0, \infty)$, and restrict ourself to show that there exists c (independent of x) such that,

$$E(\sup_{s\in[0,1]} X_s^{-k}) \le c(1 + x^{-k}).$$

The idea of the proof is to compare the process X with the power of a CIR process. Define Z as the solution of

$$dZ_t = (\alpha' Z_t + \beta')dt + \sigma' \sqrt{Z_t} dB_t, \quad Z_0 = x^{\frac{1}{p}} \tag{52}$$

with $p = (2(1-\psi))^{-1} \in (1, \infty)$ and where α', β' and σ' will be specified later.

Let us set $\tilde{X}_t = Z_t^p$. Hence, $\tilde{X}_0 = x$, and using Ito's formula and $(p - 1/2)p^{-1} = \psi$, we deduce

$$d\tilde{X}_t = \tilde{b}(\tilde{X}_t)dt + \sigma' p \tilde{X}_t^{\psi} dB_t \tag{53}$$

with $\tilde{b}(x) = \alpha' p x + (\beta' p + \frac{p(p-1)}{2}\sigma'^2)x^{1-\frac{1}{p}}$.

We set $\sigma' = \frac{\sigma}{p}$, then choose $\beta' > 0$ such that $2\beta'(\sigma')^{-2} - 1 > kp$. Finally, using $0 < 1 - p^{-1} < 1$ and $\beta > 0$, we choose $\alpha' < 0$ such that

$$\forall x > 0, \quad \tilde{b}(x) < \alpha x + \beta.$$

By (43) and (51) we deduce,

$$\forall t \ge 0, \quad \tilde{X}_t \le X_t$$

Hence, $E(\sup_{s\in[0,1]} X_s^{-k}) \le E(\sup_{s\in[0,1]} \tilde{X}_s^{-k}) = E(\sup_{s\in[0,1]} Z_s^{-kp})$.

But, we have seen in the proof of Proposition A', that for Z a CIR process solving (52) with $2\beta'(\sigma')^{-2} - 1 > kp$, we have

$$E(\sup_{s\in[0,1]} Z_s^{-kp}) \le c(1 + x^{-p})$$

The proposition is proved. □

Proposition B. *Let $f \in \mathcal{C}^1(l,r)$ satisfy:*

$$\exists c\ \forall x \in (l,r)\ \ \left|f'(x)\right| \leq c(\mathcal{B}_l^\gamma(x) + \mathcal{B}_r^\gamma(x))$$

then, for all integer $k \geq 1$, such that $k\gamma < K_l$:

$$E\left(\sup_{v\in[i\Delta_n,(i+1)\Delta_n]} |f(X_v) - f(X_{i\Delta_n})|^k \mid \mathcal{G}_i^n\right) \leq c\Delta_n^{\frac{k}{2}}(B_l^{k\gamma}(X_{i\Delta_n}) + B_r^{k(1+\gamma)}(X_{i\Delta_n})) \tag{54}$$

Proof. We start with $f(x) = x$. Let

$$\delta_{i,n} = \sup_{v\in[i\Delta_n,(i+1)\Delta_n]} |X_v - X_{i\Delta_n}|\,. \tag{55}$$

Using (45) and the Burkhölder inequality, we get:

$$E\left(\delta_{i,n} \mid \mathcal{G}_i^n\right) \leq cE\left(\left(\int_{i\Delta_n}^{(i+1)\Delta_n} a^2(X_v)dv\right)^{\frac{k}{2}} \mid \mathcal{G}_i^n\right) + cE\left(\left(\int_{i\Delta_n}^{(i+1)\Delta_n} |b|\,(X_v)dv\right)^{k} \mid \mathcal{G}_i^n\right)$$

Hence, $$E\left(\delta_{i,n}^k \mid \mathcal{G}_i^n\right) \leq c\Delta_n^{\frac{k}{2}}E\left(\sup_{s\in[i\Delta_n,(i+1)\Delta_n]} |a|^k(X_v) \mid \mathcal{G}_i^n\right) + c\Delta_n^k E\left(\sup_{s\in[i\Delta_n,(i+1)\Delta_n]} |b|^k(X_v) \mid \mathcal{G}_i^n\right)$$

Using assumption (A2), we get $|a|^k + |b|^k \leq c\mathcal{B}_r^k$, so assumption (A3) yields:

$$E\left(\delta_{i,n}^k \mid \mathcal{G}_i^n\right) \leq c\Delta_n^{\frac{k}{2}}\mathcal{B}_r^k(X_{i\Delta_n}) \tag{56}$$

Now for a general f, we set

$$\delta_{i,n}(f) = \sup_{v\in[i\Delta_n,(i+1)\Delta_n]} |f(X_v) - f(X_{i\Delta_n})| \tag{57}$$

and write, using the bound on f':

$$\delta_{i,n}(f) \leq c \sup_{v\in[i\Delta_n,(i+1)\Delta_n]} (\mathcal{B}_l^\gamma(X_v) + \mathcal{B}_r^\gamma(X_v)) \times \delta_{i,n}.$$

We choose p and q such that $\frac{1}{p} + \frac{1}{q} = 1$ and $pk\gamma < K_l$ and apply Hölder's inequality (see (55)–(57)):

$$E\left(\delta_{i,n}(f)^k \mid \mathcal{G}_i^n\right) \leq cE\left(\sup_{v\in[i\Delta_n,(i+1)\Delta_n]} (\mathcal{B}_l^{\gamma kp}(X_s) + \mathcal{B}_r^{\gamma kp}(X_s)) \mid \mathcal{G}_i^n\right)^{\frac{1}{p}} E\left(\delta_{i,n}^{qk} \mid \mathcal{G}_i^n\right)$$

Now, assumption (A3) and (56) yield:

$$E\left(\delta_{i,n}(f)^k \mid \mathcal{G}_i^n\right) \leq c(\mathcal{B}_l^{\gamma k}(X_{i\Delta_n}) + \mathcal{B}_r^{\gamma k}(X_{i\Delta_n}))(1 + \mathcal{B}_r^k(X_{i\Delta_n}))$$

So, we have the result. □

References

[1] O.E. Barndorff-Nielsen and N. Shephard. (1998). Aggregation and model construction for volatility models. Working paper series no.**10**, Center for Analytical Finance, University of Aarhus.

[2] S. Delattre and J. Jacod. (1997). A central limit theorem for normalized functions of the increments of a diffusion process, in the presence of round off errors. *Bernoulli*, **3**(1),1–28.

[3] V. Genon-Catalot, T. Jeantheau, and C. Laredo. (1998). Limit theorems for discretely observed stochastic volatility models. *Bernoulli*, **4**(3),283–303.

[4] A. Gloter. (1998). Parameter estimation for a discrete sampling of an integrated Ornstein–Uhlenbeck process. *Prepublication de l'université de Marne-la-Vallée*, **14/98**.

[5] J. Hull and A. White. (1987). The pricing of options on assets with stochastic volatilities. *J. Finance*, **42**,281–300.

[6] M. Kessler. (1997). Estimation of an ergodic diffusion from discrete observations. *Scandinavian Journal of Statistics*, **24**(2), 211–229.

[7] B. Leblanc. (1997). *Modélisation de la Volatilité d'un Actif Financier et Applications.* Thèse, Université Paris 7.

[8] M. Lefebvre. (1997). On the inverse of the first hitting time problem for bidimensional processes. *Journal of applied probability*, **34**(3),610–622.

[9] D. Revuz and M. Yor. (1994). *Continuous Martingales and Brownian Motion.* Springer–Verlag, Berlin Heidelberg, second edition.

Chapitre I.3. Estimation de paramètres pour une observation discrète de l'intégrale d'une diffusion

Parameter estimation for a discretely observed integrated diffusion process.

Abstract

We consider the estimation of unknown parameters in the drift and diffusion coefficients of a one–dimensional ergodic diffusion (X_t) when the observation is a discrete sampling of the integrated process $I_t = \int_0^t X_s ds$ at times $i\Delta, i = 1, \ldots, n$. Assuming that the sampling interval $\Delta = \Delta_n$ tends to 0 while the total length time interval $n\Delta_n$ tends to infinity, we first prove limit theorems for functionals associated with our observations. We apply these results to obtain a contrast function. The associated minimum contrast estimators are shown to be consistent and asymptotically Gaussian with different rates for drift and diffusion coefficient parameters.

The statistics of one-dimensional diffusion processes with ergodic properties and when the sample path is discretely observed has been the subject of many recent papers. More precisely, let (X_t) be given by the stochastic differential equation:

$$dX_t = b(X_t, \mu)dt + a(X_t, \sigma)dB_t, \quad X_0 = \eta \tag{1}$$

with B a standard Wiener process and η a random variable independent of B. Suppose that for some positive Δ, a sample $(X_{i\Delta}, i \leq n)$ is observed and that it is required to estimate (μ, σ). The exact likelihood of such an observation being generally untractable, other methods have been developped to obtain explicit estimators. Under the assumption of fixed sampling interval, different kinds of estimating functions have been studied (see e.g. Barndorff-Nielsen (1998), Bibby and Sørensen (1998), Kessler (1996), Kessler and Sørensen (1999), Sørensen (1998))

Another point of view which is also classical and complementary to the former one is to assume that the sampling interval $\Delta = \Delta_n$ tends to 0 as $n \to \infty$ and $n\Delta_n \to \infty$. In this framework, the likelihood of the Euler scheme of (1) is a contrast function and provides consistent and asymptotically Gaussian estimators. A noteworthy result is that drift and diffusion coefficient parameters are estimated with different rates, $(n\Delta_n)^{\frac{1}{2}}$ for the drift parameters and $n^{\frac{1}{2}}$ for diffusion coefficient parameters (see e.g. Dorogovtsev (1976), Florens-Zmirou (1989), Kessler (1997)).

In this paper, we consider a new type of observation. Our aim is to estimate the parameter (μ, σ) of (1) when we observe a discrete Δ–sampling of the integrated process

$$I_t = \int_0^t X_s ds.$$

Integrals of diffusion process have been recently considered in the field of finance in relation with stochastic volatility models (see e.g. Leblanc (1996), Barndorff–Nielsen and Sheppard (1998), Barndorff–Nielsen (1998), Genon–Catalot et *al* (1998)). Data may be obtained from option prices and their associated implied volatilities (see e.g. Pastorello et *al* (1994))

For fixed sampling interval Δ, the exact distribution of $(I_{i\Delta}, i \leq n)$ is difficult to compute except for few models. For example, the case of X a stationary Ornstein–Uhlenbeck process solving $dX_t = \mu X_t dt + \sigma dB_t$, is fully treated in Gloter (1998, a).

To deal with a general diffusion X, we shall assume that the sampling interval $\Delta = \Delta_n$ tends to 0. Now, let

$$\overline{X}_i^n = \Delta_n^{-1} \int_{i\Delta_n}^{(i+1)\Delta_n} X_s ds = \Delta_n^{-1}(I_{(i+1)\Delta_n} - I_{i\Delta_n}). \tag{2}$$

We shall base our estimation on the sample $(\overline{X}_i^n)_{i \leq n-1}$ which is in one–to–one correspondance with the observation $(I_{i\Delta_n})_{i \leq n}$. Our starting idea is that, for small Δ_n, the law of $(\overline{X}_i^n, i \leq n-1)$ may be close to the law of $(X_{i\Delta_n}, i \leq n-1)$, so that methods available for the discrete sampling $(X_{i\Delta_n})$ apply for $(\overline{X}_i^n)$. In fact such a substitution fails mainly because $(\overline{X}_i^n)$ is not Markovian.

As preliminary steps, in Gloter (1998, b) we have obtained asymptotic expansions of $\overline{X}_i^n$ as $\Delta_n \to 0$. The results of Gloter (1998, b) hold without ergodicity assumptions on model (1).

In this paper, we assume that the diffusion (X_t) has ergodic properties with invariant probability $d\nu_0(x)$. We first prove limit theorems concerning the variation and the quadratic variation of $(\overline{X}_i^n)$, which enlight the difference between $(\overline{X}_i^n)$ and the discrete sampling $(X_{i\Delta_n})$. These theorems enable us to construct a contrast by introducing the appropriate corrections on the Euler contrast of the diffusion.

The paper is organized as follows. Sections 1, 2, 3 are devoted to general limit theorems. In these sections, parameters are omitted and we set $a(x,\sigma) = a(x)$, $b(x,\mu) = b(x)$. Section 1 contains the assumptions and a recap of some expansions obtained in Gloter (1998, b). In Section 2 we introduce the following functionals of the observed process (where $\overline{X}_i = \overline{X}_i^n$ to simplify notations):

$$\overline{\nu}_n(f) = n^{-1}\sum_{i=0}^{n-1} f(\overline{X}_i) \tag{3}$$

$$\overline{\mathcal{I}}_n(f) = (n\Delta_n)^{-1}\sum_{i=0}^{n-2} f(\overline{X}_i)(\overline{X}_{i+1} - \overline{X}_i - \Delta_n b(\overline{X}_i)) \tag{4}$$

$$\overline{\mathcal{Q}}_n(f) = (n\Delta_n)^{-1}\sum_{i=0}^{n-2} f(\overline{X}_i)(\overline{X}_{i+1} - \overline{X}_i)^2 \tag{5}$$

Section 2 contains convergence in probability results and Section 3 some associated central limit theorems. The main result is that, under smoothness assumptions on f,g,

$$\left((n\Delta_n)^{\frac{1}{2}}\left(\overline{\mathcal{I}}_n(f) - \frac{1}{4}\overline{\mathcal{Q}}_n(f')\right), n^{\frac{1}{2}}\left(\frac{3}{2}\overline{\mathcal{Q}}_n(g) - \overline{\nu}_n(ga^2)\right)\right)$$

converges in distribution to a $\mathcal{N}(0, \nu_0(f^2a^2)) \otimes \mathcal{N}(0, \frac{9}{4}\nu_0(g^2a^4))$ (where ν_0 is the invariant probability of the diffusion and $\nu_0(f) = \int f(x)d\nu_0(x)$). This needs the additional (but classical) condition $n\Delta_n^2 \to 0$.

In Section 4, we give examples of diffusion models satisfying the set of assumptions introduced in Section 1.

Section 5 contains the statistical applications. We study estimators based on the minimization of

$$\mathcal{L}_n(\theta) = \sum_{i=0}^{n-1}\left\{\frac{3}{2\Delta_n}\left(\frac{\left(\overline{X}_{i+1} - \overline{X}_i - b(\overline{X}_i,\mu)\Delta_n\right)^2}{a^2(\overline{X}_i,\sigma)} + \frac{\Delta_n}{2}h(\overline{X}_i,\theta)\left(\overline{X}_{i+1} - \overline{X}_i\right)^2\right) + \log a^2(\overline{X}_i,\sigma)\right\} \tag{6}$$

with $h(x,\theta) = \frac{\partial}{\partial x}\frac{b(x,\mu)}{a^2(x,\sigma)}$. The associated minimum contrast estimator $\hat{\theta}_n = (\hat{\mu}_n, \hat{\sigma}_n)$ is shown to be consistent and $((n\Delta_n)^{\frac{1}{2}}(\hat{\mu}_n - \mu_0), n^{\frac{1}{2}}(\hat{\sigma}_n - \sigma_0))$ asymptotically normal ($\theta_0 = (\mu_0, \sigma_0)$ denotes the true value of the parameter). We compare the asymptotic variances with those obtained for estimators based on a discrete sampling of the diffusion itself (by Kessler (1997)). The only difference is a slight increase in the asymptotic variance of the estimator of σ_0. In Section 6, examples of parametric models are fully treated. Some technical results are given in the Appendix.

1 Assumptions and preliminary results

1.1 Model and Assumptions

Let (X_t) be defined as the solution on a probability space $(\Omega, \mathcal{F}, P)$ of the stochastic differential equation:

$$dX_t = a(X_t)dB_t + b(X_t)dt, \quad X_0 = \eta \tag{7}$$

where $(B_t)_{t\geq 0}$ is a standard one–dimensional Brownian motion, η is a random variable independent of $(B_t)_{t\geq 0}$. We make now some classical assumptions on functions b and a ensuring that the solution of (7) is a positive recurrent diffusion on an interval (l, r) $(-\infty \leq l < r \leq \infty)$.

To keep general notations, we introduce two positive measurable functions $\mathcal{B}_l$ and $\mathcal{B}_r$ defined on (l, r) satisfying the following property: $\forall\ \alpha,\ \beta,\ \alpha',\ \beta',\ p \geq 0,\ \exists c,\ \forall x \in (l, r)$,

$$(\mathcal{B}_l^{\alpha}(x) + \mathcal{B}_r^{\beta}(x)) \times (\mathcal{B}_l^{\alpha'}(x) + \mathcal{B}_r^{\beta'}(x)) \leq c(\mathcal{B}_l^{\alpha+\alpha'}(x) + \mathcal{B}_r^{\beta+\beta'}(x)) \tag{8}$$

$$(\mathcal{B}_l^{\alpha}(x) + \mathcal{B}_r^{\beta}(x))^p \leq c(\mathcal{B}_l^{p\alpha}(x) + \mathcal{B}_r^{p\beta}(x)) \tag{9}$$

These functions are used below to bound the growth of other functions near the boundaries l, r of the state space. For example, if $l = 0,\ r = \infty$ we may take $\mathcal{B}_l(x) = 1 + \frac{1}{x},\ \mathcal{B}_r(x) = 1 + x$; if $l = -\infty,\ r = \infty$ we may take $\mathcal{B}_l(x) = \mathcal{B}_r(x) = 1 + |x|$.

(A0) Equation (7) admits a unique strong solution such that $P(X_t \in (l, r),\ \forall t \geq 0) = 1$.

(A1) Function a and b are real valued, $\mathcal{C}^2$ on (l, r) and

$$\begin{aligned}
&\exists c,\ \alpha_1,\ \alpha_2,\ \beta_1,\ \beta_2 \geq 0,\ \forall x \in (l, r),\\
&a(x) > 0, \quad |a(x)| + |b(x)| \leq c(1 + B_r(x)),\\
&|a'(x)| \leq c(B_l^{\alpha_1}(x) + B_r^{\alpha_1}(x)), \quad |a''(x)| \leq c(B_l^{\alpha_2}(x) + B_r^{\alpha_2}(x)),\\
&|b'(x)| \leq c(B_l^{\beta_1}(x) + B_r^{\beta_1}(x)), \quad |b''(x)| \leq c(B_l^{\beta_2}(x) + B_r^{\beta_2}(x)).
\end{aligned}$$

Let $\mathcal{G}_t = \sigma(B_s, s \leq t; \eta)$.

(A2) There exists a positive constant K_l such that
$\forall\ k \in [0, K_l),\ \exists c,\ \forall t > 0,\ E\left(\sup_{s\in[t,t+1]} \mathcal{B}_l^k(X_s) \mid \mathcal{G}_t\right) \leq c\mathcal{B}_l^k(X_t)$
$\forall\ k \in [0, \infty),\ \exists c,\ \forall t > 0,\ E\left(\sup_{s\in[t,t+1]} \mathcal{B}_r^k(X_s) \mid \mathcal{G}_t\right) \leq c\mathcal{B}_r^k(X_t)$

For $x_0 \in (l, r)$, let $s(x) = \exp(-2\int_{x_0}^{x} \frac{b(u)}{a^2(u)}du)$ denote the scale density and $m(x) = \frac{1}{a^2(x)s(x)}$ the speed density.

(A3) $\int_l s(x)dx = \int^r s(x)dx = \infty, \int_l^r m(x)dx = M < \infty$.

Let

$$\nu_0(dx) = \frac{1}{M} m(x) \mathbb{1}_{\{x\in(l,r)\}} dx.$$

(A4) $\exists M_l,\ M_r > 0\ \ \nu_0(\mathcal{B}_l^{M_l}) < \infty,\ \nu_0(\mathcal{B}_r^{M_r}) < \infty.$

(A5) $\sup_{t\geq 0} E(\mathcal{B}_l^{M_l}(X_t)) < \infty, \quad \sup_{t\geq 0} E(\mathcal{B}_r^{M_r}(X_t)) < \infty$.

(A1) and (A3) imply (A0) but some results hold without (A3) under (A0)-(A2). Under (A1) and (A3), ν_0 is the unique invariant probability of model (7) and X satisfies the classical ergodic theorem

$$\forall f \in \mathbf{L}^1(d\nu_0), \frac{1}{T}\int_0^T f(X_s)ds \xrightarrow[a.s.]{T\to\infty} \nu_0(f)$$

Assumptions (A4) means that some powers of $\mathcal{B}_l$ and $\mathcal{B}_r$ are in $\mathbf{L}^1(d\nu_0)$ and hence all functions on (l,r) bounded by these powers of $\mathcal{B}_l$ and $\mathcal{B}_r$ will also be in $\mathbf{L}^1(d\nu_0)$. Assumption (A5) follows immediately from (A4) if the initial distribution is ν_0 (the process X being strictly stationary). In Section 4 we prove that (A5) follows again from (A4) when the initial condition is deterministic.

Assumption (A2) was already introduced in Kessler (1997) and in Gloter (1998, b), as a useful tool to control the behaviour of X near the endpoints l, r. In Gloter (1998, b), it is shown that (A2) holds for general diffusion processes on $\mathbb{R}$, and for some classical diffusions on $(0,\infty)$. The reason why (A2) is not symetric appears in the examples of Section 4.

1.2 Expansions for the observed process

Now, let Δ_n be a sequence of positive numbers with $\Delta_n \to 0$, as $n \to \infty$ and assume that $\Delta_n \leq 1$, for all n. We set

$$\mathcal{G}_i^n = \mathcal{G}_{i\Delta_n}.$$

We now recall the main properties of $\overline{X}_i^n = \overline{X}_i$ (see (2)) proved in Gloter (1998, b). In the following statements the constants c appearing never depend on i and n.

Proposition 1.1. *Assume (A0)-(A2) and let $f \in \mathcal{C}^1(l,r)$ satisfy:*

$$\exists \gamma \geq 0,\ \exists c > 0,\ \forall x \in (l,r)\ \ |f'(x)| \leq c(\mathcal{B}_l^\gamma(x) + \mathcal{B}_r^\gamma(x))$$

1) *For all integer $k \geq 1$, such that $k\gamma < K_l$ (with K_l given in (A2)), there exists $c > 0$ such that for all $i, n \geq 0$*

$$E\left(\sup_{v\in[i\Delta_n,(i+1)\Delta_n]} |f(X_v) - f(X_{i\Delta_n})|^k \mid \mathcal{G}_i^n\right) \leq c\Delta_n^{\frac{k}{2}}(\mathcal{B}_l^{k\gamma}(X_{i\Delta_n}) + \mathcal{B}_r^{k(1+\gamma)}(X_{i\Delta_n}))$$

2) *For all $k \geq 1$, there exists $c > 0$ such that for all $i, n \geq 0$,*

$$E\left(\left|\overline{X}_i - X_{i\Delta_n}\right|^k \mid \mathcal{G}_i^n\right) \leq c\Delta_n^{\frac{k}{2}}(1 + \mathcal{B}_r^k(X_{i\Delta_n}))$$
$$E\left(\left|\overline{X}_{i+1} - \overline{X}_i\right|^k \mid \mathcal{G}_i^n\right) \leq c\Delta_n^{\frac{k}{2}}(1 + \mathcal{B}_r^k(X_{i\Delta_n})).$$

Let us introduce

$$\xi_{i,n} = \Delta_n^{-\frac{3}{2}} \int_{i\Delta_n}^{(i+1)\Delta_n} (s - i\Delta_n) dB_s \quad \text{for } i,\ n \geq 0 \tag{10}$$

$$\xi'_{i+1,n} = \Delta_n^{-\frac{3}{2}} \int_{(i+1)\Delta_n}^{(i+2)\Delta_n} (i\Delta_n + 2\Delta_n - s) dB_s \quad \text{for } i \geq -1,\ n \geq 0 \tag{11}$$

$$U_{i,n} = \xi_{i,n} + \xi'_{i+1,n} \tag{12}$$

For all $n \geq 0$, $(\xi_{i,n})_{i\geq 0}$, $(\xi'_{i+1,n})_{i\geq 0}$ and $(U_{i,n})_{i\geq 0}$ are Gaussian processes; $\xi_{i,n}$ is $\mathcal{G}_{i+1}^n$ measurable and independent of $\mathcal{G}_i^n$; $\xi'_{i+1,n}$ is $\mathcal{G}_{i+2}^n$ measurable and independent of $\mathcal{G}_{i+1}^n$.

We easily compute the following expectations: $E\left(\xi_{i,n} \mid \mathcal{G}_i^n\right) = E\left(\xi'_{i+1,n} \mid \mathcal{G}_{i+1}^n\right) = 0$, $E\left(\xi_{i,n}^2 \mid \mathcal{G}_i^n\right) = E\left(\xi_{i+1,n}'^2 \mid \mathcal{G}_{i+1}^n\right) = \frac{1}{3}$, $E\left(\xi_{i,n}\xi'_{i,n} \mid \mathcal{G}_i^n\right) = \frac{1}{6}$.

We deduce that for $i, j \geq 0$, $var(U_{i,n}) = 2/3$, $cov(U_{i,n}, U_{i+1,n}) = 1/6$, $cov(U_{i,n}, U_{i+j,n}) = 0$ for $j \geq 2$.

Hence $(U_{i,n})_{i\geq 0}$ has the covariance structure of a MA(1) process.

The following results hold.

Theorem 1.2. *Assume (A0)-(A2).*

1) We have

$$\overline{X}_i - X_{i\Delta_n} = \Delta_n^{\frac{1}{2}} a(X_{i\Delta_n}) \xi'_{i,n} + e_{i,n},$$

with, if $2\alpha_1 < K_l$, $\forall i, n \geq 0$

$$|E\left(e_{i,n} \mid \mathcal{G}_i^n\right)| \leq \Delta_n c(1 + \mathcal{B}_r(X_{i\Delta_n})) \tag{13}$$

$$E\left(e_{i,n}^2 \mid \mathcal{G}_i^n\right) \leq \Delta_n^2 c(\mathcal{B}_l^{2\alpha_1}(X_{i\Delta_n}) + \mathcal{B}_r^{2(1+\alpha_1)}(X_{i\Delta_n})) \tag{14}$$

2) We have

$$\overline{X}_{i+1} - \overline{X}_i - b(\overline{X}_i)\Delta_n = \Delta_n^{\frac{1}{2}} a(X_{i\Delta_n}) U_{i,n} + \varepsilon_{i,n}$$

where $\varepsilon_{i,n}$ *is* $\mathcal{G}_{i+2}^n$ *measurable, and if* $\beta_1 \vee 2\beta_2 \vee 4\alpha_1 < K_l$, $\forall i, n \geq 0$

$$|E\left(\varepsilon_{i,n} \mid \mathcal{G}_i^n\right)| \leq \Delta_n^2 c(\mathcal{B}_l^{(2\alpha_1+\beta_2)\vee\beta_1}(X_{i\Delta_n}) + \mathcal{B}_r^{(2+2\alpha_2+\beta_2)\vee(1+\beta_1)}(X_{i\Delta_n})) \tag{15}$$

if $(4\alpha_1) \vee 2\alpha_2 < K_l$, $\forall i, n \geq 0$

$$E\left(\varepsilon_{i,n}^2 \mid \mathcal{G}_i^n\right) \leq \Delta_n^2 c(\mathcal{B}_l^{2\alpha_1\vee\alpha_2}(X_{i\Delta_n}) + \mathcal{B}_r^{3+2\alpha_1+\alpha_2}(X_{i\Delta_n})) \tag{16}$$

$$E\left(\varepsilon_{i,n}^4 \mid \mathcal{G}_i^n\right) \leq \Delta_n^4 c(\mathcal{B}_l^{4\alpha_1\vee 2\alpha_2}(X_{i\Delta_n}) + \mathcal{B}_r^{6+4\alpha_1+2\alpha_2}(X_{i\Delta_n})) \tag{17}$$

Furthermore, if $4\alpha_1 \vee \alpha_2 \vee 4\beta_1 < K_l$, $\forall i, n \geq 0$

$$|E\left(\varepsilon_{i,n}\xi_{i,n} \mid \mathcal{G}_i^n\right)| \leq \Delta_n^{\frac{3}{2}} c(\mathcal{B}_l^{(\alpha_1+\beta_1)\vee\alpha_2}(X_{i\Delta_n}) + \mathcal{B}_r^{(1+\alpha_1+\beta_1)\vee(2+\alpha_2)}(X_{i\Delta_n})) \tag{18}$$

$$\left|E\left(\varepsilon_{i,n}\xi'_{i+1,n} \mid \mathcal{G}_i^n\right)\right| \leq \Delta_n^{\frac{3}{2}} c(\mathcal{B}_l^{(\alpha_1+\beta_1)\vee\alpha_2)}(X_{i\Delta_n}) + \mathcal{B}_r^{(1+\alpha_1+\beta_1)\vee(2+\alpha_2)}(X_{i\Delta_n})) \tag{19}$$

2 Limit theorems for functionals of the observed process

In this section and the following one, we study the behaviour of functionals (3)–(5) for $f : (l,r) \mapsto \mathbb{R}$ satisfying some regularity assumptions. The conditions on f are expressed by the following.

- Condition $\mathbf{C}_\gamma$: $f \in \mathcal{C}^2(l,r)$ and for $\gamma \geq 0$,

$$\exists c,\ \forall x \in (l,r), \quad |f(x)| + |f'(x)| + |f''(x)| \leq c(\mathcal{B}_l^\gamma(x) + \mathcal{B}_r^\gamma(x)).$$

For statistical purposes, we also need to consider the functionals (3)–(5) for $f(x,\theta) : (l,r) \times \Theta \mapsto \mathbb{R}$ where Θ is a product of two compact intervals of $\mathbb{R}$. To obtain uniform convergences with respect to θ, the conditions on $f = f(x,\theta)$ are the following.

- Condition $\mathbf{CU}_\gamma$: $f : (l,r) \times \Theta \to \mathbb{R}$ satisfies $f(.,.) \in \mathcal{C}^2[(l,r) \times O]$ for some open set $O \supset \Theta$ and with $\gamma \geq 0$,

$$\exists c > 0,\ \forall x \in (l,r) \quad \sup_{\theta\in\Theta} |g(x,\theta)| \leq c(\mathcal{B}_l^\gamma(x) + \mathcal{B}_r^\gamma(x)), \text{ for } g = f,\ f'_x,\ f''_{x^2},\ \nabla_\theta f,\ \nabla_\theta f'_x.$$

2.1 Empirical mean

As a first application of the expansions recalled in Section 1, we give a mean theorem for the process $(\overline{X}_i)_{i\in\mathbb{N}}$.

Proposition 2.1. *Assume (A1)–(A5), let f satisfy $\mathbf{CU}_\gamma$ with $\gamma < M_l$, $1+\gamma < M_r$ and $2\gamma < K_l$, then*

$$\overline{\nu}_n(f(.,\theta)) \xrightarrow{n\to\infty} \nu_0(f(.,\theta)) \quad \textit{uniformly in } \theta, \textit{ in probability.}$$

Proof. By Lemma 7.1 of the Appendix, we only have to prove the $\mathbf{L}^1$ convergence to zero of

$$\sup_{\theta\in\Theta} n^{-1} \sum_{i=0}^{n-1} \left| f(\overline{X}_i,\theta) - f(X_{i\Delta_n},\theta) \right|.$$

By Taylor's expansion and condition $\mathbf{CU}_\gamma$ on f, we get the bound

$$\sup_{\theta\in\Theta} \left| f(\overline{X}_i,\theta) - f(X_{i\Delta_n},\theta) \right| \leq c \sup_{s\in[i\Delta_n,(i+1)\Delta_n]} (\mathcal{B}_l^\gamma(X_s) + \mathcal{B}_r^\gamma(X_s)) \left| \overline{X}_i - X_{i\Delta_n} \right|.$$

Now, the Cauchy–Schwarz inequality, (A2), Proposition 1.1 *2)* and (8) yield

$$E\left(\sup_{\theta\in\Theta} \left| f(\overline{X}_i,\theta) - f(X_{i\Delta_n},\theta) \right| \mid \mathcal{G}_i^n \right) \leq c\Delta_n^{\frac{1}{2}} (\mathcal{B}_l^\gamma(X_{i\Delta_n}) + \mathcal{B}_r^{1+\gamma}(X_{i\Delta_n})).$$

By Assumption (A5), we deduce $E(\sup_{\theta\in\Theta} \left| f(\overline{X}_i,\theta) - f(X_{i\Delta_n},\theta) \right|) \leq c\Delta_n^{\frac{1}{2}}$. So, the proposition is proved. □

2.2 Variation of the process

Our next result concerns the fonctional $\overline{\mathcal{I}}_n$ which involves the increments of the process $(\overline{X}_i)$(see (4)).

Theorem 2.2. *Assume (A1)-(A5), and let f satisfy* $\mathbf{CU}_\gamma$, *with* $(\gamma+2\alpha_1+\beta_1+\beta_2)\vee(2\gamma+\alpha_2) < M_l$, $(4+2\gamma+2\alpha_1+\alpha_2)\vee(2+\gamma+2\alpha_2+\beta_1+\beta_2) < M_r$ *and* $2\gamma\vee 4\alpha_1\vee 2\alpha_2\vee\beta_1\vee 2\beta_2 < K_l$,

$$\overline{\mathcal{I}}_n(f(.,\theta)) \xrightarrow{n\to\infty} \frac{1}{6}\nu_0\left(f'_x(.,\theta)a^2(.)\right) \text{ uniformly in } \theta\text{, in probability.} \tag{20}$$

Proof. The proof relies on the expansions of Theorem 1.2. Set for the proof

$$V_{i,n}(\theta) = f(\overline{X}_i,\theta)(\overline{X}_{i+1} - \overline{X}_i - \Delta_n b(\overline{X}_i)). \tag{21}$$

Since $V_{i,n}(\theta)$ is $\mathcal{G}^n_{i+2}$ measurable, to deal with a triangular array of martingale increments, we split $\overline{\mathcal{I}}_n(f(.,\theta))$ into the sum of terms with even index i and the sum of terms with odd index i. Now, it is enough to show that

$$(n\Delta_{2n})^{-1}\sum_{i=0}^{n-1} V_{2i,2n}(\theta) \xrightarrow{n\to\infty} \frac{1}{6}\nu_0\left(f'_x(.,\theta)a^2(.)\right) \quad \text{uniformly in probability} \tag{22}$$

(and that $(n\Delta_{2n})^{-1}\sum_{i=0}^{n-2} V_{2i+1,2n}(\theta) \xrightarrow{n\to\infty} \frac{1}{6}\nu_0\left(f'_x(.,\theta)a^2(.)\right)$, but the proof in analogous.)

By the Taylor formula, and Theorem 1.2 *2*) (recall (10)–(12)), $V_{i,n}(\theta) = \sum_{j=0}^{5} v^{(j)}_{i,n}(\theta)$, with,

$$v^{(1)}_{i,n}(\theta) = \Delta_n^{\frac{1}{2}} U_{i,n} a(X_{i\Delta_n}) f(X_{i\Delta_n},\theta) \tag{23}$$

$$v^{(2)}_{i,n}(\theta) = \Delta_n^{\frac{1}{2}} U_{i,n}(\overline{X}_i - X_{i\Delta_n}) a(X_{i\Delta_n}) f'_x(X_{i\Delta_n},\theta) \tag{24}$$

$$v^{(3)}_{i,n}(\theta) = \Delta_n^{\frac{1}{2}} U_{i,n}\frac{1}{2}(\overline{X}_i - X_{i\Delta_n})^2 a(X_{i\Delta_n}) f''_{x^2}(\hat{X}_i,\theta) \tag{25}$$

$$v^{(4)}_{i,n} = \varepsilon_{i,n} f(X_{i\Delta_n},\theta) \tag{26}$$

$$v^{(5)}_{i,n} = \varepsilon_{i,n}(X_{2i\Delta_{2n}} - \overline{X}_{2i}) f'_x(\tilde{X}_i,\theta) \tag{27}$$

where $\hat{X}_i, \tilde{X}_i \in [\overline{X}_i, X_{i\Delta_n}]$.

Now, define $\overline{\mathcal{I}}^{(j)}_n(\theta) = (n\Delta_{2n})^{-1}\sum_{i=0}^{n-1} v^{(j)}_{2i,2n}(\theta)$ for $j = 1,\dots,5$.

- Let us first study $\overline{\mathcal{I}}^{(2)}_n(\theta)$ and prove that, for each θ,

$$\overline{\mathcal{I}}^{(2)}_n(\theta) \xrightarrow[\mathbf{P}]{n\to\infty} \frac{1}{6}\nu_0\left(f'_x(.,\theta)a^2(.)\right). \tag{28}$$

By Lemma 7.3 of the Appendix, it is enough to show that:

$$(n\Delta_{2n})^{-1}\sum_{i=0}^{n-1} E\left(v^{(2)}_{2i,2n}(\theta) \mid \mathcal{G}^{2n}_{2i}\right) \xrightarrow[\mathbf{P}]{n\to\infty} \frac{1}{6}\nu_0\left(f'_x(.,\theta)a^2(.)\right), \tag{29}$$

$$(n\Delta_{2n})^{-2}\sum_{i=0}^{n-1} E\left((v^{(2)}_{2i,2n}(\theta))^2 \mid \mathcal{G}^{2n}_{2i}\right) \xrightarrow[\mathbf{L}^1]{n\to\infty} 0. \tag{30}$$

Using Theorem 1.2 *1*) and $E(U_{2i,2n}\xi'_{2i,2n} \mid \mathcal{G}_{2i}^{2n}) = \frac{1}{6}$, we get

$$E\left(v_{2i,2n}^{(2)}(\theta) \mid \mathcal{G}_{2i}^{2n}\right) = \frac{\Delta_{2n}}{6} a^2(X_{2i\Delta_{2n}}) f'_x(X_{2i\Delta_{2n}}, \theta) + r_{i,n} \tag{31}$$

with $r_{i,n} = E\left(\Delta_{2n}^{\frac{1}{2}} U_{2i,2n} f'_x(X_{2i\Delta_{2n}}, \theta) a(X_{2i\Delta_{2n}}) e_{i,n} \mid \mathcal{G}_{2i}^{2n}\right)$.

Now, (14), Assumption (A1) and condition $\mathbf{CU}_\gamma$ yield

$$|r_{i,n}| \leq \Delta_{2n}^{\frac{3}{2}} c(\mathcal{B}_l^{\gamma+\alpha_1}(X_{2i\Delta_{2n}}) + \mathcal{B}_l^{2+\gamma+\alpha_1}(X_{2i\Delta_{2n}})) \tag{32}$$

Therefore, by (A5), $(n\Delta_{2n})^{-1} \sum_{i=0}^{n-1} r_{i,n} \xrightarrow[\mathbf{L}^1]{n\to\infty} 0$.

Hence, an application of Lemma 7.1 in the Appendix to the first term of (31) yields (29).

Now, by Proposition 1.1 *2*), $E\left((v_{2i,2n}^{(2)}(\theta))^2 \mid \mathcal{G}_{2i}^{2n}\right) \leq \Delta_{2n}^2 c(\mathcal{B}_l^{2\gamma}(X_{2i\Delta_{2n}}) + \mathcal{B}_r^{4+2\gamma}(X_{2i\Delta_{2n}}))$. This gives (30), and (28) follows.

To obtain uniformity with respect to θ we shall use Proposition 7.2 of the Appendix. Let us compute

$$\nabla_\theta v_{2i,2n}^{(2)}(\theta) = \Delta_{2n}^{\frac{1}{2}} U_{2i,2n}(\overline{X}_{2i} - X_{2i\Delta_{2n}}) a(X_{2i\Delta_n}) \nabla_\theta f'_x(X_{2i\Delta_{2n}}, \theta)$$

By condition $\mathbf{CU}_\gamma$, the following bound holds

$$E(\sup_{\theta\in\Theta} \left|\nabla_\theta v_{2i,2n}^{(2)}(\theta)\right| \mid \mathcal{G}_{2i}^{2n}) \leq c\Delta_n(\mathcal{B}_l^{\gamma}(X_{2i\Delta_{2n}}) + \mathcal{B}_r^{2+\gamma}(X_{2i\Delta_{2n}})).$$

With Assumption (A5), we deduce $\sup_{n\in\mathbb{N}} E(\sup_{\theta\in\Theta} \left|\nabla_\theta v_{2i,2n}^{(2)}(\theta)\right|) \leq c\Delta_n$. Hence,

$$\sup_{n\in\mathbb{N}} E(\sup_{\theta\in\Theta} \left|\nabla_\theta \overline{\mathcal{I}}_n^{(2)}(\theta)\right|) < \infty$$

and uniformity in (28) follows.

To end the proof of the theorem, it remains to show the uniform convergence to 0 for $\overline{\mathcal{I}}_n^{(1)}(\theta)$, $\overline{\mathcal{I}}_n^{(3)}(\theta)$, $\overline{\mathcal{I}}_n^{(4)}(\theta)$ and $\overline{\mathcal{I}}_n^{(5)}(\theta)$.

• Let us study $\overline{\mathcal{I}}_n^{(1)}(\theta)$.

By (10)- (11), we have

$$E\left(v_{2i,2n}^{(1)}(\theta) \mid \mathcal{G}_{2i}^{2n}\right) = 0 \quad \text{and} \quad E\left((v_{2i,2n}^{(1)}(\theta))^2 \mid \mathcal{G}_{2i}^{2n}\right) = \frac{2\Delta_{2n}}{3} a^2(X_{2i\Delta_{2n}}) f^2(X_{2i\Delta_{2n}}, \theta).$$

We deduce, using $n\Delta_n \to \infty$, $2\gamma \leq M_l$, $2+\gamma \leq M_r$ and (A5) that

$$(n\Delta_{2n})^{-2} \sum_{i=0}^{n-1} E\left((v_{2i,2n}^{(1)}(\theta))^2 \mid \mathcal{G}_{2i}^{2n}\right) = (n\Delta_{2n})^{-1}\left(\frac{2}{3} n^{-1} \sum_{i=0}^{n-1} a^2(X_{2i\Delta_{2n}}) f^2(X_{2i\Delta_{2n}}, \theta)\right) \xrightarrow{\mathbf{L}^1} 0.$$

Now Lemma 7.3, again, yields the convergence $\overline{\mathcal{I}}_n^{(1)}(\theta) \xrightarrow{\mathbf{P}} 0$ for each θ in Θ.

To prove that this convergence is uniform we cannot use Proposition 7.2 because $E\left(\left|v_{2i,2n}^{(1)}(\theta)\right| \mid \mathcal{G}_{2i}^{2n}\right)$ is of order $\sqrt{\Delta_{2n}}$. We will instead use Theorem 20 in Ibragimov and Khas'minskii (1981) (Appendix 1). It is enough to show that there exists two constants $M \geq 0$, and $\epsilon > 0$ such that:

$$\forall\theta,\ n,\quad E\left(\left|\overline{\mathcal{I}}_n^{(1)}(\theta)\right|^{2+\epsilon}\right) \leq M \quad \text{and} \quad \forall\theta,\ \theta',\ n,\quad D_n(\theta,\theta') \leq M\left|\theta-\theta'\right|^{2+\epsilon} \tag{33}$$

with $D_n(\theta,\theta') = E\left(\left|\overline{\mathcal{I}}_n^{(1)}(\theta) - \overline{\mathcal{I}}_n^{(1)}(\theta')\right|^{2+\epsilon}\right)$.

We only prove the first inequality (the second one is similar). Using Rosenthal's inequality for martingales (see Hall and Heyde (1980) p.23), we get for any $\epsilon > 0$,

$$E\left(\left|\overline{\mathcal{I}}_n^{(1)}\right|^{2+\epsilon}\right) \leq (n\Delta_{2n})^{-2-\epsilon} E\left(\left|\sum_{i=0}^{n-1} E\left(v_{2i,2n}^2 \mid \mathcal{G}_{2i}^{2n}\right)\right|^{1+\frac{\epsilon}{2}}\right) + (n\Delta_{2n})^{-2-\epsilon}\sum_{i=0}^{n-1} E(|v_{2i,2n}|^{2+\epsilon}).$$

By the classical inequality, for $p = 1+\frac{\epsilon}{2}$, $(\sum_{i=0}^{n-1}|a_i|)^p \leq n^{p-1}\sum_{i=0}^{n-1}|a_i|^p$, we have

$$E\left(\left|\sum_{i=0}^{n-1} E\left(v_{2i,2n}^2 \mid \mathcal{G}_{2i}^{2n}\right)\right|^{1+\frac{\epsilon}{2}}\right) \leq n^{\frac{\epsilon}{2}}\sum_{i=0}^{n-1} E\left(\left|E\left(v_{2i,2n}^2 \mid \mathcal{G}_{2i}^{2n}\right)\right|^{1+\frac{\epsilon}{2}}\right)$$

But, if ϵ is small enough, $2\gamma(1+\epsilon) \leq M_l$ and $(2\gamma+1)(1+\epsilon) \leq M_r$ and by (A5) we deduce

$$\sup_{i,n} E\left(\left|E\left(v_{2i,2n}^2 \mid \mathcal{G}_{2i}^{2n}\right)\right|^{1+\frac{\epsilon}{2}}\right) \leq c\Delta_{2n}^{1+\frac{\epsilon}{2}}, \quad \sup_{i,n} E(|v_{2i,2n}|^{2+\epsilon}) \leq c\Delta_{2n}^{1+\frac{\epsilon}{2}}.$$

Hence,

$$E\left(\left|\overline{\mathcal{I}}_n^{(1)}\right|^{2+\epsilon}\right) \leq c\{(n\Delta_{2n})^{-1-\frac{\epsilon}{2}} + (n\Delta_{2n})^{-1-\frac{\epsilon}{2}} n^{-\frac{\epsilon}{2}}\}.$$

Since $(n\Delta_{2n})^{-1}$ is bounded, we obtain (33). Hence, $\overline{\mathcal{I}}_n^{(1)}(\theta) \xrightarrow[\mathbf{P}]{n\to\infty} 0$ uniformly in θ.

- Let us study $\overline{\mathcal{I}}_n^{(3)}(\theta)$. For this term we have, by (A1) and $\mathbf{CU}_\gamma$,

$$\sup_{\theta\in\Theta}\left|v_{2i,2n}^{(3)}(\theta)\right| \leq c\Delta_{2n}^{\frac{1}{2}}|U_{2i,2n}|\left(\overline{X}_{2i} - X_{2i\Delta_{2n}}\right)^2 \times \sup_{s\in[i\Delta_n,(i+1)\Delta_n]} c(\mathcal{B}_l^\gamma(X_s) + \mathcal{B}_r^{1+\gamma}(X_s)).$$

Using Proposition 1.1 *2)* (with k=8) and Assumption (A2), we get:

$$E\left(\sup_{\theta\in\Theta}\left|v_{2i,2n}^{(3)}(\theta)\right| \mid \mathcal{G}_i^n\right) \leq c\Delta_n^{\frac{3}{2}}(\mathcal{B}_l^\gamma(X_{2i\Delta_{2n}}) + \mathcal{B}_r^{3+\gamma}(X_{2i\Delta_{2n}}).$$

Hence, $\overline{\mathcal{I}}_n^{(3)}(\theta) \xrightarrow[\mathbf{L}^1]{n\to\infty} 0$ uniformly in θ.

- Now, we treat $\overline{\mathcal{I}}_n^{(4)}(\theta)$. Using (15) and (16) we show that, if $(2\gamma+2\alpha_1)\vee(2\gamma+\alpha_2)\vee(\gamma+2\alpha_1+\beta_1+\beta_2) \leq M_l$ and $(2+\gamma+2\alpha_2+\beta_1+\beta_2)\vee(3+2\gamma+2\alpha_1+\alpha_2) \leq M_r$, then

$$(n\Delta_n)^{-1}\sum_{i=0}^{n-1} E\left(\varepsilon_{2i,2n} f(X_{2i\Delta_{2n}},\theta) \mid \mathcal{G}_{2i}^{2n}\right) \xrightarrow[\mathbf{L}^1]{n\to\infty} 0, \quad (n\Delta_n)^{-1}\sum_{i=0}^{n-1} E\left(\varepsilon_{2i,2n}^2 f^2(X_{2i\Delta_{2n}},\theta) \mid \mathcal{G}_{2i}^{2n}\right) \xrightarrow[\mathbf{L}^1]{n\to\infty} 0.$$

We deduce that for all θ, $\overline{\mathcal{I}}_n^{(4)}(\theta) \xrightarrow[\mathbf{P}]{n\to\infty} 0$. Furthermore, this convergence is uniform in θ by application of Proposition 7.2 and the following inequality,

$$E\left(|\varepsilon_{2i,2n}| \sup_{\theta\in\Theta} |\nabla_\theta f(X_{2i\Delta_{2n}},\theta)| \mid \mathcal{G}_{2i}^{2n}\right) \leq \Delta_{2n} c(\mathcal{B}_l^{\gamma+\alpha_1\vee\gamma+\frac{1}{2}\alpha_2}(X_{2i\Delta_{2n}}) + \mathcal{B}_r^{\frac{5}{2}+\gamma+\alpha_1+\frac{\alpha_2}{2}}(X_{2i\Delta_{2n}})).$$

- For $\overline{\mathcal{I}}_n^{(5)}(\theta)$, using Proposition 1.1 *2)* and (17), we see that,

$$E\left(\sup_{\theta\in\Theta}\left|\varepsilon_{2i,2n}(\overline{X}_{2i} - X_{2i\Delta_{2n}})f'_x(\tilde{X}_i,\theta)\right| \mid \mathcal{G}_{2i}^{2n}\right)$$
$$\leq \Delta_{2n}^{\frac{3}{2}} \times (\mathcal{B}_l^{(\gamma+2\alpha_1)\vee(\gamma+\alpha_1+\frac{1}{2}\alpha_2)}(X_{2i\Delta_{2n}}) + \mathcal{B}_r^{\frac{5}{2}+\gamma+2\alpha_1+\frac{1}{2}\alpha_2}(X_{2i\Delta_{2n}})).$$

Thus, $\overline{\mathcal{I}}_n^{(5)}(\theta) \xrightarrow[\mathbf{L}^1]{n\to\infty} 0$ uniformly in θ. □

Remark 2.3. *To enlight the previous result, recall that*

$$(n\Delta_n)^{-1}\sum_{i=0}^{n-1} f(X_{i\Delta_n})(X_{(i+1)\Delta_n} - X_{i\Delta_n} - b(X_{i\Delta_n})\Delta_n) \xrightarrow{n\to\infty} 0.$$

The difference with our result comes from the fact that $f(X_{i\Delta_n})$ and $X_{(i+1)\Delta_n} - X_{i\Delta_n} - b(X_{i\Delta_n})\Delta_n$ have a negligible correlation (of order Δ_n) whereas $f(\overline{X}_i)$ and $\overline{X}_{i+1} - \overline{X}_i - b(\overline{X}_i)\Delta_n$ have a correlation of order $\Delta_n^{\frac{1}{2}}$.

2.3 Quadratic variation of the observed process

The following result deals with the quadratic variation of $\overline{X}_i$ (recall (5)).

Theorem 2.4. *Assume (A1)-(A5), let f satisfy* $\mathbf{CU}_\gamma$ *with $2\gamma\vee(\gamma+\alpha_1+\frac{1}{2}\alpha_2) < M_l$, $4+2\gamma+\alpha_1+\frac{1}{2}\alpha_2 < M_r$ and $2\gamma\vee 4\alpha_1\vee 2\alpha_2 < K_l$, then,*

$$\overline{\mathcal{Q}}_n(f(.,\theta)) \xrightarrow{n\to\infty} \frac{2}{3}\nu_0\left(f(.,\theta)a^2(.)\right) \text{ uniformly in } \theta, \text{ in probability.} \tag{34}$$

Proof. First, we show the pointwise in θ convergence.

Set $W_{i,n}(\theta) = \left(\overline{X}_{i+1} - \overline{X}_i\right)^2 f(\overline{X}_i,\theta)$. Since $W_{i,n}(\theta)$ is $\mathcal{G}_{i+2}^n$ measurable, to prove the convergence of $(n\Delta_{2n})^{-1}\sum_{i=0}^n W_{i,n}(\theta)$, as in the previous theorem, we deal separately with the sum of even indexes and the one of odd indexes. And it is enough to show that:

$$(n\Delta_{2n})^{-1}\sum_{i=0}^{n-1} W_{2i,2n}(\theta) \xrightarrow{n\to\infty} \frac{2}{3}\nu_0\left(f(.,\theta)a^2(.)\right) \text{ in probability.} \tag{35}$$

(and $(n\Delta_{2n})^{-1}\sum_{i=0}^{n-1} W_{2i+1,2n}(\theta) \xrightarrow{n\to\infty} \frac{2}{3}\nu_0\left(f(.,\theta)a^2(.)\right)$ in probability.)

Using Theorem 1.2 *2)* and Taylor's formula, we write: $W_{2i,2n}(\theta) = w^{(1)}_{2i,2n}(\theta) + w^{(2)}_{2i,2n}(\theta) + w^{(3)}_{2i,2n}(\theta) + w^{(4)}_{2i,2n}(\theta)$

$$\begin{aligned}
\text{with}\quad w^{(1)}_{2i,2n}(\theta) &= \Delta_{2n} U^2_{2i,2n} a^2(X_{2i\Delta_{2n}}) f(X_{2i\Delta_{2n}},\theta)\\
w^{(2)}_{2i,2n}(\theta) &= 2\Delta^{\frac{1}{2}}_{2n} U_{2i,2n} a(X_{2i\Delta_{2n}})(\varepsilon_{2i,2n} + \Delta_{2n} b(X_{2i\Delta_{2n}})) f(X_{2i\Delta_{2n}},\theta)\\
w^{(3)}_{2i,2n}(\theta) &= (\varepsilon_{2i,2n} + \Delta_{2n} b(X_{2i\Delta_{2n}}))^2 f(X_{2i\Delta_{2n}},\theta)\\
w^{(4)}_{2i,2n}(\theta) &= (\overline{X}_{2i+1} - \overline{X}_{2i})^2(\overline{X}_{2i} - X_{2i\Delta_{2n}}) f'_x(\hat{X}_i,\theta), \quad \text{where } \hat{X}_i \in [\overline{X}_{2i},\ X_{2i\Delta_{2n}}].
\end{aligned}$$

We set $\overline{\mathcal{Q}}^{(j)}_n(\theta) = (n\Delta_{2n})^{-1}\sum_{i=0}^{n-1} w^{(j)}_{2i,2n}(\theta)$ for $j = 1,2,3,4$.

- We start by studying $\overline{\mathcal{Q}}^{(1)}_n(\theta)$. Using $E\left(U^2_{2i,2n} \mid \mathcal{G}^{2n}_{2i}\right) = 2/3$ and $E\left(U^4_{2i,2n} \mid \mathcal{G}^{2n}_{2i}\right) = 4/3$ we obtain:

$$E\left(w^{(1)}_{2i,2n}(\theta) \mid \mathcal{G}^{2n}_{2i}\right) = \frac{2\Delta_{2n}}{3} f(X_{i\Delta_{2n}},\theta) a^2(X_{i\Delta_{2n}})$$
$$E\left((w^{(1)}_{2i,2n}(\theta))^2 \mid \mathcal{G}^{2n}_{2i}\right) = \frac{4\Delta^2_{2n}}{3} f^2(X_{i\Delta_{2n}},\theta) a^4(X_{i\Delta_{2n}}).$$

First, applying Lemma 7.1 we get:

$$(n\Delta_{2n})^{-1}\sum_{i=0}^{n-1} E\left(w^{(1)}_{2i,2n}(\theta) \mid \mathcal{G}^{2n}_{2i}\right) \xrightarrow{n\to\infty} \frac{2}{3}\nu_0\left(f(.,\theta)a^2(.)\right) \text{ in probability.}$$

Second, using (A5) we get

$$\sup_{i,n} E\left|E\left((w^{(1)}_{2i,2n}(\theta))^2 \mid \mathcal{G}^{2n}_{2i}\right)\right| \le c\Delta^2_{2n},$$

and therefore, $(n\Delta_{2n})^{-2}\sum_{i=0}^{n-1} E\left((w^{(1)}_{2i,2n}(\theta))^2 \mid \mathcal{G}^{2n}_{2i}\right) \xrightarrow[\mathbf{L}^1]{n\to\infty} 0$.

Hence, by Lemma 7.3, we deduce $\overline{\mathcal{Q}}^{(1)}_n(\theta) \xrightarrow{n\to\infty} \frac{2}{3}\nu_0\left(f(.,\theta)a^2(.)\right)$ in probability.

- For $\overline{\mathcal{Q}}^{(2)}_n(\theta)$, by (16) we show

$$E\left(\left|w^{(2)}_{2i,2n}(\theta)\right| \mid \mathcal{G}^{2n}_{2i}\right) \le \Delta^{\frac{3}{2}}_{2n} c(B_l^{(\gamma+\alpha_1)\vee(\gamma+\frac{1}{2}\alpha_2)}(X_{2i\Delta_{2n}}) + B_r^{\frac{5}{2}+\gamma+\alpha_1+\frac{1}{2}\alpha_2}(X_{2i\Delta_{2n}})).$$

Then, $\overline{\mathcal{Q}}^{(2)}_n(\theta) \xrightarrow[\mathbf{L}^1]{n\to\infty} 0$.

- Analogously we show that $\overline{\mathcal{Q}}^{(3)}_n(\theta) \xrightarrow[\mathbf{L}^1]{n\to\infty} 0$.
- Now, we study $\overline{\mathcal{Q}}^{(4)}_n(\theta)$. For this term using Theorem 1.2, we get

$$E\left(\left|w^{(4)}_{2i,2n}(\theta)\right| \mid \mathcal{G}^{2n}_{2i}\right) \le \Delta^{\frac{3}{2}}_{2n}(\mathcal{B}^{\gamma}_l(X_{2i\Delta_{2n}}) + \mathcal{B}^{3+\gamma}_r(X_{2i\Delta_{2n}})).$$

Hence, $\overline{\mathcal{Q}}^{(4)}_n(\theta) \xrightarrow[\mathbf{L}^1]{n\to\infty} 0$.

The pointwise in θ convergence is proved for (34). It remains to show the uniformity. Using Proposition 7.2 of the Appendix it suffices to show:

$$\sup_{n\ge 1}(n\Delta_n)^{-1}\sum_{i=0}^{n-1} E\left[\left(\overline{X}_{i+1} - \overline{X}_i\right)^2 \sup_{\theta\in\Theta}\left|\nabla_\theta f(\overline{X}_i,\theta)\right|\right] < \infty$$

But, by condition $\mathbf{CU}_\gamma$ on f, $E\left(\sup_{\theta\in\Theta}\left|\nabla_\theta f(\overline{X}_i,\theta)\right|^2 \mid \mathcal{G}_i^n\right) \le c(\mathcal{B}_l^{2\gamma}(X_{i\Delta_n})+\mathcal{B}_r^{2\gamma}(X_{i\Delta_n}))$. Joining this inequality with Proposition 1.1 and (A5) we get,

$$E\left[(\overline{X}_{i+1}-\overline{X}_i)^2\sup_{\theta\in\Theta}\left|\nabla_\theta f(\overline{X}_i,\theta)\right|\right] \le \Delta_n c,$$

and hence the result. □

Remark 2.5. *Comparing Theorem 2.4 with the well known result*

$$(n\Delta_n)^{-1}\sum_{i=0}^{n-1} f(X_{i\Delta_n})(X_{(i+1)\Delta_n}-X_{i\Delta_n})^2 \xrightarrow{n\to\infty} \nu_0\left(fa^2\right),$$

the main difference comes from the fact that the variance of $U_{i,n}$ is $\frac{2}{3}$, whereas $\Delta_n^{-\frac{1}{2}}(B_{(i+1)\Delta_n}-B_{i\Delta_n})$ has variance 1.

The quadratic variation $\overline{\mathcal{Q}}_n(f)$ based on $(\overline{X}_i)$ is therefore a biased estimator of the quadratic variation $\nu_0(fa^2)$. An analogous result was obtained in Delattre and Jacod (1997) for the quadratic variation based on discrete observations of the diffusion with round-off errors.

3 Associated central limit theorems

Now, we study some related central limit theorems. We need no more uniformity in θ.

Theorem 3.1. *Assume (A1)-(A5) and $n\Delta_n^2 \xrightarrow{n\to\infty} 0$. Let f satisfies $\mathbf{C}_\gamma$ with $(2\gamma+3\alpha_1+\alpha_2+\beta_1+\beta_2)\vee 4\gamma < M_l$, $(4+2\gamma+3\alpha_1+\alpha_2+\beta_1+\beta_2)\vee(4+4\gamma) < M_r$ and $4\gamma\vee 4\alpha_1\vee 2\alpha_2\vee\beta_1\vee 2\beta_2 < K_l$ then,*

$$\overline{N}_n(f) := \sqrt{n\Delta_n}(\overline{\mathcal{I}}_n(f)-\frac{1}{4}\overline{\mathcal{Q}}_n(f')) \xrightarrow[\mathcal{D}]{n\to\infty} \mathcal{N}\left(0,\nu_0\left(f^2a^2\right)\right) \tag{36}$$

Proof. In this proof, we use the notations (21), (23)–(27) introduced in Theorem 2.2. Furthermore, here, we set ${v'}_{i,n}^{(2)} = -\frac{1}{4}(\overline{X}_{i+1}-\overline{X}_i)^2 f'(\overline{X}_i)$.

Define $\overline{N}_n^{(j)} = (n\Delta_n)^{-\frac{1}{2}}\sum_{i=0}^{n-1} v_{i,n}^{(j)}$ for $j = 1,3,4,5$ and $\overline{N}_n^{(2)} = (n\Delta_n)^{-\frac{1}{2}}\sum_{i=0}^{n-1}(v_{i,n}^{(2)}+{v'}_{i,n}^{(2)})$. With these notations $\overline{N}_n(f) = \sum_{l=1}^{5}\overline{N}_n^{(l)}$.

• First, we study

$$\overline{N}_n^{(1)} = n^{-\frac{1}{2}}\sum_{i=0}^{n-1}\alpha(X_{i\Delta_n})U_{i,n} \text{ with } \alpha(x) = f(x)a(x).$$

In order to apply a martingale central limit theorem, we first have to reorder terms. (recall (10)–(12))

$$\overline{N}_n^{(1)} = n^{-\frac{1}{2}}\sum_{i=1}^{n-1} s_{i,n}^{(1)} + n^{-\frac{1}{2}}\left(\alpha(X_0)\xi_{0,n}+\alpha(X_{(n-1)\Delta_n})\xi'_{n-1,n}\right), \tag{37}$$

with

$$s^{(1)}_{i,n} = \alpha(X_{i\Delta_n})\xi_{i,n} + \alpha(X_{(i-1)\Delta_n})\xi'_{i,n}. \tag{38}$$

We have, $E\left(s^{(1)}_{i,n} \mid \mathcal{G}^n_i\right) = 0$ and

$$E\left((s^{(1)}_{i,n})^2 \mid \mathcal{G}^n_i\right) = \frac{1}{3}\{\alpha^2(X_{i\Delta_n}) + \alpha^2(X_{(i-1)\Delta_n}) + \alpha(X_{i\Delta_n})\alpha(X_{(i-1)\Delta_n})\}.$$

An application of Lemma 7.1, after having shown that, $n^{-1}\sum_{i=0}^{n-1}\alpha(X_{i\Delta_n})(\alpha(X_{i\Delta_n})-\alpha(X_{(i-1)\Delta_n})) = o_{\mathbf{P}}(1)$, yields:

$$n^{-1}\sum_{i=0}^{n-1} E\left((s^{(1)}_{i,n})^2 \mid \mathcal{G}^n_i\right) \xrightarrow[\mathbf{P}]{n\to\infty} \nu_0\left(f^2a^2\right). \tag{39}$$

We easily bound $E\left((s^{(1)}_{i,n})^4 \mid \mathcal{G}^n_i\right)$ and show

$$n^{-2}\sum_{i=0}^{n-1} E\left((s^{(1)}_{i,n})^4 \mid \mathcal{G}^n_i\right) \xrightarrow[\mathbf{L}^1]{n\to\infty} 0. \tag{40}$$

Using Theorem 3.2 (p. 58) in Hall and Heyde (1980), the conditions (39)–(40) are sufficient to imply

$$n^{-\frac{1}{2}}\sum_{i=0}^{n-1} s^{(1)}_{i,n} \xrightarrow[\mathcal{D}]{n\to\infty} \mathcal{N}\left(0, \nu_0\left(f^2a^2\right)\right). \tag{41}$$

Using that, by (37),

$$\overline{N}^{(1)}_n = n^{-\frac{1}{2}}\sum_{i=0}^{n-1} s^{(1)}_{i,n} + o_{\mathbf{P}}(1), \tag{42}$$

we deduce $\overline{N}^{(1)}_n \xrightarrow[\mathcal{D}]{n\to\infty} \mathcal{N}\left(0, \nu_0\left(f^2a^2\right)\right)$.

• Second, we show the convergence to 0 of $\overline{N}^{(2)}_n$. Some computations based on Therorem 1.2 *1)* and Proposition 1.1 show that

$$E\left({v'_{2i,2n}}^{(2)} \mid \mathcal{G}^n_i\right) = -\frac{\Delta_{2n}}{6}a^2(X_{2i\Delta_{2n}})f'(X_{2i\Delta_{2n}}) + r'_{2i,2n}$$

with,

$$\left|r'_{2i,2n}\right| \leq c\Delta_{2n}^{\frac{3}{2}}(\mathcal{B}_l^{\gamma+3\alpha_1\vee\gamma+2\alpha_1+\alpha_2}(X_{2i\Delta_{2n}}) + \mathcal{B}_r^{4+\gamma+3\alpha_1+\alpha_2}(X_{2i\Delta_{2n}})). \tag{43}$$

Recalling (31), we get $E\left(v^{(2)}_{2i,2n} + {v'_{2i,2n}}^{(2)} \mid \mathcal{G}^{2n}_{2i}\right) = r_{i,n} + r'_{i,n}$. Now, by (32), (43) and (A5) we deduce,

$$E\left|E\left(v^{(2)}_{2i,2n} + {v'_{2i,2n}}^{(2)} \mid \mathcal{G}^{2n}_{2i}\right)\right| \leq c\Delta_{2n}^{\frac{3}{2}}.$$

This implies $E\left|(n\Delta_{2n})^{-\frac{1}{2}}\sum_{i=0}^{n-1}E\left(v_{2i,2n}^{(2)}+v_{2i,2n}'^{(2)}\mid\mathcal{G}_{2i}^{2n}\right)\right|\leq c(n\Delta_{2n}^2)^{\frac{1}{2}}$, which tends to 0 using now the condition $n\Delta_n^2\to 0$.

As for (30), we get $(n\Delta_{2n})^{-1}\sum_{i=0}^{n-1}E\left((v_{2i,2n}^{(2)}+v_{2i,2n}'^{(2)})^2\mid\mathcal{G}_{2i}^{2n}\right)=o_{\mathbf{P}}(1)$.

This implies the convergence for the sum of term with even indexes:

$$(n\Delta_{2n})^{-\frac{1}{2}}\sum_{i=0}^{n-1}\{v_{2i,2n}^{(2)}+v_{2i,2n}'^{(2)}\}=o_{\mathbf{P}}(1).$$

Analogously we show $(n\Delta_{2n+1})^{-\frac{1}{2}}\sum_{i=0}^{n-1}\{v_{2i+1,2n}^{(2)}+v_{2i+1,2n}'^{(2)}\}=o_{\mathbf{P}}(1)$. Hence $\overline{N}_n^{(2)}\to 0$.

• We show the convergence to zero for $\overline{N}_n^{(3)}$, $\overline{N}_n^{(4)}$ and $\overline{N}_n^{(5)}$ using that $n\Delta_n^2\to 0$ (the proof is a repetition of the proof of convergence for the corresponding terms, $\overline{\mathcal{I}}_n^{(3)}$, $\overline{\mathcal{I}}_n^{(4)}$ and $\overline{\mathcal{I}}_n^{(5)}$, in Theorem 2.2).

To conclude,

$$\overline{N}_n(f)=\overline{N}_n^{(1)}+o_{\mathbf{P}}(1)=n^{-\frac{1}{2}}\sum_{i=0}^{n-1}s_{i,n}^{(1)}+o_{\mathbf{P}}(1),\tag{44}$$

and the theorem follows from (41). □

Remark 3.2. *1) Comparing with the classic convergence limit,*

$$(n\Delta_n)^{-\frac{1}{2}}\sum_{i=0}^{n-1}(X_{(i+1)\Delta_n}-X_{i\Delta_n}-\Delta_n b(X_{i\Delta_n}))f(X_{i\Delta_n})\xrightarrow{n\to\infty}\mathcal{N}\left(0,\nu_0\left(f^2a^2\right)\right),$$

it appears that, if we just replace $X_{i\Delta_n}$ by $\overline{X}_i$ above, then the expression may tend to ∞, in probability, because of the non negligeable correlation between $f(\overline{X}_i)$ and $\overline{X}_{i+1}-\overline{X}_i-b(\overline{X}_i)\Delta_n$. So we have to introduce the appropriate correction.

2) We cannot replace in the statement of the previous Theorem $\frac{1}{4}\left(\overline{X}_{i+1}-\overline{X}_i\right)^2 f'(\overline{X}_i)$, by the term $\frac{n\Delta_n}{6}\nu_0\left(f'a^2\right)$ since we can show that

$$(n\Delta_n)^{-\frac{1}{2}}\left\{\sum_{i=0}^{n-1}\frac{1}{4}f'(\overline{X}_i)\left(\overline{X}_{i+1}-\overline{X}_i\right)^2-\frac{1}{6}\nu_0\left(f'a^2\right)\right\}$$

does not tend to zero when $n\to\infty$, hence Theorem 3.1 does not provide an exact central limit theorem for Theorem 2.2.

3) The condition $n\Delta_n^2\to 0$ is classical (see Florens-Zmirou (1989)). This condition imposes that the discretization step decreases to zero fast enough, to ensure that the contribution in $\overline{N}_n(f)$ of the error terms $\varepsilon_{i,n}$ tends to 0 as $n\to\infty$.

Let us now state a central limit theorem related with the functional $\overline{\mathcal{Q}}_n(g)$ (see (5)).

Theorem 3.3. *Assume (A1)-(A5) and* $n\Delta_n^2 \xrightarrow{n\to\infty} 0$*. Let* g *satisfies* $\mathbf{C}_\gamma$ *with* $4\gamma \vee (2\gamma + 4\alpha_1 + 2\alpha_2 + \beta_1) < M_l$*,* $(8+4\gamma) \vee (6+2\gamma+4\alpha_1+2\beta_2+\beta_1) < M_r$ *and* $4\gamma \vee (2\gamma+4\alpha_1+2\alpha_2) \vee 4\beta_2 < K_l$*. Then,*

$$\overline{M}_n(g) := n^{\frac{1}{2}}(\frac{3}{2}\overline{Q}_n(g) - \overline{\nu}_n(a^2 g)) \xrightarrow[\mathcal{D}]{n\to\infty} \mathcal{N}\left(0, \frac{9}{4}\nu_0(g^2a^4)\right) \tag{45}$$

Proof. Set,

$$\delta_{i,n} = \left\{\frac{(\overline{X}_{i+1} - \overline{X}_i)^2}{\frac{2}{3}\Delta_n} - a^2(\overline{X}_i)\right\} g(\overline{X}_i) - \left\{\frac{(U_{i,n})^2\, a^2(X_{i\Delta_n})}{\frac{2}{3}} - a^2(X_{i\Delta_n})\right\} g(X_{i\Delta_n}).$$

With some computations, based on (17)–(19), we get

$$|E(\delta_{i,n} \mid \mathcal{G}_i^n)| \le c\Delta_n(\mathcal{B}_l^{(\gamma+\alpha_1)\vee(\gamma+\alpha_1+\alpha_2+\beta_1)}(X_{i\Delta_n}) + \mathcal{B}_r^{(\frac{9}{2}+\gamma+2\alpha_1+\alpha_2)\vee 2\gamma}(X_{i\Delta_n}))$$
$$E(\delta_{i,n}^2 \mid \mathcal{G}_i^n) \le \Delta_n c(\mathcal{B}_l^{2\gamma+4\alpha_1+2\alpha_2}(X_{i\Delta_n}) + \mathcal{B}_r^{6+2\gamma+2\alpha_1+\alpha_2}(X_{i\Delta_n})).$$

This implies $n^{-\frac{1}{2}}\sum_{i=0}^{n-1}\delta_{i,n} \xrightarrow{n\to\infty} 0$ in probability.

Define $\beta(x) = a^2(x)g(x)$ and

$$\overline{M}_n' = n^{-\frac{1}{2}}\sum_{i=0}^{n-1}\beta(X_{i\Delta_n})\left\{\frac{(U_{i,n})^2}{\frac{2}{3}} - 1\right\}.$$

We have thus proved that

$$\overline{M}_n(g) = \overline{M}_n' + o_{\mathbf{P}}(1). \tag{46}$$

We now deal with $\overline{M}_n'$. Reordering terms in $\overline{M}_n'$ to obtain a triangular array of martingale increments, we get (recall (12))

$$\overline{M}_n' = \frac{3}{2}n^{-\frac{1}{2}}\left\{\sum_{i=1}^{n-1} s_{i,n}^{(2)} + (\xi_{0,n}^2 - \frac{1}{3})\beta(X_0) + (\xi_{n,n}'^2 - \frac{1}{3})\beta(X_{(n-1)\Delta_n}) + 2\xi_{n-1,n}\xi_{n,n}'\beta(X_{(n-1)\Delta_n})\right\} \tag{47}$$

where

$$s_{i,n}^{(2)} = (\xi_{i,n}^2 - \frac{1}{3})\beta(X_{i\Delta_n}) + (\xi_{i,n}'^2 - \frac{1}{3})\beta(X_{(i-1)\Delta_n}) + 2\xi_{i-1,n}\xi_{i,n}'\beta(X_{(i-1)\Delta_n}) \tag{48}$$

But, now, $s_{i,n}^{(2)}$ is $\mathcal{G}_{i+1}^n$ measurable and $E\left(s_{i,n}^{(2)} \mid \mathcal{G}_i^n\right) = 0$. Furthermore, using $E\left((\xi_{i,n}^2 - 1/3)^2 \mid \mathcal{G}_i^n\right) = E\left((\xi_{i,n}'^2 - 1/3)^2 \mid \mathcal{G}_i^n\right) = 2/9$ and $E\left((\xi_{i,n}^2 - 1/3)(\xi_{i,n}'^2 - 1/3) \mid \mathcal{G}_i^n\right) = \frac{2}{6^2}$ we obtain,

$$E\left((s_{i,n}^{(2)})^2 \mid \mathcal{G}_i^n\right) = \frac{2}{9}\beta^2(X_{i\Delta_n}) + \frac{2}{9}\beta^2(X_{(i-1)\Delta_n}) + \frac{1}{9}\beta(X_{(i-1)\Delta_n})\beta(X_{i\Delta_n}) + \frac{4}{3}\xi_{i-1,n}^2\beta^2(X_{(i-1)\Delta_n}) \tag{49}$$

By Lemma 7.1 and 7.3 (in the Appendix) we show the convergence for the following array of martingale increments,

$$n^{-1}\sum_{i=0}^{n-1}\xi_{i-1,n}^2\beta^2(X_{(i-2)\Delta_n}) \xrightarrow[\mathbf{P}]{n\to\infty} \frac{1}{3}\nu_0(\beta^2) \qquad (\text{recall } \beta(x)=a^2(x)g(x)).$$

Using the bounds on the derivative of β and Proposition 1.1 *1)* we deduce $n^{-1}\sum_{i=0}^{n-1}\xi_{i-1,n}^2\beta^2(X_{(i-1)\Delta_n}) \xrightarrow[\mathbf{P}]{n\to\infty} \frac{1}{3}\nu_0(\beta^2)$. Thus we have, by (49),

$$\frac{9}{4}n^{-1}\sum_{i=1}^{n-1}E\left((s_{i,n}^{(2)})^2 \mid \mathcal{G}_i^n\right) \xrightarrow[\mathbf{P}]{n\to\infty} \frac{9}{4}\nu_0(\beta^2).$$

But, the bound on β^4 implies: $n^{-2}\sum_{i=1}^{n-1}E\left((s_{i,n}^{(2)})^4 \mid \mathcal{G}_i^n\right) \xrightarrow[\mathbf{L}^1]{n\to\infty} 0$.

Now, Theorem 3.2 (p. 58) in Hall and Heyde (1980), yields: $\frac{3}{2}n^{-\frac{1}{2}}\sum_{i=0}^{n-1}s_{i,n}^{(2)} \xrightarrow[\mathcal{D}]{n\to\infty} \mathcal{N}\left(0,\frac{9}{4}\nu_0(\beta^2)\right)$. Remarking that, by (46) and (47),

$$\overline{M}_n(g)=\frac{3}{2}n^{-\frac{1}{2}}\sum_{i=0}^{n-1}s_{i,n}^{(2)}+o_{\mathbf{P}}(1), \tag{50}$$

the theorem is proved. □

Remark 3.4. *1) Comparing with the usual property*

$$n^{-\frac{1}{2}}\sum_{i=0}^{n-1}\left\{g(X_{i\Delta_n})\frac{\left(X_{(i+1)\Delta_n}-X_{i\Delta_n}\right)^2}{\Delta_n}-g(X_{i\Delta_n})a^2(X_{i\Delta_n})\right\} \xrightarrow[\mathcal{D}]{n\to\infty} \mathcal{N}\left(0,2\nu_0(g^2a^4)\right)$$

we see that when we replace $X_{i\Delta_n}$ by $\overline{X}_i$ in the equation above the variance of the limit increases a little.

2) The previous theorem is linked with Theorem 2.4 but it does not give the central limit theorem related with it, since in fact,

$$n^{\frac{1}{2}}\left\{(n\Delta_n)^{-1}\sum_{i=0}^{n-1}g(\overline{X}_i)\left(\overline{X}_{i+1}-\overline{X}_i\right)^2-\frac{2}{3}\nu_0(ga^2)\right\}$$

is of order $\Delta_n^{-\frac{1}{2}}$.

Finally, we have:

Theorem 3.5. *Let f and g be two functions satisfying respectively $\mathbf{C}_\gamma$ and $\mathbf{C}_{\gamma'}$. Assume that the assumptions of Theorem 3.1 are valid for f, and those of Theorem 3.3 are valid for g. Suppose, furthermore, that $2\gamma+2\gamma'+\alpha_1<M_l$, $6+2\gamma+2\gamma'+\alpha_1<M_r$. Then,*

$$(\overline{N}_n(f),\overline{M}_n(g)) \xrightarrow[\mathcal{D}]{n\to\infty} \mathcal{N}\left(0,\nu_0(f^2a^2)\right)\otimes\mathcal{N}\left(0,\frac{9}{4}\nu_0(g^2a^4)\right) \tag{51}$$

Proof. We have shown, in Theorems 3.1 and 3.3 (see (44) and (50)), that $\overline{N}_n(f) - n^{-\frac{1}{2}}\sum_{i=1}^{n-1} s_{i,n}^{(1)}$ and $\overline{M}_n(g) - \frac{3}{2}n^{-\frac{1}{2}}\sum_{i=1}^{n-1} s_{i,n}^{(2)}$ tend to zero in probability. Since we deal with martingale arrays, it suffices to prove

$$n^{-1}\sum_{i=1}^{n-1} E\left(s_{i,n}^{(1)} s_{i,n}^{(2)} \mid \mathcal{G}_i^n\right) \xrightarrow[\mathbf{P}]{n\to\infty} 0. \tag{52}$$

But using (38) and (48), we get (recall $\alpha(x) = a(x)g(x)$, $\beta(x) = a^2(x)g(x)$):

$$E\left(s_{i,n}^{(1)} s_{i,n}^{(2)} \mid \mathcal{G}_i^n\right) = \xi_{i-1,n}\beta(X_{(i-1)\Delta_n})(\frac{1}{3}\alpha(X_{i\Delta_n}) + \frac{2}{3}\alpha(X_{(i-1)\Delta_n})).$$

Now, an application of Lemma 7.3 yields: $n^{-1}\sum_{i=0}^{n-1}\xi_{i-1,n}\beta(X_{(i-1)\Delta_n})\alpha(X_{(i-1)\Delta_n}) \xrightarrow[\mathbf{P}]{n\to\infty} 0$
Using Proposition 1.1 and the smoothness of α and β we get:
$n^{-1}\sum_{i=0}^{n-1}\{E\left(s_{i,n}^{(1)} s_{i,n}^{(2)} \mid \mathcal{G}_i^n\right) - \xi_{i-1,n}\beta(X_{(i-1)\Delta_n})\alpha(X_{(i-1)\Delta_n})\} \xrightarrow[\mathbf{L}^1]{n\to\infty} 0$ and hence (52). □

4 Checking Assumptions

In this Section we give examples of models for which our assumptions hold. Actually, only (A2) and (A5) have to be studied. The examples below show that they are not restrictive conditions.

The condition (A2) was studied in Gloter (1998, b) and was shown to hold for all models below.

4.1 Diffusion on $\mathbb{R}$

Here $(l, r) = (-\infty, \infty)$; $\mathcal{B}_{-\infty}(x) = 1$ and $\mathcal{B}_{\infty}(x) = 1 + |x|$; (X_t) is the solution of

$$dX_t = a(X_t)dB_t + b(X_t)dt, \quad X_0 = \eta. \tag{53}$$

Let us assume that Assumptions (A1), (A3) are satisfied, and the stationary distibution ν_0 has finite moments of every orders. This means that (A4) is satisfied for any positive constants $M_{-\infty}$ and M_{∞}. By Gloter (1998, b), Assumption (A2) is satisfied with $K_{-\infty} = \infty$.

It remains to check (A5). If η has distribution $d\nu_0$, then (A5) follows from (A4) by stationarity. The next proposition shows that (A5) follows again from (A4) if η is deterministic.

Proposition 4.1. *Let X be solution of* (53) *starting from $X_0 = y \in \mathbb{R}$. Assume that (A1), (A3) and (A4) hold with any positive constants $M_{-\infty}$ and M_{∞}, then*

$$\forall p \geq 0,\ \sup_{t\geq 0} E(|X_t|^p) < \infty \textit{ (and hence (A5) holds).}$$

Proof. Define (since $\nu_0([y,\infty)) > 0$ for all y) the probability measure: $d\pi(x) = d\nu_0(x)\nu_0([y,\infty))^{-1}\mathbb{1}_{\{x\geq y\}}$. Using that the stochastic differential equation admits strong solutions we can define on $(\Omega, \mathcal{F}, P)$ the processes (X_t^x), $x \in \mathbb{R}$ and (X_t^π), solutions of:

$$dX_t^x = a(X_t^x)dB_t + b(X_t^x)dt, \quad X_0^x = x$$
$$dX_t^\pi = a(X_t^\pi)dB_t + b(X_t^\pi)dt, \quad X_0^\pi = \Pi$$

with Π a random variable with distribution $d\pi(x)$, independent of B.

Now, since $X_0 = X_0^y \leq X_0^\pi$ a.s, using that $X = X^y$ and X^π coincide after the time $T = \inf\{s \geq 0; X_s^y = X_s^\pi\}$ we get $P(\ \forall t \geq 0, X_s \leq X_s^\pi) = 1$. So, for $p \in \mathbb{N}$ (with the notation $x^+ = x \vee 0$):

$$E\left((X_t{}^+)^p\right) \leq E\left((X_t^{\pi+})^p\right) = \int_{\mathbb{R}} E\left((X_t^{x+})^p\right) d\pi(x)$$

$$\leq \nu_0\left([y,\infty)\right)^{-1} \int_{\mathbb{R}} E\left((X_t^{x+})^p\right) d\nu_0(x) = \nu_0\left([y,\infty)\right)^{-1} \int_{\mathbb{R}} (x^+)^p \, d\nu_0(x)$$

Since by assumption, ν_0 has finite moments of every orders, we have

$$\sup_{t\geq 0} E\big((X_t^+)^p\big) < \infty.$$

We analogously show: $\sup_{t\geq 0} E\big((X_t^-)^p\big) < \infty$ and get the proposition. □

As the results, the work of Sections 2–3 encompasses a large class of diffusion models on $\mathbb{R}$, among them we can quote the Ornstein–Uhlenbeck process (for $b(x) = \mu x$ and $a(x) = \sigma$) or the Hyperbolic process (for $b(x) = \mu x(1+x^2)^{-\frac{1}{2}}$ and $a(x) = \sigma$).

4.2 Exponantial of a diffusion

When $X_t = e^{Z_t}$ is the exponantial of a diffusion, we can show that under reasonable conditions on Z, the positive diffusion X satisfies our assumptions.

4.3 The Cox–Ingersoll–Ross process

Here $(l, r) = (0, \infty)$; $\mathcal{B}_0(x) = 1 + \frac{1}{x}$, $\mathcal{B}_\infty(x) = 1 + x$ and

$$dX_t = (\mu X_t + \mu')dt + \sigma\sqrt{X_t}dB_t, \quad X_0 = \eta, \tag{54}$$

with $\mu < 0$, $\sigma, \mu' > 0$. We set $c_0 = \frac{2\mu'}{\sigma^2}, \lambda = \frac{2|\mu|}{\sigma^2}$. and suppose that $c_0 > 1$.

Assumption (A1) is clear and $c_0 > 1$ implies Assumption (A3), with the stationary probability measure, $d\nu_0(x) = \frac{\lambda^{c_0}}{\Gamma(c_0)} x^{c_0-1} e^{-\lambda x} 1\!\!1_{\{x>0\}} dx$.

Using this expression for the stationary probability, we easily check (A4) with any $M_0 < c_0$. It is shown in Gloter (1998, b) that (A2) holds with $K_0 = c_0 - 1$ (actually this model justify the introduction of the constant K_0 in (A2)).

If η has distribution ν_0, (A5) holds. If η is deterministic, we show (A5) as is Proposition 4.1.

4.4 Bilinear diffusion

Here, again $l = 0$, $r = \infty$, $\mathcal{B}_0(x) = 1 + \frac{1}{x^p}$, $\mathcal{B}_\infty(x) = 1 + x$, where p is a positive constant. X is solution of

$$dX_t = (\mu X_t + \mu')dt + \sigma X_t dB_t, \quad X_0 = \eta, \tag{55}$$

with $\mu < 0$, $\sigma, \mu' > 0$. We set $c_0 = 1 + \frac{2|\mu|}{\sigma^2}$, $\lambda = \frac{2\mu'}{\sigma^2}$.

We easily check (A1) and (A3) with $d\nu_0(x) = \frac{\lambda^{c_0}}{\Gamma(c_0)} x^{-c_0-1} e^{-\frac{\lambda}{x}} 1\!\!1_{\{x>0\}} dx$. We deduce (A4) with any $M_\infty < c_0$. Assumptions (A5) follows if η hat the stationary law or is deterministic as in Proposition 4.1. In Gloter (1998, b), we have shown that (A2) holds with $K_0 = \infty$.

5 Statistical applications. Minimum contrast estimation.

Let (X_t) be the unique solution of the equation:

$$dX_t = a(X_t, \sigma_0)dB_t + b(X_t, \mu_0)dt, \quad X_0 = \eta, \tag{56}$$

where $(B_t)_{t\geq 0}$ is a standard one dimensional Brownian motion, η is a random variable independent of $(B_t)_{t\geq 0}$; b and a are two real valued functions respectively defined on $\mathbb{R} \times \Theta_1$ and $\mathbb{R} \times \Theta_2$, where Θ_1 and Θ_2 are two compact intervals of $\mathbb{R}$. We denote by $\theta = (\mu, \sigma)$ the elements of $\Theta = \Theta_1 \times \Theta_2$, furthermore we suppose that $\theta_0 = (\mu_0, \sigma_0) \in \overset{\circ}{\Theta}$. To simplify notations and proofs we have chosen a one dimensional parameter for the drift and for the diffusion coefficient, but we could easily extend this work to multidimensional parameters.

In the case of the observation of $X_{i\Delta_n}$ a contrast may be constructed by approximating $X_{(i+1)\Delta_n} - X_{i\Delta_n}$ by a $\mathcal{N}(b(X_{i\Delta_n}, \mu_0)\Delta_n, a^2(X_{i\Delta_n}, \sigma_0)\Delta_n)$. This leads to the Euler contrast (see Kessler (1997)).

To obtain a contrast based on the observation of $\overline{X}_i$, we correct the Euler contrast by taking into account the factor $\frac{2}{3}$ in Theorem 2.4, and compensate the effect of the correlation between $\overline{X}_i$ and $\overline{X}_{i+1} - \overline{X}_i - \Delta_n b(\overline{X}_i, \mu_0)$ (Theorem 2.2).

This leads to the contrast $\mathcal{L}_n(\theta)$ given by (6) (recall $h(x,\theta) = \frac{\partial}{\partial x} \frac{b(x,\mu)}{a^2(x,\sigma)}$). Let $\hat{\theta}_n = \operatorname{arginf}_{\theta\in\Theta} \mathcal{L}_n(\theta)$ be a minimum contrast estimator

We suppose that the diffusion X satisfies Assumptions (A1)–(A5) (where $a(x)$ stands for $a(x, \sigma_0)$ and $b(x)$ for $b(x, \mu_0)$). Furthermore, to keep proofs on the behaviour of $\hat{\theta}_n$ tractable, we make the additional assumption that K_l in Assumption (A2) is equal to ∞ and that Assumptions (A4)–(A5) hold for any M_l and M_r (and hence these constants can be chosen as large as we need for the application of results of Sections 2 and 3). We have seen in Section 4 that this is true for a large class of diffusion processes (Sections 4.1–4.2). For models that do not satisfy this additional assumption, we will give specific proof of the consistency and normality of $\hat{\theta}_n$ (see Sections 6.2 and 6.3).

We suppose that the following identifiability assumption holds:

(S1) $a(x, \sigma) = a(x, \sigma_0) \quad d\nu_0(x)$ almost everywhere implies $\sigma = \sigma_0$,
$b(x, \mu) = b(x, \mu_0) \quad d\nu_0(x)$ almost everywhere implies $\mu = \mu_0$.

We need an assumption on the smoothness of $a(x, \sigma)$ and $b(x, \mu)$ with respect to the parameter.

(S2) a and b are the restrictions of functions defined on an open subset of $\mathbb{R}^2$, on which they are differentiable up to order 6, furthermore they satisfy: $\exists\ c \geq 0, \exists k \geq 0$ such that $\forall\ i, j \in \{0, \ldots, 3\}^2$, $\forall\ x \in (l, r)$:

$$\sup_{\mu\in\Theta_1} \left| \frac{\partial^{i+j}}{\partial \mu^i x^j} b(x, \mu) \right| + \sup_{\sigma\in\Theta_2} \left| \frac{\partial^{i+j}}{\partial \sigma^i x^j} a(x, \sigma) \right| + \sup_{\sigma\in\Theta_2} \left| \frac{\partial^{i+j}}{\partial \sigma^i x^j} a^{-1}(x, \sigma) \right| \leq c(\mathcal{B}_l^k(x) + \mathcal{B}_r^k(x))$$

By the previous assumption, all functions appearing below satisfy $\mathbf{CU}_\gamma$ for some $\gamma \geq 0$ and hence limit theorems of Section 2 and 3 apply for these functions.

We can now prove consistency and normality for $\hat{\theta}_n$. To maintain formulae short we denote $\partial_\sigma f = \frac{\partial}{\partial \sigma} f$, $\partial_\mu f = \frac{\partial}{\partial \mu} f$, $\partial^2_{\sigma^2} f = \frac{\partial^2}{\partial \sigma^2} f$, $\partial^2_{\sigma\mu} f = \frac{\partial^2}{\partial \sigma\mu} f$ ....

Theorem 5.1. *The estimator $\hat{\theta}_n$ is consistent,*

$$\hat{\theta}_n \xrightarrow{n\to\infty} \theta_0 \quad \textit{in probability.}$$

Proof. Following the proof of Kessler's Theorem 1 (Kessler (1997)), by Assumption (S1), it is enough to show that, uniformly in θ,

$$n^{-1}\mathcal{L}_n(\theta) \xrightarrow{n\to\infty} \nu_0\left(\frac{a^2(.,\sigma_0)}{a^2(.,\sigma)} + \log a^2(.,\sigma)\right) \quad \text{in probability} \tag{57}$$

This will ensure the convergence of $\hat{\sigma}_n$ to σ_0. Then, we prove that, uniformly in $\theta = (\mu, \sigma)$,

$$(n\Delta_n)^{-1}(\mathcal{L}_n(\mu,\sigma) - \mathcal{L}_n(\mu_0,\sigma)) \xrightarrow{n\to\infty} \frac{3}{2}\nu_0\left(\frac{(b(.,\mu) - b(.,\mu_0))^2}{a^2(.,\sigma)}\right) \text{ in probability} \tag{58}$$

This enables to obtain the convergence of $\hat{\mu}_n$ to μ_0 (for more details on why (57)–(58) imply consistency, see the Appendix 7.2).

We start the proof by (57). With the notations (3)–(5), the contrast (divided by n) writes, after easy computations,

$$n^{-1}\mathcal{L}_n(\theta) = \frac{3}{2}\overline{\mathcal{Q}}_n(a^{-2}(.,\sigma)) + \overline{\nu}_n(\log a^2(.,\sigma))$$
$$-3\Delta_n\overline{\mathcal{I}}_n(a^{-2}(.,\sigma)b(.,\mu)) + \frac{3\Delta_n}{4}\overline{\mathcal{Q}}_n(h(.,\theta)) + \frac{3\Delta_n}{2}\overline{\nu}_n(a^{-2}(.,\sigma)\{b^2(.,\mu) - 2b(.,\mu)b(.,\mu_0)\}). \tag{59}$$

Using Proposition 2.1, Theorems 2.2, 2.4 and $\Delta_n \to 0$, we easily obtain (57).

For the proof of (58), by the expression of the contrast (devided by n) above, we get

$$(n\Delta_n)^{-1}(\mathcal{L}_n(\mu,\sigma) - \mathcal{L}_n(\mu_0,\sigma)) = 3\overline{\mathcal{I}}_n\left(\frac{b}{a^2}(.,\mu_0,\sigma) - \frac{b}{a^2}(.,\mu,\sigma)\right) - \frac{3}{4}\overline{\mathcal{Q}}_n\left(h(.,\mu_0,\sigma) - h(.,\mu,\sigma)\right)$$
$$+\frac{3}{2}\overline{\nu}_n\left(\frac{(b(.,\mu) - b(.,\mu_0))^2}{a^2(.,\sigma)}\right)$$

Now, we apply Theorems 2.2, 2.4 (recall $h(x,\theta) = \partial_x(\frac{b}{a^2})(x,\theta)$ too) and Proposition 2.1 to get (58). □

We now prove that the estimator $\hat{\theta}_n$ is asymptotically normal. The scheme of the proof is classical.

Theorem 5.2. *If $n\Delta_n^2 \xrightarrow{n\to\infty} 0$, then $\left((n\Delta_n)^{\frac{1}{2}}(\hat{\mu}_n - \mu_0), n^{\frac{1}{2}}(\hat{\sigma}_n - \sigma_0)\right)$ converges in law to a*

$$\mathcal{N}\left(0, \left\{\nu_0\left(\frac{(\partial_\mu b)^2(.,\mu_0)}{a^2(.,\sigma_0)}\right)\right\}^{-1}\right) \otimes \mathcal{N}\left(0, \frac{9}{16}\left\{\nu_0\left(\frac{(\partial_\sigma a)^2(.,\sigma_0)}{a^2(.,\sigma_0)}\right)\right\}^{-1}\right)$$

Proof. Since $\theta_0 \in \overset{\circ}{\Theta}$, by Taylor's formula:

$$\int_0^1 \nabla_\theta^2\mathcal{L}_n(\theta_0 + u(\hat{\theta}_n - \theta_0))du \begin{bmatrix}\hat{\mu}_n - \mu_0\\ \hat{\sigma}_n - \sigma_0\end{bmatrix} = -\nabla_\theta\mathcal{L}_n(\theta_0) \tag{60}$$

Define:

$$\mathcal{C}_n(\theta) = \begin{bmatrix} (n\Delta_n)^{-1}\frac{\partial^2}{\partial\mu^2}\mathcal{L}_n(\theta) & n^{-1}\Delta_n^{-\frac{1}{2}}\frac{\partial^2}{\partial\sigma\mu}\mathcal{L}_n(\theta) \\ n^{-1}\Delta_n^{-\frac{1}{2}}\frac{\partial^2}{\partial\mu\sigma}\mathcal{L}_n(\theta) & n^{-1}\frac{\partial^2}{\partial\sigma^2}\mathcal{L}_n(\theta) \end{bmatrix}$$

$$\mathcal{E}_n = \begin{bmatrix} (n\Delta_n)^{\frac{1}{2}}(\hat{\mu}_n - \mu_0) \\ n^{\frac{1}{2}}(\hat{\sigma}_n - \sigma_0) \end{bmatrix}, \quad \mathcal{D}_n = \begin{bmatrix} -(n\Delta_n)^{-\frac{1}{2}}\frac{\partial}{\partial\mu}\mathcal{L}_n(\theta_0) \\ -n^{-\frac{1}{2}}\frac{\partial}{\partial\sigma}\mathcal{L}_n(\theta_0) \end{bmatrix}$$

Then (60) can be written (with some easy computations): $\int_0^1 \mathcal{C}_n(\theta_0 + u(\hat{\theta}_n - \theta_0))du\, \mathcal{E}_n = \mathcal{D}_n$. Now the proof of $\mathcal{E}_n \xrightarrow{n\to\infty} \mathcal{N}\left(0, \left\{\nu_0\left(\frac{(\partial_\mu b)^2(.,\mu_0)}{a^2(.,\sigma_0)}\right)\right\}^{-1}\right) \otimes \mathcal{N}\left(0, \frac{9}{16}\left\{\nu_0\left(\frac{(\partial_\sigma a)^2(.,\sigma_0)}{a^2(.,\sigma_0)}\right)\right\}^{-1}\right)$ consists in showing the two following points.

1) We have the convergence in law:

$$\mathcal{D}_n \xrightarrow[\mathcal{D}]{n\to\infty} \mathcal{N}\left(0, \begin{bmatrix} 9\nu_0\left(\frac{(\partial_\mu b)^2(.,\mu_0)}{a^2(.,\sigma_0)}\right) & 0 \\ 0 & 9\nu_0\left(\frac{(\partial_\sigma a)^2(.,\sigma_0)}{a^2(.,\sigma_0)}\right) \end{bmatrix}\right)$$

2) We have the uniform (with respect to θ) convergence in probability:

$$\mathcal{C}_n(\theta) \xrightarrow{n\to\infty} \begin{bmatrix} \mathcal{C}_{1,1}(\theta) & 0 \\ 0 & \mathcal{C}_{2,2}(\theta) \end{bmatrix}$$

with

$$\mathcal{C}_{1,1}(\theta) = 3\nu_0\left(\frac{(\partial_\mu b)^2(.,\mu)}{a^2(.,\sigma)} + \frac{\partial^2_{\mu^2} b}{a^2}(.,\theta)(b(.,\mu) - b(.,\mu_0))\right)$$

$$\mathcal{C}_{2,2}(\theta) = \nu_0\left((\partial_\sigma a)^2(.,\sigma)\left(\frac{6a^2(.,\sigma_0)}{a^4(.,\sigma)} - \frac{2}{a^2(.,\sigma)}\right)\right) + \nu_0\left(2\partial^2_{\sigma^2}a(.,\sigma)\left(\frac{1}{a(.,\sigma)} - \frac{a^2(.,\sigma_0)}{a^3(.,\sigma)}\right)\right).$$

Indeed, this second point immediately implies, using the consistency of $\hat{\theta}_n$,

$$\int_0^1 \mathcal{C}_n(\theta_0 + u(\hat{\theta}_n - \theta_0))du \xrightarrow[\mathbf{P}]{n\to\infty} \begin{bmatrix} 3\nu_0\left(\frac{(\partial_\mu b)^2(.,\mu_0)}{a^2(.,\sigma_0)}\right) & 0 \\ 0 & 4\nu_0\left(\frac{(\partial_\sigma a)^2(.,\sigma_0)}{a^2(.,\sigma_0)}\right) \end{bmatrix}.$$

For 1) we remark that $\overline{\nu}_n$, $\overline{\mathcal{I}}_n$ and $\overline{\mathcal{Q}}_n$ are linear functionals, therefore we can derivate the expression of $n^{-1}\mathcal{L}_n$ given in the proof of Theorem 5.1 with respect to the parameter μ and get (recall $h(x,\theta) = (\partial_x \frac{b}{a^2})(x,\theta)$),

$$(n\Delta_n)^{-\frac{1}{2}}\partial_\mu\mathcal{L}_n(\theta_0) = -3\overline{N}_n\left(\frac{\partial_\mu b}{a^2}(.,\theta_0)\right). \tag{61}$$

Analogously, we get

$$n^{-\frac{1}{2}}\partial_\sigma\mathcal{L}_n(\theta_0) = -2\overline{M}_n\left(\frac{\partial_\sigma a}{a^3}(.,\theta_0)\right) + \sqrt{n}\Delta_n\left\{3\overline{\mathcal{I}}_n\left(\partial_\sigma\left(\frac{b}{a^2}\right)(.,\theta_0)\right) + \frac{3}{4}\overline{\mathcal{Q}}_n\left(\partial_\sigma h(.,\theta_0)\right) - \frac{3}{2}\overline{\nu}_n\left(\partial_\sigma(a^{-2})(.,\sigma_0)b^2(.,\mu_0)\right)\right\}$$

By Proposition 2.1, Theorems 2.2, 2.4 and $n\Delta_n^2 \to 0$, this yields

$$n^{-\frac{1}{2}}\partial_\sigma \mathcal{L}_n(\theta_0) = -2\overline{M}_n(\frac{\partial_\sigma a}{a^3}(.,\theta_0)) + o_{\mathbf{P}}(1). \quad (62)$$

Now, 1) follows from (61)–(62) and Theorem 3.5.

To obtain 2), we derivate twice $\mathcal{L}_n$ and use results of Section 2. □

Remark 5.3. *As for estimation based on $(X_{i\Delta_n})_{0\le i\le n-1}$, the rate of convergence is different for $\hat{\mu_n}$ and $\hat{\sigma_n}$. The drift term is estimated with rate $(n\Delta_n)^{\frac{1}{2}}$ and the diffusion term is estimated with rate $n^{\frac{1}{2}}$.*

Comparing with the asymptotic variance of the estimator based on the Euler contrast (Kessler (1997), when $X_{i\Delta_n}$ itself is observed), we notice a slight increase in the variance of the estimator of the diffusion term (the constant $\frac{9}{16}$ instead of $\frac{1}{2}$ for Kessler (1997)). The estimation of μ is asymptotically efficient since $\nu_0\left(\frac{(\partial_\mu b)^2}{a^2}(.,\theta_0)\right)$ is the Fisher information of the continuous time model.

6 Examples of parametric models

We apply our statistical results to some classical models.

6.1 Example 1: Ornstein–Uhlenbeck process

The diffusion solves $dX_t = \mu X_t dt + \sigma dB_t$, with $\mu < 0$, $\sigma > 0$ and X_0 is determistic or has for distribution the stationary probability of X.

Here, we can compute explicitely the estimator $\hat{\theta}_n$ by minimizing the contrast (6). We find,

$$\begin{aligned}
\hat{\sigma}_n^2 &= \frac{3}{2}(n\Delta_n)^{-1}\sum_{i=0}^{n-1}\left(\overline{X}_{i+1}-\overline{X}_i\right)^2 \\
\hat{\mu}_n &= \frac{\Delta_n^{-1}\sum_{i=0}^{n-1}\left(\overline{X}_{i+1}-\overline{X}_i\right)\overline{X}_i}{\sum_{i=0}^{n-1}\left(\overline{X}_i\right)^2} - \frac{1}{4}\frac{\Delta_n^{-1}\sum_{i=0}^{n-1}\left(\overline{X}_{i+1}-\overline{X}_i\right)^2}{\sum_{i=0}^{n-1}\left(\overline{X}_i\right)^2}
\end{aligned}$$

(We have dropped useless terms in $\hat{\sigma}_n$).

Using results of Section 4.1, (A1)–(A5) are valid with $K_{-\infty} = \infty$, and any positive constants $M_{-\infty}$ M_∞. Thus, results of Section 5 apply:

$$(\hat{\mu}_n, \hat{\sigma^2}_n) \xrightarrow[\mathbf{P}]{n\to\infty} (\mu, \sigma^2)$$

$$\text{and if } n\Delta_n^2 \to 0, \quad \begin{bmatrix}(n\Delta_n)^{\frac{1}{2}}(\hat{\mu}_n - \mu)\\ n^{\frac{1}{2}}(\hat{\sigma}_n^2 - \sigma^2)\end{bmatrix} \xrightarrow[\mathcal{D}]{n\to\infty} \mathcal{N}\left(0, \begin{bmatrix}2|\mu| & 0\\ 0 & \frac{9}{4}\sigma^4\end{bmatrix}\right).$$

Using numerical simulations, we see that our estimator gives good results on finite samples (see Table 1, where we give mean and variance of $\hat{\theta_n}$ for different values of n and $T = n\Delta_n$ when $-\mu = \sigma^2 = 1$). It is worth noticing that if Δ is not small enough then we underestimate both σ^2 and $|\mu|$.

Table 1: Simulation results for an Ornstein–Uhlenbeck process.

n=10000,T=500,$\Delta = 1/20$		n=10000,T=20,$\Delta = 1/500$	
Mean $\hat{\sigma^2}_n = 0.9659$	n.Var $\hat{\sigma^2}_n$=2.38	Mean $\hat{\sigma^2}_n = 0.9981$	n.Var $\hat{\sigma^2}_n$=1.8677
Mean $\hat{\mu}_n$ =-0.9833	T.Var $\hat{\mu}_n$=2.11	Mean $\hat{\mu}_n$ =-1.1101	T.Var $\hat{\mu}_n$=2.08
n=2000,T=200,$\Delta = 1/10$		n=1000,T=500,$\Delta = 1/2$	
Mean $\hat{\sigma^2}_n = 0.9259$	n.Var $\hat{\sigma^2}_n$=1.64	Mean $\hat{\sigma^2}_n = 0.6986$	n.Var $\hat{\sigma^2}_n$=0.47
Mean $\hat{\mu}_n$ =-0.9835	T.Var $\hat{\mu}_n$=2.83	Mean $\hat{\mu}_n$ =-0.83087	T.Var $\hat{\mu}_n$=1.98

6.2 Example 2: The Cox–Ingersoll–Ross process

The diffusion solves (54) and starts either from the stationary distribution or from a deterministic variable. We find the following explicit expressions for the estimator $\hat{\theta}_n = (\hat{\mu}_n, \hat{\mu}'_n, \hat{\sigma}^2_n)$ by minimizing the contrast (6):

$$\begin{bmatrix} \Delta_n \sum_{i=0}^{n-1} \overline{X}_i & n\Delta_n \\ n\Delta_n & \Delta_n \sum_{i=0}^{n-1} \overline{X}_i^{-1} \end{bmatrix} \begin{bmatrix} \hat{\mu}_n \\ \hat{\mu'}_n \end{bmatrix} = \begin{bmatrix} \sum_{i=0}^{n-1} \{\overline{X}_{i+1} - \overline{X}_i\} \\ \sum_{i=0}^{n-1} \left\{ \overline{X}_i^{-1} (\overline{X}_{i+1} - \overline{X}_i) + \frac{1}{4} \overline{X}_i^{-2} (\overline{X}_{i+1} - \overline{X}_i)^2 \right\} \end{bmatrix}$$

$$\hat{\sigma}^2_n = \frac{3}{2} (n\Delta_n)^{-1} \sum_{i=0}^{n-1} \overline{X}_i^{-1} \left(\overline{X}_{i+1} - \overline{X}_i\right)^2$$

(We have dropped a useless term in $\hat{\sigma^2}_n$)

Results of Section 5 do not apply since here, $K_0 = c_0 - 1 = \frac{2\mu'}{\sigma^2} - 1 < \infty$ and we can only choose $M_0 < c_0$ (see section 4.3). We directly show the convergence of $\hat{\theta}_n$ using its expression and applying results of Sections 2 and 3. For this we have to take care that K_0 and M_0 are large enough. After some easy computation we get,

Theorem 6.1. • *If $c_0 > 9$, then $\hat{\theta}_n \xrightarrow[\mathbf{P}]{n \to \infty} \theta_0$.*

• *If $c_0 > 13$ and $n\Delta_n^2 \to 0$, then*

$$\begin{bmatrix} (n\Delta_n)^{\frac{1}{2}} (\hat{\mu}_n - \mu) \\ (n\Delta_n)^{\frac{1}{2}} (\hat{\mu'}_n - \mu') \\ n^{\frac{1}{2}} (\hat{\sigma^2}_n - \sigma^2) \end{bmatrix} \xrightarrow[\mathcal{D}]{n \to \infty} \mathcal{N} \left(0, \begin{bmatrix} 2|\mu| & \sigma^2 - 2\mu' & 0 \\ \sigma^2 - 2\mu' & (2\mu' - \sigma^2) \frac{\mu'}{|\mu|} & 0 \\ 0 & 0 & \frac{9}{4}\sigma^4 \end{bmatrix} \right)$$

6.3 Example 3: Bilinear diffusion

The diffusion X solves (55), X_0 is either a deterministic variable either the stationary law. Our estimators have the following expressions:

$$\begin{bmatrix} n\Delta_n & \Delta_n \sum_{i=0}^{n-1} \overline{X}_i^{-1} \\ \Delta_n \sum_{i=0}^{n-1} \overline{X}_i^{-1} & \Delta_n \sum_{i=0}^{n-1} \overline{X}_i^{-2} \end{bmatrix} \begin{bmatrix} \hat{\mu}_n \\ \hat{\mu'}_n \end{bmatrix} = \begin{bmatrix} \sum_{i=0}^{n-1} \left\{ \overline{X}_i^{-1} (\overline{X}_{i+1} - \overline{X}_i) + \frac{1}{4} \overline{X}_i^{-2} (\overline{X}_{i+1} - \overline{X}_i)^2 \right\} \\ \sum_{i=0}^{n-1} \left\{ \overline{X}_i^{-2} (\overline{X}_{i+1} - \overline{X}_i) + \frac{1}{2} \overline{X}_i^{-3} (\overline{X}_{i+1} - \overline{X}_i)^2 \right\} \end{bmatrix}$$

$$\hat{\sigma}^2_n = \frac{3}{2} (n\Delta_n)^{-1} \sum_{i=0}^{n-1} \overline{X}_i^{-2} \left(\overline{X}_{i+1} - \overline{X}_i\right)^2$$

By Section 4.4, (A1)–(A5) holds, with $K_0 = \infty$, with any constants $M_0 \in [0,\infty), M_\infty \in [0, c_0)$ (recall, here, $c_0 = 1 + \frac{2|\mu|}{\sigma^2}$).

We show that $\hat\theta_n$ is consistent and asymptotically normal by results of Section 2 and 3.

Theorem 6.2.

- *If $c_0 > 4$, then $\hat\theta_n \xrightarrow[\mathbf{P}]{n\to\infty} \theta_0$.*
- *If $c_0 > 8$ and $n\Delta_n^2 \to 0$, then*

$$\begin{bmatrix} (n\Delta_n)^{\frac{1}{2}}(\hat\mu_n - \mu) \\ (n\Delta_n)^{\frac{1}{2}}(\hat{\mu'}_n - \mu') \\ n^{\frac{1}{2}}(\hat{\sigma^2}_n - \sigma^2) \end{bmatrix} \xrightarrow[\mathcal{D}]{n\to\infty} \mathcal{N}\left(0, \begin{bmatrix} 2|\mu| + 2\sigma^2 & -2\mu' & 0 \\ -2\mu' & \frac{4\mu'^2}{\sigma^2 + 2|\mu|} & 0 \\ 0 & 0 & \frac{9}{4}\sigma^4 \end{bmatrix}\right)$$

7 Appendix

7.1 Technical lemmas

The following Lemma precises Lemma 8 in Kessler (1997).

Lemma 7.1. *Let $f \in \mathcal{C}^1((l,r)\times\Theta)$, satisfy*

$$\sup_{\theta\in\Theta}\{|f(x,\theta)| + |f'_x(x,\theta)| + |\nabla_\theta f(x,\theta)|\} \le c(\mathcal{B}_l^\gamma(x) + \mathcal{B}_r^{\gamma'}(x)),$$

with $\gamma \le M_l$, $1+\gamma' \le M_r$ and $\gamma < K_l$ then:

$$n^{-1}\sum_{i=0}^{n-1} f(X_{i\Delta_n},\theta) \xrightarrow{n\to\infty} \nu_0(f(.,\theta)) \text{ uniformly in } \theta\text{, in probability.} \tag{63}$$

Proof. Using Assumption (A4) and the ergodic theorem for X, we get: $(n\Delta_n)^{-1}\int_0^{n\Delta_n} f(X_s,\theta)ds \xrightarrow[\mathbf{P}]{n\to\infty} \nu_0(f(.,\theta))$.

Denote for the proof,

$$D_n(\theta) = (n\Delta_n)^{-1}\int_0^{n\Delta_n} f(X_s,\theta)ds - n^{-1}\sum_{i=0}^{n-1} f(X_{i\Delta_n},\theta) = (n\Delta_n)^{-1}\sum_{i=0}^{n-1}\int_{i\Delta_n}^{(i+1)\Delta_n}\{f(X_s,\theta) - f(X_{i\Delta_n},\theta)\}ds.$$

Using Proposition 1.1 *1)* and (A5) yields $\sup_{s\in[i\Delta_n,(i+1)\Delta_n]} E(|f(X_s,\theta) - f(X_{i\Delta_n},\theta)|) \le c\Delta_n^{\frac{1}{2}}$. We deduce that $D_n(\theta) \xrightarrow{n\to\infty} 0$ in $\mathbf{L}^1$, which gives the convergence in (63) for all θ.

To get the uniformity in θ, by Proposition 7.2, it suffices to show:

$$\sup_{n\ge0} E\left(\left|\sup_\theta \frac{1}{n}\sum_{i=0}^{n-1}\nabla_\theta f(X_{i\Delta_n},\theta)\right|\right) < \infty$$

Using $\sup_\theta |\nabla_\theta f|(x,\theta) \le c(B_l^{M_l}(x) + \mathcal{B}_r^{M_r}(x))$ and (A5) yields the result. □

Proposition 7.2. *Let $S_n(\omega, \theta)$ be a sequence of measurable real valued functions defined on $\Omega \times \Theta$, where $(\Omega, \mathcal{F}, P)$ is a probability space, and Θ is product of compact intervals of $\mathbb{R}$. We assume that $S_n(.,\theta)$ converges to zero in probability for all $\theta \in \Theta$; and that there exists an open neighboorhood of Θ on which $S_n(\omega, .)$ is continuously differentiable for all $\omega \in \Omega$. Furthermore we suppose that $\sup_{n\in\mathbb{N}} E\left(\sup_{\theta\in\Theta} |\nabla_\theta S_n(\theta)|\right) < \infty$, then*

$$S_n(\theta) \xrightarrow{n\to\infty} 0 \text{ uniformly in } \theta, \text{ in probability}$$

Proof. Let $\epsilon > 0$ and $\eta > 0$, let us show that for n large enough: $P(\sup_{\theta\in\Theta} |S_n(\theta)| > \eta) < \epsilon$. Denote $Z_n = \sup_{\theta\in\Theta} |\nabla_\theta S_n(\theta)|$, and let M such that $\sup_{n\in\mathbb{N}} E(Z_n) M^{-1} < \frac{\epsilon}{3}$.

Using that Θ is compact we can find an integer d and $(\alpha_1, \ldots, \alpha_d) \in \Theta^d$, such that for all θ in Θ: $\inf_{i\in\{1,\ldots,d\}} |\alpha_i - \theta| < \frac{\eta}{2M}$. Define n_0 such that, $n \geq n_0$ implies $P\left(|S_n(\alpha_i)| > \frac{\eta}{2}\right) < \frac{\epsilon}{3d}$. Using that Θ is convex:

$$\sup_{\theta\in\Theta} |S_n(\theta)| \leq Z_n \frac{\eta}{2M} + \sup_{i=1,\ldots,d} |S_n(\alpha_i)|$$

We deduce:

$$\begin{aligned} P(\sup_{\theta\in\Theta} |S_n(\theta)| \geq \eta) &\leq P(Z_n \frac{\eta}{2M} \geq \frac{\eta}{2}) + P(\sup_{i=1,\ldots,d} |S_n(\alpha_i)| \geq \frac{\eta}{2}) \\ &\leq \frac{E(Z_n)}{M} + \sum_{i=0}^{d} P\left(|S_n(\alpha_i)| > \frac{\eta}{2}\right) \leq \frac{\epsilon}{3} + \frac{\epsilon}{3} d^{-1} d < \epsilon. \end{aligned}$$

□

We recall, the useful lemma which is given in Genon-Catalot and Jacod (1993).

Lemma 7.3. *Let χ_i^n, U be random variables, with χ_i^n being $\mathcal{G}_i^n$-measurable. The following two conditions imply $\sum_{i=1}^n \chi_i^n \xrightarrow{\mathbf{P}} U$:*

$$\sum_{i=1}^{n} E(\chi_i^n \mid \mathcal{G}_{i-1}^n) \xrightarrow{\mathbf{P}} U$$

$$\sum_{i=1}^{n} E((\chi_i^n)^2 \mid \mathcal{G}_{i-1}^n) \xrightarrow{\mathbf{P}} 0$$

7.2 Details on the proof of Theorem 5.1

To maintain a short proof of consistency, we have asserted that, according to Kessler (1997), the following conditions are sufficient to ensure the consistency of $\hat{\theta}_n = (\hat{\mu}_n, \hat{\sigma}_n) = \text{arginf}_{(\mu,\sigma)\in\Theta_1\times\Theta_2} \mathcal{L}_n(\mu, \sigma)$ (see the proof of Theorem 5.1).

- The parameter set $\Theta = \Theta_1 \times \Theta_2$ is compact.
- The convergence, uniformly in (μ, σ), in probability (see (57))

$$n^{-1}\mathcal{L}_n(\mu, \sigma) \xrightarrow{n\to\infty} K_1(\sigma_0, \sigma) \tag{64}$$

  where $\sigma \to K_1(\sigma_0, \sigma)$ is a continuous function and has a unique minimum at $\sigma = \sigma_0$.

- The convergence, uniformly in (μ, σ), in probability (see (58))

$$(n\Delta_n)^{-1}(\mathcal{L}_n(\mu,\sigma) - \mathcal{L}_n(\mu_0,\sigma)) \xrightarrow{n\to\infty} K_2(\mu_0,\sigma_0,\mu,\sigma) \tag{65}$$

where $K_2(\mu_0, \sigma_0, ., .)$ is continuous and $K_2(\mu_0, \sigma_0, ., \sigma_0)$ is a non negative function with a unique minimum, equal to 0, reached at $\mu = \mu_0$.

This proof of consistency is non standard, due to the fact that μ and σ are estimated with different rate, and we do not know any other references than Kessler (1997) dealing with such a problem. Hence, we give more details, here, for the proof.

First, we show the consistency of $\hat{\sigma_n}$. For $\eta > 0$,

$$\{|\hat{\sigma}_n - \sigma_0| \geq \eta\} \subset \left\{ \inf_{(\mu,\sigma)\in\Theta_1\times B^c(\sigma_0,\eta)} \mathcal{L}_n(\mu,\sigma) = \inf_{(\mu,\sigma)\in\Theta_1\times\Theta_2} \mathcal{L}_n(\mu,\sigma) \right\} \subset A_n$$

where $B^c(\sigma_0, \eta) = \Theta_2 \cap \{|\sigma - \sigma_0| \geq \eta\}$ and $A_n = \left\{\inf_{(\mu,\sigma)\in\Theta_1\times B^c(\sigma_0,\eta)} n^{-1}(\mathcal{L}_n(\mu,\sigma) - \mathcal{L}_n(\mu,\sigma_0)) \leq 0\right\}$.

But, by uniform convergence of $n^{-1}(\mathcal{L}_n(\mu,\sigma) - \mathcal{L}_n(\mu,\sigma_0))$ to $K_1(\sigma_0,\sigma) - K_1(\sigma_0,\sigma_0)$, we deduce $\overline{\lim}_{n\to\infty} P(A_n) \leq P(A)$ where

$$A = \left\{ \inf_{(\mu,\sigma)\in\Theta_1\times B^c(\sigma_0,\eta)} K_1(\sigma_0,\sigma) - K_1(\sigma_0,\sigma_0) \leq 0 \right\}.$$

The unique minimum of $K_1(\sigma_0, \sigma)$ is reached at $\sigma = \sigma_0$. Hence $P(A) = 0$ and we have proved

$$\overline{\lim}_{n\to\infty} P(|\hat{\sigma}_n - \sigma_0| \geq \eta) = 0.$$

The second step of the proof is the consistency of $\hat{\mu}_n$. Let $h, \eta > 0$, then

$$\{|\hat{\mu}_n - \mu_0| \geq h\} \cap \{|\hat{\sigma}_n - \sigma_0| < \eta\} \subset B_n$$

where $B_n = \left\{\inf_{(\mu,\sigma)\in B^c(\mu_0,h)\times B(\sigma_0,\eta)} (n\Delta_n)^{-1}(\mathcal{L}_n(\mu,\sigma) - \mathcal{L}_n(\mu_0,\sigma)) \leq 0\right\}$ and $B^c(\mu_0,h)\times B(\sigma_0,\eta) = \Theta \cap \{|\mu - \mu_0| \geq h,\ |\sigma - \sigma_0| < \eta\}$.

But, by (65), $\overline{\lim}_{n\to\infty} P(B_n) \leq P(B)$, with

$$B = \left\{ \inf_{(\mu,\sigma)\in B^c(\mu_0,h)\times B(\sigma_0,\eta)} K_2(\mu_0,\sigma_0,\mu,\sigma) \leq 0 \right\}.$$

Using that $\inf_{\mu\in B^c(\mu_0,h)} K_2(\mu_0,\sigma_0,\mu,\sigma_0) > 0$, we get by compacity of Θ, for h small enough $P(B) = 0$. This implies

$$\overline{\lim}_{n\to\infty} P(|\hat{\mu}_n - \mu_0| \geq h,\ |\hat{\sigma}_n - \sigma_0| < \eta) = 0.$$

To conclude, we write

$$P(|\hat{\mu}_n - \mu_0| \geq h) \leq P(|\hat{\mu}_n - \mu_0| \geq h,\ |\hat{\sigma}_n - \sigma_0| < \eta) + P(|\hat{\sigma}_n - \sigma_0| \geq \eta)$$

and the consistency of $\hat{\mu}_n$ follows.

References

- Barndorff-Nielsen Ole E. (1998). Processes of normal inverse gaussian type. *Finance Stoch*, **2**(1), 41–68.

- Barndorff-Nielsen O.E. and Shephard N. (1998). Aggregation and model construction for volatility models. Working paper series no.**10**, Center for Analytical Finance, University of Aarhus.

- Bibby B.M. and Sørensen M. (1995). Martingale estimations functions for discretely observed diffusion processes. *Bernoulli*, **1**, 17–39.

- Bibby B.M. and Sørensen M. (1998). Simplified Esimating Functions for diffusion Models with a High–dimensional Parameter Preprint No. **10**, Dept. of Theor. Statist. Univ. of Copenhagen.

- Delattre S. and Jacod J. (1997). A central limit theorem for normalized functions of the increments of a diffusion process, in the presence of round–off errors. *Bernoulli,* **3***(1)*, 1–28.

- Dorogovtsev A. (1976). The consistency of an estimate of a parameter of a stochastic differential equation. *Theory Probab. Math. Stat.*, **10**, 73–82.

- Florens-Zmirou D. (1989). Approximate discrete–time schemes for statistics of diffusion processes. *Statistics*, **20**(4), 547–557.

- Genon-Catalot V. and Jacod J. (1993). On the estimation of the diffusion coefficient for multidimensional diffusion processes. *Annales de l'I.H.P.,Probabilité-Statistiques*, **29**, 119–151.

- Genon-Catalot V., Jeantheau T. and Laredo C. (1998). Parameter estimation for a discretely observed stochastic volatility models. *Bernoulli*, **4**(3), 283–303.

- Gloter A. (1998, a). Parameter estimation for a discrete sampling of an integrated Ornstein-Uhlenbeck process. *Prépublication de l'université de Marne–la–Vallée*, **14/98**.

- Gloter A. (1998, b). Discrete sampling of an integrated diffusion process and a comparison with the Euler scheme. *Prépublication de l'université de Marne–la–Vallée*, **23/98**.

- Hall P. and Heyde C.C. (1980) *Martingale limit theory and its application.* Academic Press.

- Ibragimov and Khas'minskii R.Z. (1981). *Statistical Estimation, Asymptotic Theory.* Springer-Verlag.

- Kessler M. (1996). Simple and explicit estimating functions for a discretely observed diffusion process. *Research Report No.* **336**, *Dept. Theor. Statist. Univ. of Aarhus.*

- Kessler M. (1997). Estimation of an ergodic diffusion from discrete observations. *Scandinavian Journal of Statistics No.* **24***(2)*, 211–229.

- Kessler M. and Sørensen M. (1999). Estimating equations based on eigenfunctions for a discretely observed diffusion process. *Research Report No.* **332**, *Dept. Theor. Statist. Univ. of Aarhus. To appear in Bernoulli.*

- Leblanc B. (1996). Une aproche unifiée pour une forme exacte du prix d'une option dans les différents modèles a volatilité stochastique. *Stochastics Stochastics Rep.*, **57**(1-2), 1–35.

- Pastorello S., Renault E. and Touzi N. (1994). Statistical inference for random variance option pricing. *Southern European Economics Discussion Series*, D.P.136.

- Sørensen M. (1998). On asymptotics of estimating functions. *Preprint No.* **6**, *Dept. of Theor. Statist. Univ. of Copenhagen.*

- Taylor S.J. and Xu X. (1995). The incremental volatility information in one million foreign exchange quotations. *Preprint, Conference in high frequency data in finance, Zürich.*

Partie II. Modèles à volatilité stochastique

Chapitre II.1. Discrétisation du modèle à volatilité stochastique sur un intervalle de temps fini, approximation de la volatilité intégrée et applications statistiques.

Discrete sampling of a stochastic volatility model on a fixed length time interval, approximations of the integrated volatility and statistical applications.

Abstract

This paper deals with some approximation and estimation problems for stochastic volatility models. Consider a two-dimensional diffusion process (Y_t, X_t), where X_t is a positive diffusion and Y_t has diffusion coefficient $X_t^{\frac{1}{2}}$. Assume that only (Y_t) is observed according to a double discretization, at times $(i + \frac{j}{m})\Delta_n$, $i \leq n$, $j \leq m$, for some small sampling interval Δ_n. Within each subinterval $[i\Delta_n, (i+1)\Delta_n]$, the quadratic variation of the observations defines an approximation of the integral $\int_{i\Delta_n}^{(i+1)\Delta_n} X_s ds$, which, in turn, can be compared with $X_{i\Delta_n}$.

We explore the properties of these approximations. When the diffusion coefficient of X_t depends on a linear unknown parameter θ, we construct an explicit estimator of θ^2 and prove its consistency and asymptotic normality when $\Delta_n = \frac{1}{n}$, as n and m tend to infinity simultaneously. Examples and numerical simulation results are given.

The continuous time stochastic volatility models have been the subject of many recent contributions. These models, introduced by Hull and White ([9]) assume that, if $Y_t = \log S_t$ is the logarithm of the price process (S_t) of some asset, then:

$$dY_t = \rho(\sigma_t^2, t)dt + \sigma_t dW_t. \tag{1}$$

where $X_t = \sigma_t^2$ is the so–called stochastic volatility and follows a stochastic differential equation driven by another Brownian motion:

$$dX_t = b(X_t)dt + a(X_t)dB_t. \tag{2}$$

In (1) and (2), the Brownian motions B and W may be correlated.

Such models have raised several new problems in the field of finance and of statistics of random processes (see e.g. for a review of these models, Ghysels and al. [4], for related models and statistical issues, Barndorff–Nielsen and Shephard [1], Drost and Werker [3], Genon–Catalot and al. [7], Sørensen [13]).

First, the computation of option prices based on the asset (S_t) depends on the exact distribution of the quadratic variation process $< Y >_t = \int_0^t X_s ds$, which is not easily tractable (see e.g. Leblanc [10] and references therein). Moreover, the volatility being unobservable, approximations or estimations have to rely on a discrete sampling of the process (Y_t) only.

In a previous paper (Gloter [8]), we have obtained approximations of the distribution of $(J_i^n)_{i \geq 0}$ with:

$$J_i^n = \int_{i\Delta_n}^{(i+1)\Delta_n} X_s ds,$$

for small Δ_n. In particular, $(\Delta_n^{-1} J_i^n)$ is compared with $(X_{i\Delta_n})$.

Here, we give further approximations of (J_i^n) based on a discrete sampling of (Y_t). For this, we assume that the sample path (Y_t) is observed, according to a double discretization, at times $(i + \frac{j}{m})\Delta_n$, $i = 0, 1, \ldots, n-1$, $j = 0, 1, \ldots, m$. The appoximation of J_i^n is naturally given by:

$$\widehat{J}_i^{n,m} = \sum_{j=0}^{m-1} (Y_{(i+\frac{j+1}{m})\Delta_n} - Y_{(i+\frac{j}{m})\Delta_n})^2$$

We study the exact rate of the difference $\Delta_n^{-1}(\widehat{J}_i^{n,m} - J_i^n)$. In relation with our work we must quote the papers of Nelson [11] and Foster and Nelson [12]. There, a double discretization as above is used for filtering purposes, i.e. volatility prediction. Our concern is different. We have in view statistical applications to the estimation of unknown parameters in the volatility model (2) based on a discrete sampling of (Y_t) within a fixed length time interval. For this, we set $\Delta_n = \frac{1}{n}$ and consider the equally spaced observations $(Y_{\frac{i}{n}+\frac{j}{mn}}, i = 0, \ldots, n-1, j = 0, \ldots, m)$ of (Y_t) on $[0, 1]$.

With such an observation, we can only expect consistent estimators of the diffusion coefficient of model (2). Indeed, if we had a discrete observation $(X_{\frac{i}{n}}, i = 0, 1, \ldots, n)$ of the volatility itself, then well known results show that drift parameters cannot be estimated, but diffusion coefficient parameters can be estimated without any knowledge of the drift (see e.g. Dohnal [2], Genon–Catalot and Jacod [6]). In particular if (X_t) has the special form:

$$dX_t = b(X_t)dt + \theta a(X_t)dB_t \tag{3}$$

with $b(.)$ unknown and $a(.)$ known, then the classical (and optimal) estimator of θ^2 is given by:

$$\theta_n^{*2} = \sum_{i=0}^{n-1} \frac{(X_{\frac{i+1}{n}} - X_{\frac{i}{n}})^2}{a^2(X_{\frac{i}{n}})} \tag{4}$$

It is consistent as $n \to \infty$ and, due to the multiplicative form of the diffusion coefficient, satisfies:

$$\sqrt{n}(\theta_n^{*2} - \theta^2) \xrightarrow[\mathcal{D}]{n\to\infty} \mathcal{N}(0, 2\theta^4). \tag{5}$$

In the case of model (3), we construct an estimator $\hat{\theta}^2_{n,m}$ of θ^2 based on $(Y_{\frac{i}{n}+\frac{j}{mn}})_{\substack{0\le i\le n-1\\ 0\le j\le m}}$, expressed as a function of $(n\widehat{J_i}^{n,m})$. Its asymptotic properties as $n \to \infty$ and $m = m_n \to \infty$ are investigated and we specify the relationship between m_n and n ensuring consistency and asymptotic normality with rate $\sqrt{n}$. Setting $N = m_n n + 1$, we can interpret this rate of convergence in terms of N, the total number of observations. A crucial assumption here is that (X_t) must be positive.

The paper is organized as follows. In Section 1 we present the model, the assumptions and recall some expansions of $\Delta_n^{-1} J_i^n$ obtained in Gloter [8] which are needed for the sequel (Proposition 1.2 and Theorem 1.3).

Section 2 is devoted to the expansion of $\Delta_n^{-1}\widehat{J_i}^{n,m}$ in terms of $\Delta_n^{-1}J_i^n$ and $X_{i\Delta_n}$. Distinct results are obtained whether the two Brownians motions are correlated or not (Theorem 2.1 and Theorem 2.2). We end this section with an application, when $\Delta_n = n^{-1}$, to the approximated quadratic variation of $(\Delta_n^{-1}J_i^n)$, $\widehat{V}_n = \sum_{i=0}^{n-2}(\widehat{nJ_{i+1}}^{n,m_n} - n\widehat{J_i}^{n,m_n})^2$. We show that $\widehat{V}_n$ behaves differently according to the rate of convergence of m_n to ∞, as $n \to \infty$, and give the possible limits.

In Section 3, we study some preliminary steps for the statistical application. We obtain bounds for the conditional moments of $\Delta_n^{-1}\widehat{J_i}^{n,m}$ of positive and negative orders. With $\mathcal{G}_i^n = \sigma(W_s, B_s, s \le i\Delta_n, Y_0, X_0)$, we prove that $E\left((\Delta_n^{-1}\widehat{J_i}^{n,m})^k \mid \mathcal{G}_i^n\right)$ can be bounded for $k \ge 0$, (resp. $k \le 0$), uniformly in m by a polynomial function of $X_{i\Delta_n}$ (resp. $X_{i\Delta_n}^{-1}$). The main difficulty is for $k \le 0$. As a key tool, we obtain the following general result which has its own interest. If $(Z_i, i \ge 1)$ are independent real random variables, with densities $\phi_i(x)$ uniformly bounded by M on the neighbourhood of 0, $[-\eta, \eta]$, then for $k \ge 0$ and $m \ge 2k+3$,

$$E\left((Z_1^2 + \cdots + Z_m^2)^{-k}\right) \le c(k, M, \eta)m^{-k},$$

where $c(k, M, \eta)$ is a constant independent of m. It is worth noting that the above bound is non asymptotic and sharp since we can compare the bound with the exact value when the Z_i's are $\mathcal{N}(0,1)$.

Section 4 is devoted to the estimation of θ^2 in model (3). The estimator is given by:

$$\hat{\theta}^2{}_{n,m} = \frac{3}{2}\sum_{i=0}^{n-2}\left\{\frac{(\widehat{nJ_{i+1}}^{n,m} - n\widehat{J_i}^{n,m})^2}{a^2(n\widehat{J_i}^{n,m})} - \frac{4}{m}\frac{(n\widehat{J_i}^{n,m})^2}{a^2(n\widehat{J_i}^{n,m})}\right\}$$

First, we prove that

$$\tilde{\theta^2}_n = \sum_{i=0}^{n-1} \frac{(nJ_{i+1}^n - nJ_i^n)^2}{\frac{2}{3}a^2(nJ_i^n)}$$

is such that $\sqrt{n}(\tilde{\theta}_n^2 - \theta^2) \xrightarrow[\mathcal{D}]{n\to\infty} \mathcal{N}(0, \frac{9}{4}\theta^4)$. This result has to be compared with (4)–(5): a correction factor $\frac{2}{3}$ is needed for the change of $X_{\frac{i}{n}}$ by nJ_i^n and there is a slight increase of asymptotic variance.

Then, a further bias correction is needed for the change of nJ_i^n by $n\widehat{J}_i^{n,m}$. The asymptotic properties of $\widehat{\theta^2}_{n,m}$ are obtained as $n \to \infty$ and $m = m_n \to \infty$. We prove that if $n^{\frac{2}{3}}/m_n \to 0$, $\widehat{\theta^2}_{n,m_n}$ is consistent.

If, further $n/m_n \to 0$, then $\sqrt{n}(\widehat{\theta^2}_{n,m_n} - \theta^2) \xrightarrow{n\to\infty} \mathcal{N}\left(0, \frac{9}{4}\theta^4\right)$.

And if $m_n = n$, we obtain the tightness of the sequence $\sqrt{n}(\widehat{\theta^2}_{n,n} - \theta^2)$. So, setting $N = nm_n + 1$ for the total number of observations, we get a rate of convergence slower than $N^{\frac{1}{4}}$ or exactly $N^{\frac{1}{4}}$ when $m_n = n$.

Finally, we study through numerical simulations the estimation of θ^2 when $X_t = e^{\theta B_t}$.

In the Appendix, some auxiliary results are proved.

1 Framework and preliminary results

1.1 Model, notations and assumptions

Let (Y_t, X_t) be the two-dimensional diffusion process defined as the solution on a probability space $(\Omega, \mathcal{A}, P)$ of

$$dY_t = \rho(X_t, t)dt + \sigma_t dW_t, \qquad Y_0 = \eta' \tag{6}$$

$$\sigma_t^2 = X_t, \quad dX_t = b(X_t)dt + a(X_t)dB_t, \qquad X_0 = \eta \tag{7}$$

where $(B_t)_{t\geq 0}$ and $(W_t)_{t\geq 0}$ are two possibly correlated Brownian motions. We write

$$W_t = \chi B_t + \sqrt{1-\chi^2}\tilde{W}_t$$

where $(B_t, \tilde{W}_t)_{t\geq 0}$ is a two-dimensional standard Brownian motion and $\chi \in (-1,1)$ is a known constant. We assume that the initial random variables (η, η') are independent of $(\tilde{W}_t, B_t)_{t\geq 0}$. We shall say that we are in the independent case if W and B are independent ($\chi = 0$) and in the general case if $\chi \in (-1,1)$.

Let us set

$$\mathcal{G}_t = \sigma((W_s, B_s), s \leq t;\ \eta;\ \eta'), \quad \mathcal{G}_{t,B} = \sigma(W_s,\ s \leq t; B_s,\ s \geq 0;\ \eta;\ \eta'). \tag{8}$$

Let us notice that, by (6), in the independent case, conditionally on $\mathcal{G}_{0,B}$, the random variable $Y_{t+h} - Y_t$ is Gaussian, with mean $\int_t^{t+h} \rho(X_s, s)ds$ and variance $\int_t^{t+h} X_s ds$.

Let us now introduce our assumptions:

(A1) Equation (7) admits a unique strong solution (X_t) satisfying $P(\forall t, X_t > 0) = 1$.

Below, it will be necessary to control the behaviour of (X_t) near 0 and ∞. To deal separately with each boundaries, we introduce the two following functions defined on $(0, \infty)$.

$$\mathcal{B}_0(x) = 1 + \frac{1}{x}, \quad \mathcal{B}_\infty(x) = 1 + x^k \ (k > 0). \tag{9}$$

In our examples, k will be either 1 or 2. The properties of these functions that we use are the following:

For all five non negative real numbers $\alpha, \beta, \alpha', \beta', p$, there exists a constant c such that for all $x \in (0, \infty)$:

$$(\mathcal{B}_0^\alpha(x) + \mathcal{B}_\infty^\beta(x)) \times (\mathcal{B}_0^{\alpha'}(x) + \mathcal{B}_\infty^{\beta'}(x)) \leq c(\mathcal{B}_0^{\alpha+\alpha'}(x) + \mathcal{B}_\infty^{\beta+\beta'}(x))$$
$$(\mathcal{B}_0^\alpha(x) + \mathcal{B}_\infty^\beta(x))^p \leq c(\mathcal{B}_0^{p\alpha}(x) + \mathcal{B}_\infty^{p\beta}(x)).$$

The other assumptions are:

(A2) $a,\ b \in \mathcal{C}^2(0, \infty)$ and $\exists c > 0,\ \exists \alpha_1, \alpha_2, \beta_1, \beta_2 \geq 0,\ \forall x \in (0, \infty)$,

$$|a(x)| + |b(x)| \leq c\mathcal{B}_\infty(x),$$
$$|a'(x)| \leq c(\mathcal{B}_0^{\alpha_1}(x) + \mathcal{B}_\infty^{\alpha_1}(x)), \quad |a''(x)| \leq c(\mathcal{B}_0^{\alpha_2}(x) + \mathcal{B}_\infty^{\alpha_2}(x)),$$
$$|b'(x)| \leq c(\mathcal{B}_0^{\beta_1}(x) + \mathcal{B}_\infty^{\beta_1}(x)), \quad |b''(x)| \leq c(\mathcal{B}_0^{\beta_2}(x) + \mathcal{B}_\infty^{\beta_2}(x)).$$

(A3) There exists $K_0 > 0$ such that,
$\forall k \in [0, K_0),\ \exists c,\ \forall t \geq 0,\ E\left(\sup_{s\in[t,t+1]} \mathcal{B}_0^k(X_s) \mid \mathcal{G}_t\right) \leq c\mathcal{B}_0^k(X_t)$
$\forall k \in [0, \infty),\ \exists c,\ \forall t \geq 0,\ E\left(\sup_{s\in[t,t+1]} \mathcal{B}_\infty^k(X_s) \mid \mathcal{G}_t\right) \leq c\mathcal{B}_\infty^k(X_t)$

(A4) The function $\rho(x, s) : (0, \infty) \times \mathbb{R}_+ \mapsto \mathbb{R}$ is Borel and satisfies:

$$\exists\ c, \quad \forall x > 0, \quad \forall s \geq 0, \quad |\rho(x, s)| \leq c\mathcal{B}_\infty(x)$$

Let us point out that only (A3) is difficult to check. For the following examples of positive diffusions, Assumption (A3) was checked in [8] (proofs are given in Section 4 of [8] and are not repeated here).

• Example 1: Exponantial of a diffusion on $\mathbb{R}$.
Here $\mathcal{B}_0(x) = 1 + \frac{1}{x}$ and $\mathcal{B}_\infty(x) = 1 + x^2$. Let $X_t = \exp(Z_t)$ with (Z_t) a diffusion on $\mathbb{R}$ solving: $dZ_t = \tilde{a}(Z_t)dB_t + \tilde{b}(Z_t)dt$. Assume that the functions $\tilde{a}$, $\tilde{b}$ belong to $C^2(\mathbb{R})$ and that their second derivatives have polynomial growth. Assume also that $\tilde{a}$, $\tilde{a}'$ are bounded, that $\tilde{b}$ has linear growth and $\limsup_{z\to\infty} \tilde{b}(z) < \infty$, $\liminf_{z\to-\infty} \tilde{b}(z) > -\infty$.

Then, (X_t) satisfies all assumptions, and $K_0 = \infty$.

• Example 2: Bilinear process.
$dX_t = (\alpha X_t + \beta)dt + \sigma X_t dB_t$, with $\alpha < 0$, σ, $\beta > 0$, $K_0 = \infty$.

• Example 3: Square–root process.
$dX_t = (\alpha X_t + \beta)dt + \sigma\sqrt{X_t}dB_t$, with $\alpha < 0$, σ, $\beta > 0$ $\frac{2\beta}{\sigma^2} > 1$ and $K_0 = \frac{2\beta}{\sigma^2} - 1$.

This model justifies the constant K_0 in (A3).

1.2 Integrated volatility and related expansions

Now, let Δ_n be a sequence of positive numbers tending to 0 as $n \to \infty$. For convenience, we suppose too that $\Delta_n \leq 1$, for all n. Let

$$J_i^n = \int_{i\Delta_n}^{(i+1)\Delta_n} X_s ds. \tag{10}$$

To simplify notations, we omit the superscript n and simply write $J_i = J_i^n$.

In this section we recall some properties of (J_i) proved in [8].

Consider the following variables:

$$\xi_{i,n} = \Delta_n^{-\frac{3}{2}} \int_{i\Delta_n}^{(i+1)\Delta_n} (s - i\Delta_n) dB_s \quad \text{for } i,\ n \geq 0 \tag{11}$$

$$\xi'_{i+1,n} = \Delta_n^{-\frac{3}{2}} \int_{(i+1)\Delta_n}^{(i+2)\Delta_n} (i\Delta_n + 2\Delta_n - s) dB_s \quad \text{for } i \geq -1,\ n \geq 0 \tag{12}$$

$$U_{i,n} = \xi_{i,n} + \xi'_{i+1,n} \tag{13}$$

Let us set (see (8))

$$\mathcal{G}_i^n = \mathcal{G}_{i\Delta_n} \tag{14}$$

For all $n \geq 0$, $(\xi_{i,n})_{i\geq 0}$, $(\xi'_{i+1,n})_{i\geq -1}$ and $(U_{i,n})_{i\geq 0}$ are Gaussian processes; $\xi_{i,n}$ is $\mathcal{G}_{i+1}^n$ measurable and independent of $\mathcal{G}_i^n$; $\xi'_{i+1,n}$ is $\mathcal{G}_{i+2}^n$ measurable and independent of $\mathcal{G}_{i+1}^n$. Straightforward computations yield

$$E\left(\xi_{i,n} \mid \mathcal{G}_i^n\right) = E\left(\xi'_{i+1,n} \mid \mathcal{G}_{i+1}^n\right) = 0 \tag{15}$$

$$E\left(\xi_{i,n}^2 \mid \mathcal{G}_i^n\right) = E\left(\xi'^2_{i+1,n} \mid \mathcal{G}_{i+1}^n\right) = \frac{1}{3},\ E\left(\xi_{i,n}\xi'_{i,n} \mid \mathcal{G}_i^n\right) = \frac{1}{6}. \tag{16}$$

So, for $i \geq 0$, $Var(U_{i,n}) = \frac{2}{3}$, $Cov(U_{i,n}, U_{i+1,n}) = \frac{1}{6}$ and $Cov(U_{i,n}, U_{i+j,n}) = 0$ if $j \geq 2$.

Hence, $(U_{i,n})_{i\geq 0}$ has the covariance structure of a $MA(1)$ process.

In all our statements below, we make use of a generic constant c which may change from a line to another but never depends on i or n. On the contrary, we pay a careful attention to all exponants and relate them with the exponants given in (A2)–(A4).

First, we recall a useful result on the regularity of the diffusion X itself and its direct application to $\Delta_n^{-1} J_i$.

Proposition 1.1. *Let $k \geq 1$, then there exists c such that, for all $i, n \geq 0$*

$$E\left(\sup_{s\in[i\Delta_n,(i+1)\Delta_n]} |X_s - X_{i\Delta_n}|^k \mid \mathcal{G}_i^n\right) \leq c\Delta_n^{\frac{k}{2}} \mathcal{B}_\infty^k(X_{i\Delta_n}) \tag{17}$$

$$E\left(\left|\Delta_n^{-1} J_i - X_{i\Delta_n}\right|^k \mid \mathcal{G}_i^n\right) \leq c\Delta_n^{\frac{k}{2}} \mathcal{B}_\infty^k(X_{i\Delta_n}) \tag{18}$$

$$E\left(\left|\Delta_n^{-1}(J_{i+1} - J_i)\right|^k \mid \mathcal{G}_i^n\right) \leq c\Delta_n^{\frac{k}{2}} \mathcal{B}_\infty^k(X_{i\Delta_n}) \tag{19}$$

Now, we give the first order expansion for the difference $\Delta_n^{-1} J_i - X_{i\Delta_n}$.

Proposition 1.2. *Assume that $2\alpha_1 < K_0$, (with α_1 and K_0 given in (A2)-(A3)) then:*

$$\Delta_n{}^{-1} J_i - X_{i\Delta_n} = a(X_{i\Delta_n})\Delta_n^{\frac{1}{2}}\xi'_{i,n} + e_{i,n}$$

$$\textit{where} \qquad E\left(e_{i,n}^2 \mid \mathcal{G}_i^n\right) \le \Delta_n^2 c(\mathcal{B}_0^{2\alpha_1}(X_{i\Delta_n}) + \mathcal{B}_\infty^{2(1+\alpha_1)}(X_{i\Delta_n})) \tag{20}$$

Proposition 1.2 shows that the difference between $\Delta_n^{-1} J_i$ and $X_{i\Delta_n}$ is of order $\Delta_n^{\frac{1}{2}}$.

Now, the following theorem gives an asymptotic description of $\Delta_n^{-1}(J_{i+1} - J_i) - b(\Delta_n^{-1} J_i)\Delta_n$. In the special case of a constant, its law may be approximated by the one of a MA(1) process, whereas the law of $(X_{(i+1)\Delta_n} - X_{i\Delta_n} - b(X_{i\Delta_n})\Delta_n)$ is known to be close to the one of a sequence of independent Gaussian variables.

Theorem 1.3. *We have (see (11)-(13))*

$$\Delta_n^{-1}(J_{i+1} - J_i) - b(\Delta_n^{-1} J_i)\Delta_n = \Delta_n^{\frac{1}{2}} a(X_{i\Delta_n})U_{i,n} + \varepsilon_{i,n}$$

where $\varepsilon_{i,n}$ is $\mathcal{G}_{i+2}^n$ measurable, and there exists a positive constant c such that for all i, n,

if $(4\alpha_1) \vee 2\alpha_2 < K_0$, then,

$$E\left(\varepsilon_{i,n}^2 \mid \mathcal{G}_i^n\right) \le \Delta_n^2 c(\mathcal{B}_0^{2\alpha_1 \vee \alpha_2}(X_{i\Delta_n}) + \mathcal{B}_\infty^{3+2\alpha_1+\alpha_2}(X_{i\Delta_n})) \tag{21}$$

$$E\left(\varepsilon_{i,n}^4 \mid \mathcal{G}_i^n\right) \le \Delta_n^4 c(\mathcal{B}_0^{4\alpha_1 \vee 2\alpha_2}(X_{i\Delta_n}) + \mathcal{B}_\infty^{6+4\alpha_1+2\alpha_2}(X_{i\Delta_n})) \tag{22}$$

Moreover, the remainder term $\varepsilon_{i,n}$ satisfies, if $4\alpha_1 \vee \alpha_2 \vee 4\beta_1 < K_0$,

$$|E\left(\varepsilon_{i,n} U_{i,n} \mid \mathcal{G}_i^n\right)| \le \Delta_n^{\frac{3}{2}} c(\mathcal{B}_0^{(\alpha_1+\beta_1)\vee\alpha_2}(X_{i\Delta_n}) + \mathcal{B}_\infty^{(1+\alpha_1+\beta_1)\vee(2+\alpha_2)}(X_{i\Delta_n})) \tag{23}$$

2 Approximations of the integrated volatility

Here, we assume that the sample path (Y_t) is observed at times $(i + \frac{j}{m})\Delta_n$, with sampling interval $\frac{\Delta_n}{m}$ depending on a double index (n, m).

We define

$$\widehat{J}_i^{n,m} = \widehat{J}_i^m = \sum_{j=0}^{m-1} (Y_{(i+\frac{j+1}{m})\Delta_n} - Y_{(i+\frac{j}{m})\Delta_n})^2 \tag{24}$$

We will now investigate the properties of $\Delta_n^{-1} \widehat{J}_i^m$ as an approximation of $\Delta_n^{-1} J_i$ and of $X_{i\Delta_n}$.

Let us introduce the following notations, for $i \ge 0$, $0 \le j \le m-1$:

$$\alpha_{i,n,j,m} = 2\int_{(i+\frac{j}{m})\Delta_n}^{(i+\frac{j+1}{m})\Delta_n} (Y_s - Y_{(i+\frac{j}{m})\Delta_n})\rho(X_s, s)ds \tag{25}$$

$$\beta_{i,n,j,m} = 2\int_{(i+\frac{j}{m})\Delta_n}^{(i+\frac{j+1}{m})\Delta_n} (Y_s - Y_{(i+\frac{j}{m})\Delta_n})\sigma_s dW_s \tag{26}$$

$$D_{i,n,m} = \Delta_n^{-1} \sum_{j=0}^{m-1} \alpha_{i,n,j,m}, \quad V_{i,n,m} = \Delta_n^{-1} \sum_{j=0}^{m-1} \beta_{i,n,j,m} \tag{27}$$

$$E_{i,n,m} = D_{i,n,m} + V_{i,n,m} \tag{28}$$

Notice that $\alpha_{i,n,j,m}$, $\beta_{i,n,j,m}$, $D_{i,n,m}$, $V_{i,n,m}$ and $E_{i,n,m}$ are $\mathcal{G}_{i+1}^n$ measurable. In the Appendix, we prove bounds for $\alpha_{i,n,m}$ and $\beta_{i,n,m}$ which are useful for this section.

2.1 The general case

Our first step is to give an estimate of the difference $\Delta_n^{-1}\widehat{J}_i^m - \Delta_n^{-1}J_i$ in the general case.

Recall that the general case is when the two Brownian motions in (1) and (2) may be correlated ($\chi \neq 0$). Then, we have the following result.

Theorem 2.1. *We have*

$$\Delta_n^{-1}\widehat{J}_i^m = \Delta_n^{-1}J_i + E_{i,n,m} \qquad \textit{(see (10), (24)-(28))} \tag{29}$$

with, for all $i \geq 0$, $1 \leq m$:

$$E\left(E_{i,n,m}^2 \mid \mathcal{G}_i^n\right) \leq cm^{-1}\mathcal{B}_\infty^4(X_{i\Delta_n}) \tag{30}$$

$$E\left(E_{i,n,m}^4 \mid \mathcal{G}_i^n\right) \leq cm^{-1}\mathcal{B}_\infty^8(X_{i\Delta_n}). \tag{31}$$

Proof. Using Ito's formula, (6), (25), (26),

$$(Y_{(i+\frac{j+1}{m})\Delta_n} - Y_{(i+\frac{j}{m})\Delta_n})^2 = \alpha_{i,n,j,m} + \beta_{i,n,j,m} + \int_{(i+\frac{j}{m})\Delta_n}^{(i+\frac{j+1}{m})\Delta_n} X_s ds.$$

Hence (29) follows from (24) and (28).

Using Lemma 5.4 (in the Appendix) and $E\left(\alpha_{i,n,j,m}^2 \mid \mathcal{G}_i^n\right) \leq E\left(\alpha_{i,n,j,m}^4 \mid \mathcal{G}_i^n\right)^{\frac{1}{2}}$, we have:

$$E\left(\left(\sum_{j=0}^{m-1}\alpha_{i,n,j,m}\right)^2 \mid \mathcal{G}_i^n\right) \leq m\sum_{j=0}^{m-1}E\left(\alpha_{i,n,j,m}^2 \mid \mathcal{G}_i^n\right) \leq cm^{-1}\Delta_n^3\mathcal{B}_\infty^4(X_{i\Delta_n}) \tag{32}$$

Using again Lemma 5.4, $E\left(\beta_{i,n,j,m}^2 \mid \mathcal{G}_i^n\right) \leq E\left(\beta_{i,n,j,m}^4 \mid \mathcal{G}_i^n\right)^{\frac{1}{2}}$ and $E\left(\beta_{i,n,m,j}\beta_{i,n,m,j'} \mid \mathcal{G}_i^n\right) = 0$ for $j \neq j'$, we have:

$$E\left(\left(\sum_{j=0}^{m-1}\beta_{i,n,j,m}\right)^2 \mid \mathcal{G}_i^n\right) = \sum_{j=0}^{m-1}E\left(\beta_{i,n,j,m}^2 \mid \mathcal{G}_i^n\right) \leq cm^{-1}\Delta_n^2\mathcal{B}_\infty^3(X_{i\Delta_n}).$$

Now, (27)–(28) yields (30).

By Lemma 5.4,

$$E\left(\left(\sum_{j=0}^{m-1}\alpha_{i,n,j,m}\right)^4 \mid \mathcal{G}_i^n\right) \leq m^3\sum_{j=0}^{m-1}E\left(\alpha_{i,n,j,m}^4 \mid \mathcal{G}_i^n\right) \leq cm^{-2}\Delta_n^6\mathcal{B}_\infty^8(X_{i\Delta_n}). \tag{33}$$

Using $E\left(\prod_{u=1}^4 \beta_{i,n,j_u,m} \mid \mathcal{G}_i^n\right) \leq \prod_{u=1}^4 E\left(\beta_{i,n,j_u,m}^4 \mid \mathcal{G}_i^n\right)^{\frac{1}{4}}$ and Lemma 5.4, we get:

$$\begin{aligned} E\left(\left(\sum_{j=0}^{m-1} \beta_{i,n,j,m}\right)^4 \mid \mathcal{G}_i^n\right) &= \sum_{\substack{0\leq j_1,j_2,j_3,j_4\leq m-1 \\ j_u\neq j_{u'} \text{ for } u\neq u'}} E\left(\prod_{u=1}^4 \beta_{i,n,j_u,m} \mid \mathcal{G}_i^n\right) + \sum_{\substack{0\leq j_1,j_2,j_3,j_4\leq m-1 \\ \#\{j_1,j_2,j_3,j_4\}\leq 3}} E\left(\prod_{u=1}^4 \beta_{i,n,j_u,m} \mid \mathcal{G}_i^n\right) \\ &= \sum_{\substack{0\leq j_1,j_2,j_3,j_4\leq m-1 \\ \#\{j_1,j_2,j_3,j_4\}\leq 3}} E\left(\prod_{u=1}^4 \beta_{i,n,j_u,m} \mid \mathcal{G}_i^n\right) \leq cm^3\frac{\Delta_n^4}{m^4}\mathcal{B}_\infty^6(X_{i\Delta_n}) \end{aligned} \tag{34}$$

Joining (33) and (34) gives (31). □

It appears from (30) that $E_{i,n,m}^2$ is of order m^{-1}. This gives an error term of order $m^{\frac{1}{2}}$ for the difference $\Delta_n^{-1}(\widehat{J}_i^m - J_i)$. However, from (31), we see that the fourth conditional moment of $E_{i,n,m}$ is only of order m^{-1} and not m^{-2} as could be expected. This comes from the correlation between the two Brownian motions. Also, from the above proof, we see that the main term in $E_{i,n,m}$ comes from $V_{i,n,m}$, which is the error term due to $\sigma_t dW_t$ (see (26)). The term $D_{i,n,m}$ is null when the drift term $\rho(x,s)$ is equal to 0.

2.2 The independent case

Along this section $\chi = 0$, and the two Brownian motions B and W are independent. In this case, the expansions of the previous section can be improved. And, we can draw some consequences concerning the estimation of the quadratic variation of (X_t) by means of $\widehat{J}_i^m$.

2.2.1 Expansions

The following theorem precises Theorem 2.1. Let us set (see (8))

$$\mathcal{G}_{i,B}^n = \mathcal{G}_{i\Delta_n,B} \tag{35}$$

Theorem 2.2. *We have*

$$\Delta_n^{-1}\widehat{J}_i^m = \Delta_n^{-1}J_i + E_{i,n,m}$$

with, for all $i \geq 0$, $1 \leq m$:

$$\left|E\left(E_{i,n,m} \mid \mathcal{G}_{i,B}^n\right)\right| \leq cm^{-\frac{1}{2}}\Delta_n^{\frac{1}{2}} \sup_{s\in[i\Delta_n,(i+1)\Delta_n]} \mathcal{B}_\infty^2(X_s) \tag{36}$$

$$E\left(E_{i,n,m}^2 \mid \mathcal{G}_{i,B}^n\right) \leq cm^{-1} \sup_{s\in[i\Delta_n,(i+1)\Delta_n]} \mathcal{B}_\infty^4(X_s) \tag{37}$$

$$E\left(E_{i,n,m}^4 \mid \mathcal{G}_{i,B}^n\right) \leq cm^{-2} \sup_{s\in[i\Delta_n,(i+1)\Delta_n]} \mathcal{B}_\infty^8(X_s) \tag{38}$$

Proof. We have $E\left(\sum_{j=0}^{m-1}\beta_{i,n,j,m} \mid \mathcal{G}_{i,B}^n\right)=0$. And, using Lemma 5.4 (see the Appendix)

$$E\left(\left|\sum_{j=0}^{m-1}\alpha_{i,n,j,m}\right| \mid \mathcal{G}_{i,B}^n\right)\leq m\left(m^{-1}\Delta_n\right)^{\frac{3}{2}}\sup_{s\in[i\Delta_n,(i+1)\Delta_n]}\mathcal{B}_\infty^2(X_s)$$

This gives (36).

We prove (37) in a way similar to the proof of (30).

For the proof of (38), first we obtain, similarly to (33),

$$E\left(\left(\sum_{j=0}^{m-1}\alpha_{i,n,j,m}\right)^4 \mid \mathcal{G}_{i,B}^n\right)\leq cm^{-2}\Delta_n^6\sup_{s\in[i\Delta_n,(i+1)\Delta_n]}\mathcal{B}_\infty^8(X_s).$$

Second, we write

$$E\left(\left(\sum_{j=0}^{m-1}\beta_{i,n,j,m}\right)^4 \mid \mathcal{G}_{i,B}^n\right)=\sum_{\substack{0\leq j_1,j_2,j_3,j_4\leq m-1\\ j_u\neq j_{u'} \text{ for } u\neq u'}}E\left(\prod_{u=1}^4\beta_{i,n,j_u,m} \mid \mathcal{G}_{i,B}^n\right)+$$

$$12\sum_{\substack{0\leq j_1,j_2,j_3\leq m-1\\ j_u\neq j_{u'} \text{ for } u\neq u'}}E\left(\beta_{i,n,j_1,m}^2\beta_{i,n,j_2,m}\beta_{i,n,j_3,m} \mid \mathcal{G}_{i,B}^n\right)+\sum_{\substack{0\leq j_1,j_2,j_3,j_4\leq m-1\\ \#\{j_1,j_2,j_3,j_4\}\leq 2}}E\left(\prod_{u=1}^4\beta_{i,n,j_u,m} \mid \mathcal{G}_{i,B}^n\right)$$

Now, the first sum in the above sum satisfies $\sum_{\substack{0\leq j_1,j_1,j_3,j_4\leq m-1\\ j_u\neq j_{u'} \text{ for } u\neq u'}}E\left(\prod_{u=1}^4\beta_{i,n,j_u,m} \mid \mathcal{G}_{i,B}^n\right)=0$. By Lemma 5.4, the third term satisfies

$$\sum_{\substack{0\leq j_1,j_2,j_3,j_4\leq m-1\\ \#\{j_1,j_2,j_3,j_4\}\leq 2}}E\left(\prod_{u=1}^4\beta_{i,n,j_u,m} \mid \mathcal{G}_{i,B}^n\right)\leq cm^{-2}\Delta_n^4\sup_{s\in[i\Delta_n,(i+1)\Delta_n]}\mathcal{B}_r^6(X_s).$$

So, it is now enough to prove that, for all distinct $0\leq j_1,j_2,j_3\leq m-1$,

$$E\left(\beta_{i,n,j_1,m}^2\beta_{i,n,j_2,m}\beta_{i,n,j_3,m} \mid \mathcal{G}_{i,B}^n\right)=0. \tag{39}$$

If j_2 or $j_3>j_1$, then (39) is clear since

$$E\left(\beta_{i,n,j_2,m} \mid \mathcal{G}_{(i+\frac{j_2}{m})\Delta_n,B}^n\right)=0 \text{ and } E\left(\beta_{i,n,j_3,m} \mid \mathcal{G}_{(i+\frac{j_3}{m})\Delta_n,B}^n\right)=0.$$

So, we may assume that $j_1>j_2>j_3$. Let us set,

$$\tilde{\beta^2}_{i,n,j_1,m}=E\left(\beta_{i,n,j_1,m}^2 \mid \mathcal{G}_{(i+\frac{j_1}{m})\Delta_n,B}\right).$$

Using (26) and the independence of B and W, we have

$$\tilde{\beta^2}_{i,n,j_1,m}=4\int_{(i+\frac{j_1}{m})\Delta_n}^{(i+\frac{j_1+1}{m})\Delta_n}E\left((Y_s-Y_{(i+\frac{j_1}{m})\Delta_n})^2X_s \mid \mathcal{G}_{(i+\frac{j_1}{m})\Delta_n,B}\right)ds.$$

But conditionally on B, $(Y_s - Y_{(i+\frac{j_1}{m})\Delta_n})$ is Gaussian, and we deduce that

$$\tilde{\beta^2}_{i,n,j_1,m} = 4\int_{(i+\frac{j_1}{m})\Delta_n}^{(i+\frac{j_1+1}{m})\Delta_n} \phi_{i,n,j_1,m}(s)X_s ds, \text{ with}$$

$$\phi_{i,n,j,m}(s) = \left(\int_{(i+\frac{j}{m})\Delta_n}^{s} \rho(X_v, v)dv\right)^2 + \int_{(i+\frac{j}{m})\Delta_n}^{s} X_v dv. \tag{40}$$

Thus, $\tilde{\beta^2}_{i,n,j_1,m}$ is only a functional of X and hence is $\mathcal{G}_{0,B}$ measurable.

Now, using that $\beta_{i,n,j_2,m}$ and $\beta_{i,n,j_3,m}$ are $\mathcal{G}_{(i+\frac{j_1}{m})\Delta_n,B}$ measurable (since $j_1 > j_2 > j_3$):

$$E\left(\beta^2_{i,n,j_1,m}\beta_{i,n,j_2,m}\beta_{i,n,j_3,m} \mid \mathcal{G}^n_{i,B}\right) = E\left(\tilde{\beta^2}_{i,n,j_1,m}\beta_{i,n,j_2,m}\beta_{i,n,j_3,m} \mid \mathcal{G}^n_{i,B}\right)$$

But we have seen that $\tilde{\beta^2}_{i,n,j_1,m}$ is $\mathcal{G}_{0,B}$ measurable, therefore it is $\mathcal{G}^n_{i,B}$ measurable too. This gives:

$$E\left(\tilde{\beta^2}_{i,n,j_1,m}\beta_{i,n,j_2,m}\beta_{i,n,j_3,m} \mid \mathcal{G}^n_{i,B}\right) = \tilde{\beta^2}_{i,n,j_1,m}E\left(\beta_{i,n,j_2,m}\beta_{i,n,j_3,m} \mid \mathcal{G}^n_{i,B}\right)$$

Now, using $j_2 \neq j_3$ we have $E\left(\beta_{i,n,j_2,m}\beta_{i,n,j_3,m} \mid \mathcal{G}^n_{i,B}\right) = 0$. This yields (39), and so (41) is proved. □

Notice, that conditionally on B, the expectation of $E_{i,n,m}$ is of order $\Delta_n^{\frac{1}{2}}m^{-\frac{1}{2}}$ whereas its standard deviation is of order $m^{-\frac{1}{2}}$ (see (36)–(37)). This means that errors $E_{i,n,m}$ are almost centered. Furthermore, comparing (31) and (38), it appears that in the independent case we obtain a better bound for the fourth moment of $E_{i,n,m}$.

Recall that $E_{i,n,m} = V_{i,n,m} + D_{i,n,m}$ (see (28)). We give a more precise result for the second moment each term. We start with $V_{i,n,m}$.

Proposition 2.3. *The following expansions hold (see (27))*

$$E\left(V^2_{i,n,m} \mid \mathcal{G}^n_{i,B}\right) = 2m^{-1}X^2_{i\Delta_n} + r_{i,n,m}, \quad \text{where } r_{i,n,m} \text{ is } \mathcal{G}_{0,B} \text{ measurable, and}$$
$$E\left(r^2_{i,n,m} \mid \mathcal{G}^n_i\right) \leq cm^{-2}\Delta_n\mathcal{B}^6_\infty(X_{i\Delta_n}). \tag{41}$$
$$E\left(V^2_{i+1,n,m} \mid \mathcal{G}^n_{i,B}\right) = 2m^{-1}X^2_{i\Delta_n} + s_{i,n,m}, \quad \text{where } s_{i,n,m} \text{ is } \mathcal{G}_{0,B} \text{ measurable, and}$$
$$E\left(s^2_{i,n,m} \mid \mathcal{G}^n_i\right) \leq cm^{-2}\Delta_n\mathcal{B}^6_\infty(X_{i\Delta_n}). \tag{42}$$
$$E\left(V_{i,n,m}V_{i+1,n,m} \mid \mathcal{G}^n_{i,B}\right) = 0 \tag{43}$$

Proof. By (26) and (27), we write:

$$E\left({V_{i,n,m}}^2 \mid \mathcal{G}^n_{i,B}\right) = 4\Delta_n^{-2}\sum_{j=0}^{m-1}\int_{(i+\frac{j}{m})\Delta_n}^{(i+\frac{j+1}{m})\Delta_n} E\left((Y_s - Y_{(i+\frac{j}{m})\Delta_n})^2 \mid \mathcal{G}^n_{i,B}\right)X_s ds.$$

But, conditionally on B, $(Y_s - Y_{(i+\frac{j}{m})\Delta_n})$ is Gaussian, so we deduce (see (40)):

$$E\left(V^2_{i,n,m} \mid \mathcal{G}^n_{i,B}\right) = 4\Delta_n^{-2}\sum_{j=0}^{m-1}\int_{(i+\frac{j}{m})\Delta_n}^{(i+\frac{j+1}{m})\Delta_n} \phi_{i,n,j,m}(s)X_s ds.$$

Hence,

$$r_{i,n,m} = 4\Delta_n^{-2} \sum_{i=0}^{m-1} \int_{(i+\frac{j}{m})\Delta_n}^{(i+\frac{j+1}{m})\Delta_n} \{\phi_{i,n,j,m}(s)X_s - X_{i\Delta_n}^2 (s - (i + \frac{j}{m})\Delta_n)\} ds,$$

which is $\mathcal{G}_{0,B}$ measurable.

To obtain (41), by (40), we write $r_{i,n,m} = z_{i,n,m} + z'_{i,n,m}$ with:

$$z_{i,n,m} = 4\Delta_n^{-2} \sum_{i=0}^{m-1} \int_{(i+\frac{j}{m})\Delta_n}^{(i+\frac{j+1}{m})\Delta_n} \left(\int_{(i+\frac{j}{m})\Delta_n}^{s} \rho(X_v, v) dv \right)^2 X_s ds,$$

$$z'_{i,n,m} = 4\Delta_n^{-2} \sum_{i=0}^{m-1} \int_{(i+\frac{j}{m})\Delta_n}^{(i+\frac{j+1}{m})\Delta_n} \int_{(i+\frac{j}{m})\Delta_n}^{s} (X_v X_s - X_{i\Delta_n}^2) dv ds.$$

By Assumption (A4), we have:

$$\left(\int_{(i+\frac{j}{m})\Delta_n}^{(i+\frac{j+1}{m})\Delta_n} \left(\int_{(i+\frac{j}{m})\Delta_n}^{s} \rho(X_v) dv \right)^2 X_s ds \right)^2 \leq c m^{-6} \Delta_n^6 \sup_{s \in [i\Delta_n, (i+1)\Delta_n]} \mathcal{B}_\infty^6(X_s).$$

Using (A3), this yields:

$$E\left({z_{i,n,m}}^2 \mid \mathcal{G}_i^n \right) \leq c m^{-4} \Delta_n^2 \mathcal{B}_\infty^6(X_{i\Delta_n}). \tag{44}$$

By (17), (A3) and with some computations, we get:

$$E\left(\sup_{v,s \in [(i+\frac{j}{m})\Delta_n, (i+\frac{j+1}{m})\Delta_n]} (X_v X_s - X_{i\Delta_n}^2)^2 \mid \mathcal{G}_i^n \right) \leq c \Delta_n \mathcal{B}_\infty^4(X_{i\Delta_n}). \tag{45}$$

We deduce:

$$E\left({z'_{i,n,m}}^2 \mid \mathcal{G}_i^n \right) \leq c m^{-2} \Delta_n \mathcal{B}_\infty^4(X_{i\Delta_n}). \tag{46}$$

We join (44) and (46) to obtain (41).

We show analogously (42), and (43) is immediate. □

We give the result for $D_{i,n,m}$.

Proposition 2.4. *The following inequality holds:*

$$\forall i \geq 0,\ \forall n \geq 1,\ \forall m \geq 1, \quad E\left(D_{i,n,m}^2 \mid \mathcal{G}_{i,B}^n \right) \leq c m^{-1} \Delta_n \sup_{s \in [i\Delta_n, (i+1)\Delta_n]} \mathcal{B}_\infty^4(X_s) \tag{47}$$

Proof. Proceeding as for (32), we show: $E\left(\left(\sum_{j=0}^{m-1} \alpha_{i,n,j,m} \right)^2 \mid \mathcal{G}_{i,B}^n \right) \leq c m^{-1} \Delta_n \sup_{s \in [i\Delta_n, (i+1)\Delta_n]} \mathcal{B}_\infty^4(X_s)$. This yields the result. □

Now, we may join results of Section 1 and Section 2.

Corollary 2.5. *Assume that $4\alpha_1 \vee 2\alpha_2 < K_0$, then the following expansion holds:*

$$\Delta_n^{-1}(\widehat{J_{i+1}}^m - \widehat{J}_i^m) = \Delta_n^{\frac{1}{2}} a(X_{i\Delta_n})U_{i,n} + E_{i+1,n,m} - E_{i,n,m} + \eta_{i,n,m} \quad (see (13)), \tag{48}$$

with $E\left(\eta_{i,n,m}^2 \mid \mathcal{G}_i^n\right) \leq c\Delta_n^2(\mathcal{B}_0^{2\alpha_1 \vee \alpha_2}(X_{i\Delta_n}) + \mathcal{B}_\infty^{3+2\alpha_1+\alpha_2}(X_{i\Delta_n}))$.

Proof. By Theorems 1.3 and 2.1 we have the expansion (48) with $\eta_{i,n} = \varepsilon_{i,n} + \Delta_n b(\Delta_n^{-1}J_i)$. Now, (21), (A2) and (A3) yield the desired bound for $E\left(\eta_{i,n,m}^2 \mid \mathcal{G}_i^n\right)$. □

2.2.2 Approximation of the quadratic variation in a fixed length time interval

Let us now give an application of the results of Section 2.2.1 when the observations are taken within the time interval $[0, 1]$. For this, we set $\Delta_n = \frac{1}{n}$, and assume that $m = m_n$ depends on n and tends to ∞ as $n \to \infty$.

Consider the quadratic variation associated with $(n\widehat{J}_i^{m_n})$

$$\widehat{V}_n = \sum_{i=0}^{n-2} (n\widehat{J_{i+1}}^{m_n} - n\widehat{J}_i^{m_n})^2 \tag{49}$$

This random variable should be the natural estimator of $\int_0^1 a^2(X_s)ds$ based on the observations. Actually, the proposition below shows that this is not true.

Proposition 2.6. *Assume that $K_0 = \infty$ in (A3), and that X_0 is deterministic. Then, if $\frac{\sqrt{n}}{m_n} \xrightarrow[n\to\infty]{} 0$,*

$$\widehat{V}_n = \frac{2}{3}\int_0^1 a^2(X_s)ds + \frac{4n}{m_n}\int_0^1 X_s^2 ds + o_{\mathbf{P}}(1).$$

Proof. By (A3), the assumptions imply

$$\forall k \geq 0, \quad \sup_{t\in[0,1]} E(\mathcal{B}_0^k(X_t) + \mathcal{B}_\infty^k(X_t)) < \infty. \tag{50}$$

Using (48), $E\left(U_{i,n}^2 \mid \mathcal{G}_i^n\right) = \frac{2}{3}$, (36), (41)-(43), (47) and (50), we get

$$E\left((n\widehat{J_{i+1}}^{m_n} - n\widehat{J}_i^{m_n})^2 - 4m_n^{-1}X_{\frac{i}{n}}^2 \mid \mathcal{G}_i^n\right) = \frac{2}{3}n^{-1}a^2(X_{\frac{i}{n}}) + a_{i,n} \tag{51}$$

with, if $\frac{\sqrt{n}}{m_n} \to 0$,

$$\sum_{i=0}^{n-2} a_{i,n} \xrightarrow[\mathbf{L}^1]{n\to\infty} 0.$$

We deduce

$$\sum_{i=0}^{n-2}\{(n\widehat{J_{i+1}}^{m_n} - n\widehat{J}_i^{m_n})^2 - 4m_n^{-1}X_{\frac{i}{n}}^2\} \to \frac{2}{3}\int_0^1 a^2(X_s)ds.$$

Futhermore, by (19) and (38), we have

$$E\left(\left(\left(\widehat{nJ_{i+1}}^{m_n} - n\widehat{J}_i^{m_n}\right)^2 - 4m_n^{-1}X^2_{\frac{i}{n}}\right)^2 \mid \mathcal{G}_i^n\right) \leq c(n^{-2} + m_n^{-2})\mathcal{B}^8_\infty(X_{i\Delta_n}).$$

Using Lemma (5.5) (in the Appendix), we obtain

$$\sum_{i=0}^{n-2}\{\left(\widehat{nJ_{i+1}}^{m_n} - n\widehat{J}_i^{m_n}\right)^2 - 4m_n^{-1}X^2_{\frac{i}{n}}\} \xrightarrow[\mathbf{P}]{n\to\infty} \frac{2}{3}\int_0^1 a^2(X_s)ds.$$

The proposition follows easily. □

In Gloter [8] (Proposition 2.9), we have proved that

$$\sum_{i=0}^{n-1}(nJ_{i+1} - nJ_i)^2 = \frac{2}{3}\int_0^1 a^2(X_s)ds + o_{\mathbf{P}}(1)$$

So, we already know that the quadratic variation of (nJ_i) underestimates the quadratic variation of (X_t). With $(n\widehat{J}_i^m)$, a further bias term appears. According to $\frac{n}{m_n}$, three different behaviours are possible.

- If $\frac{n}{m_n} \to 0$, $\widehat{V}_n$ is a consistent estimator of $\frac{2}{3}\int_0^1 a^2(X_s)ds$.
- If $m_n = n$, a new bias term is added.
- If $\frac{n}{m_n} \to \infty$ (with the restriction $\frac{\sqrt{n}}{m_n} \to 0$), then $\widehat{V_n}$ explodes due to $E_{i,n,m}$.

3 Conditional moments of positive and negative orders

For the statistical applications of Section 4, we need limit theorems for expressions containing $f\left(\Delta_n^{-1}\widehat{J}_i^m\right)$ with $f(x) = x^k$ or $f(x) = x^{-k}$. This justify the study of bounds for the conditional moments of $\Delta_n^{-1}\widehat{J}_i^m$ and $(\Delta_n^{-1}\widehat{J}_i^m)^{-1}$.

First, remark that we can easily bound conditional moments of $\Delta_n^{-1}J_i$. Using (A3), we have, for $k \geq 0$,

$$E\left((\Delta_n^{-1}J_i)^k \mid \mathcal{G}_i^n\right) \leq E\left(\sup_{s\in[i\Delta_n,(i+1)\Delta_n]} X_s^k \mid \mathcal{G}_i^n\right) \leq \mathcal{B}^k_\infty(X_{i\Delta_n}),$$

and similarly, for $k \in [0, K_0)$, $E\left((\Delta_n^{-1}J_i)^{-k} \mid \mathcal{G}_i^n\right) \leq \mathcal{B}_0^k(X_{i\Delta_n})$.

Now, we intend to obtain, uniformly in m, bounds for $E\left((\Delta_n^{-1}\widehat{J}_i^m)^{-k} \mid \mathcal{G}_i^n\right)$ and $E\left((\Delta_n^{-1}\widehat{J}_i^m)^{k} \mid \mathcal{G}_i^n\right)$. For the negative orders, a difficulty appears. Conditionally on B, $\Delta_n^{-1}\widehat{J}_i^m$ is a sum of m squares of independent normal variables, therefore $E\left((\Delta_n^{-1}\widehat{J}_i^m)^{-k} \mid \mathcal{G}_{i,B}^n\right)$ may be infinite (this will be the case if $m = 1$, for all $k \geq \frac{1}{2}$, since the inverse of a normal law is not integrable). Furthermore, when $m \to \infty$ each variable $\Delta_n^{-1}\left(Y_{(i+\frac{i+1}{m})\Delta_n} - Y_{(i+\frac{i}{m})\Delta_n}\right)^2$ (see (24)) tends to zero. Hence, at first glance, a uniform bound for negative moments of seems difficult to achieve.

To obtain it, we start by proving a general result on negative moments of sums of positive variables (Section 3.1). Then, we apply it to square of Gaussian variables. (Section 3.2).

Positive moments are easier to obtain.

3.1 A general inverse moment result

Theorem 3.1. *Let $Z_1, \dots, Z_m$ be m independent random variables. We suppose that Z_i has a density $\phi_i(x)$ and that there exists $\eta > 0$, $M > 0$ such that for all $x \in [-\eta, \eta]$ and for all $i \in \{1, \dots, m\}$, $\phi_i(x) \leq M$ (i.e. these densities are uniformly bounded near zero). Then, for all $k \geq 0$, there exists $c(k, M, \eta)$, a constant depending on k, $, M$, η, such that for all $m \geq 2k+3$:*

$$E\left(\left(\sum_{i=1}^{m} Z_i^2\right)^{-k}\right) \leq c(k, M, \eta) m^{-k} \tag{52}$$

Proof. By increasing M, we may assume that $\frac{1}{2M} \leq \eta$. Let $(U_i)_{i=1,\dots,m}$, be m independent real variables with uniform distribution on $[0, \frac{1}{2M}]$. First we show that:

$$M_k := E\left(\left(\sum_{i=1}^{m} Z_i^2\right)^{-k}\right) \leq E\left(\left(\sum_{i=1}^{m} U_i^2\right)^{-k}\right) := N_k. \tag{53}$$

Denote by ϕ_i the distribution function of $|Z_i|$ and by ψ the distribution function of U_i. Then, for $x \geq 0$,

$$\phi_i(x) = P(|Z_i| \leq x) = \int_{-x}^{x} \phi_i(s) ds \leq (2Mx \wedge 1) = \psi(x) \tag{54}$$

This implies that, for all $i = 1, \dots, m$ and $t \in (0,1)$,

$$\psi^{-1}(t) \leq \phi^{-1}(t). \tag{55}$$

Now, consider m independent real variables, $A_1, \dots, A_m$, with uniform law on $(0,1)$, and set, for all i,

$$Z_i^* = \phi_i^{-1}(A_i), \quad U_i^* = \psi^{-1}(A_i).$$

We know that $(Z_i^*)_{i=1,\dots,m}$ and $(U_i^*)_{i=1,\dots,m}$ have respectively the same law as $(|Z_i|)_{i=1,\dots,m}$ and $(U_i)_{i=1,\dots,m}$. Furthermore by (55), we have

$$\forall i, \quad U_i^* \leq Z_i^* \quad a.s.$$

So,

$$M_k = E\left(\left(\sum_{i=1}^{m} Z_i^{*2}\right)^{-k}\right) \leq E\left(\left(\sum_{i=1}^{m} U_i^{*2}\right)^{-k}\right) = N_k.$$

And (53) is proved.

It remains to obtain an upper bound for N_k. For this note that

$$\begin{aligned} P(U_1^2 + \dots + U_m^2 \leq \varepsilon) &= \int_{[0, \frac{1}{2M}]^m} \mathbb{1}_{\{u_1^2 + \dots + u_m^2 \leq \varepsilon\}} (2M)^m du_1 \dots du_m \\ &\leq (2M)^m \int_{\mathbb{R}^m} \mathbb{1}_{\{u_1^2 + \dots + u_m^2 \leq \varepsilon\}} du_1 \dots du_m \end{aligned}$$

We denote by $\sigma_m = 2\frac{\pi^{\frac{m}{2}}}{\Gamma(\frac{m}{2})}$ the area of the unit sphere in the Euclidian space $\mathbb{R}^m$. By a change of variable, we get:

$$P(U_1^2 + \cdots + U_m^2 \leq \varepsilon) \leq (2M)^m \int_0^{\sqrt{\varepsilon}} \rho^{m-1} d\rho \sigma_m = (2M)^m \frac{\varepsilon^{\frac{m}{2}}}{m} \sigma_m. \tag{56}$$

Let ρ_m be a positive constant depending on m, that will be specified later. Define

$$N'_k = E\left((U_1^2 + \cdots + U_m^2)^{-k} 1\!\!1_{\{U_1^2+\cdots+U_m^2 \leq \rho_m^2\}}\right).$$

We have

$$N'_k \leq \sum_{q=1}^{\infty} u_q,$$

where

$$u_q = \frac{(q+1)^k}{\rho_m^{2k}} P(U_1^2 + \cdots + U_m^2 \in [\frac{\rho_m^2}{q+1}, \frac{\rho_m^2}{q}]).$$

Using (56) with $\varepsilon = \frac{\rho_m^2}{q}$, we get

$$u_q \leq \frac{(q+1)^k}{q^{\frac{m}{2}}} (2M)^m \rho_m^{m-2k} \frac{\sigma_m}{m}.$$

Since $\frac{m}{2} - k \geq \frac{3}{2}$, we can set

$$s = \sum_{q=1}^{\infty} \frac{(q+1)^k}{q^{\frac{m}{2}}} < \infty.$$

So,

$$N'_k \leq s(2M)^m \rho_m^{m-2k} \frac{\sigma_m}{m}.$$

Let us recall the exact inequality connected with the Stirling formula (see Feller p.54 [5]):

$$\Gamma(\frac{m}{2}) > \left(\frac{\frac{m}{2}-1}{e}\right)^{\frac{m}{2}-1} \sqrt{2\pi(\frac{m}{2}-1)}.$$

Using this inequality and the value of σ_m yields

$$N'_k \leq \frac{s(2M)^m \rho_m^{m-2k} \sqrt{2}}{m(\frac{\frac{m}{2}-1}{e\pi})^{\frac{m}{2}-1} \sqrt{\frac{m}{2}-1}}.$$

Now, we set $\rho_m = \left(\frac{\frac{m}{2}-1}{4e\pi M^2}\right)^{\frac{1}{2}}$ and the previous inequality reduces to

$$N'_k \leq \frac{s\sqrt{2}(4M)^k}{\left(\frac{\frac{m}{2}-1}{e\pi}\right)^{k-1} m\sqrt{\frac{m}{2}-1}} \leq Cm^{-k}$$

where C only depends on k, M, η.

Since $E\left((U_1^2 + \cdots + U_m^2)^{-k} 1\!\!1_{\{U_1^2+\cdots+U_m^2 > \rho_m^2\}}\right) \leq \rho_m^{-2k} = (\frac{8e\pi M^2}{m-2})^k$, we have

$$E((U_1^2 + \cdots + U_m^2)^{-k}) \leq c(k, M, \eta) m^{-k}.$$

Hence Theorem 3.1 is proved. □

This result can be applied to Gaussian variables.

Corollary 3.2. *Let $Z_1, \dots, Z_m$ be m independent random variables such that the law of Z_i is $\mathcal{N}(\mu_i, v_i)$. Then, for all $k \geq 0$, there exists $c(k)$, a constant depending only of k, such that for all $m \geq 2k+3$:*

$$E\left((Z_1^2 + \cdots + Z_m^2)^{-k}\right) \leq c(k)(\min_{i=1,\dots,m} v_i)^{-k} m^{-k}$$

Proof. If there exists i such that $v_i = 0$, the result is obvious. Otherwise, for all $i = 1, \dots, m$ we set $\tilde{Z}_i = \frac{Z_i}{\sqrt{v_i}}$. Then,

$$E\left((Z_1^2 + \cdots + Z_m^2)^{-k}\right) \leq (\min_{i=1,\dots,m} v_i)^{-k} E\left((\tilde{Z}_1^2 + \cdots + \tilde{Z}_m^2)^{-k}\right)$$

Since the density of $\tilde{Z}_i$ is bounded on $\mathbb{R}$ by $\frac{1}{\sqrt{2\pi}}$ we can apply Theorem 3.1, hence we get the result. □

Remark 3.3. *To make sure that the bound in Corollary 3.2 is sharp, we compute, in the case where for all i, $\mu_i = 0$ and $v_i = v$, by using that $\sum_{i=1}^m Z_i^2$ is a chi square variable, with $m > 2k$,*

$$E\left((\sum_{i=1}^m Z_i^2)^{-k}\right) = v^{-k} \frac{\Gamma(\frac{m}{2} - k)}{\Gamma(\frac{m}{2})} \sim_{m \to \infty} 2^k v^{-k} m^{-k}.$$

3.2 Bounds for conditional moments

Recall $\mathcal{B}_0(x) = 1 + \frac{1}{x}$ and $\mathcal{B}_\infty(x) = 1 + x$ or $\mathcal{B}_\infty(x) = 1 + x^2$.

Theorem 3.4. *1) Let $0 \leq k < K_0$. There exists c such that, for all $m \geq 2k+3$,*

$$E\left(\mathcal{B}_0(\Delta_n^{-1} \widehat{\mathcal{J}}_i^m)^k \mid \mathcal{G}_i^n\right) \leq c\mathcal{B}_0^k(X_{i\Delta_n})$$

2) Let $k \geq 1$. There exists c such that, for all m,

$$E\left(\mathcal{B}_\infty(\Delta_n^{-1} \widehat{\mathcal{J}}_i^m)^k \mid \mathcal{G}_i^n\right) \leq c\mathcal{B}_\infty^{2k}(X_{i\Delta_n}).$$

Proof. Let us prove 1). We have (see (24)) $\widehat{\mathcal{J}}_i^m = \sum_{j=0}^{m-1} Z_j^2$, with

$$Z_j = \int_{(i+\frac{j}{m})\Delta_n}^{(i+\frac{j+1}{m})\Delta_n} \rho(X_s, s) ds + \int_{(i+\frac{j}{m})\Delta_n}^{(i+\frac{j+1}{m})\Delta_n} \sigma_s dW_s.$$

Conditionally on $\mathcal{G}_{i,B}^n$ (recall (8) and (35)), Z_j has law $\mathcal{N}(\mu_j, v_j)$ with

$$\mu_j = \int_{(i+\frac{j}{m})\Delta_n}^{(i+\frac{j+1}{m})\Delta_n} \rho(X_s, s) ds + \chi \int_{(i+\frac{j}{m})\Delta_n}^{(i+\frac{j+1}{m})\Delta_n} \sigma_s dB_s,$$

$$v_j = (1-\chi^2) \int_{(i+\frac{j}{m})\Delta_n}^{(i+\frac{j+1}{m})\Delta_n} X_s ds.$$

By Corollary 3.2, we deduce for $m \geq 2k+3$:

$$E\left(\left(\widehat{J}_i^m\right)^{-k} \mid \mathcal{G}_{i,B}^n\right) \leq c(k)(\min_{i=0,\dots,m-1} v_i)^{-k} m^{-k}$$

But, $v_j \geq (1-\chi^2)\frac{\Delta_n}{m} \inf_{s\in[i\Delta_n,(i+1)\Delta_n]} X_s$ yields (recall that $\chi^2 \neq 1$),

$$E\left(\left(\widehat{J}_i^m\right)^{-k} \mid \mathcal{G}_{i,B}^n\right) \leq c(k)\Delta_n^{-k}(1-\chi^2)^{-k} \sup_{s\in[i\Delta_n,(i+1)\Delta_n]} X_s^{-k}$$

Now, using the inclusion $\mathcal{G}_i^n \subset \mathcal{G}_{i,B}^n$ and Assumption (A3) we obtain

$$E\left(\left(\Delta_n^{-1}\widehat{J}_i^m\right)^{-k} \mid \mathcal{G}_i^n\right) \leq c\mathcal{B}_0^k(X_{i\Delta_n})$$

This yields 1).

Let us prove 2). Using that $(\sum_{j=1}^m Z_j^2)^k \leq m^{k-1}\sum_{j=1}^m Z_j^{2k}$, we obtain

$$E\left(\left(\widehat{J}_i^m\right)^{k} \mid \mathcal{G}_{i,B}^n\right) \leq cm^k \sup_{0\leq j\leq m-1} (v_j^k + \mu_j^{2k}) \tag{57}$$

By the Burkhölder inequality, we get

$$\begin{aligned} E\left(\mu_j^{2k} \mid \mathcal{G}_i^n\right) &\leq c(m^{-1}\Delta_n)^{2k} E\left(\sup_{s\in[i\Delta_n,(i+1\Delta_n)]} \rho^{2k}(X_s,s) \mid \mathcal{G}_i^n\right) + cE\left(\left(\int_{(i+\frac{j}{m})\Delta_n}^{(i+\frac{j+1}{m})\Delta_n} X_s ds\right)^k \mid \mathcal{G}_i^n\right) \\ &\leq c(m^{-1}\Delta_n)^{2k} E\left(\sup_{s\in[i\Delta_n,(i+1)\Delta_n]} \rho^{2k}(X_s,s) \mid \mathcal{G}_i^n\right) + c(m^{-1}\Delta_n)^k E\left(\sup_{s\in[i\Delta_n,(i+1)\Delta_n]} X_s^k \mid \mathcal{G}_i^n\right) \end{aligned}$$

Using (A4), then (A3) yields

$$E\left(\mu_j^{2k} \mid \mathcal{G}_i^n\right) \leq c\left(m^{-1}\Delta_n\right)^k \mathcal{B}_\infty^{2k}(X_{i\Delta_n}).$$

Analogously, we get

$$E\left(v_i^k \mid \mathcal{G}_i^n\right) \leq c\left(m^{-1}\Delta_n\right)^k \mathcal{B}_\infty^{k}(X_{i\Delta_n}).$$

We deduce from (57) that

$$E\left(\left(\widehat{J}_i^m\right)^{k} \mid \mathcal{G}_i^n\right) \leq c\Delta_n^k \mathcal{B}_\infty^{2k}(X_{i\Delta_n}).$$

Hence 2) is proved. □

Remark 3.5. *From the above proof, we see that if we strengthen (A4) by supposing that ρ is bounded on $(0,\infty)\times\mathbb{R}$ then we have:* $E\left((\Delta_n^{-1}\widehat{J}_i^m)^k \mid \mathcal{G}_i^n\right) \leq c\mathcal{B}_\infty^k(X_{i\Delta_n})$.

4 Estimation of the diffusion coefficient of the volatility: A special model.

Now, we assume that (X_t) solves the stochastic differential equation

$$dX_t = \theta a(X_t)dB_t + b(X_t)dt.$$

and study the estimation of θ based on the observation of (Y_t) on the fixed length time interval $[0,1]$. Note that the drift function is unknown and hence may depend on θ.

As in Section 2.2.2, we set $\Delta_n = \frac{1}{n}$, and let $m = m_n$ depend on n.

The choice of a multiplicative form for the diffusion coefficient, $a(x,\theta) = \theta a(x)$, yields two significant simplifications in the standard problem of inference based on $(X_{\frac{i}{n}})_{i=0,\dots,n}$: it enables to find an explicit estimator of θ^2, which is asymptotically normal (see (4)–(5)).

To simplify proofs, we shall set $\rho(x,s) = 0$ (hence $dY_t = \sqrt{X_t}dW_t$), but the results in this section remain valid for any $\rho(x,s)$ satisfying (A4).

We assume the two Brownian motions B and W are independent, and that (A1)–(A4) are satisfied with $K_0 = \infty$. In addition, we suppose that

$$\exists\gamma,\ c \geq 0,\ \forall x \in (0,\infty),\ \frac{1}{a(x)} \leq c(\mathcal{B}_0^\gamma(x) + \mathcal{B}_\infty^\gamma(x)) \tag{58}$$

and that

$$\forall k \geq 0, \quad E(X_0^{-k} + X_0^k) < \infty.$$

By $(A3)$, the above assumption implies that

$$\forall k \in \mathbb{N},\ \sup_{t\in[0,1]} E(\mathcal{B}_0^k(X_t)) < \infty \quad \text{and} \quad \sup_{t\in[0,1]} E(\mathcal{B}_\infty^k(X_t)) < \infty. \tag{59}$$

4.1 Estimator and rate of subsamplings.

We start by constructing an approximation of θ^2 based on the unobserved variables $J_i = \int_{\frac{i}{n}}^{\frac{i+1}{n}} X_s ds$, for $i = 0, \dots, n-1$.

Proposition 4.1. *We set* $\tilde{\theta^2}_n = \frac{3}{2}\sum_{i=0}^{n-2} \frac{(nJ_{i+1}-nJ_i)^2}{a^2(nJ_i)}$, *then:*

$$Z_n = \sqrt{n}(\tilde{\theta^2}_n - \theta^2) \xrightarrow[\mathcal{D}]{n\to\infty} \mathcal{N}(0, \frac{9}{4}\theta^4)$$

Proof. Using Theorem 1.3, we split

$$\tilde{\theta^2}_n = \sum_{i=0}^{n-2}(u_{i,n}^{(1)} + u_{i,n}^{(2)} + u_{i,n}^{(3)} + u_{i,n}^{(4)}) \quad \text{where (see (11)–(13))},$$

$$
\begin{aligned}
u_{i,n}^{(1)} &= \frac{3\theta^2}{2} n^{-1} U_{i,n}^2 \\
u_{i,n}^{(2)} &= \frac{3\theta}{a(X_{\frac{i}{n}})} n^{-\frac{1}{2}} U_{i,n}(\varepsilon_{i,n} + n^{-1} b(nJ_i)) \\
u_{i,n}^{(3)} &= \frac{3(\varepsilon_{i,n} + n^{-1} b(nJ_i))^2}{2a^2(X_{\frac{i}{n}})} \\
u_{i,n}^{(4)} &= \frac{3}{2}(nJ_{i+1} - nJ_i)^2 (a^{-2}(nJ_i) - a^{-2}(X_{\frac{i}{n}}))
\end{aligned}
$$

By (11)–(13), $u_{i,n}^{(1)}$ is the square of a Gaussian variable, and $E(u_{i,n}^{(1)}) = n^{-1}\theta^2$. Using that $cov(U_{i,n}^2, U_{j,n}^2) = 2cov^2(U_{i,n}, U_{j,n})$, we obtain $var(u_{i,n}^{(1)}) = n^{-1}2\theta^4$ and $cov(u_{i,n}^{(1)}, u_{i+1,n}^{(1)}) = (8n)^{-1}\theta^4$. We easily deduce that

$$\sqrt{n}\left(\sum_{i=0}^{n-2} u_{i,n}^{(1)} - \theta^2\right) \xrightarrow[\mathcal{D}]{n\to\infty} \mathcal{N}\left(0, \frac{9\theta^4}{4}\right).$$

By (23), (58) there exists $c \geq 0$ such that

$$\left| E\left(\sqrt{n} u_{i,n}^{(2)} \mid \mathcal{G}_i^n\right) \right| \leq c n^{-\frac{3}{2}} (\mathcal{B}_0^c(X_{\frac{i}{n}}) + \mathcal{B}_\infty^c(X_{\frac{i}{n}})),$$

and using (22), for another constant c, $E\left(n(u_{i,n}^{(2)})^2 \mid \mathcal{G}_i^n\right) \leq c n^{-2}(\mathcal{B}_0^c(X_{\frac{i}{n}}) + \mathcal{B}_\infty^c(X_{\frac{i}{n}}))$.

By (59), we deduce $\sum_{i=0}^{n-2} E\left(\sqrt{n} u_{i,n}^{(2)} \mid \mathcal{G}_i^n\right) \xrightarrow[\mathbf{L}^1]{n\to\infty} 0$ and $\sum_{i=0}^{n-2} E\left(n(u_{i,n}^{(2)})^2 \mid \mathcal{G}_i^n\right) \xrightarrow[\mathbf{L}^1]{n\to\infty} 0$. This implies, by Lemma 5.5 in the Appendix, $\sqrt{n}\sum_{i=0}^{n-2} u_{i,n}^{(2)} \xrightarrow[\mathbf{P}]{n\to\infty} 0$.

Analogously, we show,

$$E\left(\left|u_{i,n}^{(3)}\right| \mid \mathcal{G}_i^n\right) \leq c n^{-2}(\mathcal{B}_0^c(X_{\frac{i}{n}}) + \mathcal{B}_\infty^c(X_{\frac{i}{n}})).$$

Then, we deduce, by (59), that $\sqrt{n}\sum_{i=0}^{n-2} u_{i,n}^{(3)} \xrightarrow[\mathbf{L}^1]{n\to\infty} 0$.

It remains to prove $\sqrt{n}\sum_{i=0}^{n-2} u_{i,n}^{(4)} \xrightarrow[\mathbf{P}]{n\to\infty} 0$. By a Taylor expansion for $a^{-2}(nJ_i) - a^{-2}(X_{\frac{i}{n}})$, Proposition 1.2 and Theorem 1.3, we write,

$$u_{i,n}^{(4)} = n^{-\frac{3}{2}} g(X_{\frac{i}{n}}) U_{i,n}^2 \xi'_{i,n} + \tilde{u}_{i,n}^{(4)}$$

where $g(x) = a^2(x)(a^{-2})'(x)$, and $\tilde{u}_{i,n}^{(4)}$ satisfies for some constant c, $E\left(\left|\tilde{u}_{i,n}^{(4)}\right| \mid \mathcal{G}_i^n\right) \leq c n^{-2}(\mathcal{B}_0^c(X_{\frac{i}{n}}) + \mathcal{B}_\infty^c(X_{\frac{i}{n}}))$.

This implies $\sqrt{n}\sum_{i=0}^{n-2} \tilde{u}_{i,n}^{(4)} \xrightarrow[\mathbf{L}^1]{n\to\infty} 0$.

We conclude by proving

$$\sqrt{n}\sum_{i=0}^{n-2} a^2(X_{\frac{i}{n}}) g(X_{\frac{i}{n}}) U_{i,n}^2 \xi'_{i,n} n^{-\frac{3}{2}} \xrightarrow[\mathbf{P}]{n\to\infty} 0.$$

This is done since $E\left(U_{i,n}^2 \xi'_{i,n} \mid \mathcal{G}_i^n\right) = 0$. □

Remark 4.2. *By the previous result, we see that if we replace* $X_{\frac{i}{n}}$ *by* nJ_i *in the expression of the standard approximation of* θ^2, $\theta_n^{2*} = \sum_{i=0}^{n-2} \frac{\left(X_{\frac{i+1}{n}} - X_{\frac{i}{n}}\right)^2}{a^2(X_{\frac{i}{n}})}$, *then we underestimate* θ^2 *(by a factor* $\frac{2}{3}$*). The asymptotic variance of* $\tilde{\theta^2}_n$ *is slightly bigger than the one of* θ_n^{2*} *(recall (5)).*

Now, we deduce from $\tilde{\theta^2}_n$ an estimator based on (Y_t).

The results of Section 2.2.2 lead us to define:

$$\hat{\theta^2}_{n,m} = \frac{3}{2} \sum_{i=0}^{n-2} \left\{ \frac{(\widehat{nJ_{i+1}}^m - n\widehat{J_i}^m)^2}{a^2(n\widehat{J_i}^m)} - \frac{4}{m} \frac{(n\widehat{J_i}^m)^2}{a^2(n\widehat{J_i}^m)} \right\} \quad \text{for } n, m \geq 1. \tag{60}$$

The next theorem gives the asymptotic behaviour of $\hat{\theta^2}_{n,m}$ when n tends to ∞, and $m = m_n$ depends on n and tends to ∞ as $n \to \infty$.

Theorem 4.3. 1) *Assume that* m_n *is a sequence of integers such that* $n^{\frac{2}{3}} m_n^{-1} = o(1)$, *then* $\hat{\theta^2}_{n,m_n}$ *is consistent:*

$$\hat{\theta^2}_{n,m_n} \xrightarrow[\mathbf{P}]{n \to \infty} \theta^2$$

2) *Furthermore, we have*

$$\sqrt{n}(\hat{\theta^2}_{n,m_n} - \tilde{\theta}_n^2) = \left((m_n n^{-1})^{\frac{1}{2}} + (m_n n^{-1})^{\frac{3}{2}} \right) B_{n,m_n},$$

where (B_{n,m_n}) *is a sequence bounded in* $\mathbf{L}^1$, *for* n *large enough.*

Proof. We write

$$\sqrt{n}(\hat{\theta^2}_{n,m} - \tilde{\theta^2}_n) = T_{n,m}^{(1)} + T_{n,m}^{(2)} + T_{n,m}^{(3)} \tag{61}$$

with:

$$T_{n,m}^{(1)} = 6m^{-1} n^{\frac{1}{2}} \sum_{i=0}^{n-2} \left\{ \frac{(n\widehat{J_i}^m)^2}{a^2(n\widehat{J_i}^m)} - \frac{X_{\frac{i}{n}}^2}{a^2(X_{\frac{i}{n}})} \right\} \tag{62}$$

$$T_{n,m}^{(2)} = \frac{3}{2} n^{\frac{1}{2}} \sum_{i=0}^{n-2} \left\{ \frac{(E_{i+1,n,m} - E_{i,n,m})^2}{a^2(X_{\frac{i}{n}})} - \frac{4}{m} \frac{X_{\frac{i}{n}}^2}{a^2(X_{\frac{i}{n}})} \right\} \tag{63}$$

$$T_{n,m}^{(3)} = \frac{3}{2} n^{\frac{1}{2}} \sum_{i=0}^{n-2} \left\{ \frac{(\widehat{nJ_{i+1}}^m - n\widehat{J_i}^m)^2}{a^2(n\widehat{J_i}^m)} - \frac{(nJ_{i+1} - nJ_i)^2}{a^2(nJ_i)} - \frac{(E_{i+1,n,m} - E_{i,n,m})^2}{a^2(X_{i\Delta_n})} \right\} \tag{64}$$

• First, we bound $T_{n,m}^{(1)} = \frac{6\sqrt{n}}{m} \sum_{i=0}^{n-2} (f(n\widehat{J_i}^m) - f(nJ_i) + f(nJ_i) - f(X_{\frac{i}{n}}))$, with $f(x) = \frac{x^2}{a^2(x)}$.
By Taylor's formula

$$f(n\widehat{J_i}^m) - f(nJ_i) = f'(\xi) E_{i,n,m}, \quad \text{with } \xi \in [nJ_i,\ n\widehat{J_i}^m].$$

Using the bound $|f'(\xi)| \leq c(\mathcal{B}_0^c(\xi) + \mathcal{B}_\infty^c(\xi))$, and the monotonicity of $\mathcal{B}_0$ and $\mathcal{B}_\infty$, we deduce:

$$|f'(\xi)| \leq c(\mathcal{B}_0^c(nJ_i) + \mathcal{B}_\infty^c(nJ_i) + \mathcal{B}_0^c(n\widehat{J_i}^m) + \mathcal{B}_\infty^c(n\widehat{J_i}^m))$$

By (A3), Theorem 3.4 and (58), we deduce that for m large enough $E(|f'(\xi)|^2) \leq c$. Using Cauchy-Schwartz's inequality and (37), we get

$$E\left(\left|f(n\widehat{J}_i^{m}) - f(nJ_i)\right|\right) \leq cm^{-\frac{1}{2}} \tag{65}$$

By (18), we show analogously that

$$E(\left|f(nJ_i) - f(X_{\frac{i}{n}})\right|) \leq cn^{-\frac{1}{2}} \tag{66}$$

Joining (65) and (66) one gets

$$E(\left|T^{(1)}_{n,m}\right|) \leq c(m^{-1}n + (m^{-1}n)^{\frac{3}{2}}).$$

- Now, we deal with $T^{(2)}_{n,m}$. Using that, by $\rho(x,s) = 0$, $E_{i,n,m} = V_{i,n,m}$ we have

$$T^{(2)}_{n,m} = \sum_{i=0}^{n-2} u_{i,n,m}$$

with

$$u_{i,n,m} = \frac{3}{2}n^{\frac{1}{2}}\left\{\frac{(V_{i+1,n,m} - V_{i,n,m})^2}{a^2(X_{\frac{i}{n}})} - 4m^{-1}\frac{X^2_{\frac{i}{n}}}{a^2(X_{\frac{i}{n}})}\right\}$$

Let us give a bound for $\left(\sum_{i=0}^{n-2} u_{2i,2n,m}\right)^2$. For $i < i'$, using that $u_{2i,2n,m}$ is $\mathcal{G}^n_{2i+2,B}$ $(\subset \mathcal{G}^n_{2i',B})$ measurable, and Proposition 2.3, we obtain

$$E(u_{2i,2n,m}u_{2i',2n,m}) = E(u_{2i,2n,m}E(u_{2i',2n,m} \mid \mathcal{G}^{2n}_{2i',B})) = E\left(u_{2i,2n,m}\frac{n^{\frac{1}{2}}(r_{2i',2n,m} + s_{2i',2n,m})}{\frac{2}{3}a^2(X_{\frac{i}{n}})}\right)$$

But $r_{2i',2n,m} + s_{2i',2n,m}$ is $\mathcal{G}_{0,B}$ measurable, hence

$$\begin{aligned}E(u_{2i,2n,m}u_{2i',2n,m}) &= E\left(E(u_{2i,2n,m} \mid \mathcal{G}^{2n}_{2i,B})\frac{n^{\frac{1}{2}}(r_{2i',2n,m} + s_{2i',2n,m})}{\frac{2}{3}a^2(X_{\frac{i'}{n}})}\right)\\ &= E\left(\frac{n^{\frac{1}{2}}(r_{2i,2n,m} + s_{2i,2n,m})}{\frac{2}{3}a^2(X_{\frac{i}{n}})}\frac{n^{\frac{1}{2}}(r_{2i',2n,m} + s_{2i',2n,m})}{\frac{2}{3}a^2(X_{\frac{i'}{n}})}\right).\end{aligned}$$

We deduce (with (41) and (42)), $|E(u_{2i,2n,m}u_{2i',2n,m}))| \leq cm^{-2}$. Now, $\left|E(u^2_{2i,2n,m})\right| \leq cm^{-2}n$ gives $E\left((\sum_{i=0}^{n-1} u_{2i,2n,m})^2\right) \leq cm^{-2}n^2$.

Proceeding analogously with the odd indexes, we obtain

$$E((T^{(2)}_{n,m})^2) \leq c(m^{-1}n)^2.$$

- For $T^{(3)}_{n,m}$, some computations based on (18), (19), (36) and (38) yield

$$E(\left|T^{(3)}_{n,m}\right|) \leq c((m^{-1}n)^{\frac{1}{2}} + m^{-1}n + (m^{-1}n)^{\frac{3}{2}}).$$

Hence, 2) is proved

Under the condition $n^{\frac{2}{3}}/m_n = o(1)$, 1) follows from the consistency of $\tilde{\theta}_n$ and 2). □

Remark 4.4. • *If we set $m_n = n$, then $\sqrt{n}(\hat{\theta^2}_{n,n} - \theta^2) = \sqrt{n}(\tilde{\theta^2}_n - \theta_0^2) + 2B_{n,n}$ hence*

$$\sup_n P(\sqrt{n}\left|\hat{\theta^2}_{n,n} - \theta^2\right| > M) \xrightarrow{M\to\infty} 0.$$

Now taking into account that $\hat{\theta^2}_{n,n}$ uses $N = n^2 + 1$ data, we have constructed an estimator with rate $N^{\frac{1}{4}}$.

- *If $\frac{m_n}{n} \xrightarrow{n\to\infty} \infty$ then the estimator is asymptotically normal: $\sqrt{n}(\hat{\theta^2}_{n,m_n} - \theta^2) \xrightarrow{n\to\infty} \mathcal{N}(0, \frac{9\theta^4}{4})$.*
- *We know nothing on B_{n,m_n} and especially on how it depends on θ.*

4.2 Some numerical simulation results.

We now investigate, by numerical simulations, the quality of the estimator $\hat{\theta^2}_{n,m}$ for the following model

$$\begin{cases} dY_t = \sqrt{X_t}dW_t, \ Y_0 = 0 \\ X_t = \exp(\theta B_t), \ X_0 = 1, \ B \text{ and } W \text{ are two independent Brownian motions.} \end{cases}$$

Since $dX_t = \theta X_t dB_t + \frac{\theta^2}{2}X_t dt$, the estimator of θ is given by

$$\hat{\theta^2}_{n,m} = \sum_{i=0}^{n-2} \frac{(n\widehat{J_{i+1}}^m - n\widehat{J_i}^m)^2}{\frac{2}{3}(n\widehat{J_i}^m)^2} - \frac{6(n-1)}{m}.$$

Tables 2, 4 and 6 show for different values of θ, n and $\frac{m}{n}$, the empirical mean and standard deviation of $\hat{\theta^2}_{n,m}$ (with 200 replications).

To evaluate the effect of $B_{n,m}$ we give in Tables 1, 3 and 5 the empirical mean and standard deviation of $\tilde{\theta^2}_n$ for the same values of θ^2 and n (with 200 replications).

Table 1: (Mean; Standard deviation) of $\tilde{\theta^2}_n$ for $\theta^2 = 1$.

n=10	n=20	n=50	n=100	n=200
1.01; 0.76	0.99; 0.37	0.98; 0.23	1.00; 0.14	1.01; 0.10

Table 2: (Mean; Standard deviation) of $\hat{\theta^2}_{n,m}$ for $\theta^2 = 1$.

	n=10	n=20	n=50	n=100	n=200
$\frac{m}{n}$=1	10.2; 25	5.0; 7.07	2.39; 2.63	1.59; 1.58	1.38; 0.91
$\frac{m}{n}$=4	1.67; 2.41	1.38; 1.17	1.13; 0.60	1.02; 0.37	1.02; 0.27
$\frac{m}{n}$=10	1.19; 1.2	1.11; 0.70	1.00; 0.36		
$\frac{m}{n}$=20	0.94; 0.68	1.01; 0.50	1.00; 0.27		
$\frac{m}{n}$=50	0.97; 0.70	0.96; 0.34			
$\frac{m}{n}$=100	1.10; 0.76				

From Table 1, Table 2, we see that, when $\frac{m}{n}$ is large, standard deviations of $\hat{\theta^2}_{n,m}$ and of $\tilde{\theta^2_n}$ are close. However, $\hat{\theta^2}_{n,m}$ only achieves a reasonable standard deviation for $(n, \frac{m}{n}) = (20, 50)$,

Table 3: (Mean; Standard deviation) of $\tilde{\theta^2}_n$ for $\theta^2 = 0.1$.

n=10	n=20	n=50	n=100
0.093;0.048	0.095;0.035	0.096; 0.021	0.099; 0.015

Table 4: (Mean; Standard deviation) of $\hat{\theta^2}_{n,m}$ for $\theta^2 = 0.1$.

	n=10	n=20	n=50	n=100
$\frac{m}{n}$=1	10.7; 31	3.85; 6.2	1.12; 2.1	0.51; 1.7
$\frac{m}{n}$=4	0.36; 1.40	0.26; 0.83	0.11; 0.43	0.13; 0.27
$\frac{m}{n}$=10	0.12; 0.43	0.062; 0.27	0.10; 0.16	0.095; 0.12
$\frac{m}{n}$=20	0.08; 0.21	0.11; 0.144	0.088; 0.087	
$\frac{m}{n}$=50	0.089; 0.11	0.092; 0.072	0.098; 0.049	
$\frac{m}{n}$=100	0.079; 0.067	0.093; 0.051		

Table 5: (Mean; Standard deviation) of $\tilde{\theta^2}_n$ for $\theta^2 = 0.01$.

n=10	n=20	n=50	n=100
0.0089;0.0047	0.0094;0.0034	0.0097; 0.002	0.0098; 0.0014

Table 6: (Mean; Standard deviation) of $\hat{\theta^2}_{n,m}$ for $\theta^2 = 0.01$.

	n=10	n=20	n=50	n=100
$\frac{m}{n}$=1	7.75; 16	2.75; 4.7	0.60; 1.82	0.50; 1.24
$\frac{m}{n}$=4	0.042; 1.0	0.11; 0.76	0.086; 0.43	0.029; 0.27
$\frac{m}{n}$=10	0.040;0.41	0.035; 0.26	0.014; 0.16	-0.00052;0.10
$\frac{m}{n}$=20	-0.026; 0.15	0.0013;0.13	0.0054; 0.075	0.0083; 0.05
$\frac{m}{n}$=50	0.0021; 0.073	0.0068; 0.044	0.013; 0.034	
$\frac{m}{n}$=100	0.0067; 0.038	0.0038; 0.025	0.0093; 0.015	

$(n, \frac{m}{n}) = (50, 20)$ and $(n, \frac{m}{n}) = (200, 4)$. For these values, the total number of observations N is respectively about 20000, 50000 and 160000, which are huge numbers. Notice that $\hat{\theta^2}_{n,n}$, although consistent, does not provide a valid estimate (at least if $n \leq 200$).

By Table 4, the estimation of $\theta^2 = 0.1$ by $\hat{\theta^2}_{n,m}$ is less accurate than for $\theta^2 = 1$, whereas $\tilde{\theta^2}_n$ remains good. (Table 3).

In Table 6, the lowest standard deviation is 0.015 when $n = 50$ and $\frac{m}{n} = 100$. This deviation is still bigger than the true value of the parameter $\theta^2 = 0.01$. And in particular for 28% of the replications, we obtained a negative value for the estimate of θ^2.

From Table 5–6, we see that the standard deviation of $\hat{\theta^2}_{n,m}$ is much higher than the one of $\tilde{\theta^2}_n$. Therefore, the high variance for $\hat{\theta^2}_{n,m}$ seems to be due to $B_{n,m}$.

We observe too that for, $\frac{m}{n} \leq 10$, standard deviations of $\hat{\theta^2}_{n,m}$ are almost equal if $\theta^2 = 0.01$ and if $\theta^2 = 0.1$. Hence the standard deviation of $B_{n,m}$ appears almost independent of θ^2 for small values of θ^2.

5 Appendix

5.1 Why small values of the parameter are badly estimated

Our estimator $\hat{\theta^2}_{n,m}$ failed to provide good estimates of small values of θ^2, because its variance is not reducing to zero when $\theta \to 0$. We recall, that on the contrary, the asymptotic variance of $\tilde{\theta^2}_n$ tends to zero when $\theta \to 0$. Now, the theoretical analysis of the asymptotic law is much harder for $\hat{\theta^2}_{n,m}$ than for $\tilde{\theta^2}_n$.

Therefore, for simplification, we only study $\hat{\theta^2}_{n,m}$ for $\theta = 0$.

The following proposition gives an equivalent, of the standard deviation of $\hat{\theta^2}_{n,m}$.

Proposition 5.1. *If $\theta = 0$, we have*

$$\exists c, m_0 \geq 0,\ \forall n \geq 1,\ \forall m \geq m_0, \quad \left|E(\hat{\theta^2}_{n,m})\right| \leq c m^{-\frac{3}{2}} n^{\frac{1}{2}}$$

$$Assume\ that\ n, m, \frac{m}{n} \to \infty\ then,\ Var^{\frac{1}{2}}(\hat{\theta^2}_{n,m}) \sim \sqrt{180} m^{-1} n^{\frac{1}{2}}$$

Proof. Since $\theta = 0$, X is a constant process, thus $X_{\frac{i}{n}} = nJ_i = X_0,\ \forall i, n \geq 0$.

We deduce, by (61),

$$\sqrt{n}\, \hat{\theta^2}_{n,m} = T^{(2)}_{n,m} + T^{(3)}_{n,m} \tag{67}$$

Now, the proof consists in computing expectations and variances of $T^{(2)}_{n,m}$ and $T^{(3)}_{n,m}$.

Before this, let us set $Z_i = \sum_{j=0}^{m-1} nm \left(\int_{\frac{i}{n}+\frac{j}{nm}}^{\frac{i}{n}+\frac{j+1}{nm}} dW_s\right)^2$. The Z_i's are independent for $i = 0, \ldots, n-2$ and have $\chi^2(m)$ distribution.

By (62), (64), $a(x) = x$, and remarking that $n\widehat{J_i}^m = m^{-1} X_0 Z_i$, we have

$$T^{(2)}_{n,m} = -3m^{-3} n^{\frac{1}{2}} \sum_{i=0}^{n-2} \{(Z_{i+1} - Z_i)^2 - 4m\}$$

$$T^{(3)}_{n,m} = \frac{3}{2} m^{-2} n^{\frac{1}{2}} \sum_{i=0}^{n-2} (Z_{i+1} - Z_i)^2 \{\frac{X_0^2}{(n\widehat{J_i}^m)^2} - 1\}$$

- We start with studing $T^{(2)}_{i,n}$. Set $v_{i,n,m} = (Z_{i+1} - m - (Z_i - m))^2 - 4m$.

By Lemma 5.2, $E(v_{i,n,m}) = 0$. Hence,

$$E(T^{(1)}_{i,n,m}) = 0. \tag{68}$$

Computations based on Lemma 5.2, imply

$$E(v_{i,n,m} v_{i',n,m}) = \begin{cases} 48m^2 + 96m & \text{if } i = i' \\ 16m^2 + 48m & \text{if } |i - i'| = 1 \\ 0 & \text{if } |i - i'| \geq 2 \end{cases}$$

We deduce for $n, m \to \infty$,

$$E(T^{(2)}_{n,m}) \sim 180 m^{-2} n^2. \tag{69}$$

• Now, we deal with $T^{(3)}_{n,m}$. By Taylor's formula, we write $T^{(3)}_{n,m} = \sum_{i=0}^{n-2} w^{(1)}_{i,n,m} + \sum_{i=0}^{n-2} w^{(2)}_{i,n,m}$, where,

$$w^{(1)}_{i,n,m} = -3m^{-3}\sqrt{n}(Z_{i+1} - m - (Z_i - m))^2(Z_i - m)$$
$$w^{(2)}_{i,n,m} = m^{-4}\frac{9}{4}\sqrt{n}X_0^4\xi^{-4}_{i,n,m}(Z_{i+1} - Z_i)^2(Z_i - m)^2,$$

and $\xi_{i,n,m} \in [nJ_i, n\widehat{J}_i^{(m)}]$.

Using Lemma 5.2, if $|i - i'| \geq 2$, $\left|E(w^{(1)}_{i,n,m}w^{(1)}_{i',n,m})\right| = \left|E(w^{(1)}_{i,n,m})E(w^{(1)}_{i',n,m})\right| = 72m^{-4}n$.

By Cauchy–Schwarz's inequality, and (72) (with $k = 6$), if $|i - i'| \leq 1$, $\left|E(w^{(1)}_{i,n,m}w^{(1)}_{i',n,m})\right| \leq cm^{-3}n$.

We deduce,

$$E\left((\sum_{i=0}^{n-2} w^{(1)}_{i,n,m})^2\right) \leq c(m^{-4}n^3 + m^{-3}n^2). \tag{70}$$

By application of Theorem 3.4, $\sup_{i,n\geq 0,m\geq 17} E(\xi^{-8}_{i,n,m}) \leq \sup_{i,n\geq 0,m\geq 17} E(X_0^{-8} \vee (n\widehat{J}_i^{m})^{-8}) < \infty$. Using (72), we deduce $E((w^{(2)}_{i,n,m})^2) \leq cm^{-4}$.

Hence,

$$E\left((\sum_{i=0}^{n-2} w^{(2)}_{i,n,m})^2\right) \leq cn^2m^{-4}.$$

It follows

$$E\left((T^{(3)}_{n,m})^2\right) \leq c(m^{-3}n^4 + m^{-3}n^2) \tag{71}$$

Joining (67)–(69), we prove the Proposition. □

The previous result implies, that for fixed n, m (with large n, $\frac{n}{m}$) and for $\theta^2 = 0$, our estimator $\hat{\theta^2}_{n,m}$ may often take values greater than $\frac{\sqrt{180n}}{m}$. So, in practice, $\hat{\theta^2}_{n,m}$ can not be used to estimate a parameter θ^2, which is known to be of magnitude lower than $\frac{\sqrt{180n}}{m}$.

In Table 7, we compute $\frac{\sqrt{180n}}{m}$ for some values of n, m. Comparing with Tables 4 and 6, it appears that the theoretical equivalent of the deviation for $\theta^2 = 0$ is really close to the empirical deviation for $\theta^2 = 0.01$ and $\theta^2 = 0.1$.

Table 7 shows that if n and $\frac{m}{n}$ are lower than 100, we can not estimate a parameter $\theta^2 \leq 0.01$ by using $\hat{\theta^2}_{n,m}$.

We end this section, with the following lemma which was used to prove Proposition 5.1.

Lemma 5.2. *Assume that Z has a chi-square distribution with m degrees of freedom, then*

$$E(Z) = m$$
$$E((Z - m)^2) = 2m$$
$$E((Z - m)^3) = 8m$$
$$E((Z - m)^4) = 12m^2 + 48m, \quad E\left(\{(Z - m)^2 - 2m\}^2\right) = 16m^2 + 48m$$
$$\forall k \geq 0, \exists c(k) > 0, \forall m \geq 1, \quad E((Z - m)^k) \leq c(k)m^{\frac{k}{2}} \tag{72}$$

Table 7: Equivalent of the standard deviation of $\hat{\theta^2}_{n,m}$ for $\theta = 0$.

	n=10	n=20	n=50	n=100
$\frac{m}{n}$=1	4.24	3	1.90	1.34
$\frac{m}{n}$=4	1.06	0.75	0.47	0.34
$\frac{m}{n}$=10	0.42	0.30	0.19	0.13
$\frac{m}{n}$=20	0.21	0.15	0.09	0.07
$\frac{m}{n}$=50	0.085	0.06	0.038	0.027
$\frac{m}{n}$=100	0.042	0.030	0.019	0.013

Proof. We compute the generating moment function of $X - m$:

$$\psi(\lambda) = E(e^{\lambda(X-m)}) = \frac{e^{-\lambda m}}{(1-2\lambda)^{\frac{m}{2}}} = \left(\frac{e^{-\lambda}}{\sqrt{1-2\lambda}}\right)^m$$

Now, combining the Taylor expansion of $\frac{e^{-\lambda}}{\sqrt{1-2\lambda}} = 1 + \lambda^2 + \dots$, and the one of $(1+u)^m = 1 + mu + \frac{m(m-1)}{2}u^2 + \dots$, we see that the k-th term in the expansion of $\psi(\lambda)$ is a polynomial function of m with degree $\leq \frac{m}{2}$. This implies (72).

With some computations we show the others inequalities. □

5.2 Some technicals lemmas

Lemma 5.3. • *Assume that we are in the independent case ($\chi = 0$). Let $k \geq 1$, then $\exists c$ such that for all s, $H \geq 0$, with $H \leq 1$:*

$$E\left(\sup_{h\in[0,H]} |Y_{s+h} - Y_s|^k \mid \mathcal{G}_{s,B}\right) \leq cH^{\frac{k}{2}} \sup_{h\in[s,s+H]} \mathcal{B}^k_\infty(X_s) \tag{73}$$

$$E\left(\sup_{h\in[0,H]} |Y_{s+h} - Y_s|^k \mid \mathcal{G}_s\right) \leq cH^{\frac{k}{2}}\mathcal{B}^k_\infty(X_s) \tag{74}$$

• *In the dependent case ($\chi \neq 0$) (74) remains valid.*

Proof. For $\chi = 0$, W is a brownian motion independent of B, we can apply Burkhölder's inequality to $Y_{s+h} - Y_s = \int_s^{s+h} \rho(X_s)ds + \int_s^{s+h} \sigma_s dW_s$:

$$E\left(\sup_{h\in[0,H]} (Y_{s+h} - Y_s)^k \mid \mathcal{G}_{s,B}\right) \leq cH^k E\left(\sup_{h\in[0,H]} \rho(X_{s+h})^k \mid \mathcal{G}_{s,B}\right) + cE\left((\int_s^{s+H} X_s ds)^{\frac{k}{2}} \mid \mathcal{G}_{s,B}\right)$$

Using (A4), we bound $\rho(X_{s+h})$ and X_{s+h} by $c\mathcal{B}_\infty(X_{s+h})$, and use that (X_t) is $\mathcal{G}_{s,B}$ measurable to get (73). We deduce (74) by Assumption (A3).

In the dependent case we show inequality (74) by applying Burkhölder's inequality to bound $E\left(\sup_{h\in[0,H]}(Y_{s+h} - Y_s)^k \mid \mathcal{G}_s\right)$. □

Lemma 5.4. • *Assume that we are in the independent case. There exists c such that for all $0 \le i \le n-1$, $0 \le j \le m-1$:*

$$E\left(\alpha^4_{i,n,j,m} \mid \mathcal{G}^n_{i,B}\right) \le c(m^{-1}\Delta_n)^6 \sup_{s\in[i\Delta_n,(i+1)\Delta_n]} \mathcal{B}^8_\infty(X_s) \tag{75}$$

$$E\left(\alpha^4_{i,n,j,m} \mid \mathcal{G}^n_{i}\right) \le c(m^{-1}\Delta_n)^6 \mathcal{B}^8_\infty(X_{i\Delta_n}) \tag{76}$$

$$E\left(\beta^4_{i,n,j,m} \mid \mathcal{G}^n_{i,B}\right) \le c(m^{-1}\Delta_n)^4 \sup_{s\in[i\Delta_n,(i+1)\Delta_n]} \mathcal{B}^6_\infty(X_s) \tag{77}$$

$$E\left(\beta^4_{i,n,j,m} \mid \mathcal{G}^n_{i}\right) \le c(m^{-1}\Delta_n)^4 \mathcal{B}^6_\infty(X_{i\Delta_n}) \tag{78}$$

• *In the dependent case* (76) *and* (78) *remains valid.*

Proof. By (25) and (A4),

$$\alpha^4_{i,n,j,m} \le 2^4\left(\frac{\Delta_n}{m}\right)^4 \sup_{s\in[(i+\frac{j}{m})\Delta_n,(i+\frac{j+1}{m})\Delta_n]} (Y_s - Y_{(i+\frac{j}{m})\Delta_n})^4 \sup_{s\in[i\Delta_n,(i+1)\Delta_n]} \rho^4(X_s,s)$$

$$\alpha^4_{i,n,j,m} \le 2^4\left(\frac{\Delta_n}{m}\right)^4 \sup_{s\in[(i+\frac{j}{m})\Delta_n,(i+\frac{j+1}{m})\Delta_n]} (Y_s - Y_{(i+\frac{j}{m})\Delta_n})^4 \sup_{s\in[i\Delta_n,(i+1)\Delta_n]} \mathcal{B}^4_\infty(X_s).$$

So we get in the independent case (75), by application of (73) with $k=4$; and we get in both cases (76) by using the Cauchy–Schwarz inequality in $E\left(\alpha^4_{i,n,j,m} \mid \mathcal{G}^n_i\right)$ and then using (74) and (A3) with $k=8$.

Using, in the independent case, the Burkhölder inequality:

$$E\left(\beta^4_{i,n,j,m} \mid \mathcal{G}^n_{i,B}\right) \le cE\left(\left(\int_{(i+\frac{j}{m})\Delta_n}^{(i+\frac{j+1}{m})\Delta_n} (Y_s - Y_{(i+\frac{j}{m})\Delta_n})^2 X_s ds\right)^2 \mid \mathcal{G}^n_{i,B}\right)$$

$$\le c(m^{-1}\Delta_n)^2\left(\sup_{s\in[i\Delta_n,(i+1)\Delta_n]} \mathcal{B}^2_\infty(X_s)\right) E\left(\sup_{s\in[(i+\frac{j}{m})\Delta_n,(i+\frac{j+1}{m})\Delta_n]} (Y_s - Y_{(i+\frac{j}{m})\Delta_n})^4 \mid \mathcal{G}^n_{i,B}\right)$$

Now, (73) yields (77).

We show (78) in both cases by similar computations. □

We recall, a useful lemma which is given in [6].

Lemma 5.5. *Let χ^n_i, U be random variables, with χ^n_i being $\mathcal{G}^n_i$-measurable. The following two conditions imply $\sum_{i=1}^n \chi^n_i \xrightarrow{\mathbf{P}} U$:*

$$\sum_{i=1}^n E(\chi^n_i \mid \mathcal{G}^n_{i-1}) \xrightarrow{\mathbf{P}} U$$

$$\sum_{i=1}^n E((\chi^n_i)^2 \mid \mathcal{G}^n_{i-1}) \xrightarrow{\mathbf{P}} 0$$

References

[1] O.E. Barndorff-Nielsen and N. Shephard. (1998), Aggregation and model construction for volatility models. Working paper series no.**10**, Center for Analytical Finance, University of Aarhus.

[2] G. Dohnal. (1987), On estimating the diffusion coefficient. *Journal of Applied Probability*, **24**,105–114.

[3] F. Drost and B. Werker. (1996), Closing the GARCH gap: Continuous time GARCH modeling. *J. Econom.*, **74(1)**,31–57.

[4] E. Ghysels, A. Harvey, E. Renault (1996), Stochastic volatility. *Handbook of statistics*, **14**,119–192.

[5] W. Feller. (1969), *An introduction to probability theory and its application*, volume 1. third edition.

[6] V. Genon-Catalot and J. Jacod. (1993), On the estimation of the diffusion coefficient for multidimensional diffusion processes. *Annales de l'I.H.P.,Probabilité-Statistiques*, **29**, 119–151, 1993.

[7] V. Genon-Catalot, T. Jeantheau, and C. Laredo. (1999), Parameter estimation for discretely observed stochastic volatility models. *To appear in Bernoulli.*

[8] A. Gloter. (1998), Discrete sampling of an integrated diffusion process and a comparison with the Euler scheme. *Prépublication de l'université de Marne-la-Vallée*, **23/98**.

[9] J. Hull and A. White. (1987), The pricing of options on assets with stochastic volatilities. *J. Finance*, **42**, 281–300.

[10] B. Leblanc. (1997), *Modélisation de la Volatilité d'un Actif Financier et Applications.* PhD thesis, Université Paris 7.

[11] D. Nelson. (1992), Filtering and forcasting with misspecified ARCH model I. getting the right variance with the wrong model. *J. Econom.*, **52(1/2)**,61–90.

[12] D. Nelson and D. Foster. (1994), Asymptotic filtering theory for univariate ARCH model. *Econometrica*, **62(1)**,1–41.

[13] M. Sørensen. (1999), Prediction–based estimating functions. *Preprint.*

Chapitre II.2. Estimation de paramètres pour une diffusion cachée

Parameter estimation for a hidden diffusion

Abstract

Assume that (Y_t, X_t) is a two-dimensional diffusion, where (X_t) is an autonomous positive diffusion, with ergodic properties, and (Y_t) has for diffusion coefficient $X_t^{\frac{1}{2}}$. We observe (Y_t) according to a double discretization at times $(i + \frac{j}{m})\Delta_n$, $i = 0, \ldots, n-1$, $j = 0, \ldots, m$, where $\Delta_n \to 0$ and $n\Delta_n \to \infty$. Introducing $\hat{X}_i = \Delta_n^{-1} \sum_{j=0}^{m-1} \left(Y_{(i+\frac{j+1}{m})\Delta_n} - Y_{(i+\frac{j}{m})\Delta_n} \right)^2$, the quadratic variation of the observations within the subinterval $[i\Delta_n, (i+1)\Delta_n]$, we prove limit theorems for some functionals of $(\hat{X}_i, i = 0, \ldots, n-1)$. These enable us to construct an appropriate contrast which yields consistent and asymptotically Gaussian estimators of parameters governing X, as $m = m_n$ tends to infinity with n. Drift and diffusion coefficient parameters of X are estimated with different rates under conditions on m_n and Δ_n. Some examples are presented with a numerical study.

The aim of this paper is to estimate unknown parameters of a hidden diffusion in the framework of the continuous time stochastic volatility models introduce by Hull and White (1987). Consider the two–dimensional diffusion process (Y_t, X_t) given by:

$$dY_t = \rho(X_t, t)dt + \sigma_t dW_t,\ Y_0 = \eta' \tag{1}$$

$$X_t = \sigma_t^2, \quad dX_t = b(X_t, \mu_0, \vartheta_0)dt + a(X_t, \vartheta_0)dB_t,\ X_0 = \eta. \tag{2}$$

where (B, W) is a Brownian motion of $\mathbb{R}^2$, (η', η) are random variables independent of (B, W) (with $\eta > 0$) and $(\mu_0, \vartheta_0) \in \mathbb{R}^2$ are unknown parameters. The function $\rho(x,t)$ may be known or unknown. In such a model, $Y_t = \ln S_t$ represents the logarithm of some asset price S_t, and X_t is the stochastic volatility of the asset.

This model has two noteworthy features. First, stochastic volatility models appear as continuous time limits of discrete time ARCH models introduced by Engle (1982) (see Nelson (1990)). Second, computation of option prices in stochastic volatility models are consistent with the empirical "smile" of volatility effect (see Chesney and Scott (1989)).

A practical use of these models for options pricing requires to estimate the unknown parameters governing the process X, while the knowledge of ρ is useless. However this volatility process is unobservable.

Some authors use observations of option prices in the market to infer parameters of X (see e.g. Pastorello *et al* (1994), Renault and Touzi (1996), Moreaux, Navatte and Villa (1996)).

Others use a discrete sampling of the trajectory of Y. Indeed, high frequency data for the process (Y_t) are generally available.

The problem of estimating the unknown parameters of the volatility process X_t from a discrete observation of the coordinate (Y_t) has been the subject of several recent contributions. As it is always the case for discretely observed diffusion process, a direct likelihood approach is hardly tractable (see Genon–Catalot *et al.* (1999, a)).

Assuming that the volatility process is ergodic, several kinds of explicit estimators have been proposed. In the case of observations with fixed sampling interval, empirical moments estimators are possible (see e.g. Wiggins (1987), Scott (1987) , Genon–Catalot *et al.* (1998)). However, such estimators may be strongly biased (see e.g. Ruiz (1994)). Otherwise the prediction–based estimating equations constructed by Sørensen (1999) also provide explicit estimators. Nevertheless, this method may require high computation time even for low dimensional parameters.

Other authors have suggested inference methods based on simulation (see e.g. Gouriéroux and Monfort (1993), Gouriéroux *et al.* (1993)) or a non linear filtering approach (see e.g. Del Moral *et al* (1998), Frey and Runggaldier (1999), Genon–Catalot *et al.* (1999, b)).

On the other hand, an approach which is well–fitted to high–frequency data, is to assume that $\Delta = \Delta_n$ tends to 0 while $n\Delta_n \to \infty$. For instance, an explicit method based on the observation of a discrete sample $(Y_{i\Delta_n}, i \le n)$ is proposed in Genon–Catalot *et al.* (1999, a). This method is able to estimate the unknown parameters present in the stationary distribution of the unobserved diffusion X. Our purpose, here, is to infer all the parameters of the (X_t) model, under the same ergodicity assumption on X.

To this aim, we assume that the sample path (Y_t) is observed according to a double discretization at times $(i+\frac{j}{m})\Delta_n$, with $i=0,\ldots,n-1$, $j=0,1,\ldots,m$. So, we have $N=nm+1$ regularly spaced observations with sampling interval $m^{-1}\Delta_n$.

Our idea is the following. In a previous paper (Gloter (1999, b)), we have studied the problem of estimating the unknown parameters of the (X_t) model when the observation is given by the sample $(\overline{X}_{i,n})$ with

$$\overline{X}_{i,n}=\Delta_n^{-1}\int_{i\Delta_n}^{(i+1)\Delta_n} X_s ds.$$

Now, for $m\geq 1$, we consider the natural approximation of $\overline{X}_{i,n}$, based on the observations:

$$\hat{X}_{i,n,m}=\Delta_n^{-1}\sum_{j=0}^{m-1}\left(Y_{(i+\frac{j+1}{m})\Delta_n}-Y_{(i+\frac{j}{m})\Delta_n}\right)^2.$$

In another previous paper (Gloter (1999, a)), we have investigated the properties of $(\hat{X}_{i,n,m})$ as an approximation of $\overline{X}_{i,n}$ and $X_{i\Delta_n}$.

Here, relying on the results of both papers, we prove further results on the variation and quadratic variation of $(\hat{X}_{i,n,m})$. These are obtained by letting $m=m_n$ tend to infinity and Δ_n tend to 0 with appropriate rates. Then, we deduce explicit estimators of ϑ_0, μ_0 which are expressed as functions of $(\hat{X}_{i,n,m})$ and specify their asymptotic behaviour as $n\to\infty$.

To maintain proofs shorter, we suppose that $\rho(x,t)=0$ (see (1)). The case $\rho\neq 0$ is discussed in the last section.

The paper is organized as follows. In Sections 1–2, which are devoted to probabilistic properties, parameters are omitted and we set $a(x)=a(x,\vartheta_0)$, $b(x)=b(x,\mu_0,\vartheta_0)$. Section 1 contains the assumptions and a recap of the results proved in Gloter (1999, a, b) which are needful for the comprehension of this paper. In Section 2, we study the following functionals of the observations (f is a function):

$$\hat{\nu}_{n,m}(f)=n^{-1}\sum_{i=0}^{n-1} f(\hat{X}_{i,n,m}) \tag{3}$$

$$\hat{\mathcal{I}}_{n,m}(f)=(n\Delta_n)^{-1}\sum_{i=0}^{n-1} f(\hat{X}_{i,n,m})(\hat{X}_{i+1,n,m}-\hat{X}_{i,n,m}-\Delta_n b(\hat{X}_{i,n,m})) \tag{4}$$

$$\hat{\mathcal{Q}}_{n,m}(f)=(n\Delta_n)^{-1}\sum_{i=0}^{n-1} f(\hat{X}_{i,n,m})(\hat{X}_{i+1,n,m}-\hat{X}_{i,n,m})^2. \tag{5}$$

First, we have convergence in probability results. If ν_0 denotes the stationary distribution of the diffusion X, then, as $m=m_n\to\infty$, $\hat{\nu}_{n,m_n}\to\nu_0(f)$ (Proposition 2.5). For the two other functionals, the results appear more surprising but coincide with the analogous results obtained in Gloter (1999, b) for the same functionals with $\overline{X}_{i,n}$ instead of $\hat{X}_{i,n,m}$: if $m_n^{-1}=o(\Delta_n)$, then (Theorem 2.6):

$$\hat{\mathcal{I}}_{n,m_n}(f)\xrightarrow{n\to\infty} 1/6\nu_0(a^2f')$$
$$\hat{\mathcal{Q}}_{n,m_n}(f)\xrightarrow{n\to\infty} 2/3\nu_0(a^2f).$$

A related central limit theorem is obtained (Theorem 2.7), under the additional conditions $n\Delta_n^2 \to 0$ and $m_n^{-1} = o(n^{-1}\Delta_n)$.

In Section 3, we introduce a contrast which is a modification of the Euler contrast of the diffusion X taking into account the limit theorems of Section 2. The associated minimum contrast estimators $(\hat{\mu}_{n,m_n}, \hat{\vartheta}_{n,m_n})$ are proved to be consistent and asymptotically Gaussian with respective rates $((n\Delta_n)^{\frac{1}{2}}, n^{\frac{1}{2}})$. The estimator of the drift parameters has the same rate and asymptotic variance of any efficient drift estimator based on $(X_t, t \in [0, n\Delta_n])$. The combination of the condition on n, Δ_n, m_n yields a rather slow rate of convergence $n^{\frac{1}{2}}$ for the diffusion coefficient parameters with respect to the total number of observations (nm_n). Explicit examples and numerical simulations results are presented in Section 4. In Section 5, we state the extension of our results to the case where the two Brownians motion B and W are correlated and $\rho \neq 0$.

1 Framework and preliminary results

1.1 Model and assumptions

Let (Y_t, X_t) be the two–dimensional diffusion process defined as the solution on a probability space $(\Omega, \mathcal{A}, P)$ of:

$$dY_t = \sigma_t dW_t, \qquad Y_0 = \eta' \tag{6}$$

$$\sigma_t^2 = X_t, \quad dX_t = b(X_t)dt + a(X_t)dB_t, \qquad X_0 = \eta \tag{7}$$

where $(B_t)_{t\geq 0}$ and $(W_t)_{t\geq 0}$ are two independent Brownian motions. We assume that the initial random variables (η, η') are independent of $(W_t, B_t)_{t\geq 0}$ and that η is positive.

We make classical assumptions on a and b which ensure that the solution of (7) is positive recurrent on $(0, \infty)$.

First, we introduce the two following functions defined on $(0, \infty)$:

$$\mathcal{B}_0(x) = 1 + x^{-k_1} \text{ and } \mathcal{B}_\infty(x) = 1 + x^2,$$

with $k_1 \geq 0$. The choice of the constant k_1 will be specified in each of the explicit models (7). These functions enable us to bound separately the behaviour of other functions near 0 and ∞. We shall use the following properties of these functions.

For all five non negative real numbers $\alpha, \beta, \alpha', \beta', p$, there exists a constant c such that for all $x \in (0, \infty)$:

$$(\mathcal{B}_0^\alpha(x) + \mathcal{B}_\infty^\beta(x)) \times (\mathcal{B}_0^{\alpha'}(x) + \mathcal{B}_\infty^{\beta'}(x)) \leq c(\mathcal{B}_0^{\alpha+\alpha'}(x) + \mathcal{B}_\infty^{\beta+\beta'}(x)) \tag{8}$$

$$(\mathcal{B}_0^\alpha(x) + \mathcal{B}_\infty^\beta(x))^p \leq c(\mathcal{B}_0^{p\alpha}(x) + \mathcal{B}_\infty^{p\beta}(x)). \tag{9}$$

Our set of assumptions is:

(A0) Equation (7) admits a unique strong solution such that $P(X_t > 0, \forall t \geq 0) = 1$.

(A1) $a,\ b \in \mathcal{C}^2(0,\infty)$ and $\exists c > 0,\ \exists \alpha_1, \alpha_2, \beta_1, \beta_2 \geq 0,\ \forall x \in (0,\infty)$,

$$\begin{aligned}
&|a(x)| + |b(x)| \leq c\mathcal{B}_\infty(x), \quad a(x) > 0, \\
&\left|a'(x)\right| \leq c(\mathcal{B}_0^{\alpha_1}(x) + \mathcal{B}_\infty^{\alpha_1}(x)), \quad \left|a''(x)\right| \leq c(\mathcal{B}_0^{\alpha_2}(x) + \mathcal{B}_\infty^{\alpha_2}(x)), \\
&\left|b'(x)\right| \leq c(\mathcal{B}_0^{\beta_1}(x) + \mathcal{B}_\infty^{\beta_1}(x)), \quad \left|b''(x)\right| \leq c(\mathcal{B}_0^{\beta_2}(x) + \mathcal{B}_\infty^{\beta_2}(x)).
\end{aligned}$$

Let us set

$$\mathcal{G}_t = \sigma((W_s, B_s), s \leq t;\ \eta;\ \eta'), \quad \mathcal{G}_{t,B} = \sigma(W_s,\ s \leq t; B_s,\ s \geq 0;\ \eta;\ \eta'). \tag{10}$$

(A2) There exists a positive constant K_0 such that
$\forall\ k \in [0, K_0),\ \exists c,\ \forall t \geq 0,\ E\left(\sup_{s\in[t,t+1]} \mathcal{B}_0^k(X_s) \mid \mathcal{G}_t\right) \leq c\mathcal{B}_0^k(X_t)$
$\forall\ k \in [0, \infty),\ \exists c,\ \forall t \geq 0,\ E\left(\sup_{s\in[t,t+1]} \mathcal{B}_\infty^k(X_s) \mid \mathcal{G}_t\right) \leq c\mathcal{B}_\infty^k(X_t)$

For $x_0 \in (0,\infty)$, we define the scale density $s(x) = \exp(-2\int_{x_0}^x \frac{b(u)}{a^2(u)}du)$, and the speed density $m(x) = \frac{1}{a^2(x)s(x)}$.

(A3) $\int_0 s(x)dx = \int^\infty s(x)dx = \infty, \int_0^\infty m(x)dx = M < \infty$.

Let

$$\nu_0(dx) = \frac{1}{M} m(x) \mathbb{1}_{\{x\in(0,\infty)\}} dx. \tag{11}$$

(A4) $\exists M_0,\ M_\infty > 0\ \ \nu_0(\mathcal{B}_0^{M_0}) < \infty,\ \nu_0(\mathcal{B}_\infty^{M_\infty}) < \infty$.

(A5) $\sup_{t\geq 0} E(\mathcal{B}_0^{M_0}(X_t)) < \infty, \quad \sup_{t\geq 0} E(\mathcal{B}_\infty^{M_\infty}(X_t)) < \infty$.

Let us remark that (A1) and (A3) imply (A0), but in the sequel some results hold without (A3). Under (A3), ν_0 is the unique invariant probability of model (7), and X satisfies the classical ergodic theorem:

$$\forall f \in \mathbf{L}^1(d\nu_0), \frac{1}{T}\int_0^T f(X_s)ds \xrightarrow[a.s.]{T\to\infty} \nu_0(f).$$

Using Assumption (A4), any Borel function which satisfies the condition, $\exists c \geq 0,\ \forall x,\ |f(x)| \leq c(\mathcal{B}_0^{M_0}(x) + \mathcal{B}_\infty^{M_\infty}(x))$, is an element of $\mathbf{L}^1(d\nu_0)$.

Assumptions (A2) and (A5) are crucial tools in proofs. The checking of these assumptions is treated in full details in Gloter (1998). It is there shown that (A2) holds for all classical diffusion processes used in finance to model the volatility. Assumption (A5) follows immediately from (A4) when η has distribution ν_0, but also when η is deterministic. In Section 4, we give examples of parametric models satisfying (A3) and (A5). The reason why (A2) is not symetric appears in Example 4.2.

1.2 Approximation of the integrated volatility (hidden diffusion)

Let Δ_n be a sequence of positive numbers tending to 0 as $n \to \infty$ (and for convenience $\forall n,\ \Delta_n \leq 1$).

We define for $i = 0, \ldots, n-1$ the (unobserved) variables:

$$\overline{X}_i = \overline{X}_{i,n} = \Delta_n^{-1} \int_{i\Delta_n}^{(i+1)\Delta_n} X_s ds. \tag{12}$$

Now, let m be an integer; we assume that the sample path $(Y_t)_{t\in[0,n\Delta_n]}$ is observed with a regular sampling interval of length $\frac{\Delta_n}{m}$ (hence, the total number of observations N is $nm+1$). We define the observed approximation of $\overline{X}_i$, for $i = 0, \ldots, n-1$,

$$\hat{X}_i = \hat{X}_{i,n,m} = \Delta_n^{-1} \sum_{j=0}^{m-1} (Y_{(i+\frac{j+1}{m})\Delta_n} - Y_{(i+\frac{j}{m})\Delta_n})^2 \tag{13}$$

In a previous paper (Gloter (1999, a)), we have studied the properties of $\hat{X}_i$ as an approximation of $\overline{X}_i$ or $X_{i\Delta_n}$. Some of these results are needed for the limit theorems of Section 2 on which rely the statistical applications of Section 3. We recall these now. Additional notations are useful.

First, we introduce the following σ–algebras,

$$\mathcal{G}_i^n = \mathcal{G}_{i\Delta_n}, \quad \mathcal{G}_{i,B}^n = \mathcal{G}_{i\Delta_n,B} \qquad \text{(recall (10))}. \tag{14}$$

Second, we define the following random variables

$$E_{i,n,m} = \Delta_n^{-1} \sum_{j=0}^{m-1} 2 \int_{(i+\frac{j}{m})\Delta_n}^{(i+\frac{j+1}{m})\Delta_n} (Y_s - Y_{(i+\frac{j}{m})\Delta_n})\sigma_s dW_s \tag{15}$$

Clearly,

$$\forall i, j \geq 0 \text{ with } i \neq j, \quad E\left(E_{i,n,m} \mid \mathcal{G}_{i,B}^n\right) = 0, \quad E\left(E_{i,n,m}E_{j,n,m} \mid \mathcal{G}_{0,B}\right) = 0. \tag{16}$$

These random variables are connected to the $\hat{X}_i$'s by the following proposition.

Proposition 1.1. *We have*

$$\hat{X}_i = \overline{X}_i + E_{i,n,m}, \ \textit{(recall (12)-(13))} \tag{17}$$

Furthermore, under (A0), (A1), (A2), the following bounds hold for the error term $E_{i,n,m}$:

$$\exists c, \forall i, n, m \geq 0, E\left(E_{i,n,m}^2 \mid \mathcal{G}_i^n\right) \leq cm^{-1}\mathcal{B}_\infty^4(X_{i\Delta_n}), \quad E\left(E_{i,n,m}^4 \mid \mathcal{G}_i^n\right) \leq cm^{-2}\mathcal{B}_\infty^8(X_{i\Delta_n}). \tag{18}$$

The inequalities (18) yield that the error $E_{i,n,m}$ has a decay of order (at least) $m^{-\frac{1}{2}}$. Furthermore, this error is centered conditionally on (B) (see (16)).

To end this section, we recall a technical result on the integrability of $\hat{X}_i$ and $\hat{X}_i^{-1}$ (again proofs may be found in Gloter (1999, a)).

Proposition 1.2. *Assume (A0), (A1), (A2), then the following holds:*

$$\forall k \in [0, K_0) \text{ (recall (A2))}, \ \exists c \geq 0, \ \forall i, n \geq 0, \ \forall m \geq 2k+3, \quad E\left(\mathcal{B}_0^k(\hat{X}_i) \mid \mathcal{G}_i^n\right) \leq c\mathcal{B}_0^k(X_{i\Delta_n}) \tag{19}$$

$$\forall k \geq 1, \ \exists c, \forall i, n, m \geq 0, \quad E\left(\mathcal{B}_\infty^k(\hat{X}_i) \mid \mathcal{G}_i^n\right) \leq c\mathcal{B}_\infty^{2k}(X_{i\Delta_n}) \tag{20}$$

These inequalities are powerful tools in proofs to bound $f(\hat{X}_i)$ when f is a function which tends to ∞ near 0 or ∞.

1.3 Expansions for the integrated volatility and related limit theorems

In another previous paper (Gloter (1998)), we have investigated the properties of $\overline{X}_i$ as an approximation of $X_{i\Delta_n}$. As a consequence, we have obtained general limit theorems for functionals of the sample $(\overline{X}_i)$ identical to those given in the introduction but with $\overline{X}_i$ instead of $\hat{X}_i$. We briefly recall these results.

First, under Assumptions (A0)–(A2), the following inequality holds:

$$\forall k > 0, \ \exists c > 0, \ \forall i, n \geq 0, \quad E\left(\left|\overline{X}_{i+1} - \overline{X}_i\right|^k \mid \mathcal{G}_i^n\right) \leq c\Delta_n^{\frac{k}{2}}\mathcal{B}_\infty^k(X_{i\Delta_n}) \tag{21}$$

This means that $\overline{X}_{i+1} - \overline{X}_i$ is of order $\Delta_n^{\frac{1}{2}}$. Now, we state the two main expansions of the process $(\overline{X}_i)$ proved in Gloter (1998).

Theorem 1.3. *Assume (A0)-(A2), then the following holds.*

$$\overline{X}_{i+1} - \overline{X}_i - b(\overline{X}_i)\Delta_n = \Delta_n^{\frac{1}{2}}a(X_{i\Delta_n})(\xi_{i,n} + \xi'_{i+1,n}) + \varepsilon_{i,n} \tag{22}$$

$$\overline{X}_i - X_{i\Delta_n} = \Delta_n^{\frac{1}{2}}a(X_{i\Delta_n})\xi'_{i,n} + e_{i,n} \tag{23}$$

with if $4\alpha_1 \vee 2\alpha_2 < K_0$, $E\left(\varepsilon_{i,n}^2 + e_{i,n}^2 \mid \mathcal{G}_i^n\right) \leq \Delta_n^2 c(\mathcal{B}_0^{2\alpha_1 \vee 2\alpha_2}(X_{i\Delta_n}) + \mathcal{B}_\infty^{3+2\alpha_1+\alpha_2}(X_{i\Delta_n}))$. *The variables* $\xi_{i,n}$, $\xi'_{i,n}$ *are Gaussian and have the explicit expressions*

$$\xi_{i,n} = \Delta_n^{-\frac{3}{2}}\int_{i\Delta_n}^{(i+1)\Delta_n}(s - i\Delta_n)dB_s, \quad \xi'_{i,n} = \Delta_n^{-\frac{3}{2}}\int_{i\Delta_n}^{(i+1)\Delta_n}(i\Delta_n + \Delta_n - s)dB_s.$$

And $var(\xi_{i,n}) = var(\xi'_{i,n}) = 1/3$, $cov(\xi_{i,n}, \xi'_{i,n}) = 1/6$.

Remark 1.4. *We can deduce from Theorem 1.3 and* (17) *the following expansions for the quadratic variations* $\hat{X}_i$ *(see* (13)*).*

$$\hat{X}_{i+1} - \hat{X}_i - b(\hat{X}_i)\Delta_n = \Delta_n^{\frac{1}{2}}a(X_{i\Delta_n})(\xi_{i,n} + \xi'_{i+1,n}) + E_{i+1,n,m} - E_{i,n,m} + \Delta_n\{b(\overline{X}_i) - b(\hat{X}_i)\} + \varepsilon_{i,n},$$

$$\hat{X}_i - X_{i\Delta_n} = \Delta_n^{\frac{1}{2}}a(X_{i\Delta_n})\xi'_{i,n} + E_{i,n,m} + e_{i,n}.$$

These expansions will enable us in Section 2 to find the appropriate rate at which m must tend to infinity with n.

Now, we shall make use of Assumptions (A3)–(A5), and assume that $n\Delta_n \to \infty$. Let us introduce for $f : (0,\infty) \to \mathbb{R}$,

$$\overline{\nu}_n(f) = n^{-1}\sum_{i=0}^{n-1} f(\overline{X}_i) \tag{24}$$

$$\overline{\mathcal{I}}_n(f) = (n\Delta_n)^{-1}\sum_{i=0}^{n-2} f(\overline{X}_i)(\overline{X}_{i+1} - \overline{X}_i - \Delta_n b(\overline{X}_i)), \tag{25}$$

$$\overline{\mathcal{Q}}_n(f) = (n\Delta_n)^{-1}\sum_{i=0}^{n-2} f(\overline{X}_i)(\overline{X}_{i+1} - \overline{X}_i)^2. \tag{26}$$

Smoothness conditions on f are required:

• A function f satisfies $\mathbf{C}_\gamma$, for some $\gamma \geq 0$, if:
$f : (0,\infty) \to \mathbb{R}$ belongs to $\mathcal{C}^2(0,\infty)$ and

$$\exists c > 0,\ \forall x \in (0,\infty) \quad \sup_{\theta\in\Theta} |g(x,\theta)| \leq c(\mathcal{B}_l^\gamma(x) + \mathcal{B}_r^\gamma(x)),\ \text{for } g = f,\ f'_x,\ f''_{x^2}.$$

For statistical purposes, we introduce a parameter set Θ equal to a product of two compact intervals of $\mathbb{R}$, and a more precise condition on the functions.

• A function f satisfies $\mathbf{CU}_\gamma$, for some $\gamma \geq 0$, if:
$f : (0,\infty) \times \Theta \to \mathbb{R}$ belongs to $\mathcal{C}^2[(0,\infty) \times O]$ for some open set $O \supset \Theta$ and

$$\exists c > 0,\ \forall x \in (0,\infty) \quad \sup_{\theta\in\Theta} |g(x,\theta)| \leq c(\mathcal{B}_l^\gamma(x) + \mathcal{B}_r^\gamma(x)),\ \text{for } g = f,\ f'_x,\ f''_{x^2},\ \nabla_\theta f,\ \nabla_\theta f'_x.$$

Now, we state results of Gloter (1999, b).

Theorem 1.5. *Assume (A1)–(A5), with $K_0 = \infty$ and that M_0, M_∞ (in (A2), (A4), (A5)) can be chosen as large as we want.*

- *Let f and g satisfy $\mathbf{CU}_\gamma$, with some $\gamma \geq 0$, then*

$$\overline{\nu}_n(f(.,\theta)) \xrightarrow{n\to\infty} \nu_0(f(.,\theta)), \quad \textit{uniformly in } \theta, \textit{ in probability} \tag{27}$$

$$\overline{\mathcal{I}}_n(f(.,\theta)) \xrightarrow{n\to\infty} \frac{1}{6}\nu_0(f'_x(.,\theta)a^2(.)), \ \textit{uniformly in } \theta, \textit{ in probability} \tag{28}$$

$$\overline{\mathcal{Q}}_n(f(.,\theta)) \xrightarrow{n\to\infty} \frac{2}{3}\nu_0(f(.,\theta)a^2(.)), \quad \textit{uniformly in } \theta, \textit{ in probability} \tag{29}$$

- *Assume that f and g satisfy $\mathbf{C}_\gamma$, with some $\gamma \geq 0$, and that $n\Delta_n^2 \xrightarrow{n\to\infty} 0$. Then,*

$$\begin{bmatrix} \sqrt{n\Delta_n}(\overline{\mathcal{I}}_n(f) - \frac{1}{4}\overline{\mathcal{Q}}_n(f')) \\ \sqrt{n}(\frac{3}{2}\overline{\mathcal{Q}}_n(g) - \overline{\nu}_n(ga^2)) \end{bmatrix} \xrightarrow[\mathcal{D}]{n\to\infty} \mathcal{N}\left(0, \begin{bmatrix} \nu_0(f^2a^2) & 0 \\ 0 & \frac{9}{4}\nu_0(g^2a^4) \end{bmatrix}\right) \tag{30}$$

Remark 1.6. 1) *In the case where M_0, M_∞ or K_0 can not be chosen as large as we need, we give in Gloter (1999, b) conditions linking γ with α_1, α_2, β_1, β_2 (see (A2)), M_0, M_∞ and K_0 which ensure again the validity of* (27)–(30).

2) *The rather unexpected results* (27)-(30) *are due to the expansions* (22)-(23) *and explain the analogous result for the functionals* $\hat{\mathcal{I}}_n$ *and* $\hat{\mathcal{Q}}_n$ *in Section 2 (see Remark 2.8 3)).*

3) *The condition* $n\Delta_n^2 \to 0$ *in the central limit theorem is classical (see e.g. Kessler (1997), Genon-Catalot, Jeantheau, Larédo (1999)). But, it is worth noting that* (30) *is not the central limit theorem that could be expected from* (28)-(29) *(see Gloter (1999, b) for details). This also explains the analogous theorem for* $\hat{\mathcal{I}}_n$ *and* $\hat{\mathcal{Q}}_n$.

2 Limit theorems.

Now, suppose that m_n is a sequence of integers tending to ∞ as $n \to \infty$. We aim to give conditions on the speed at which m_n tend to infinity which ensure the convergence of $\hat{\nu}_{n,m_n}(f)$, $\hat{\mathcal{I}}_{n,m_n}(f)$ and $\hat{\mathcal{Q}}_{n,m_n}(f)$ (see (3)-(5)) and compute the limits. Our results rely on Theorem 1.5 and preliminary lemmas which establish the exact rates of convergence of the differences between the "hat" and the "bar" functionals ((24)-(26)).

2.1 Preliminary lemmas

For the need of proofs, we introduce a condition less restrictive than $\mathbf{CU}_\gamma$.

• A function f satisfies $\mathbf{CU}_\gamma^1$, for some $\gamma \geq 0$, if:
$f : (0,\infty) \times \Theta \to \mathbb{R}$ belongs to $\mathcal{C}^2[(0,\infty) \times O]$ for some open set $O \supset \Theta$ and

$$\exists c > 0,\ \forall x \in (0,\infty) \quad \sup_{\theta\in\Theta} |g(x,\theta)| \leq c(\mathcal{B}_l^\gamma(x) + \mathcal{B}_r^\gamma(x)), \text{ for } g = f,\ f'_x,\ \nabla_\theta f.$$

Our first result is a lemma on the difference $f(\hat{X}_i) - f(\overline{X}_i)$. This lemma is used in the proof of other lemmas of this section.

Lemma 2.1. *Assume (A0)-(A2). Let* f *satisfy* $\mathbf{CU}_\gamma^1$ *and suppose* $k = 1$ *or* $k = 2$. *If* $2k\gamma < K_0$, *then* $\exists c,\ \forall i \geq 0,\ n \geq 0,\ m \geq 2k\gamma + 3$,

$$E\left(\sup_{\theta\in\Theta}\left|f(\hat{X}_i,\theta) - f(\overline{X}_i,\theta)\right|^k \mid \mathcal{G}_i^n\right) \leq cm^{-\frac{k}{2}}(\mathcal{B}_0^{k\gamma}(X_{i\Delta_n}) + \mathcal{B}_\infty^{2k\gamma+2k}(X_{i\Delta_n})) \tag{31}$$

Proof. Using Taylor's expansion, and (17), we get

$$\left|f(\hat{X}_i,\theta) - f(\overline{X}_i,\theta)\right|^k \leq \left|f'_x(\xi,\theta)\right|^k |E_{i,n,m}|^k \text{ where } \xi \in [\overline{X}_i, \hat{X}_i].$$

By condition $\mathbf{CU}_\gamma^1$ and the monotonicity of $\mathcal{B}_0$ and $\mathcal{B}_\infty$, we have

$$|f'_x(\xi,\theta)|^{2k} \leq c\left(\mathcal{B}_0^{2\gamma k}(\xi) + \mathcal{B}_\infty^{2\gamma k}(\xi)\right) \leq c\left(\mathcal{B}_0^{2\gamma k}(\overline{X}_i) + \mathcal{B}_0^{2\gamma k}(\hat{X}_i) + \mathcal{B}_\infty^{2\gamma k}(\overline{X}_i) + \mathcal{B}_\infty^{2\gamma k}(\hat{X}_i)\right) \tag{32}$$

Using Assumption (A2), (19)-(20), we deduce (for $m \geq 2\gamma + 3$),

$$E\left(\sup_{\theta\in\Theta}|f'(\xi,\theta)|^{2k} \mid \mathcal{G}_i^n\right) \leq c(\mathcal{B}_0^{2\gamma k}(X_{i\Delta_n}) + \mathcal{B}_\infty^{4\gamma k}(X_{i\Delta_n})).$$

Recalling that $k = 1$ or 2, (18) writes

$$E\left(|E_{i,n,m}|^{2k} \mid \mathcal{G}_i^n\right) \leq cm^{-k}\mathcal{B}_\infty^{4k}(X_{i\Delta_n}). \tag{33}$$

By application of the Cauchy–Schwarz inequality, (32)–(33) and (8) we deduce inequality (31). □

Lemma 2.2. *Assume (A1)–(A5) and let f satisfy $\mathbf{CU}_\gamma^1$ with $2\gamma < K_0$, $\gamma \leq M_0$, $2+2\gamma \leq M_\infty$. Let $m = m_n$ tend to infinity and consider a sequence (r_n) of real numbers. The condition $r_n m_n^{-\frac{1}{2}} \to 0$ implies:*

$$r_n(\hat{\nu}_{n,m_n}(f(.,\theta)) - \overline{\nu}_n(f(.,\theta))) \xrightarrow{n\to\infty} 0, \text{ uniformly in } \theta \text{ in probability.}$$

Proof. Set $\mathbf{A}_{i,n,m}(\theta) = f(\hat{X}_i,\theta) - f(\overline{X}_i,\theta)$. By Lemma 2.1, for $m \geq 2\gamma + 3$,

$$E\left(\sup_{\theta\in\Theta}|\mathbf{A}_{i,n,m}(\theta)| \mid \mathcal{G}_i^n\right) \leq cm^{-\frac{1}{2}}(\mathcal{B}_0^\gamma(X_{i\Delta_n}) + \mathcal{B}_\infty^{2+2\gamma}(X_{i\Delta_n})).$$

Using (A5), $\gamma \leq M_0$ and $2 + 2\gamma \leq M_\infty$, we deduce for m large enough:

$$E\left(\sup_{\theta\in\Theta}|\mathbf{A}_{i,n,m}(\theta)|\right) \leq cm^{-\frac{1}{2}}. \tag{34}$$

By the definition of $\mathbf{A}_{i,n,m}(\theta)$, (3), (24), and (34), we obtain the result. □

Lemma 2.3. *Assume (A1)–(A5) and let f satisfy $\mathbf{CU}_\gamma^1$ with $4\gamma \vee 4\beta_1 < K_0$, $2\gamma + \beta_1 < M_0$, $4 + 2\gamma + 2\beta_1 < M_\infty$. Let $m = m_n$ tend to infinity and consider a sequence (r_n) of real numbers such that r_n^{-1} is bounded. The condition $r_n\Delta_n^{-\frac{1}{2}}m_n^{-\frac{1}{2}} \to 0$ implies,*

$$r_n(\hat{\mathcal{I}}_{n,m_n}(f(.,\theta)) - \overline{\mathcal{I}}_n(f(.,\theta))) \xrightarrow{n\to\infty} 0, \text{ uniformly in } \theta \text{ in probability.}$$

Proof. We denote

$$\mathbf{B}_{i,n}(\theta) = \mathbf{B}_{i,n,m_n}(\theta) = (\hat{X}_{i+1} - \hat{X}_i - b(\hat{X}_i)\Delta_n)f(\hat{X}_i,\theta) - (\overline{X}_{i+1} - \overline{X}_i - b(\overline{X}_i)\Delta_n)f(\overline{X}_i,\theta)$$

and split $\mathbf{B}_{i,n}(\theta) = \sum_{k=1}^4 \mathbf{B}_{i,n}^{(k)}(\theta)$, where (recall (17)):

$$\begin{aligned}
\mathbf{B}_{i,n}^{(1)}(\theta) &= (\overline{X}_{i+1} - \overline{X}_i - b(\overline{X}_i)\Delta_n)(f(\hat{X}_i,\theta)) - f(\overline{X}_i,\theta)) \\
\mathbf{B}_{i,n}^{(2)}(\theta) &= (E_{i+1,n,m_n} - E_{i,n,m_n})f(\overline{X}_i,\theta) \\
\mathbf{B}_{i,n}^{(3)}(\theta) &= (E_{i+1,n,m_n} - E_{i,n,m_n})(f(\hat{X}_i,\theta) - f(\overline{X}_i,\theta)) \\
\mathbf{B}_{i,n}^{(4)}(\theta) &= \Delta_n(b(\overline{X}_i) - b(\hat{X}_i))f(\hat{X}_i,\theta)
\end{aligned}$$

The proof consists in showing the uniform convergence to 0 of $r_n(n\Delta_n)^{-1}\sum_{i=0}^{n-1}\mathbf{B}_{i,n}^{(l)}(\theta)$ for $l = 1,\ldots,4$.

Using Lemma 2.1 (with k=2) and (21), we obtain

$$E\left(\sup_{\theta\in\Theta}\left|\mathbf{B}_{i,n}^{(1)}\right| \mid \mathcal{G}_i^n\right) \leq cm_n^{-\frac{1}{2}}\Delta_n^{\frac{1}{2}}(\mathcal{B}_0^\gamma(X_{i\Delta_n}) + \mathcal{B}_\infty^{3+2\gamma}(X_{i\Delta_n}))$$

By Assumption (A5), we deduce

$$E(\sup_{\theta\in\Theta}\left|\mathbf{B}^{(1)}_{i,n}(\theta)\right|) \le c m_n^{-\frac{1}{2}}\Delta_n^{\frac{1}{2}} \tag{35}$$

Hence, the condition on m_n implies $r_n(n\Delta_n)^{-1}\sum_{i=0}^{n-1}\mathbf{B}^{(1)}_{i,n}(\theta) \xrightarrow[\mathrm{L}^1]{n\to\infty} 0$, uniformly in θ.

In order to prove the uniform convergence to zero of $r_n(n\Delta_n)^{-1}\sum_{i=0}^{n-1}\mathbf{B}^{(2)}_{i,n}(\theta)$, we have to split the expression into the sum of even indexes and odd indexes. Indeed, as we shall see now, this device implies that we deal with a sum of terms which are independent conditionally on $(X_t, t\ge 0)$. We only show the uniform convergence to zero of $r_{2n}(2n\Delta_{2n})^{-1}\sum_{i=0}^{n-2}\mathbf{B}^{(2)}_{2i,2n}(\theta)$. The proof for the sum of odd terms $r_{2n+1}(2n\Delta_{2n+1})^{-1}\sum_{i=0}^{n-2}\mathbf{B}^{(2)}_{2i+1,2n+1}(\theta)$ is the same.

For this, using Theorem 20 of the Appendix 1 in Ibragimov and Khas'minskii (1981), it is enough to show that there exists $\epsilon>0$ such that the two following properties hold:

$$\sup_{\theta\in\Theta,n\ge 0} E\left(\left(\frac{r_{2n}}{2n\Delta_{2n}}\right)^{2+\epsilon}\left|\sum_{i=0}^{n-2}\mathbf{B}^{(2)}_{2i,2n}(\theta)\right|^{2+\epsilon}\right) \xrightarrow{n\to\infty} 0 \tag{36}$$

$$\exists M,\ \forall(\theta,\theta')\in\Theta^2,\quad \sup_{n\ge 0} E\left(\left(\frac{r_{2n}}{2n\Delta_{2n}}\right)^{2+\epsilon}\left|\sum_{i=0}^{n-2}(\mathbf{B}^{(2)}_{2i,2n}(\theta)-\mathbf{B}^{(2)}_{2i,2n}(\theta'))\right|^{2+\epsilon}\right) \le M\left|\theta-\theta'\right|^{2+\epsilon} \tag{37}$$

((36) implies pointwise L^2–convergence and (36)–(37) gives the uniformity.)

We only show (36), since (37) can be obtained by a similar proof.

Using Rosenthal's inequality for martingales (see Hall and Heyde (1980) p.23), we get for any $\epsilon>0$.

$$E\left(\left|\sum_{i=0}^{n-2}\mathbf{B}^{(2)}_{2i,2n}(\theta)\right|^{2+\epsilon}\right) \le cE\left(\left|\sum_{i=0}^{n-2}E\left((\mathbf{B}^{(2)}_{2i,2n}(\theta))^2\mid\mathcal{G}^{2n}_{2i}\right)\right|^{1+\frac{\epsilon}{2}}\right)+c\sum_{i=0}^{n-2}E\left(\left|\mathbf{B}^{(2)}_{2i,2n}(\theta)\right|^{2+\epsilon}\right) \tag{38}$$

Applying the classical inequality $(\sum_{i=0}^{n-2}|a_i|)^p \le n^{p-1}\sum_{i=0}^{n-2}|a_i|^p$, with $p=1+\frac{\epsilon}{2}$, we deduce

$$E\left(\left|\sum_{i=0}^{n-2}E\left((\mathbf{B}^{(2)}_{2i,2n}(\theta))^2\mid\mathcal{G}^{2n}_{2i}\right)\right|^{1+\frac{\epsilon}{2}}\right) \le n^{\frac{\epsilon}{2}}\sum_{i=0}^{n-2}E\left(\left|E\left((\mathbf{B}^{(2)}_{2i,2n}(\theta))^2\mid\mathcal{G}^{2n}_{2i}\right)\right|^{1+\frac{\epsilon}{2}}\right) \tag{39}$$

Using (18) and (A5), we can show that if ϵ is small enough,

$$\sup_{i,n}E\left(\left|E\left((\mathbf{B}^{(2)}_{2i,2n}(\theta))^2\mid\mathcal{G}^{2n}_{2i}\right)\right|^{1+\frac{\epsilon}{2}}\right)\le cm_{2n}^{-1-\frac{\epsilon}{2}},\quad \sup_{i,n}E\left(\left|\mathbf{B}^{(2)}_{2i,2n}(\theta)\right|^{2+\epsilon}\right)\le cm_{2n}^{-1-\frac{\epsilon}{2}}. \tag{40}$$

From (38)–(40), we deduce

$$E\left(\left(\frac{r_{2n}}{2n\Delta_{2n}}\right)^{2+\epsilon}\left|\sum_{i=0}^{n-2}\mathbf{B}^{(2)}_{2i,2n}(\theta)\right|^{2+\epsilon}\right)\le c(r_{2n}(m_{2n}\Delta_{2n})^{-\frac{1}{2}})^{2+\epsilon}(n\Delta_{2n})^{-1-\frac{\epsilon}{2}}$$

But $(n\Delta_n)^{-1} \to 0$ and $r_n(\Delta_n m_n)^{-\frac{1}{2}} \to 0$. So, we get (36).

The convergence of the other sums is easier.

Using Lemma 2.1 and (18), we obtain

$$E(\sup_{\theta\in\Theta} \left|\mathbf{B}_{i,n}^{(3)}(\theta)\right|) \leq cm_n^{-1}. \tag{41}$$

At last, by Lemma 2.1, condition $\mathbf{CU}_\gamma$ and (19)–(20), we get:

$$E(\sup_{\theta\in\Theta} \left|\mathbf{B}_{i,n}^{(4)}(\theta)\right|) \leq cm_n^{-\frac{1}{2}}\Delta_n. \tag{42}$$

Now, equations (41)–(42), and the fact that (r_n^{-1}) is bounded imply, for $l = 4, 5$,

$$r_n(n\Delta_n)^{-1} \sup_{\theta\in\Theta} \sum_{i=0}^{n-2} \mathbf{B}_{i,n}^{(l)}(\theta) \xrightarrow[\mathbf{L}^1]{n\to\infty} 0 \text{ uniformly in } \theta. \tag{43}$$

□

Lemma 2.4. *Assume (A1)–(A5) and let f satisfy $\mathbf{CU}_\gamma^1$ with $4\gamma < K_0$, $\gamma < M_0$, $4 + 2\gamma < M_\infty$. Let $m = m_n$ and r_n be as in Lemma 2.3. Then the conditions $r_n\Delta_n^{-\frac{1}{2}}m_n^{-\frac{1}{2}} \to 0$ implies,*

$$r_n(\hat{\mathcal{Q}}_{n,m_n}(f(.,\theta)) - \overline{\mathcal{Q}}_n(f(.,\theta))) \xrightarrow{n\to\infty} 0, \text{ uniformly in } \theta \text{ in probability.}$$

Proof. We set

$$\mathbf{C}_{i,n}(\theta) = \mathbf{C}_{i,n,m_n}(\theta) = (\hat{X}_{i+1} - \hat{X}_i)^2 f(\hat{X}_i,\theta) - (\overline{X}_{i+1} - \overline{X}_i)^2 f(\overline{X}_i,\theta)$$

and split $\mathbf{C}_{i,n}(\theta) = \sum_{k=1}^3 \mathbf{C}_{i,n}^{(k)}(\theta)$ with:

$$\begin{aligned}
\mathbf{C}_{i,n}^{(1)}(\theta) &= (E_{i+1,n,m_n} - E_{i,n,m_n})^2 f(\hat{X}_i,\theta)\\
\mathbf{C}_{i,n}^{(2)}(\theta) &= 2(E_{i+1,n,m_n} - E_{i,n,m_n})(\overline{X}_{i+1} - \overline{X}_i) f(\hat{X}_i,\theta)\\
\mathbf{C}_{i,n}^{(3)}(\theta) &= (\overline{X}_{i+1} - \overline{X}_i)^2\{f(\hat{X}_i,\theta) - f(\overline{X}_i,\theta)\}
\end{aligned}$$

First, using condition $\mathbf{CU}_\gamma^{(1)}$ on f, (19), (20) and (18), we obtain

$$E(\sup_{\theta\in\Theta} \left|\mathbf{C}_{i,n}^{(1)}(\theta)\right|) \leq cm_n^{-1}(\mathcal{B}_0^\gamma(X_{i\Delta_n}) + \mathcal{B}_\infty^{4+2\gamma}(X_{i\Delta_n})). \tag{44}$$

By Assumption (A5),

$$E(\sup_{\theta\in\Theta} \left|\mathbf{C}_{i,n}^{(1)}(\theta)\right|) \leq cm_n^{-1}. \tag{45}$$

Second, by similar computations based on (21) and (18), we get:

$$E(\sup_{\theta\in\Theta} \left|\mathbf{C}_{i,n}^{(2)}(\theta)\right|) \leq cm_n^{-\frac{1}{2}}\Delta_n^{\frac{1}{2}}. \tag{46}$$

Last, by Lemma 2.1 and (21)

$$E(\sup_{\theta\in\Theta}\left|\mathbf{C}_{i,n}^{(3)}(\theta)\right|) \leq c m_n^{-\frac{1}{2}}\Delta_n. \tag{47}$$

By (45)–(47),

$$r_n(n\Delta_n)^{-1}E\left(\sum_{i=0}^{n-1}\sup_{\theta\in\Theta}|\mathbf{C}_{i,n}(\theta)|\right) \leq r_n((m_n\Delta_n)^{-\frac{1}{2}} + (m_n\Delta_n)^{-1}) \xrightarrow{n\to\infty} 0.$$

Recalling the definition of $\mathbf{C}_{i,n}$, (5) and (26), the lemma is proved. □

2.2 Main results

In the following statements, we assume that (A1)–(A5) are valid with, for simplicity, the stronger property that M_0, M_∞ and K_0 (in (A2), (A4)–(A5)) may be chosen as large as we want.

Proposition 2.5. *Let f satisfy $\mathbf{CU}_\gamma$ for some $\gamma \geq 0$ and suppose that $m_n \xrightarrow{n\to\infty} \infty$, then*

$$\hat{\nu}_{n,m_n}(f(.,\theta)) \xrightarrow[\mathbf{P}]{n\to\infty} \nu_0\left(f(.,\theta)\right) \text{ uniformly in } \theta.$$

Proof. We join (27) and Lemma 2.2 with $r_n = 1$. □

Theorem 2.6. *Let f satisfy $\mathbf{CU}_\gamma$ for some $\gamma \geq 0$ and suppose that $m_n^{-1} = o(\Delta_n)$. Then,*

$$\hat{\mathcal{I}}_{n,m_n}(f(.,\theta)) \xrightarrow[\mathbf{P}]{n\to\infty} \frac{1}{6}\nu_0\left(f'_x(.,\theta)a^2(.)\right) \text{ uniformly in } \theta.$$
$$\hat{\mathcal{Q}}_{n,m_n}(f(.,\theta)) \xrightarrow[\mathbf{P}]{n\to\infty} \frac{2}{3}\nu_0\left(f(.,\theta)a^2(.)\right) \text{ uniformly in } \theta.$$

Proof. For the convergence of $\hat{\mathcal{I}}_{n,m_n}(f)$, we join (28) and Lemma 2.3 with $r_n = 1$. To obtain the convergence $\hat{\mathcal{Q}}_{n,m_n}(f)$, we join (29) and Lemma 2.4 (with $r_n = 1$ again). □

Theorem 2.7. *Let f and g satisfy $\mathbf{C}_\gamma$ for some $\gamma \geq 0$, and suppose that $n\Delta_n^2 \to 0$. Then,*

$$\text{if } m_n^{-1} = o(n^{-1}), \quad \hat{N}_n(f) := \sqrt{n\Delta_n}(\hat{\mathcal{I}}_{n,m_n}(f) - \frac{1}{4}\hat{\mathcal{Q}}_{n,m_n}(f')) \xrightarrow[\mathcal{D}]{n\to\infty} \mathcal{N}\left(0, \nu_0\left(f^2a^2\right)\right) \tag{48}$$

$$\text{if } m_n^{-1} = o(n^{-1}\Delta_n), \quad \hat{M}_n(g) := \sqrt{n}(\frac{3}{2}\hat{\mathcal{Q}}_{n,m_n}(g) - \hat{\nu}_{n,m_n}(a^2g)) \xrightarrow[\mathcal{D}]{n\to\infty} \mathcal{N}\left(0, \frac{9}{4}\nu_0\left(g^2a^4\right)\right) \tag{49}$$

and under the second condition on m_n (which contains the first one), the vector $(\hat{N}_n(f), \hat{M}_n(g))$ converges to a Gaussian vector with independent coordinates.

Proof. Recalling (30), we use Lemma 2.3 with $r_n = (n\Delta_n)^{\frac{1}{2}}$ and Lemma 2.4 with f', $r_n = (n\Delta_n)^{\frac{1}{2}}$ to get (48).

By lemma 2.2 with a^2g, $r_n = n^{\frac{1}{2}}$ and Lemma 2.4 with $r_n = n^{\frac{1}{2}}$), we deduce (48). The independence is immediate by the diagonal form of the covariance matrix in (30). □

Remark 2.8. 1) *The condition of Theorem 2.6, $m_n^{-1} = o(\Delta_n)$, can be naturally interpreted. It means that in the expansions of Remark 1.4, the error terms $E_{i,n,m}$ are negligible versus the terms $\Delta_n^{\frac{1}{2}} a(X_{i\Delta_n})(\xi_{i,n} + \xi'_{i+1,n})$ and $\Delta_n^{\frac{1}{2}} a(X_{i\Delta_n})\xi'_{i,n}$.*

2) If the diffusion X satisfies (A1)–(A5) but, now, not for any positive M_0, M_∞ and K_0 then analogous results to those of Proposition 2.5 and Theorems 2.6–2.7 can be obtained but only for $f \in \mathbf{CU}_\gamma$ with certain values of γ.

3) If we set

$$\mathcal{I}_n(f) = (n\Delta_n)^{-1} \sum_{i=0}^{n-2} f(X_{i\Delta_n})(X_{(i+1)\Delta_n} - X_{i\Delta_n} - \Delta_n b(X_{i\Delta_n}))$$

$$\mathcal{Q}_n(f) = (n\Delta_n)^{-1} \sum_{i=0}^{n-2} f(X_{i\Delta_n})(X_{(i+1)\Delta_n} - X_{i\Delta_n})^2,$$

then it is well known that $\mathcal{I}_n(f) \underset{\mathbf{P}}{\longrightarrow} 0$ and $\mathcal{Q}_n(f) \underset{\mathbf{P}}{\longrightarrow} \nu_0\left(fa^2\right)$ (see e.g. Kessler (1997)). Hence, the asymptotic behaviour of the "hat" and the "without hat" functionals are strongly different.

3 An explicit contrast for the stochastic volatility model

In this section, we consider a parametric model for X with parameter $\theta_0 = (\mu_0, \vartheta_0)$ belonging to $\Theta = \Theta_1 \times \Theta_2$,

$$dX_t = a(X_t, \vartheta_0)dB_t + b(X_t, \mu_0, \vartheta_0)dt \tag{50}$$

where $a(x,\vartheta) : (0,\infty) \times \Theta_2 \to \mathbb{R}$, $b(x,\mu,\vartheta) : (0,\infty) \times \Theta_1 \times \Theta_2 \to \mathbb{R}$ and Θ_1, Θ_2 are two compact intervals of $\mathbb{R}$. We assume that the true value θ_0 of the parameter is in the interior of Θ.

Our goal is to estimate this parameter θ_0, from the observations $(Y_{(i+\frac{j}{m})\Delta_n})_{\substack{0\le i\le n \\ 0\le j\le m}}$ (recall (6)). For this, we define the contrast

$$\hat{U}_{n,m}(\theta) = \frac{3}{2} \sum_{i=0}^{n-2} \frac{(\hat{X}_{i+1} - \hat{X}_i - b(\hat{X}_i,\theta)\Delta_n)^2}{n\Delta_n a^2(\hat{X}_i,\vartheta)} + \hat{\nu}_{n,m}(\log a^2(.,\vartheta)) + \frac{3\Delta_n}{4}\hat{\mathcal{Q}}_{n,m}(h(.,\theta)) \tag{51}$$

where $h(x,\theta) = \partial_x\left(b(x,\theta)a^{-2}(x,\vartheta)\right)$ (∂_x stands for $\frac{\partial}{\partial x}$).

Remark that this contrast is a modification of the contrast based on the Euler scheme,

$$U_n(\theta) = \sum_{i=0}^{n-2} \frac{(X_{(i+1)\Delta_n} - X_{i\Delta_n} - b(X_{i\Delta_n},\theta)\Delta_n)^2}{n\Delta_n a^2(X_{i\Delta_n},\vartheta)} + \nu_n(\log a^2(.,\vartheta)),$$

which takes into account the differences between $\hat{X}_i$ and $X_{i\Delta_n}$ (see Remark 2.8 3)).

We set $\hat{\theta}_{n,m} = \operatorname{arginf}_{\theta\in\Theta} \hat{U}_{n,m}(\theta)$ for a minimum contrast estimator.

Let us assume that the diffusion X satisfies (A1)–(A5) with $a(x) = a(x,\vartheta_0)$, $b(x) = b(x,\mu_0,\vartheta_0)$ and the stronger hypothesis that M_0, M_∞ and K_0 can be chosen as large as we want. The corresponding stationary distribution will still be denoted by ν_0.

We add the two following assumptions

(S1) An identifiability assumption,

$a(x,\vartheta) = a(x,\vartheta_0) \quad d\nu_0(x)$ *a.e.* implies $\vartheta = \vartheta_0$,

$b(x,\vartheta_0,\mu) = b(x,\vartheta_0,\mu_0) \quad d\nu_0(x)$ *a.e.* implies $\mu = \mu_0$.

(S2) A smoothness assumption in the parameter θ for a and b,

a and b are the restrictions to $(0,\infty)\times\Theta$ of functions defined on an open subset of $(0,\infty)\times\mathbb{R}^2$, on which they are differentiable up to order 6. Furthermore, $\forall\ \vartheta \in \Theta_2,\ x > 0, \quad a(x,\vartheta) > 0$ and $\exists\ c,\ \gamma \geq 0$ such that $\forall\ i,j,k\ \in \{0,\dots,3\}^3$, with $i+j \leq 3,\ \forall\ x > 0$:

$$\sup_{(\mu,\vartheta)\in\Theta}\left|\frac{\partial^{i+j+k}}{\partial\mu^i\partial\vartheta^j\partial x^k}b(x,\mu,\vartheta)\right| \leq c(1+|x|)^\gamma$$

$$\sup_{\vartheta\in\Theta_2}\left|\frac{\partial^{i+k}}{\partial\vartheta^i\partial x^k}a(x,\vartheta)\right| + \sup_{\vartheta\in\Theta_2}\left|\frac{\partial^{i+k}}{\partial\vartheta^i\partial x^k}a^{-1}(x,\vartheta)\right| \leq c(1+|x|)^\gamma$$

By (S2), all functions appearing below satisfy $\mathbf{C}_\gamma$ or $\mathbf{CU}_\gamma$ for some $\gamma \geq 0$.

Our first theorem gives conditions on the rate of m_n in order to get the consistency of $\hat{\theta}_{n,m_n}$.

Theorem 3.1. *Assume that m_n is a sequence of integers such that, $m_n^{-1} = o(\Delta_n)$. Then, the estimator is consistent, $\hat{\theta}_{n,m_n} \xrightarrow[\mathbf{P}]{n\to\infty} \theta_0$.*

Proof. To prove the result, we cannot follow the standard proof of consistency of minimum contrast estimators (such as given in Dacunha–Castelle and Duflo (1983) Chapter 3). Actually, the two parameters μ_0 and ϑ_0 are not estimated at the same rate. This situation is analogous to the one investigated in Kessler (1997) and Gloter (1999, b). Proceeding as in the proof of Kessler's theorem 1, the consistency of $\hat{\theta}_{n,m_n}$ is obtained, using (S1), from the two following steps.

1) First, to get the consistency for the diffusion parameter ϑ, it is enough to show that, uniformly in θ,

$$n^{-1}\hat{U}_{n,m_n}(\theta) \xrightarrow{n\to\infty} \nu_0\left(\frac{a^2(.,\vartheta_0)}{a^2(.,\vartheta)} + \log a^2(.,\vartheta)\right) \quad \text{in probability.}$$

2) Then,the consistency for $\hat{\mu}_{n,m_n}$ is obtained by proving that uniformly in $\theta = (\mu,\vartheta)$, in probability,

$$(n\Delta_n)^{-1}(\hat{U}_{n,m_n}(\mu,\vartheta) - \hat{U}_{n,m_n}(\mu_0,\vartheta)) \xrightarrow{n\to\infty} \frac{3}{2}\nu_0\left(\frac{(b(.,\mu,\vartheta)-b(.,\mu_0,\vartheta))^2}{a^2(.,\vartheta)}\right) + \alpha(\mu,\vartheta),$$

where

$$\alpha(\mu,\vartheta) = \frac{3}{2}\nu_0\left(\frac{\{b(.,\mu_0,\vartheta_0)-b(.,\mu_0,\vartheta)\}\{2b(.,\mu_0,\vartheta_0)-2b(.,\mu,\vartheta)\}}{a^2(.,\vartheta)}\right)$$

tends to 0 uniformly in μ as $\theta \to \theta_0$.

We start the proof by 1). To apply the results of Section 2, we write $n^{-1}\hat{U}_{n,m_n}(\theta)$ using notations (3), (5) and (4) with $b(x) = b(x,\mu_0,\vartheta_0)$. This gives $(\theta = (\mu,\vartheta))$,

$$n^{-1}\hat{U}_{n,m_n}(\theta) = \frac{3}{2}\hat{\mathcal{Q}}_{n,m_n}(a^{-2}(.,\vartheta)) + \hat{\nu}_{n,m_n}(\log a^2(.,\vartheta)) + \Delta_n Z_n(\theta) \tag{52}$$

with

$$Z_n(\theta) = -3\hat{\mathcal{I}}_{n,m_n}(\frac{b(.,\theta)}{a^2(.,\vartheta)}) + \frac{3}{4}\hat{\mathcal{Q}}_{n,m_n}(h(.,\theta)) + \frac{3}{2}\hat{\nu}_{n,m_n}(\frac{\{b^2(.,\theta) - 2b(.,\theta)b(.,\theta_0)\}}{a^2(.,\vartheta)}). \tag{53}$$

Now, 1) follows from Proposition 2.5 and Theorem 2.6 and $\Delta_n \to 0$.

Now, we prove 2). By easy computations on the above formula, we get

$$\begin{aligned}(n\Delta_n)^{-1}(\hat{U}_{n,m_n}(\mu,\vartheta) - \hat{U}_{n,m_n}(\mu_0,\vartheta)) &= 3\hat{\mathcal{I}}_{n,m_n}\left(\frac{b}{a^2}(.,\mu_0,\vartheta) - \frac{b}{a^2}(.,\mu,\vartheta)\right)\\ &- \frac{3}{4}\hat{\mathcal{Q}}_{n,m_n}\left(h(.,\mu_0,\vartheta) - h(.,\mu,\vartheta)\right)\\ &+ \frac{3}{2}\hat{\nu}_{n,m_n}\left(\frac{b^2(.,\theta) - 2b(.,\theta)b(.,\theta_0) - b^2(.,\mu_0,\vartheta) + 2b(.,\mu_0,\vartheta)b(.,\theta_0)}{a^2(.,\vartheta)}\right)\end{aligned}$$

Now, by Theorem 2.6 (recall $h(x,\theta) = \partial_x(\frac{b}{a^2})(x,\theta)$)

$$3\hat{\mathcal{I}}_{n,m_n}\left(\frac{b}{a^2}(.,\mu_0,\vartheta) - \frac{b}{a^2}(.,\mu,\vartheta)\right) \xrightarrow{n\to\infty} \frac{1}{2}\nu_0\left((h(.,\mu_0,\vartheta) - h(.,\mu,\vartheta))a^2(.,\vartheta_0)\right)$$

and

$$\frac{3}{4}\hat{\mathcal{Q}}_{n,m_n}\left(h(.,\mu_0,\vartheta) - h(.,\mu,\vartheta)\right) \xrightarrow{n\to\infty} \frac{1}{2}\nu_0\left((h(.,\mu_0,\vartheta) - h(.,\mu,\vartheta))a^2(.,\vartheta_0)\right).$$

So the difference of the above two terms tends to 0.

For the third term note that:

$$\begin{aligned}b^2(.,\theta) - 2b(.,\theta)b(.,\theta_0) - b^2(.,\mu_0,\vartheta) + 2b(.,\mu_0,\vartheta)b(.,\theta_0) &= (b(.,\mu,\vartheta) - b(.,\mu_0,\vartheta))^2\\ &+ \{b(.,\mu_0,\vartheta_0) - b(.,\mu_0,\vartheta)\}\{2b(.,\mu_0,\vartheta_0) - 2b(.,\mu,\vartheta)\}.\end{aligned}$$

Applying Proposition 2.5 we get 2). □

For the asymptotic normality more conditions on Δ_n, m_n are required.

Theorem 3.2. *Assume that $n\Delta_n^2 \to 0$, and that $m_n^{-1} = o(n^{-1}\Delta_n)$. Then, we have the convergence in law of $\left((n\Delta_n)^{\frac{1}{2}}(\hat{\mu}_{n,m_n} - \mu_0), n^{\frac{1}{2}}(\hat{\vartheta}_{n,m_n} - \vartheta_0)\right)$ to a*

$$\mathcal{N}\left(0, \left\{\nu_0\left(\frac{(\partial_\mu b)^2(.,\theta_0)}{a^2(.,\vartheta_0)}\right)\right\}^{-1}\right) \otimes \mathcal{N}\left(0, \frac{9}{16}\left\{\nu_0\left(\frac{(\partial_\vartheta a)^2(.,\vartheta_0)}{a^2(.,\vartheta_0)}\right)\right\}^{-1}\right)$$

Proof. Since $\theta_0 \in \overset{\circ}{\Theta}$, using Taylor's formula:

$$\int_0^1 \nabla_\theta^2 \hat{U}_{n,m_n}(\theta_0 + u(\hat{\theta}_{n,m_n} - \theta_0))du \begin{bmatrix}\hat{\mu}_{n,m_n} - \mu_0\\ \hat{\vartheta}_{n,m_n} - \vartheta_0\end{bmatrix} = -\nabla_\theta \hat{U}_{n,m_n}(\theta_0) \tag{54}$$

In the equation above, we force the rates of convergence $(n\Delta_n)^{\frac{1}{2}}$ for $\hat{\mu}_{n,m_n}$ and $n^{\frac{1}{2}}$ for $\hat{\vartheta}_{n,m_n}$ to appear. After easy computations, (54) writes

$$\int_0^1 \mathcal{C}_n(\theta_0 + u(\hat{\theta}_{n,m_n} - \theta_0))du\ \mathcal{E}_n = \mathcal{D}_n$$

with

$$\mathcal{E}_n = \begin{bmatrix} (n\Delta_n)^{\frac{1}{2}}(\hat{\mu}_{n,m_n} - \mu_0) \\ n^{\frac{1}{2}}(\hat{\vartheta}_{n,m_n} - \vartheta_0) \end{bmatrix}, \quad \mathcal{D}_n = \begin{bmatrix} -(n\Delta_n)^{-\frac{1}{2}}\frac{\partial}{\partial\mu}\hat{U}_{n,m_n}(\theta_0) \\ -n^{-\frac{1}{2}}\frac{\partial}{\partial\vartheta}\hat{U}_{n,m_n}(\theta_0) \end{bmatrix}$$

$$\mathcal{C}_n(\theta) = \begin{bmatrix} (n\Delta_n)^{-1}\frac{\partial^2}{\partial\mu^2}\hat{U}_{n,m_n}(\theta) & n^{-1}\Delta_n^{-\frac{1}{2}}\frac{\partial^2}{\partial\vartheta\partial\mu}\hat{U}_{n,m_n}(\theta) \\ n^{-1}\Delta_n^{-\frac{1}{2}}\frac{\partial^2}{\partial\mu\partial\vartheta}\hat{U}_{n,m_n}(\theta) & n^{-1}\frac{\partial^2}{\partial\vartheta^2}\hat{U}_{n,m_n}(\theta) \end{bmatrix}.$$

Now the proof of $\mathcal{E}_n \xrightarrow{n\to\infty} \mathcal{N}\left(0, \left\{\nu_0\left(\frac{(\partial_\mu b)^2(.,\mu_0)}{a^2(.,\vartheta_0)}\right)\right\}^{-1}\right) \otimes \mathcal{N}\left(0, \frac{9}{16}\left\{\nu_0\left(\frac{(\partial_\vartheta a)^2(.,\vartheta_0)}{a^2(.,\vartheta_0)}\right)\right\}^{-1}\right)$ reduces to the two following points.

1) We have the convergence in law:

$$\mathcal{D}_n \xrightarrow[\mathcal{D}]{n\to\infty} \mathcal{N}\left(0, \begin{bmatrix} 9\nu_0\left(\frac{(\partial_\mu b)^2(.,\mu_0)}{a^2(.,\vartheta_0)}\right) & 0 \\ 0 & 9\nu_0\left(\frac{(\partial_\vartheta a)^2(.,\vartheta_0)}{a^2(.,\vartheta_0)}\right) \end{bmatrix}\right)$$

2) We have the uniform (with respect to θ) convergence in probability:

$$\mathcal{C}_n(\theta) \xrightarrow{n\to\infty} \begin{bmatrix} \mathcal{C}_{1,1}(\theta) & 0 \\ 0 & \mathcal{C}_{2,2}(\theta) \end{bmatrix}$$

where $\mathcal{C}_{1,1}(\theta)$ and $\mathcal{C}_{2,2}(\theta)$ are such that

$$\mathcal{C}_{1,1}(\theta) \xrightarrow{\theta\to\theta_0} 3\nu_0\left(\frac{(\partial_\mu b)^2(.,\mu)}{a^2(.,\vartheta)}\right) \text{ and } \mathcal{C}_{2,2}(\theta) \xrightarrow{\theta\to\theta_0} 4\nu_0\left(\frac{(\partial_\vartheta a)^2(.,\vartheta_0)}{a^2(.,\vartheta_0)}\right).$$

Indeed, this second point immediately implies, using the consistency of $\hat{\theta}_{n,m_n}$,

$$\int_0^1 \mathcal{C}_n(\theta_0 + u(\hat{\theta}_n - \theta_0))du \xrightarrow[\mathbf{P}]{n\to\infty} \begin{bmatrix} 3\nu_0\left(\frac{(\partial_\mu b)^2(.,\mu_0)}{a^2(.,\vartheta_0)}\right) & 0 \\ 0 & 4\nu_0\left(\frac{(\partial_\vartheta a)^2(.,\vartheta_0)}{a^2(.,\vartheta_0)}\right) \end{bmatrix}.$$

We prove 1). The functionals $\hat{\nu}_{n,m_n}$, $\hat{\mathcal{I}}_{n,m_n}$ and $\hat{\mathcal{Q}}_{,m_n}$ are linear in f. It enables us to derivate the expression (52) given in the proof of Theorem 3.1 with respect to the parameter and get (using the notations of Theorem 2.7 and $h(x,\theta) = (\partial_x \frac{b}{a^2})(x,\theta)$),

$$(n\Delta_n)^{-\frac{1}{2}}\partial_\mu\hat{U}_{n,m_n}(\theta_0) = -3\hat{N}_{n,m_n}\left(\frac{\partial_\mu b}{a^2}(.,\theta_0)\right). \tag{55}$$

Analogously, we get

$$n^{-\frac{1}{2}}\partial_\vartheta\hat{U}_{n,m_n}(\theta_0) = -2\hat{M}_{n,m_n}\left(\frac{\partial_\vartheta a}{a^3}(.,\vartheta_0)\right) + \beta_n \tag{56}$$

where (see (53))

$$\beta_n = n^{\frac{1}{2}}\Delta_n\frac{\partial}{\partial\vartheta}Z_n(\theta_0).$$

The expression of $Z_n(\theta)$ and the linearity of the three functionals allow to apply Proposition 2.5 and Theorem 2.6. And we obtain $\frac{\partial}{\partial\vartheta}Z_n(\theta_0) = O_{\mathbf{P}}(1)$. Since $n\Delta_n^2 \to 0$, we deduce

$$\beta_n = o_{\mathbf{P}}(1). \tag{57}$$

Now, 1) follows from (55)–(57) and Theorem 2.7.

To obtain 2), we derivate twice $\hat{U}_{n,m_n}$ and use Proposition 2.5 and Theorem 2.6. This yields $\mathcal{C}_n(\theta) \xrightarrow{n\to\infty} \begin{bmatrix} \mathcal{C}_{1,1}(\theta) & 0 \\ 0 & \mathcal{C}_{2,2}(\theta) \end{bmatrix}$ with,

$$\mathcal{C}_{1,1}(\theta) = \nu_0\left(-2\frac{\partial^2_{\vartheta^2}a(.,\vartheta)a^2(.,\vartheta_0)}{a^3(.,\vartheta_0)} + 6\frac{\partial_\vartheta a(.,\vartheta)a^2(.,\vartheta_0)}{a^4(.,\vartheta_0)} + 2\frac{\partial^2_{\vartheta^2}a(.,\vartheta)}{a(.,\vartheta_0)} - 2\frac{\partial_\vartheta a(.,\vartheta)}{a^2(.,\vartheta_0)}\right)$$

$$\mathcal{C}_{2,2}(\theta) = 3\nu_0\left(\frac{(\partial_\mu b(.,\theta))^2 + 2\partial^2_{\mu^2}b(.,\theta)(b(.,\theta) - b(.,\theta_0))}{a^2(.,\vartheta)}\right)$$

We easily obtain the limits of $\mathcal{C}_{1,1}(\theta)$ and $\mathcal{C}_{2,2}(\theta)$ as $\theta \to \theta_0$. Thus, the theorem is proved. □

Remark 3.3. *1. It is important to note that our identifiabity assumption (S1) is identical to the identifiability assumption given in Kessler (1997). This means that the parameters of model (50) which are identifiable by the discrete observation $(X_{i\Delta_n},\ i \le n)$ and those which are identifiable in our framework are the same. Thus, by our approach, we are able to estimate more parameters than in Genon-Catalot et al. (1999, a) where only parameters present in the stationary distribution ν_0 were reached. Of course, our set of observation is more rich and we benefit from the double discretization procedure.*

2. The estimator $\hat{\mu}_{n,m_n}$ has rate $\sqrt{n\Delta_n}$ with asymptotic variance $\left(\int \frac{(\partial_\mu b)^2(x,\mu_0)}{a^2(x,\vartheta_0)}d\nu_0(x)\right)^{-1}$. It is remarkable that this is the same asymptotic behaviour as any efficient estimator based on the direct observation of the hidden trajectory $(X_t)_{t\in[0,n\Delta_n]}$.

3. Our number of observations is $N = nm_n + 1$, and the rate is only $n^{\frac{1}{2}}$ for the estimation of ϑ. Of course, it is interesting to choose $m_n \to \infty$ as slow as possible, but the two conditions $n\Delta_n^2 = o(1)$ and $m_n^{-1} = o(n^{-1}\Delta_n)$ imply that $n^{\frac{3}{2}} = o(m_n)$. We deduce that necessarily $n = o(N^{\frac{2}{5}})$, thus the rate of estimation is slower than $N^{\frac{1}{5}}$. This is is much slower than $N^{\frac{1}{2}}$, the speed which is obtained from a direct sample, with size N, of X. However, let us stress the point that we know nothing about the rate of convergence of the exact maximum likelihood in this problem.

4 Examples of hidden diffusion models

4.1 Exponantial of a diffusion

Here, $\mathcal{B}_0(x) = 1 + x^{-1}$ and $\mathcal{B}_\infty(x) = 1 + x^2$. The process X being positive, it is natural to consider the exponantial of a real diffusion. Assume that we are given the model

$$X_t = \exp(Z_t), \quad dZ_t = \tilde{a}(Z_t,\vartheta_0)dB_t + \tilde{b}(Z_t,\mu_0)dt.$$

where Z is a diffusion on $\mathbb{R}$. The diffusion X satisfies the stochastic differential equation

$$dX_t = a(X_t, \vartheta_0)dB_t + b(X_t, \mu_0, \vartheta_0)dt, \tag{58}$$

with $a(x, \vartheta) = x\tilde{a}(\ln x, \vartheta)$ and $b(x, \mu, \vartheta) = x\tilde{b}(\ln x, \mu) + \frac{1}{2}x\tilde{a}^2(\ln x, \vartheta)$.

In Gloter (1998), we have shown that the following conditions on $\tilde{a}$ and $\tilde{b}$ imply the validity of (A2) with $K_0 = \infty$ for X: $\tilde{a}(., \vartheta_0)$ and $\tilde{a}'(., \vartheta_0)$ are bounded, $\limsup_{x\to\infty} \tilde{b}(., \mu_0) < \infty$ and $\liminf_{x\to-\infty} \tilde{b}(., \mu_0) > -\infty$.

The other difficult assumption to check is (A5), we can deduce it from (A4) in most models by a coupling argument (as in Proposition 4.1 of Gloter (1999, b)).

As an explicit example, consider the Ornstein–Uhlenbeck process Z given by

$$dZ_t = \mu_0 Z_t dt + \vartheta_0 dB_t \quad (\mu_0 < 0) \text{ and } X_t = e^{Z_t}.$$

Then (A1)-(A5) are valid with $K_0 = \infty$ and any positive constants M_0, M_∞. Furthermore, here we can find an explicit minimum for the contrast (51).

$$\hat{\vartheta^2}_{n,m} = \frac{3}{2n\Delta_n} \sum_{i=0}^{n-2} \hat{X}_i^{-2}(\hat{X}_{i+1} - \hat{X}_i)^2 \tag{59}$$

$$\hat{\mu}_{n,m} = \frac{\left\{ \sum_{i=0}^{n-2} \frac{(\hat{X}_{i+1} - \hat{X}_i - \frac{\hat{\vartheta^2}_{n,m}}{2}\hat{X}_i\Delta_n)(\ln \hat{X}_i)}{\hat{X}_i} + \frac{1}{4}\sum_{i=0}^{n-2} \frac{\ln(\hat{X}_i) - 1}{\hat{X}_i^2}(\hat{X}_{i+1} - \hat{X}_i)^2 \right\}}{\sum_{i=0}^{n-2} \Delta_n (\ln \hat{X}_i)^2}. \tag{60}$$

(We have dropped a negligible term in $\hat{\vartheta^2}_{n,m_m}$ for simplification.)

Results of Section 3 apply.

If $m^{-1} = m_n^{-1} = o(\Delta_n)$ then, $\hat{\theta}_{n,m_n} \xrightarrow[\mathbf{P}]{n\to\infty} \theta_0$.

If $n\Delta_n^2 \to 0$, and $m_n^{-1} = o(n^{-1}\Delta_n)$ then,

$$\begin{bmatrix} \sqrt{n\Delta_n}(\hat{\mu}_{n,m_n} - \mu_0) \\ \sqrt{n}(\hat{\vartheta^2}_{n,m_n} - \vartheta_0^2) \end{bmatrix} \xrightarrow[\mathcal{D}]{n\to\infty} \mathcal{N}\left(0, \begin{bmatrix} 2\,|\mu_0| & 0 \\ 0 & \frac{9}{4}\vartheta_0^4 \end{bmatrix}\right).$$

To study how $\hat{\theta}_{n,m}$ behaves on finite sample, numerical simulation results are presented in Tables 1–6.

In Tables 1–2, we have $n = 50$, $m = 100$ (hence the total number of observations N is 5000) and $\Delta = 0.3$. Table 1 gives the estimator $\hat{\vartheta^2}_{n,m}$ of the diffusion coefficient for different values of μ_0 and ϑ_0. The estimator overestimates the true value and the better results are for $\vartheta_0^2 = 1$. Table 2 gives $\hat{\mu}_{n,m}$. We know that the rates of convergence of both estimators are slow, which explains the biases and standard deviations. Also recall that we have to combine $n\Delta_n$ big with $n\Delta_n^2$, $n(m_n\Delta_n)^{-1}$ small.

When $n = 100$, $m = 500$ and $\Delta = 0.3$ ($N = 50000$), (Tables 3–4) results are much better. In Tables 5–6, with the same number of observations ($N = 50000$), $\Delta = 0.05$ is smaller. It is not surprising that the results are not improved when Δ decreases, because $n\Delta$ becomes too small and $n(m_n\Delta_n)^{-1}$ becomes too large.

4.2 C.I.R process

Here, $\mathcal{B}_0(x) = 1 + x^{-1}$ and $\mathcal{B}_\infty(x) = 1 + x$. We suppose that:

$$dX_t = (\alpha X_t + \beta)dt + \vartheta\sqrt{X_t}dB_t, \quad X_0 = \eta. \tag{61}$$

with $\alpha < 0$, $\vartheta, \beta > 0$, $c_0 = \frac{2\beta}{\vartheta^2} > 1$.

We assume that the initial distribution is either deterministic, or the stationary distribution of X, which exists since $c_0 > 1$. We have an explicit expression for the estimator,

$$\begin{bmatrix} \Delta_n \sum_{i=0}^{n-2} \hat{X}_i & n\Delta_n \\ n\Delta_n & \Delta_n \sum_{i=0}^{n-2} \hat{X}_i^{-1} \end{bmatrix} \begin{bmatrix} \hat{\alpha}_{n,m_n} \\ \hat{\beta}_{n,m_n} \end{bmatrix} = \begin{bmatrix} \sum_{i=0}^{n-2} (\hat{X}_{i+1} - \hat{X}_i) \\ \sum_{i=0}^{n-2} \left\{ \hat{X}_i^{-1}(\hat{X}_{i+1} - \hat{X}_i) + \frac{1}{4}\hat{X}_i^{-2}(\hat{X}_{i+1} - \hat{X}_i)^2 \right\} \end{bmatrix}$$

$$\hat{\vartheta^2}_{n,m_n} = \frac{3}{2}(n\Delta_n)^{-1} \sum_{i=0}^{n-2} \hat{X}_i^{-1} \left(\hat{X}_{i+1} - \hat{X}_i \right)^2$$

We know by Gloter (1998), (1999, b) that (A1)-(A5) are satisfied with $K_0 = c_0 - 1$ and any $(M_0, M_\infty) \in [0, c_0) \times [0, \infty)$. So, the results of Section 3 do not apply directly. Nevertheless, we have consistency and asymptotic normality by a specific proof on the explicit expressions of the estimator.

Theorem 4.1. • *If $c_0 > 13$, and $m_n^{-1} = o(\Delta_n)$, then, $\hat{\theta}_{n,m_n} \xrightarrow[\mathbf{P}]{n\to\infty} \theta$.*

• *Furthermore, if $n\Delta_n^2 \to 0$, and $m_n^{-1} = o(n^{-1}\Delta_n)$, then,*

$$\begin{bmatrix} \sqrt{n\Delta_n}(\hat{\alpha}_{n,m_n} - \alpha) \\ \sqrt{n\Delta_n}(\hat{\beta}_{n,m_n} - \beta) \\ \sqrt{n}(\hat{\vartheta^2}_{n,m_n} - \vartheta^2) \end{bmatrix} \xrightarrow[\mathcal{D}]{n\to\infty} \mathcal{N}\left(0, \begin{bmatrix} 2|\alpha| & \vartheta^2 - 2\beta & 0 \\ \vartheta^2 - 2\beta & (2\beta - \vartheta^2)\frac{\beta}{|\alpha|} & 0 \\ 0 & 0 & \frac{9}{4}\vartheta^4 \end{bmatrix}\right)$$

Proof. For the proof, we need to recall the adaptation of Theorem 1.5, when X is a C.I.R. process (see Gloter (1999, b) for proofs).

Let f satisfy $\mathbf{CU}_\gamma$ and g satisfy $\mathbf{CU}_{\gamma'}$.

- If $1 + 2\gamma < c_0$, then (27) holds.
- If $4 \vee 2\gamma + \frac{3}{2} < c_0$, then (28) holds.
- If $4 \vee 2\gamma + 1 < c_0$, then (29) holds.
- If $4 + 2\gamma \vee 1 + 4\gamma < c_0$ and $6 + 2\gamma' \vee 1 + 4\gamma' < c_0$ then (30) holds.

We deduce, for instance, consistency of $\hat{\vartheta^2}_{n,m_n} = \frac{3}{2}\hat{\mathcal{Q}}_{n,m_n}(x^{-1})$ when $c_0 > 13$, by application of (29) (since x^{-1} satisfy $\mathbf{CU}_3$) and Lemma 2.4 (with $r_n = 1$.)

The consistency of $\hat{\alpha}_{n,m_n}, \hat{\beta}_{n,m_n}$ and the asymptotic normality for $\hat{\theta}_{n,m_n}$ follow by analogous computations. □

4.3 Bilinear process

Again, $\mathcal{B}_0(x) = 1 + x^{-1}$ and $\mathcal{B}_\infty(x) = 1 + x$. We suppose that, $dX_t = (\alpha X_t + \beta)dt + \vartheta X_t dB_t$, with $\alpha < 0$, $\vartheta, \beta > 0$; we set $c_0 = 1 + \frac{2\alpha}{\vartheta^2}$. The initial distribution is either a Dirac mass, or the stationary distribution of X. We have explicit expression for the estimator:

$$\begin{bmatrix} n\Delta_n & \Delta_n \sum_{i=0}^{n-2} \hat{X}_i^{-1} \\ \Delta_n \sum_{i=0}^{n-2} \hat{X}_i^{-1} & \Delta_n \sum_{i=0}^{n-2} \hat{X}_i^{-2} \end{bmatrix} \begin{bmatrix} \hat{\alpha}_{n,m_n} \\ \hat{\beta}_{n,m_n} \end{bmatrix} = \begin{bmatrix} \sum_{i=0}^{n-2} \left\{ \hat{X}_i^{-1}(\hat{X}_{i+1} - \hat{X}_i) + \frac{1}{4}\hat{X}_i^{-2}(\hat{X}_{i+1} - \hat{X}_i)^2 \right\} \\ \sum_{i=0}^{n-2} \left\{ \hat{X}_i^{-2}(\hat{X}_{i+1} - \hat{X}_i) + \frac{1}{2}\hat{X}_i^{-3}(\hat{X}_{i+1} - \hat{X}_i) \right\} \end{bmatrix}$$

$$\hat{\vartheta^2}_{n,m_n} = \frac{3}{2}(n\Delta_n)^{-1} \sum_{i=0}^{n-2} \hat{X}_i^{-2}(\hat{X}_{i+1} - \hat{X}_i)$$

This process satisfies (A1)–(A5) with $K_0 = \infty$, and any $(M_0, M_\infty) \in [0, \infty) \times [0, c_0)$. As in Section 4.2, by an adaptation of Theorem 1.5, we can prove

Theorem 4.2. • *If $c_0 > 4$, and $m_n^{-1} = o(\Delta_n)$, then,* $\hat{\theta}_{n,m_n} \xrightarrow[\mathbf{P}]{n\to\infty} \theta$.

- *If $c_0 > 8$, $n\Delta_n^2 \to 0$, and $m_n^{-1} = o(n^{-1}\Delta_n)$, then,*

$$\begin{bmatrix} \sqrt{n\Delta_n}(\hat{\alpha}_{n,m_n} - \alpha) \\ \sqrt{n\Delta_n}(\hat{\beta}_{n,m_n} - \beta) \\ \sqrt{n}(\hat{\vartheta^2}_{n,m_n} - \vartheta^2) \end{bmatrix} \xrightarrow[\mathcal{D}]{n\to\infty} \mathcal{N}\left(0, \begin{bmatrix} 2|\alpha| + 2\vartheta^2 & -2\beta & 0 \\ -2\beta & \frac{4\beta^2}{\vartheta^2 + 2|\alpha|} & 0 \\ 0 & 0 & \frac{9}{4}\vartheta^4 \end{bmatrix}\right)$$

5 Extension

In this section, we present some extensions of our work. Our first step is to consider $\rho \neq 0$ (recall (1)). Then, we consider a correlation between the two Brownian motions B and W.

5.1 Presence of a drift term for Y

Here, we assume that (Y) is solution of $dY_t = \rho(X_t, t)dt + \sigma_t dW_t$. We make the following assumption on ρ.

(A6) The function $\rho(x, s) : (0, \infty) \times \mathbb{R}_+ \mapsto \mathbb{R}$ is Borel and satisfies:

$$\exists\, c, \quad \forall x > 0, \quad \forall s \geq 0, \quad |\rho(x, s)| \leq c\mathcal{B}_\infty(x).$$

In Gloter (1999, a), we have shown that the error term $E_{i,n,m} = \hat{X}_i - \overline{X}_i$, now, writes:

$$E_{i,n,m} = \Delta_n^{-1} \sum_{j=0}^{m-1} 2 \int_{(i+\frac{j}{m})\Delta_n}^{(i+\frac{j+1}{m})\Delta_n} (Y_s - Y_{(i+\frac{j}{m})\Delta_n})\sigma_s dW_s + \Delta_n^{-1} \sum_{j=0}^{m-1} 2 \int_{(i+\frac{j}{m})\Delta_n}^{(i+\frac{j+1}{m})\Delta_n} (Y_s - Y_{(i+\frac{j}{m})\Delta_n})\rho(X_s, s)ds.$$

Hence, a new term of error, due to ρ, has appeared (compare with (15)).

But, under (A6), this new term $\Delta_n^{-1} \sum_{j=0}^{m-1} 2 \int_{(i+\frac{j}{m})\Delta_n}^{(i+\frac{j+1}{m})\Delta_n} (Y_s - Y_{(i+\frac{j}{m})\Delta_n})\rho(X_s, s)ds$ is negligible versus $\Delta_n^{-1} \sum_{j=0}^{m-1} 2 \int_{(i+\frac{j}{m})\Delta_n}^{(i+\frac{j+1}{m})\Delta_n} (Y_s - Y_{(i+\frac{j}{m})\Delta_n})\sigma_s dW_s$.

Therefore, under the additional assumption (A6), we may prove that results of Sections 2–3 still hold without any modifications.

5.2 Case of correlated brownian motions

In the financial literature concerned with stochastic volatility models (1)–(2), it is not generally supposed that the two Brownian motions B and W are independent. We have investigated the same problem of estimation as above in the case where $d < B, W >_t = \chi dt$, with $\chi \in (-1, 1)$. The correlation parameter χ may be known or unknown.

Results on the error $E_{i,n,m}$ are different when a correlation is considered. We still have $E\left(E^2_{i,n,m} \mid \mathcal{G}^n_i\right) \leq cm^{-1}\mathcal{B}^4_\infty(X_{i\Delta_n})$, but now $E\left(E^4_{i,n,m} \mid \mathcal{G}^n_i\right) \leq cm^{-1}\mathcal{B}^8_\infty(X_{i\Delta_n})$ (compare with (18)). On the contrary, results (19)–(20) are still valid. (see Gloter (1999, a) for proofs of these results)

These differences for the error yield that the condition on m_n for the validity of theorems of Section 2 is more restrictive. We give these results, but omit their proofs since they are quite similar to those of Section 2.

Theorem 5.1. *Assume that (A1)–(A6) are valid with any positive constant M_0, M_∞ and K_0. Let f satisfy* $\mathbf{CU}_\gamma$ *for some $\gamma \geq 0$ and suppose that m_n is a sequence of integer.*

- *If $m_n^{-1} = o(\Delta_n^2)$, then $\hat{\nu}_{n,m_n}(f(.,\theta)) \xrightarrow[\mathbf{P}]{n\to\infty} \nu_0(f(.,\theta))$ uniformly in θ.*
- *If $m_n^{-1} = o(\Delta_n^2)$, then $\hat{\mathcal{I}}_{n,m_n}(f(.,\theta)) \xrightarrow[\mathbf{P}]{n\to\infty} \frac{1}{6}\nu_0\left(f'_x(.,\theta)a^2(.)\right)$ uniformly in θ.*
- *If $m_n^{-1} = o(\Delta_n^2)$, then $\hat{\mathcal{Q}}_{n,m_n}(f(.,\theta)) \xrightarrow[\mathbf{P}]{n\to\infty} \frac{2}{3}\nu_0\left(g(.,\theta)a^2(.)\right)$ uniformly in θ.*

Let f and g satisfy $\mathbf{C}_\gamma$ *for some $\gamma \geq 0$ and suppose that $n\Delta_n^2 \to 0$.*

- *If $m_n^{-1} = o(n^{-1}\Delta_n)$ then $\hat{N}_n(f) = \sqrt{n\Delta_n}(\hat{\mathcal{I}}_{n,m_n}(f) - \frac{1}{4}\hat{\mathcal{Q}}_{n,m_n}(f')) \xrightarrow[\mathcal{D}]{n\to\infty} \mathcal{N}\left(0, \nu_0\left(f^2a^2\right)\right)$*
- *If $m_n^{-1} = o(n^{-1}\Delta_n^2)$, then, $\hat{M}_n(g) = \sqrt{n}(\frac{3}{2}\hat{\mathcal{Q}}_{n,m_n}(g) - \hat{\nu}_{n,m_n}(a^2g)) \xrightarrow[\mathcal{D}]{n\to\infty} \mathcal{N}\left(0, \frac{9}{4}\nu_0\left(g^2a^4\right)\right)$*

and the vector $(\hat{N}_n(f), \hat{M}_n(g))$ converges to a vector with independent coordinates.

Now, we may deduce consistency and normality for the estimator introduced in Section 3.

Theorem 5.2. *Assume that (A1)–(A6) are valid with any positive constant M_0, M_∞ and K_0.*

1) If $m_n^{-1} = o(\Delta_n^2)$, the estimator is consistent $\hat{\theta}_{n,m_n} \to \theta_0$.

2) Assume that $n\Delta_n^2 \to 0$, and that $m_n^{-1} = o(n^{-1}\Delta_n^2)$. Then, we have the convergence in law of $\left((n\Delta_n)^{\frac{1}{2}}(\hat{\mu}_{n,m_n} - \mu_0), n^{\frac{1}{2}}(\hat{\vartheta}_{n,m_n} - \vartheta_0)\right)$ to a

$$\mathcal{N}\left(0, \left\{\nu_0\left(\frac{(\partial_\mu b)^2(.,\theta_0)}{a^2(.,\vartheta_0)}\right)\right\}^{-1}\right) \otimes \mathcal{N}\left(0, \frac{9}{16}\left\{\nu_0\left(\frac{(\partial_\vartheta a)^2(.,\vartheta_0)}{a^2(.,\vartheta_0)}\right)\right\}^{-1}\right)$$

References

- Chesney M. and Scott L.O. (1989), Pricing european options: a comparison of the modified black–scholes model and a random variance model. *Journal of Financial and quantitative Analysis*, **24**, 267–284.

- Dacunha–Castelle D. and Duflo M. (1983), *Probabilité et statistiques. Tome 2: Problèmes à temps mobile.* Masson.

- Del Moral P., Jacod J. and Protter Ph. (1998), The Monte–Carlo method for filtering with discrete–time observations. *Preprint, laboratoire de probabilités de l'université Paris 6*, **453**.
- Engle R.F. (1982), Autoregressive conditional heteroscedasticity with estimates of the variance of U.K. inflation. *Econometrica*, **50**, 987–1008.
- Frey R. and Runggaldier W.J. (1999), A nonlinear filtering approach to volatility estimation with a view towards high frequency data. *Preprint.*
- Genon-Catalot V., Jeantheau T., and Laredo C. (1998), Stochastic volatility models as hidden markov models and statistical applications. *Prépublication de l'université de Marne–la–Vallée* **22/98**.
- Genon-Catalot V., Jeantheau T., and Laredo C. (1999, a), Parameter estimation for discretely observed stochastic volatility models. *To appear in Bernoulli* **5**(5).
- Genon-Catalot V., Jeantheau T., and Laredo C. (1999, b), Consistency of conditional likelihood estimators for stochastic volatility models. *Preprint.*
- Gloter A. (1998), Discrete sampling of an integrated diffusion process and a comparison with the Euler scheme. *Prépublication de l'université de Marne–la–Vallée*, **23/98**.
- Gloter A. (1999, a), Discrete sampling of a stochastic volatility model on a fixed length time interval, approximations of the integrated volatility and statistical applications. *Prépublication de l'université de Marne–la–Vallée*, **09/99**.
- Gloter A. (1999, b), Parameter estimation for a discretely observed integrated diffusion process. *Prépublication de l'université de Marne–la–Vallée.*
- Gourieroux C., Montfort A., and Renault E. (1993), Indirect inference. *Journal of applied econometrics*, **8**, 85–118.
- Gourieroux C., Montfort A. (1993), Simulation based inference: a survey with special reference to panel data models. *Journal of econometrics.*
- Hall P. and Heyde C.C. (1980) *Martingale limit theory and its application.* Academic Press.
- Ibragimov I.A. and Khas'minskii R.Z. (1981), *Statistical Estimation, Asymptotic Theory.* Springer-Verlag.
- Kessler M. (1997), Estimation paramétrique des coefficients d'une diffusion ergodique à partir d'observations discrètes. *Scand. J. Stat.*, **24**(2), 211-229.
- Moreaux F., Navatte P., and Villa C. (1996), Market volatility index and implicit maximum likelihood estimation of stochastic volatility models. *Preprint.*

- Nelson D.B. (1990), ARCH models as diffusion approximations. *Journal of Econometrics*, **45**, 7–38.

- Pastorello S., Renault E. and Touzi N. (1994), Statistical inference for random variance option pricing. *Southern European Economic Discussion Series*, D.P. 136.

- Renault E. and Touzi N. (1996), Option hedging and implied volatilities in a stochastic volatility model. *Mathematical Finance*, **6**(3), 279–302 .

- Ruiz E. (1994), Quasi–maximum likelihood estimation of stochastic volatility models. *Journal of econometrics*, **63**, 289–306.

- Scott L.O. (1987), Option pricing when the variance changes randomly: theory, estimation, and an application. *Journal of financial and quantitative analysis*, **22**(4).

- Sørensen M. (1999), Prediction–based estimating functions. *Preprint.*

- Wiggins J.B. (1987), Option values under stochastic volatility. Theory and empirical estimates *Journal of financial economics*, **19**, 351–372.

Table 1: Mean, standard deviation of $\hat{\vartheta}^2{}_{n,m}$ when $n = 50$, $m = 100$ and $\Delta = 0.3$ for different values of (ϑ_0^2, μ_0). $Z_t = \ln X_t$ is an Ornstein–Uhlenbeck process and $N = 5000$, $n\Delta_n = 15$, $n\Delta_n^2 = 4.5$, $n(\Delta_n m_n)^{-1} = 1.67$ (recall conditions for Theorem 2.7 (49)).

	$\mu_0 = -2$	$\mu_0 = -1$	$\mu_0 = -0.5$	$\mu_0 = -0.1$
$\vartheta_0^2 = 0.01$		0.22, 0.064		
$\vartheta_0^2 = 0.1$		0.30, 0.082		
$\vartheta_0^2 = 0.5$		0.75, 0.22		
$\vartheta_0^2 = 1$	1.09, 0.36	1.41, 0.63	1.54, 0.52	1.75, 0.59
$\vartheta_0^2 = 2$		3.29, 1.59		

Table 2: Mean, standard deviation of $\hat{\mu}_{n,m}$ when $n = 50$, $m = 100$ and $\Delta = 0.3$ for different values of (ϑ_0^2, μ_0). $Z_t = \ln X_t$ is an Ornstein–Uhlenbeck process and $N = 5000$, $n\Delta_n = 15$, $n\Delta_n^2 = 4.5$, $nm_n^{-1} = 0.5$ (recall conditions for Theorem 2.7 (48)).

	$\mu_0 = -2$	$\mu_0 = -1$	$\mu_0 = -0.5$	$\mu_0 = -0.1$
$\vartheta_0^2 = 0.01$		-4.65, 0.91		
$\vartheta_0^2 = 0.1$		-2.43, 0.79		
$\vartheta_0^2 = 0.5$		-1.69, 0.66		
$\vartheta_0^2 = 1$	-2.63, 0.90	-1.63, 0.85	-0.97, 0.57	-0.38, 0.35
$\vartheta_0^2 = 2$		-1.94, 1.00		

Table 3: Mean, standard deviation of $\hat{\vartheta}^2{}_{n,m}$ when $n = 100$, $m = 500$ and $\Delta = 0.3$ for different values of (ϑ_0^2, μ_0). $Z_t = \ln X_t$ is an Ornstein–Uhlenbeck process and $N = 50000$, $n\Delta_n = 30$, $n\Delta_n^2 = 9$, $n(\Delta_n m_n)^{-1} = 0.67$ (recall conditions for Theorem 2.7 (49)).

	$\mu_0 = -2$	$\mu_0 = -1$	$\mu_0 = -0.5$	$\mu_0 = -0.1$
$\vartheta_0^2 = 0.01$		0.048, 0.007		
$\vartheta_0^2 = 0.1$		0.128, 0.018		
$\vartheta_0^2 = 0.5$		0.509, 0.10		
$\vartheta_0^2 = 1$	0.901, 0.22	1.107, 0.24	1.597, 0.49	1.42, 0.36
$\vartheta_0^2 = 2$		2.80 , 0.84		

Table 4: Mean, standard deviation of $\hat{\mu}_{n,m}$ when $n = 100$, $m = 500$ and $\Delta = 0.3$ for different values of (ϑ_0^2, μ_0). $Z_t = \ln X_t$ is an Ornstein–Uhlenbeck process and $N = 50000$, $n\Delta_n = 30$, $n\Delta_n^2 = 9$, $nm_n^{-1} = 0.2$ (recall conditions for Theorem 2.7 (48)).

	$\mu_0 = -2$	$\mu_0 = -1$	$\mu_0 = -0.5$	$\mu_0 = -0.1$
$\vartheta_0^2 = 0.01$		-2.95, 0.51		
$\vartheta_0^2 = 0.1$		-1.32, 0.35		
$\vartheta_0^2 = 0.5$		-1.15, 0.30		
$\vartheta_0^2 = 1$	-2.10, 0.52	-1.23, 0.36	-0.87, 0.38	-0.23, 0.17
$\vartheta_0^2 = 2$		-1.57, 0.56		

Table 5: Mean, standard deviation of $\hat{\vartheta}_{n,m}$ when $n = 100$, $m = 500$ and $\Delta = 0.05$ for different values of (ϑ_0^2, μ_0). $Z_t = \ln X_t$ is an Ornstein–Uhlenbeck process and $N = 50000$, $n\Delta_n = 5$, $n\Delta_n^2 = 0.25$, $n(\Delta_n m_n)^{-1} = 4$ (recall conditions for Theorem 2.7 (49)).

	$\mu_0 = -2$	$\mu_0 = -1$	$\mu_0 = -0.5$	$\mu_0 = -0.1$
$\vartheta_0^2 = 0.01$		0.25, 0.042		
$\vartheta_0^2 = 0.1$		0.34, 0.058		
$\vartheta_0^2 = 0.5$		0.73, 0.11		
$\vartheta_0^2 = 1$	1.25, 0.22	1.27, 0.22	1.28, 0.23	1.32, 0.22
$\vartheta_0^2 = 2$		2.40 , 0.45		

Table 6: Mean, standard deviation of $\hat{\mu}_{n,m}$ when $n = 100$, $m = 500$ and $\Delta = 0.05$ for different values of (ϑ_0^2, μ_0). $Z_t = \ln X_t$ is an Ornstein–Uhlenbeck process and $N = 50000$, $n\Delta_n = 5$, $n\Delta_n^2 = 0.25$, $nm_n^{-1} = 0.2$ (recall conditions for Theorem 2.7 (48)).

	$\mu_0 = -2$	$\mu_0 = -1$	$\mu_0 = -0.5$	$\mu_0 = -0.1$
$\vartheta_0^2 = 0.01$		-16.5, 4.8		
$\vartheta_0^2 = 0.1$		-4.60, 2.39		
$\vartheta_0^2 = 0.5$		-2.09, 1.22		
$\vartheta_0^2 = 1$	-2.80, 1.29	-1.27, 1.12	-1.19, 0.94	-0.55, 0.83
$\vartheta_0^2 = 2$		-1.67, 1.09		

Appendice. Un problème d'existence de densité

Notons X une diffusion sur $\mathbb{R}$, $\Delta > 0$ un pas de discrétisation et pour $i \in \mathbb{N}$

$$J_i = \int_{i\Delta}^{(i+1)\Delta} X_s ds.$$

Le problème de l'existence d'une densité pour $(J_0, \dots, J_q)$ est important pour les applications statistiques, car cette existence est la condition minimale pour pouvoir considérer la vraisemblance exacte dans les deux parties de cette thèse. Ce problème est lié à celui de l'existence de densités de transition pour la diffusion bi-dimensionnelle $(I_t, X_t) = (\int_0^t X_s ds, X_t)$.

Rappelons que le problème d'existence de densité de transition pour une diffusion multidimensionnelle Z solution d'une équation différentielle stochastique a été longuement étudié. En utilisant le calcul de Malliavin, on peut montrer que sous les conditions dites de Hörmander (rappelées dans le premier paragraphe) et sous des conditions de régularité des coefficients de l'équation différentielle, cette densité existe,

$$E(f(Z_{h+t}) \mid Z_h = z) = \int f(z')p_t(z,z')dz'.$$

De plus cette densité $p_t(z,z')$ est régulière en z' (Voir Bell (1987), Nualart (1995), Watanabe (1984)). La régularité de cette densité en (z,t) ainsi que des majorations pour $|\partial_t \partial_z \partial_{z'} p_t(z,z')|$ sont étudiées dans Kusuoka–Stroock (1985), sous des conditions plus restrictives que la condition de Hörmander.

Dans le premier paragraphe de cet Appendice, nous rappelons les conditions d'existence de densité de transition pour une diffusion multidimensionnelle. Dans le second, nous montrons que l'existence d'une densité pour $(J_0, \dots, J_q)$ peut être obtenue sous les mêmes conditions que celles qui assurent l'existence de densités de transition pour la diffusion X.

1 Condition de Hörmander

Supposons que (Z_t) soit une diffusion sur $\mathbb{R}^m$ solution de

$$Z_t = z_0 + \sum_{j=1}^{d} \int_0^t \sigma_j(Z_s)dW_s^j + \int_0^t \delta(Z_s)ds, \tag{1}$$

où $\sigma_1, \dots, \sigma_d, \delta$ appartiennent a $\mathcal{C}^\infty(\mathbb{R}^m, \mathbb{R}^m)$ et $(W^1, \dots, W^d)$ est un mouvement brownien de dimension d. De plus, on suppose que l'hypothèse suivante est vérifiée.

B : Toutes les derivées d'ordre supérieur ou égal à 1 de $\sigma_1, \dots, \sigma_d$ et δ sont bornées.

Avant d'écrire la condition de Hörmander rappelons ce qu'est le crochet de deux champs de vecteurs. Si U et V sont deux champs de vecteurs sur $\mathbb{R}^m$ (i.e. U, V $\mathbb{R}^m \to \mathbb{R}^m$ sont $\mathcal{C}^\infty$) alors on definit le champ de vecteurs sur $\mathbb{R}^m$, $[U,V] = DV.U - DU.V$. La i–eme composante de ce champ est donc:

$$[U,V]_i = \sum_{j=1}^{m} \frac{\partial V_i}{\partial z_j} U_j - \frac{\partial U_i}{\partial z_j} V_j.$$

Notons $A_0 = \delta - \frac{1}{2}\sum_{j=1}^{d} D\sigma_j.\sigma_j$ et $A_j = \sigma_j$ pour $j = 1, \dots, d$. De plus introduisons une famille de champs de vecteurs indexée par $\alpha \in \mathcal{M}$, où $\mathcal{M} = \bigcup_{k\geq 0}\{0, 1, \dots, d\}^k$ est l'ensemble des multi–indices à valeurs dans $\{0, \dots, d\}$. Cette définition se fait par récurrence sur la longueur $|\alpha|$ du multi–indice.

Pour $i \in \{1, \dots, d\}$ on pose $A_i^\alpha = A_i$, si $|\alpha| = 0$ et pour $i \in \{1, \dots, d\}$, $j \in \{0, \dots, d\}$, $\alpha \in \mathcal{M}$, on pose $A_i^{(j,\alpha)} = [A_j, A_i^\alpha]$, où (j, α) désigne le multi–indice $(j, \alpha_1, \dots, \alpha_{|\alpha|})$.

On dit que la condition de Hörmander est vérifiée en z_0 si la condition suivante est réalisée.

$$\mathbf{H}(z_0): \quad \bigcup_{i=1}^{d} \bigcup_{\alpha \in \mathcal{M}} A_i^\alpha \text{ engendre } \mathbb{R}^m.$$

Pour illustrer cette condition, prenons le cas d'une diffusion unidimensionnelle X solution de

$$dX_t = a(X_t)dW_t + b(X_t)dt.$$

Alors $A_0(x) = b(x) - \frac{1}{2}a'(x)a(x)$ et $A_1 = a(x)$ et la condition de Hörmander en x se réduit à la condition $\mathbf{h}(x)$ suivante,

$$\mathbf{h}(x): \quad \begin{cases} a(x) \neq 0 \\ \text{ou bien } b(x) \neq 0 \text{ et } \exists n \geq 1 \text{ tel que } a^{(n)}(x) \neq 0 \end{cases}.$$

Rappelons le résultat principal d'existence de densités.

Théorème 1.1. *Soit Z solution de* (1) *et supposons que* **B** *soit verifiée et que* **H**(*z*) *soit vraie pour tout z dans $\mathbb{R}^m$. Alors le processus Z admet des densités de transitions $p_t(z, z')$ pour tout $t > 0$ et $(z, z') \to p_t(z, z')$ est mesurable pour $t > 0$ et C^∞ en z' à z fixé.*

Preuve. Il est demontré dans Nualart (1995) que sous **B** et **H**(z), il existe pour tout $t > 0$ une densité de classe C^∞ pour Z_t si Z est solution de (1), issue de $Z_0 = z$. Notons $p_t(z, z')$ cette densité. Il nous reste à montrer la mesurabilité en (z, z'). Pour cela considérons $Z(z, t, \omega)$ solution de (1) issue de $Z_0 = z$ mesurable en (z, t, ω) où ω designe une réalisation du mouvement Brownien $(W^1, \dots, W^d)$. Et soit ϕ une fonction mesurable de $\mathbb{R}^m$ à valeurs dans $\mathbb{R}$ d'intégrale 1. Alors, par continuité en z', pour tout z fixé, de $p_t(z, z')$, nous déduisons

$$p_t(z, z') = \lim_{n \to \infty} E(\phi_n(z' - Z(z, t, .))), \text{ où } \phi_n(x) = n^m \phi(nz).$$

La mesurabilité en (z, z') (par le theorème de Fubini) de $E(\phi_n(z' - Z(z, t, .)))$ implique donc la mesurabilité de $p_t(z, z')$. □

2 Densité pour les integrés

On suppose que X est une diffusion réelle solution de

$$dX_t = a(X_t)dW_t + b(X_t)dt, \quad X_0 = x_0, \tag{2}$$

où W est un mouvement brownien unidimensionel.

Nous voulons montrer que si cette diffusion X satisfait aux conditions de Hörmander alors le vecteur $(J_0, \dots, J_q)$ admet une densité. Désignons désormais par Z la diffusion sur $\mathbb{R}^2$,

$$Z_t = (I_t, X_t) := (\int_0^t X_s ds, X_t).$$

Ainsi, $d = 1$, $\sigma_1(z) = \sigma(z) = \begin{bmatrix} 0 \\ a(z_2) \end{bmatrix}$ et $\delta(z) = \begin{bmatrix} z_2 \\ b(z_2) \end{bmatrix}$. Montrons que cette diffusion admet des densités de transition.

Proposition 2.1. *Supposons que a et b soient C^∞ avec toutes leurs derivées d'ordre superieur ou égal à 1 bornées. De plus supposons que $\mathbf{h}(x)$ est verifiée pour tout x réel. Alors Z admet des densités de transitions notées $p_t(z, z')$ pour tout $t > 0$. De plus $(z, z') \to p_t(z, z')$ est mesurable, pour $t > 0$ et C^∞ en z' à z fixé.*

Preuve. La diffusion $Z = (I, X)$ verifie

$$dZ_t = \begin{bmatrix} X_t \\ b(X_t) \end{bmatrix} dt + \begin{bmatrix} 0 \\ a(X_t) \end{bmatrix} dW_t.$$

Par le Théorème 1.1, il suffit de prouver que la diffusion Z satisfait $\mathbf{H}(z_0)$ pour tout $z_0 = (i_0, x_0) \in \mathbb{R}^2$. Ici $A_0(i, x) = \begin{bmatrix} x \\ \tilde{b}(x) \end{bmatrix}$, où $\tilde{b}(x) = b(x) - \frac{1}{2}a(x)a'(x)$ et $A_1(i, x) = \begin{bmatrix} 0 \\ a(x) \end{bmatrix}$.

• **1$^{\text{er}}$ cas.** Supposons que $a(x_0) \neq 0$. Par simple calcul de crochet de champs de vecteurs

$$A_1^{(0)} = [A_0, A_1](i_0, x_0) = \begin{bmatrix} -a(x_0) \\ -a(x_0)\tilde{b}'(x_0) + a'(x_0) \end{bmatrix} \tag{3}$$

et donc $\left\{A_1(i_0, x_0), A_1^{(0)}(i_0, x_0)\right\}$ est une base de $\mathbb{R}^2$. $\mathbf{H}(i_0, x_0)$ est verifiée.

• **2$^{\text{nd}}$ cas.** Supposons que $b(x_0) \neq 0$ et qu'il existe $n \geq 1$ tel que $a(x_0) = \dots = a^{(n-1)}(x_0) = 0$, $a^{(n)}(x_0) \neq 0$.

Definissons par récurrence les champs de vecteurs suivants: $V^{(1)} = [A_0, A_1]$, $V^{(p+1)} = [A_0, V^{(p)}]$ pour $p \geq 1$ (c'est à dire $V^{(p)} = A_1^\alpha$ où α est le p-uplet $(0, \dots, 0)$). Nous avons besoin du lemme suivant qui calcule ces champs de vecteurs.

Lemme 2.2. *Pour $p \geq 1$,*

$$V^{(p)}(i, x) = \begin{bmatrix} -pa^{(p-1)}(x)\tilde{b}^{p-1}(x) + F(p-2, p-1) \\ a^{(p)}(x)\tilde{b}^p(x) + G(p-1, p) \end{bmatrix} \tag{4}$$

où $F(-1, 0) = 0$ et pour $p \geq 2$, $F(p-2, p-1)$ a la forme suivante

$$F(p-2, p-1) = a(x)P_{1,p-1}(x) + \dots + a^{(p-2)}(x)P_{p-2,p-1}(x) \tag{5}$$

et les $P_{i,p-1}(x)$ sont des polynômes en $(\tilde{b}(x), \tilde{b}'(x), \dots, \tilde{b}^{(p-1)}(x))$. Pour $p \geq 1$

$$G(p-1, p) = a(x)Q_{1,p}(x) + \dots + a^{(p-1)}(x)Q_{p-1,p}(x) \tag{6}$$

et les $Q_{i,p}(x)$ sont des polynômes en $(\tilde{b}(x), \tilde{b}'(x), \dots, \tilde{b}^{(p)}(x))$.

Preuve du lemme. On procède par récurrence. Pour $p = 1$, par (3), on a le resultat avec $G(0,1) = -a(x)\tilde{b}'(x)$.

Ensuite, on voit que si l'expression (4) est vraie au rang p, alors par un calcul de crochet

$$V^{(p+1)}(i,x) = \begin{bmatrix} -pa^{(p)}(x)\tilde{b}^p(x) + F(p-1,p) \\ a^{(p+1)}(x)\tilde{b}^{p+1}(x) + G(p,p+1) \end{bmatrix}$$

avec $F(p-1,p) = -G(p-1,p) - pa^{(p-1)}(x)(\tilde{b}^{p-1}(x))' + \tilde{b}(x)(F(p-2,p-1))'$

$$G(p,p+1) = \{-a^{(p)}(x)\tilde{b}^p(x) - G(p-1,p)\}\tilde{b}'(x) + a^{(p)}(x)\tilde{b}(x)(\tilde{b}^p(x))' + \tilde{b}(x)(G(p-1,p))'$$

Ces deux formules de récurrence pour $(F(p-1,p), G(p,p+1))$ en fonction de $(F(p-2,p-1), G(p-1,p))$ impliquent (5) et (6). Le lemme est donc prouvé. □

En utilisant $a(x_0) = \cdots = a^{(n-1)}(x_0) = 0$ et le Lemme 2.2, on voit que

$$V^{(n)}(i_0,x_0) = \begin{bmatrix} 0 \\ a^{(n)}(x_0)\tilde{b}^n(x_0) \end{bmatrix}, \quad V^{(n+1)}(i_0,x_0) = \begin{bmatrix} -(n+1)a^{(n)}(x_0)\tilde{b}^n(x_0) \\ a^{(n+1)}(x_0)\tilde{b}^{n+1}(x_0) + G(n,n+1) \end{bmatrix}$$

Comme $b(x_0) \neq 0$, $a(x_0) = 0$ et $a^{(n)}(x_0) \neq 0$, les vecteurs $V^{(n)}(i_0,x_0)$ et $V^{(n+1)}(i_0,x_0)$ engendrent $\mathbb{R}^2$. $\mathbf{H}(i_0,x_0)$ est donc vérifiée. □

Théorème 2.3. *Supposons que X soit solution de (2) où a et b sont C^∞ avec toutes leurs derivées d'ordre superieur ou égal à 1 bornées. De plus on suppose que $\mathbf{h}(x)$ est verifiée pour tout x réel. Alors, pour tout $q \in \mathbb{N}$, le vecteur $(J_0,\dots,J_q)$ admet une densité.*

Preuve. Le processus $(J_l, X_{(l+1)\Delta})_{l\geq 0}$ est une chaîne de Markov. La probabilité de transition de cette chaîne est la loi conditionnelle de $(J_0 = I_\Delta, X_\Delta)$ sachant $I_0 = 0$, $X_0 = x_0$, dont la densité existe d'après la Proposition 2.1. Notons–la $p_\Delta(x_0;(j,x))$.

Donc, le vecteur $(J_0, X_\Delta, J_1, X_{2\Delta}, \dots, J_q, X_{(q+1)\Delta})$ admet la densité sur $\mathbb{R}^{2q+2}$:

$$p(j_0,x_1,j_1,x_2,\dots,j_q,x_{q+1}) = \prod_{l=0}^{q} p_\Delta(x_l;(j_l,x_{l+1})). \tag{7}$$

Le vecteur $(J_0,\dots,J_q)$ admet donc la densité

$$\int_{\mathbb{R}^{q+1}} p(j_0,x_1,j_1,x_2,\dots,j_q,x_{q+1})dx_1\dots dx_{q+1}. \tag{8}$$

□

Remarque 2.4. *Le théorème précédent ne permet pas de traiter le cas d'un processus de Cox-Ingersoll-Ross pour X (à cause de la condition de derivée bornée pour a). Cependant, l'existence de densité pour les intégrées de ce processus est établie dans Cox-Ingersoll-Ross (1985).*

References

- D. R. Bell (1987). *The Malliavin calculus.* Pitman monographs and surveys in pure and applied mathematics 34. Longman and Wiley.

- Cox J.C., Ingersoll J.E., and Ross S.A. (1985). A theory of term structure of interest rates. *Econometrica*, **53**(2), 385–407.

- S. Kusuoka and D.W. Stroock (1985). Application of the Malliavin calculus II. *J. Fac. Sci. Univ. Sect IA math.*, **32**, 1–76.

- D. Nualart (1995). *The Malliavin calculus and related topics.* Springer–Verlag.

- S. Watanabe (1984). *Lectures on stochastic differential equation and Malliavin calculus.* Springer–Verlag.