

# Application of the Kusuoka approximation with a tree-based branching algorithm to the pricing of interest-rate derivatives under the HJM model

Mariko Ninomiya

## ABSTRACT

This paper demonstrates the application of a new higher-order weak approximation, called the Kusuoka approximation, with discrete random variables to non-commutative multi-factor models. Our experiments show that using the Heath–Jarrow–Morton model to price interest-rate derivatives can be practically feasible if the Kusuoka approximation is used along with the tree-based branching algorithm.

## 1. Introduction

In [4], Kusuoka proposed a new scheme for higher-order weak approximation of diffusion processes, which we shall refer to as the Kusuoka approximation. Lyons and Victoir [6] extensively developed the idea of this scheme by applying the notion of the free Lie algebra to establish the cubature formula on Wiener space. There have been some studies on the Kusuoka approximation with continuous random variables [9–11]; it was shown that low-discrepancy sequences effectively solve the partial sampling problem. Implementation of the Kusuoka approximation scheme with discrete random variables has been demonstrated in [7, 8]; these studies dealt with stochastic differential equations (SDEs) that are driven by one-dimensional Brownian motion.

In this paper, we focus on SDEs that are driven by multi-dimensional Brownian motions and whose vector fields are non-commutative. One of the major problems we face here is the selection of an appropriate sampling method. We show the effectiveness of the tree-based branching algorithm (TBBA) introduced by Crisan and Lyons [2] in overcoming this problem. We apply our scheme in an experiment that involves pricing interest-rate derivatives under the Heath–Jarrow–Morton (HJM) model, and use the results to verify practicability of the HJM model.

The HJM model has the form

$$dF(t, T) = \left\{ \sum_{j=1}^d s^j(t, T, F(t, T)) \int_t^T s^j(t, u, F(t, T)) du \right\} dt + \sum_{j=1}^d s^j(t, T, F) dB^j(t),$$

where  $F(t, T)$  is the forward rate seen at time  $t$  for a contract maturing at time  $T$ ,  $s^1, \dots, s^d$  are the standard deviations of  $F(t, T)$ , and  $(B^1(t), \dots, B^d(t))$  denotes a  $d$ -dimensional Brownian motion [1, 3].

Theoretically, the HJM model has been regarded as one of the most general interest-rate models. However, it has been claimed that a multi-factor implementation of the HJM framework for practical use is impossible at the present stage of computer technology [12].

---

Received 29 December 2008; revised 26 January 2010.

2000 Mathematics Subject Classification 65C30, 65C05.

This research was partially supported by the 21st Century COE Program at the Graduate School of Mathematical Sciences, University of Tokyo; it is also closely related to a study conducted during the MSc course at Imperial College London.

When we theoretically price interest-rate derivatives under the HJM model, we also need  $F(0, s)$  where  $s \in \mathbb{R}_{>0}$ . The prices of the derivatives obtained at this stage are therefore functions of  $F(0, s)$  and  $(s^1, \dots, s^n)$ , say  $J(s^1, \dots, s^n, F(0, s))$ . We seek  $(s^1, \dots, s^n)$  that decreases the difference between  $T(s^1, \dots, s^n, F(0, s))$  and  $M$ , a vector of actual prices of the derivatives observed in the market. More precisely, we try to obtain the point  $(s^1, \dots, s^n)$  which solves the following minimization problem:

$$\min_{s^1, \dots, s^n} \|M - T(\hat{s}^1, \dots, \hat{s}^n, F(0, s))\|_{L_2}. \quad (1.1)$$

This minimization procedure is called model calibration. Since model calibration is a typical optimization problem,  $T(s^1, \dots, s^n, F(0, s))$  must be calculated a number of times (as seen above). The efficacy of adopting the HJM model is therefore determined by the speed of computing  $T(s^1, \dots, s^n, F(0, s))$ . In other words, if we can calculate  $T(s^1, \dots, s^n, F(0, s))$  sufficiently fast, then the HJM model can be regarded as a practical model for pricing interest-rate derivatives. Therefore, the attempt to achieve faster computation of  $T(s^1, \dots, s^n, F(0, s))$  is an important area of focus in this problem.

The author would like to point out that Kusuoka [5] and Lyons and Victoir [6] explain how one can develop a Lie/upstairs approximation that is parallel to the Taylor/downstairs scheme used in this paper. One would expect the local error estimation to be similar in the two cases.

## 2. The Kusuoka approximation

NOTATION 2.1. Let  $B^0(t) = t$ , and let  $(B^1(t), \dots, B^d(t))$  be a  $d$ -dimensional Brownian motion. We denote by  $C_b^\infty(\mathbb{R}^N; \mathbb{R}^N)$  the set of  $\mathbb{R}^N$ -valued infinitely differentiable functions defined on  $\mathbb{R}^N$  whose derivatives are all bounded.

Let  $X(t, x)$  be a solution of the Stratonovich stochastic integral equation

$$X(t, x) = x + \sum_{i=0}^d \int_0^t V_i(X(s, x)) \circ dB^i(s), \quad (2.1)$$

where  $V_i \in C_b^\infty(\mathbb{R}^N; \mathbb{R}^N)$  for  $i = 0, 1, \dots, d$  and  $\circ dB^j(s)$  denotes the Stratonovich integral. We are interested in weak approximation, that is, approximation of  $(P_T^X f)(x) = E[f(X(T, x))]$  where  $f \in C_b^\infty(\mathbb{R}^N; \mathbb{R})$ .

Each  $V_i \in C_b^\infty(\mathbb{R}^N; \mathbb{R}^N)$  is regarded as a vector field in the following way:

$$V_i f(x) = \sum_{j=1}^N V_i^j(x) \frac{\partial f}{\partial x_j}(x) \quad (2.2)$$

for  $f \in C_b^\infty(\mathbb{R}^N; \mathbb{R})$ .

DEFINITION 2.1. Let  $\mathcal{A}_0 = \bigcup_{k=1}^\infty \{0, 1, \dots, d\}^k$  and  $\mathcal{A} = \{\emptyset\} \cup \mathcal{A}_0$ . Then, for  $\alpha \in \mathcal{A}$ ,  $|\alpha|$  and  $\|\alpha\|$  are defined as

$$|\alpha| = \begin{cases} 0 & \text{if } \alpha = \emptyset, \\ k & \text{if } \alpha = (a^1, \dots, a^k) \text{ where } a^i \in \{0, 1, \dots, d\} \text{ for } i = 1, 2, \dots, k \end{cases}$$

and

$$\|\alpha\| = |\alpha| + \text{card}(\{i : 1 \leq i \leq |\alpha|, a^i = 0\}).$$

DEFINITION 2.2. For  $\alpha, \beta \in \mathcal{A}$ , let  $\alpha * \beta$  denote  $(a^1, \dots, a^k, b^1, \dots, b^l)$  for  $\alpha = (a^1, \dots, a^k)$  and  $\beta = (b^1, \dots, b^l)$ . We define the multiple Stratonovich integral as follows:

for  $i \in \{0, 1, \dots, d\}$ ,

$$\begin{aligned} B^{\circ 0}(t) &= 1, \\ B^{\circ(i)}(t) &= B^i(t), \\ B^{\circ\alpha^*i}(t) &= \int_0^t B^{\circ\alpha}(s) \circ dB^i(s). \end{aligned}$$

DEFINITION 2.3. For two operators  $V$  and  $W$ , the Lie bracket is defined by

$$[V, W] = VW - WV.$$

Also, we define  $V_{[\alpha]}$  for  $\alpha \in \mathcal{A}$  and  $i = 0, 1, \dots, d$  by

$$\begin{aligned} V_{[\emptyset]} &= 0, \\ V_{[i]} &= V_i, \\ V_{[\alpha^*(i)]} &= [V_{[\alpha]}, V_i]. \end{aligned}$$

DEFINITION 2.4. A set of random variables  $\{Z_\alpha \mid \alpha \in \mathcal{A}_0\}$  is called an  $m$ -moment similar family if it satisfies the following three conditions:

- (i)  $Z_{(0)} = 1$ ;
- (ii)  $E[|Z_\alpha^n|] < \infty$  for all  $\alpha \in \mathcal{A}_0$  and  $n \geq 1$ ;
- (iii)  $E[Z_{\alpha_1} \dots Z_{\alpha_k}] = E[B^{\circ\alpha_1}(1) \dots B^{\circ\alpha_k}(1)]$ .

We introduce a new functional  $Q_{(s)}$  which is a Markov operator.

DEFINITION 2.5. Let  $m \in \mathbb{N}$  and let  $\{Z_\alpha \mid \alpha \in \mathcal{A}_0\}$  be an  $m$ -moment similar family. For  $0 \leq s \leq 1$ , we define

$$\begin{aligned} (Q_{(s)}f)(x) &= E \left[ f \left( \sum_{k=0}^m \frac{1}{k!} \sum_{\substack{\alpha_1, \dots, \alpha_k \in \mathcal{A} \\ \|\alpha_1\| + \dots + \|\alpha_k\| \leq m}} s^{\frac{1}{2}(\|\alpha_1\| + \dots + \|\alpha_k\|)} \right. \right. \\ &\quad \left. \left. \times (P_{\alpha_1}^0 \dots P_{\alpha_k}^0)(V_{[\alpha_1]} \dots V_{[\alpha_k]}H)(x) \right) \right], \end{aligned} \tag{2.3}$$

where  $H : \mathbb{R}^N \rightarrow \mathbb{R}^N$  is the identity map and

$$P_\alpha^0 = \frac{1}{|\alpha|} \sum_{k=1}^{|\alpha|} \frac{(-1)^{k+1}}{k} \sum_{\substack{\beta_1, \dots, \beta_k \in \mathcal{A}_0 \\ \beta_1 * \dots * \beta_k = \alpha}} Z_{\beta_1} \dots Z_{\beta_k}.$$

THEOREM 2.1 (Kusuoka [4, 5]). Let

$$\mathcal{A}^+ = \left\{ \alpha \in \mathcal{A} : \alpha \in \bigcup_{k=1}^\infty \{1, \dots, d\}^k \text{ if } |\alpha| \geq 2 \text{ and } \alpha \neq (0) \text{ if } |\alpha| = 1 \right\}.$$

Suppose that the following condition is satisfied.

(UFG) There exists an integer  $l$  and  $\{\varphi_{\alpha, \beta}\}_{\beta \in \mathcal{A}^+} \subset C_b^\infty(\mathbb{R}^N; \mathbb{R})$  such that

$$\begin{aligned} &[V_{\alpha_1}, [V_{\alpha_2}, [\dots, [V_{\alpha_{j-1}}, V_{\alpha_j}] \dots]]] \\ &= \sum_{\substack{\beta = (\beta_1, \dots, \beta_i) \in \mathcal{A}^+ \\ \|\beta\| \leq l}} \varphi_{\alpha, \beta} [V_{\beta_1}, [V_{\beta_2}, [\dots, [V_{\beta_{i-1}}, V_{\beta_i}] \dots]]] \quad \text{for } \alpha \in \mathcal{A}^+. \end{aligned}$$

Then

$$\|P_s^X f - Q_{(s)}f\|_\infty \leq Cs^{(m+1)/2} \|\text{grad}(f)\|_\infty, \tag{2.4}$$

where  $s \in (0, 1]$ ,  $f \in C_b^\infty(\mathbb{R}^N; \mathbb{R})$ , and  $C$  is a positive constant.

A  $d$ -dimensional 5-moment similar family  $\{Z_\alpha \mid \alpha \in \mathcal{A}_0\}$  can be defined as follows:

$$\begin{aligned} Z_{(0)} &= 1, \\ Z_{(i)} &= \eta_i \quad \text{for } i = 1, \dots, d, \\ Z_{(i,j)} &= \frac{1}{2}(\eta_i \eta_j + \sigma_{ij}), \quad Z_{(j,i)} = \frac{1}{2}(\eta_i \eta_j - \sigma_{ij}) \quad \text{for } 1 \leq i < j \leq d, \\ Z_{(i,i)} &= \frac{\eta_i^2}{2}, \quad Z_{(i,0)} = Z_{(0,i)} = \frac{\eta_i}{2}, \quad Z_{(i,i,i)} = \frac{\eta_i^3}{6} \quad \text{for } i = 1, \dots, d, \\ Z_{(i,i,j)} &= Z_{(j,i,i)} = \frac{\eta_j}{4}, \quad Z_{(j,j,i)} = Z_{(i,j,j)} = \frac{\eta_i}{4} \quad \text{for } 1 \leq i < j \leq d, \\ Z_{(0,0)} &= \frac{1}{2}, \\ Z_{(i,i,j,j)} &= Z_{(j,j,i,i)} = \frac{1}{8} \quad \text{for } 1 \leq i < j \leq d, \\ Z_{(0,i,i)} &= Z_{(i,i,0)} = \frac{1}{4} \quad \text{for } i = 1, \dots, d, \\ Z_{(w)} &= 0 \quad \text{otherwise,} \end{aligned}$$

where the  $\sigma_{ij}$  and  $\eta_i$  are independent random variables which satisfy  $P(\sigma_{ij} = \pm 1) = 1/2$  for  $1 \leq i < j \leq d$  and  $P(\eta_i = 0) = 2/3$ ,  $P(\eta_i = \pm\sqrt{3}) = 1/6$  for  $i = 1, \dots, d$ .

### 3. Partial sampling problem and TBBA

#### 3.1. Partial sampling problem

When we use  $\{Z_\alpha\}$ , as defined in the previous section, in the implementation of the Kusuoka approximation with  $d = 2$ , each step requires 18 points, which means that there are  $18^n$  possible points for  $n$  steps. In practice, it is impossible to calculate every point, and thus we rely on partial calculation.

#### 3.2. Tree-based branching algorithm

The TBBA enables us to choose a small number of sample points from a discrete set when approximating the expectation of some function of random variables. Following [2, 8], we define the TBBA as follows.

Let  $I$  be a countable set and let  $\mathcal{M}^P(I)$  be the set of probability measures on  $I$ , that is,

$$\mathcal{M}^P(I) = \left\{ x = (x(i))_{i \in I} \in (\mathbb{R}_{\geq 0})^I \mid \sum_{i \in I} x(i) = 1 \right\}$$

with the product topology. We denote by  $\mathcal{B}(\mathcal{M}^P(I))$  the Borel  $\sigma$ -algebra on  $\mathcal{M}^P(I)$  generated by the cylinder sets, as usual. Let  $(\Omega, \mathcal{F}, P)$  be another probability space with no atoms. Let  $a \in \mathcal{M}^P(I)$  be a probability with finite entropy  $H(a) = \sum_{i \in I} a(i) \log a(i) < \infty$ , and let  $\mathcal{A}_n^a$ ,  $n \in \mathbb{N}$ , be a collection of  $\mathcal{M}^P(I)$ -valued random variables of  $(\Omega, \mathcal{F}, P)$ , defined as follows:

$$\mathcal{A}_n^a = \left\{ \tilde{a} : \Omega \rightarrow \mathcal{M}^P(I) \mid \begin{array}{l} E_\Omega^P[\tilde{a}] = a \text{ and } \tilde{a} = \frac{1}{n} \sum_{k=1}^n \delta_{V_k} \text{ for some} \\ I\text{-valued } \mathcal{F}\text{-measurable random variables} \\ V_1, \dots, V_n \text{ of } (\Omega, \mathcal{F}, P). \end{array} \right\}, \quad (3.1)$$

where  $\delta_x$  denotes the Dirac distribution concentrated at  $x$ . We can regard  $\mathcal{A}_n^a$  as a set of unbiased estimators requiring  $n$  samples or as a set of measures realized by  $n$  samples.

Our objective is to find a ‘good’  $\tilde{a} \in \mathcal{A}_n^a$ , that is, an  $\tilde{a} \in \mathcal{A}_n^a$  whose discrepancy with  $a$  is small. To measure the discrepancy between an element  $\tilde{a}$  of  $\mathcal{A}_n^a$  and  $a$ , we introduce  $H(\tilde{a}|a)$ , the mean relative entropy of  $\tilde{a}$  with respect to  $a$ :

$$\begin{aligned}
 H(\tilde{a}|a) &= E_\Omega^P \left[ \sum_{i \in I} \tilde{a}(i) \log \left( \frac{\tilde{a}(i)}{a(i)} \right) \right] \\
 &= \sum_{i \in I} E_\Omega^P [\tilde{a}(i) \log(\tilde{a}(i))] - \sum_{i \in I} a(i) \log(a(i)).
 \end{aligned}
 \tag{3.2}$$

For  $x \in \mathbb{R}$ , let  $\lfloor x \rfloor$  denote the integer part of  $x$ , and write  $\text{fp}(x) = x - \lfloor x \rfloor$ .

DEFINITION 3.1. If for any  $i \in I$ ,  $\tilde{a} \in \mathcal{A}_n^a$  satisfies the condition that

$$\tilde{a}(i) = \begin{cases} \lfloor na(i) \rfloor / n & \text{with probability } 1 - \text{fp}(na(i)), \\ (\lfloor na(i) \rfloor + 1) / n & \text{with probability } \text{fp}(na(i)), \end{cases}
 \tag{3.3}$$

then we say that  $\tilde{a}$  has the property  $MV(n)$ .

THEOREM 3.1 [2]. Let  $\tilde{a} \in \mathcal{A}_n^a$ . Then  $H(\tilde{a}|a)$  is minimal if and only if  $\tilde{a}$  has the property  $MV(n)$ .

The TBBA consists of the following two steps. In the first step, we construct a binary tree  $T$  which is consistent with  $I$  and  $a$ ; this means that the binary tree  $T$  must satisfy the following conditions.

- (T1) Each edge  $e$  of  $T$  has positive weight  $w(e)$ .
- (T2) Each edge  $e$  has one upper node  $u(e)$  and one lower node  $l(e)$ .
- (T3) The set of lowest nodes of  $T$  is equal to  $I$ .
- (T4) For any two edges  $e_1$  and  $e_2$ , if  $e_1 \neq e_2$  and  $u(e_1) = u(e_2)$ , then  $w(e_1) + w(e_2) = 1$ .
- (T5) Let  $n_r$  be the root node of  $T$ . If  $\{e_1, e_2, \dots, e_k\}$  is a set of edges such that  $u(e_1) = n_r$ ,  $l(e_h) = u(e_{h+1})$  for  $h = 1, \dots, k - 1$  and  $l(e_k) \in I$ , then  $\prod_{h=1}^k w(e_h) = a(l(e_k))$ .

In the second step, we construct a random variable  $\xi_m$  for each node  $m$  of  $T$ . For an arbitrary node  $m$  of  $T$ , we define  $I_m$  to be the set of lowest nodes which are the descendants of  $m$  and set  $a(m) = \sum_{i \in I_m} a(i)$ . We proceed with the construction of  $\xi_m$  by applying the following two rules recursively.

Rule 1.  $\xi_{n_r} \equiv 1$ .

Rule 2. If a node  $m$  of  $T$  has two lower nodes  $m_1$  and  $m_2$ , then  $\xi_{m_1}$  and  $\xi_{m_2}$  are obtained from  $\xi_m$  by using the following algorithm. We remark that  $a(m) = a(m_1) + a(m_2)$ , and then either  $\lfloor na(m) \rfloor = \lfloor na(m_1) \rfloor + \lfloor na(m_2) \rfloor$  or  $\lfloor na(m) \rfloor = \lfloor na(m_1) \rfloor + \lfloor na(m_2) \rfloor + 1$  holds.

Case 1:  $\lfloor na(m) \rfloor = \lfloor na(m_1) \rfloor + \lfloor na(m_2) \rfloor$ . In this case, we define

$$\begin{aligned}
 \xi_{m_1} &= \lfloor na(m_1) \rfloor / n + (\xi_m - \lfloor na(m) \rfloor / n) \eta_m, \\
 \xi_{m_2} &= \lfloor na(m_2) \rfloor / n + (\xi_m - \lfloor na(m) \rfloor / n) (1 - \eta_m),
 \end{aligned}$$

where

$$\eta_m = \begin{cases} 0 & \text{with probability } \text{fp}(na(m_2)) / \text{fp}(na(m)), \\ 1 & \text{with probability } \text{fp}(na(m_1)) / \text{fp}(na(m)). \end{cases}$$

Case 2:  $\lfloor na(m) \rfloor = \lfloor na(m_1) \rfloor + \lfloor na(m_2) \rfloor + 1$ . In this case, we define

$$\begin{aligned}
 \xi_{m_1} &= (\lfloor na(m_1) \rfloor + 1) / n + (\xi_m - (\lfloor na(m) \rfloor + 1) / n) \eta_m, \\
 \xi_{m_2} &= (\lfloor na(m_2) \rfloor + 1) / n + (\xi_m - (\lfloor na(m) \rfloor + 1) / n) (1 - \eta_m),
 \end{aligned}$$

where

$$\eta_m = \begin{cases} 0 & \text{with probability } (1 - \text{fp}(na(m_2)))/(1 - \text{fp}(na(m))), \\ 1 & \text{with probability } (1 - \text{fp}(na(m_1)))/(1 - \text{fp}(na(m))). \end{cases}$$

**THEOREM 3.2.** *For an arbitrary node  $m$  of  $T$ , the random variable  $\xi_m$  defined by the above algorithm satisfies*

$$\xi_m = \begin{cases} \lfloor na(m) \rfloor / n & \text{with probability } 1 - \text{fp}(na(m)), \\ (\lfloor na(m) \rfloor + 1) / n & \text{with probability } \text{fp}(na(m)). \end{cases}$$

From Theorem 3.1 and this theorem, it is clear that by using the TBBA, we can obtain the random measure  $\tilde{a}$  which attains the minimal mean relative entropy with respect to  $a$ .

We remark that the TBBA is a well-known stratified sampling technique.

#### 4. The HJM model

We now introduce the HJM model; our treatment is based on [1, 3].

**NOTATION 4.1.** Let  $P(t, T)$  be the price at time  $t$  of a zero-coupon bond paying 1 at time  $T$ . Let  $f(t, T_1, T_2)$  be a forward rate as seen at time  $t$  for the period between time  $T_1$  and time  $T_2$ . Also, let  $F(t, T)$  be an instantaneous forward rate as seen at time  $t$  for a contract that matures at time  $T$ , and let  $r(t)$  be a short-term risk-free interest rate at time  $t$ . In addition,  $v^j(t, T, P(t, T))$  will denote the  $j$ th factor of the volatility of  $P(t, T)$ .

The risk-neutral process for  $P(t, T)$  has the form

$$dP(t, T) = r(t)P(t, T) dt + \sum_{j=1}^d v^j(t, T, P(t, T))P(t, T) dB^j(t). \quad (4.1)$$

Here, we should note that the drift of  $P(t, T)$  equals  $r(t)P(t, T)$ , which is equivalent to the arbitrage-free condition [1].

Let

$$dF(t, T) = m(t, T, P(t, T)) dt + \sum_{j=1}^d s^j(t, T, P(t, T)) dB^j(t). \quad (4.2)$$

From equations (4.1) and (4.2) and the definition of an instantaneous forward rate, we obtain the following HJM drift conditions:

$$m(t, T, P(t, T)) = \sum_{j=1}^d s^j(t, T, P(t, T)) \int_t^T s^j(t, u, P(t, u)) du$$

and

$$s^j(t, T, P(t, T)) = v_2^j(t, T, P(t, T)),$$

where  $v_2^1, \dots, v_2^d$  indicate, respectively, the partial derivatives of  $v^1, \dots, v^d$  with respect to the second variable [1].

#### 5. Application

In this section, we apply the Kusuoka approximation with discrete random variables to the pricing of a caplet according to the HJM model.

5.1. Preliminaries

In this experiment, we use the Nelson–Siegel family for the initial yield curve  $F(0, s)$ , which then has the form

$$F(0, s) = z_1 + z_2e^{-z_4s} + z_3se^{-z_4s} \tag{5.1}$$

for  $s \in \mathbb{R}_{>0}$ , where  $z_i \in \mathbb{R}$  for  $i = 1, 2, 3, 4$ .

Also, we consider a two-factor model in which

$$s^1(t, T, P(t, T)) = C_1$$

and

$$s^2(t, T, P(t, T)) = C_2 - \lambda(T - t),$$

where  $s^1$  and  $s^2$  denote the standard deviations of  $F(t, T)$ , as in Section 1, and  $C_1, C_2$  and  $\lambda$  are constants.

5.1.1. *Caplets.* We give a brief introduction to caplets, which constitute a very basic block of interest-rate derivatives [1, 3].

NOTATION 5.1. Let  $L$  be the LIBOR rate set at time  $T_1$  and paid at time  $T_2$ , and let  $K$  be a fixed rate. Let  $\theta_{T_1T_2}$  denote an accrual factor between  $T_1$  and  $T_2$ .

The payoff of the caplet at time  $T_2$  is given by

$$\theta_{T_1T_2}[L - K]^+.$$

Then, the price is known to be as follows (see [1]):

$$P(0, T_2)E\left[\left[\frac{1}{P(T_1, T_2)} - 1 - (T_2 - T_1)K\right]^+\right].$$

From (5.1), it is clear that

$$\begin{aligned} P(0, T_2) &= \exp\left(\int_0^{T_2} F(0, u) du\right) \\ &= \exp\left\{-z_2T_2 + \frac{z_2}{z_4}(e^{-z_4T_2} - 1) + \frac{z_3}{z_4}T_2e^{-z_4T_2} + \frac{z_3}{z_4}(e^{-z_4T_2} - 1)\right\}. \end{aligned}$$

Therefore, in order to price the caplet, we need to investigate the process of  $P(t, s)$ , where  $t < s$ . In addition, the process of  $r(t)$  has to be considered, since the drift of  $P(t, T)$  includes  $r(t)$ , as shown by equation (4.1).

5.1.2. *Kusuoka approximation.* Since  $r(t) = F(t, t)$  and  $F(t, t) = F(0, t) + \int_0^t dF(u, t)$ , we have

$$r(t) = F(0, t) + \int_0^t dF(u, t). \tag{5.2}$$

Upon substituting (4.2) into (5.2) and carrying out some calculations, we obtain

$$r(t) = F(0, 0) + \int_0^t \left(A(u) + \sum_{j=1}^d Y^j(u)\right) du + \sum_{j=1}^d \int_0^t v_2^j(u, u, P(u, u)) dB^j(u) \tag{5.3}$$

with

$$A(u) = F_2(0, u) + \sum_{j=1}^d \int_0^u \{v^j(s, u, P(s, u))v_{22}^j(s, u, P(s, u)) + (v_2^j(s, u, P(s, u)))^2\} ds, \tag{5.4}$$

$$Y^j(u) = \int_0^u v_{22}^j(s, u) dB^j(s),$$

where  $F_2(0, u)$  is the partial derivative of  $F(0, u)$  with respect to the second variable and  $v_{22}^1, \dots, v_{22}^d$  are, respectively, the second-order derivatives of  $v^1, \dots, v^d$  with respect to the second variable.

From the definitions of  $s^1$  and  $s^2$ , we have the following:

$$\begin{aligned} v^1(t, T, P(t, T)) &= C_1(T - t), \\ v^2(t, T, P(t, T)) &= (T - t) \left( C_2 - \frac{\lambda}{2}(T - t) \right), \\ v_{22}^1(t, T, P(t, T)) &= C_1, \\ v_{22}^2(t, T, P(t, T)) &= C_2 - \lambda(T - t), \\ v_{22}^1(t, T, P(t, T)) &= 0 \quad \text{and} \quad v_{22}^2(t, T, P(t, T)) = -\lambda. \end{aligned} \tag{5.5}$$

On substituting (5.5) into the equation for  $A(u)$ , we get

$$A(u) = F_2(0, u) + (C_1^2 + C_2^2)u - \frac{3}{2}C_2\lambda u^2 + \frac{1}{2}\lambda^2 u^3$$

and  $Y^1(u) = 0$ , from the definition of  $Y^1(u)$ .

All things considered, the diffusion process  $X$  discussed here has to have four dimensions with the following elements:

$$x_1(u) = P(u, T), \quad x_2(u) = r(u), \quad x_3(u) = Y^2(u) \quad \text{and} \quad x_4(u) = u.$$

Therefore,  $X$  has the form

$$X(t, \hat{x}) = \hat{x} + \sum_{j=0}^2 \int_0^t \hat{V}_j(x_1(u), x_2(u), x_3(u), x_4(u)) dB^j(u),$$

where

$$\begin{aligned} \hat{x} &= \begin{pmatrix} P(0, T) \\ r(0) \\ 0 \\ 0 \end{pmatrix}, \\ \hat{V}_0(x_1(u), x_2(u), x_3(u), x_4(u)) &= \begin{pmatrix} x_1(u)x_2(u) \\ F_2(0, x_4(u)) + (C_1^2 + C_2^2)x_4(u) - \frac{3}{2}C_2\lambda((x_4(u))^2) + \frac{1}{2}\lambda^2(x_4(u))^3 + x_3(u) \\ 0 \\ 1 \end{pmatrix}, \end{aligned}$$

$$\hat{V}_1(x_1(u), x_2(u), x_3(u), x_4(u)) = \begin{pmatrix} C_1(T - x_4(u))x_1(u) \\ C_1 \\ 0 \\ 0 \end{pmatrix}$$

and

$$\hat{V}_2(x_1(u), x_2(u), x_3(u), x_4(u)) = \begin{pmatrix} (T - x_4(u)) \left( C_2 - \frac{\lambda}{2}(T - x_4(u)) \right) x_1(u) \\ C_2 \\ -\lambda \\ 0 \end{pmatrix}.$$

When we apply the Kusuoka approximation, it is convenient to transform the process into an expression involving the Stratonovich integral instead of the Ito integral.



The process  $X$  in Stratonovich form becomes

$$X(t, x) = x + \sum_{j=0}^2 \int_0^t V_j(x_1(u), x_2(u), x_3(u), x_4(u)) \circ dB^j(u),$$

where

$$\begin{aligned}
 & x = \hat{x}, \\
 & V_0(x_1(u), x_2(u), x_3(u), x_4(u)) \\
 &= \begin{pmatrix} x_1(u)x_2(u) - \frac{1}{2}x_1(u) \left\{ C_1^2 + \left( C_2 - \frac{\lambda}{2}(T - x_4(u)) \right)^2 \right\} (T - x_4(u))^2 \\ F_2(0, x_4(u)) + (C_1^2 + C_2^2)x_4(u) - \frac{3}{2}C_2\lambda(x_4(u))^2 + \frac{1}{2}\lambda^2(x_4(u))^3 + x_3(u) \\ 0 \\ 1 \end{pmatrix},
 \end{aligned}$$

$$V_1(x_1(u), x_2(u), x_3(u), x_4(u)) = \hat{V}_1(x_1(u), x_2(u), x_3(u), x_4(u))$$

and

$$V_2(x_1(u), x_2(u), x_3(u), x_4(u)) = \hat{V}_2(x_1(u), x_2(u), x_3(u), x_4(u)).$$

### 5.2. Implementation

In this experiment, we consider a 5-moment similar family, as defined in Section 2, with  $d = 2$ .

Here, we focus on the model which has the following  $V_0, V_1$  and  $V_2$ :

$$\begin{aligned}
 V_0(x) &= \left[ x_1x_2 - \frac{1}{2}x_1 \left\{ C_1^2 + \left( C_2 - \frac{\lambda}{2}(T - x_4) \right)^2 \right\} (T - x_4^2) \right] \frac{\partial}{\partial x_1} \\
 &\quad + \left[ F_2(0, x_4) + (C_1^2 + C_2^2)x_4 - \frac{3}{2}C_2\lambda x_4^2 + \frac{1}{2}\lambda^2 x_4^3 + x_3 \right] \frac{\partial}{\partial x_2}, \\
 V_1(x) &= C_1(T - x_4)x_1 \frac{\partial}{\partial x_1} + C_1 \frac{\partial}{\partial x_2}, \\
 V_2(x) &= (T - x_4) \left( C_2 - \frac{\lambda}{2}(T - x_4) \right) x_1 \frac{\partial}{\partial x_1} + C_2 \frac{\partial}{\partial x_2} - \lambda \frac{\partial}{\partial x_3}.
 \end{aligned}$$

The approximation operator  $Q_{(s)}$  introduced in Definition 2.5 can be expressed in four dimensions as follows:

$$(Q_{(s)}f)(x) = E[f(G(s, x, \sigma, \eta_1, \eta_2))],$$

where

$$G(s, x, \sigma, \eta_1, \eta_2) = \begin{pmatrix} G_1(s, x, \sigma, \eta_1, \eta_2) \\ G_2(s, x, \sigma, \eta_1, \eta_2) \\ G_3(s, x, \sigma, \eta_1, \eta_2) \\ G_4(s, x, \sigma, \eta_1, \eta_2) \end{pmatrix}.$$

We have not provided the details of the four elements of  $G$ , since this would require 24 pages, as can be seen in [13]. We use a particle method for the implementation of the TBBA; see [2, 8].

### 5.3. Results

In this experiment, we set the fixed rate for the caplet to be  $K = 0.04$ ; we take the maturity time of the option as  $T_1 = 5.0$  and the maturity time of the reference rate as  $T_2 = 6.0$ ; the parameters of the volatility functions are  $C_1 = 0.01$ ,  $C_2 = 0.01$  and  $\lambda = 0.003$ ; the parameters of the initial yield curve are  $z_1 = 0.03$ ,  $z_2 = -0.01$ ,  $z_3 = 0.009$  and  $z_4 = 0.15$ .

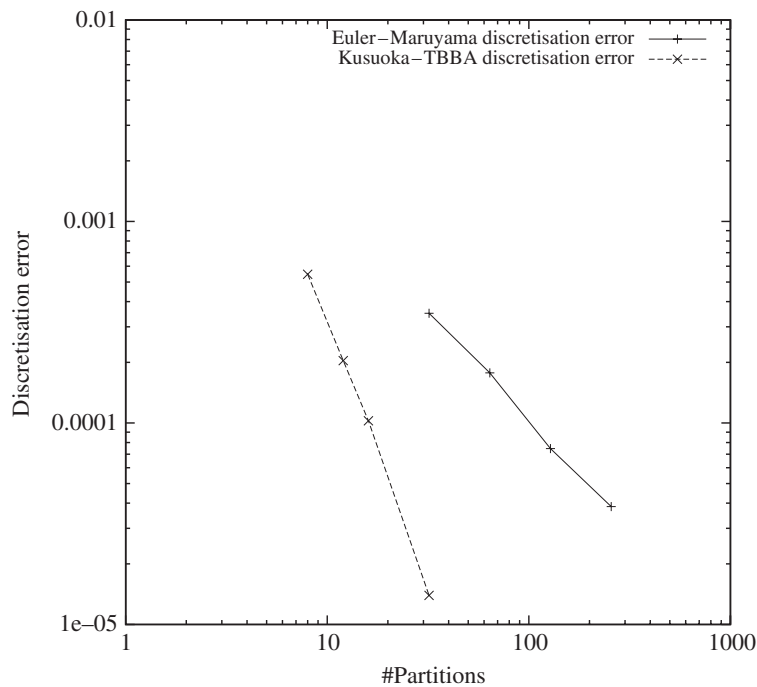


FIGURE 5.1. Discretisation error of the Euler–Maruyama and Kusuoka–TBBA methods.

5.3.1. *Two types of errors.* Two types of errors, discretisation error and integration error, arise in the calculation of  $E[f(X(T_1, x))]$ .

We define  $Y$  as a discrete random variable obtained by applying (2.3) repeatedly; specifically,

$$(Q_{(s_n)}Q_{(s_{n-1})} \cdots Q_{(s_1)}f)(x) = E[f(Y)],$$

where  $s_n + \cdots + s_1 = T_1$ .

The discretisation error is defined by

$$|P_{T_1}^X f - E[f(Y)]|. \quad (5.6)$$

The integration error is the error that arises in numerical calculations of  $E[f(Y)]$ .

5.3.2. *Discretisation error.* Figure 5.1 shows the relation between the number of partitions and the discretisation error (with both quantities plotted on a logarithmic scale). It should be noted that we took  $6.324997544 \times 10^{-2}$ , which was obtained from the Kusuoka–TBBA method with  $1.0 \times 10^8$  samples and  $n = 64$ , to be the true value. For the Euler–Maruyama approximation, discretisation errors were plotted with  $n = 32, 64, 128$  and  $256$ . The number of partitions required for comparable discretisation errors in the Kusuoka–TBBA case were just  $n = 8, 12, 16$  and  $32$ .

From Figure 5.1, we can see that the slope obtained in the case of the Euler–Maruyama scheme is  $-1$  as expected, whereas the slope obtained from the Kusuoka–TBBA approximation is steeper than  $-((5+1)/2 - 1) = -2$ , which is consistent with Theorem 2.1. Moreover, it is almost certain that to achieve 3-digit accuracy, it is sufficient to take  $n = 12$  in the Kusuoka approximation, while  $n$  must be at least  $64$  in the Euler–Maruyama approximation.

5.3.3. *Naive MC error.* The integration error for the naive Monte Carlo method is called the naive MC error and is defined by

$$\left| \frac{\sum_{k=1}^M f(Y_k)}{M} - E[f(Y)] \right|, \quad (5.7)$$

where the  $Y_k$  are independent random variables whose distributions are identical to that of  $Y$ .

Let  $h$  and  $g$  be functions such that  $h(x) = x^2$  and  $g = h \circ f$ . Then we can easily obtain

$$\text{Var}[f(Y)] = \text{Var}[f(X(T_1, x))] + O(n^{-2} \|\text{grad}(g)\|_\infty).$$

Therefore, the fact stated in the next remark follows.

REMARK 5.1. As long as we use the naive Monte Carlo method (that is, the Monte Carlo method without any variance reduction techniques) for the numerical approximation of  $E[f(X(T_1, x))]$ , the number of sample points required to attain a given accuracy is independent of both the number of partitions and the order of the approximation scheme.

In order to compare integration errors from using the Euler–Maruyama approximation (Euler–Maruyama–MC) with those from using the Kusuoka approximation with naive Monte Carlo (Kusuoka–MC) or the Kusuoka approximation with TBBA (Kusuoka–TBBA), we selected from Figure 5.1 three combinations of the number of partitions, shown in the following table.

Kusuoka	8	12	16
Euler–Maruyama	32	64	128

REMARK 5.2.

- For the Euler–Maruyama–MC scheme, we generated ten batches of Monte Carlo simulations with  $1.0 \times 10^7$  samples each and calculated the standard deviation of these batches around the convergence limit, which we took to be the average of a large number of Monte Carlo evaluations.
- For the Kusuoka–MC scheme, we followed the same procedure as in the Euler–Maruyama–MC case; here, the result of a Kusuoka–TBBA approximation using a large number of samples was taken as the convergence limit.
- For the Kusuoka–TBBA scheme, we performed the approximation ten times and calculated the standard deviation for each around the convergence limit; this limit was the result obtained from using a large number of samples.

The results obtained from using each combination are shown in Figures 5.2–5.4. These figures show the following noteworthy points.

(1) Kusuoka–TBBA gives much better results than the other two approximations for all three combinations of the number of partitions. To achieve 3-digit accuracy in the Kusuoka–TBBA approximation, we need only  $1.0 \times 10^4$  samples when  $n = 8$  and  $2.0 \times 10^4$  samples when  $n = 12$  or 16; in contrast, with the other two approximations we need four to ten times the number of samples required for Kusuoka–TBBA.

(2) When the ratio of the number of samples to the number of possible paths of random variables is not sufficiently high, the effect of the TBBA tends to be small.

5.3.4. *Evaluation of the effect of Kusuoka approximation with TBBA.* As observed in the figures, the speedup ratio depends on the required accuracy. For instance, to achieve a discretisation error of  $2.0 \times 10^{-4}$ ,  $n$  has to be approximately 64 in Euler–Maruyama–MC and 12 in Kusuoka–TBBA. From Figure 5.3, we observe that for an integration error of  $2.0 \times 10^{-4}$ ,

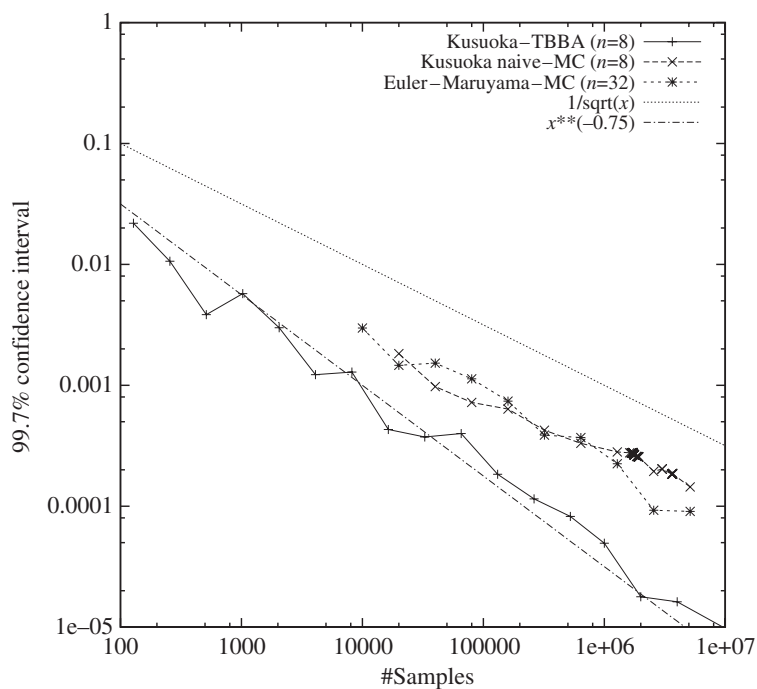


FIGURE 5.2. Convergence speed of the  $n = 32$  Euler-Maruyama and  $n = 8$  Kusuoka approximations.

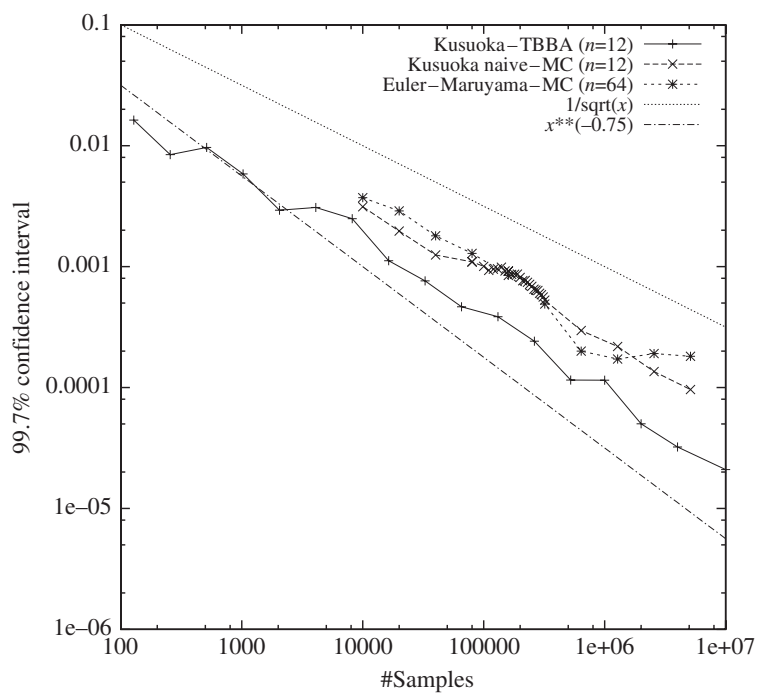


FIGURE 5.3. Convergence speed of the  $n = 64$  Euler-Maruyama and  $n = 12$  Kusuoka approximations.

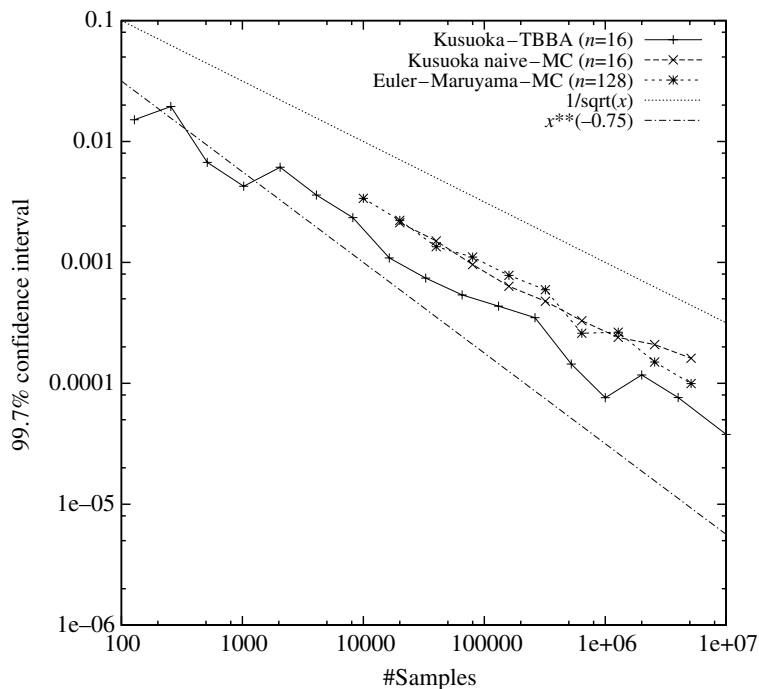


FIGURE 5.4. Convergence speed of the  $n = 128$  Euler–Maruyama and  $n = 16$  Kusuoka approximations.

we require  $1.0 \times 10^7$  samples in Euler–Maruyama–MC and  $3.0 \times 10^5$  samples in Kusuoka–TBBA. In this case, the computation time on our computer (with a Pentium 4, 1.7 GHz processor) was 725 seconds using Euler–Maruyama–MC and 24 seconds using Kusuoka–TBBA. Since the slope of the discretisation error in Kusuoka–TBBA is much steeper than that in Euler–Maruyama–MC, the speedup ratio would increase if we tried to increase the accuracy.

## 6. Discussion

We have shown that using the Kusuoka approximation with discrete random variables, particularly the Kusuoka–TBBA scheme, makes a significant contribution to computational performance. The improvements observed experimentally indicate the feasibility of using the HJM model as a practical model for pricing interest-rate derivatives.

However, we have not been able to overcome the following problems.

(1) In general, it is extremely difficult to construct  $G(s, x, \sigma, \eta_1, \eta_2)$ . In our case, although the Lie algebra generated by the vector fields  $V_0$ ,  $V_1$  and  $V_2$  was not very complicated, as it was possible to select the forms of  $s^1$  and  $s^2$ , generating  $G(s, x, \sigma, \eta_1, \eta_2)$  was still a laborious task. Allowing any  $s^j$  to be non-polynomial would lead to a formidable problem.

It ought to be noted again that there are some other implementation methods (see, for example, [9–11]) in which continuous random variables are used instead of discrete random variables as in this study. In such methods, symbolic calculation is successfully avoided, and the use of low-discrepancy sequences results in the same effect as that of TBBA in our study.

(2) The TBBA is a stratified sampling technique. We have not, however, arrived at a convincing theoretical explanation for the ratio that suppresses the effect of the TBBA, as observed in the comparison of convergence speeds of the different approximation methods.

*References*

1. D. BRIGO and F. MERCURIO, *Interest rate models – theory and practice: with smile, inflation and credit*, 2nd edn (Springer, Berlin, 2007).
2. D. CRISAN and T. LYONS, ‘Minimal entropy approximations and optimal algorithms for the filtering problem’, *Monte Carlo Methods Appl.* 8 (2002) 343–355.
3. J. HULL, *Options, futures, and other derivatives* (Prentice Hall, Upper Saddle River, NJ, 2000).
4. S. KUSUOKA, ‘Approximation of expectation of diffusion process and mathematical finance’, *Taniguchi Conference on Mathematics Nara ’98*, Advanced Studies in Pure Mathematics 31 (eds M. Maruyama and T. Sunada; Mathematical Society of Japan, Tokyo, 2001) 147–165.
5. S. KUSUOKA, ‘Approximation of expectation of diffusion processes based on Lie algebra and Malliavin calculus’, *Adv. Math. Econ.* 6 (2004) 69–83.
6. T. LYONS and N. VICTOIR, ‘Cubature on Wiener space’, *Proc. R. Soc. Lond. Ser. A Math. Phys. Sci.* 460 (2004) 169–198.
7. S. NINOMIYA, ‘A new simulation scheme of diffusion processes: application of the Kusuoka approximation to finance problems’, *Math. Comput. Simulation* 62 (2003) no. 3–6, 479–486.
8. S. NINOMIYA, ‘A partial sampling method applied to the Kusuoka approximation’, *Monte Carlo Methods Appl.* 9 (2003) 27–38.
9. M. NINOMIYA, ‘A new weak approximation scheme of stochastic differential equations by using the Runge–Kutta method’, PhD Thesis, The University of Tokyo, Tokyo, 2008.
10. M. NINOMIYA and S. NINOMIYA, ‘A new weak approximation scheme of stochastic differential equations by using the Runge–Kutta method’, *Finance Stoch.* 13 (2009) no. 3, 415–443.
11. S. NINOMIYA and N. VICTOIR, ‘Weak approximation of stochastic differential equations and application to derivative pricing’, *Appl. Math. Finance* 15 (2008) 107–121.
12. R. REBONATO, *Interest-rate option models*, 2nd edn (John Wiley & Sons, Chichester, 1998).
13. M. SHIMIZU, ‘Application of the Kusuoka approximation with tree based branching algorithm to pricing interest-rate derivatives with the HJM model’, Master Thesis, Imperial College of Science, Technology and Medicine, London, 2002.

Mariko Ninomiya  
3-1, Hongo 7-chome  
Bunkyo-ku, Tokyo, 113-0033  
Japan

[mariko.nnmy@gmail.com](mailto:mariko.nnmy@gmail.com)