國立臺灣大學理學院應用數學科學研究所

碩士論文

Institute of Applied Mathematical Sciences

College of Science

National Taiwan University

Master Thesis

閾值擴散過程的近似最大概似估計法

Approximate Maximum Likelihood Estimation of a Threshold Diffusion Process

余定宏

Ting-Hung Yu

指導教授：蔡恆修博士

Advisor: Henghsiu Tsai, Ph.D.

中華民國 106 年 6 月

June, 2017

doi:10.6342/NTU201701414

doi:10.6342/NTU201701414

誌謝

能夠完成此份碩論，我想先謝謝我的指導教授-蔡恆修博士，在其擔任中研院統計所研究員忙碌的研究時光中，願意撥空擔任我的碩論指導教授，並且也在研究進行的過程裡，給予我許多嘗試的機會和犯錯的空間，甚至也在異常忙碌的時間內，特別將時間排給我討論，也常在討論的過程中給予駑鈍的我直接的建議或是幫助，或是在半夜中回覆信件解答我的疑惑，論文才得以完成，取得學位。(我猜想老師的 H0: 至少可以畢業，然後在顯著水準 (Significant Level) 設定為 0.05 的情況下，決策為不拒絕虛無假設 (Null Hypothesis)，才能讓我順利畢業拿到碩士學位，希望自己未來別讓老師覺得犯了型 II 錯誤 (Type II Error) XDD)

其次，也非常感謝擔任我口試委員的銀慶剛老師和翁久幸老師，非常感謝老師們願意撥空聽我報告，並且也在口試過程中給予我關於這份論文的提醒和建議，雖然目前在理論結果部分還有許多不足，但仍認同我的努力，並且也相信我還會繼續進步，再次感謝老師們正面的鼓勵和指導。

另外，非常感謝家人的支持，若沒有你們在背後所給予的不論是心靈上或是財務上的幫助，我想憑我一個人可能也不會走到今天；特別是從小看著我到大的奶奶，雖然您認同我在碩二下這半年沒回花蓮努力在臺北趕論文的決心，雖然您還是時常會忍不住在電話或視訊中罵我不孝孫，但我知道您還是愛我的，雖然以後只會更忙碌，相聚時光也會更少，但至少這個階段，我又完成了一件不怎麼偉大但卻可以讓您很開心的事了，希望我的努力也能讓你 (們) 繼續感到光榮。

在臺大匆匆兩年寒暑，這段幾乎每天都要去研究室 (天數 546 室) 或唸書或寫作業的日子，這個過程中，也謝謝認識的每一位大家，不論是教過我的教授：陳宏教授、江金倉教授、張志中教授、黃啟瑞教

doi:10.6342/NTU201701414

授、管中閔教授、銀慶剛教授；讓我擔任微積分助教的鄭明燕教授；時常幫助我的系辦助教們：稚敏姐、雅真助教、怡茹助教、···；平時聊天或修課討論的大家：紹宣學長、融昇學長、立榜哥、膺任姐、儲輔哥、書妤姐、耿弘哥、偉宏哥、唯植哥、致凱哥、逸安哥、宗堂哥、亦行哥、柏緯哥、俊昇哥、俊緯、子模、奎佑、以恆、太維、學涵、芷瑄、啟樺、世緯、智涵、林真、福軒、易達、睿耆、尚儒、君宥、皓翔、庭翰、植皓、泉楷、書賢、凱婷、···; 教會和團契的朋友們：鄭牧師、葉牧師、佩儀姐、盈昌哥、懷力哥、惠恒哥、志碩哥、彥安、乃瑜、鼎方、嫚翎、思堯、恩萱、淵傑、珊儀、昱宏、···；在臺大查經班的大家：思豪哥、百達老師、宏驛哥、世偉學長、以達學長、淑英學姐、麗英學姐、麒儒學長、鈴珮學姐、淑君學姐、···；以及任何我未在這提及但在碩士班期間中幫助我甚巨的人們！

雖然天數館座落在醉月湖畔，但卻鮮少真有心思駐足欣賞她的美···，畢業以後，應該也更難有此閒情雅致去玩味了吧！收拾回憶的最後，坐在白牆環繞的研究室中回想這一切，雖偶有上五樓的快活，或許還難以領略望盡天涯路的求學氣魄，更別論為伊消得人憔悴的煎熬，或甚那人卻在燈火闌珊處的渾然自得，但這兩年的求學時光，也教人習得相忘於江湖的豪氣，這或許也稱的上是略有小成，哈哈...，最後，期許自己能繼續向著標竿直跑，將榮耀歸與上帝。

. 筆　於臺大醉月湖畔天數 546 室

doi:10.6342/NTU201701414

摘要

本論文係根據 Su 和 Chan 於 2015 年 [23] 所提出的 quasi-likelihood estimator(QLE) 所發想而成，其中我們關注在如何計算單一閾值擴散過程 (a threshold diffusion process) 其最大概似估計量 (maximum likelihood estimator) 的問題。起因是現實世界的觀測的資料是離散而且可能是不規則區間 (irregularly-spaced)，且針對閾值擴散過程其最大概似估計量 (maximum likelihood estimator) 也因其概似函數 (likelihood function) 為非線性的結構，所以閾值擴散過程的最大概似估計量只有以隨機積分 (stochastic integrals) 表示的隱形式 (implicit form)，也因此產生一個問題：如何使用離散而不規則的資料去“近似”單一閾值擴散過程的最大概似估計量？針對這個問題，我們提出了所謂“近似最大概似估計法 (approximate maximum likelihood method)”去估計單一閾值擴散過程上的參數，而根據此法而得的估計量則稱為“近似最大概似估計量 (approximate maximum likelihood estimator；AMLE)”；更進一步，我們利用模擬的結果去給出近似最大概似估計量的大樣本性質，並且也利用這個方法針對長期的利率結構進行一些判讀，而使用的利率資料為 Federal Reserve Bank’s H15 資料集中的 three-month US treasury rate 和 10-year treasury constant maturity rate-3-month treasury bill: secondary market rate 。

關鍵字：不規則區間資料, 閾值擴散過程, 非線性連續時間序列, 隨機微分方程, 最大概似估計。

doi:10.6342/NTU201701414

doi:10.6342/NTU201701414

Abstract

Based on the idea of quasi-likelihood estimator(QLE) in Su and Chan(2015), we focus on a problem arisen from estimating the maximum likelihood estimators(MLEs) for a threshold diffusion process. Since the data observed are discrete in the real world, and MLE for a threshold diffusion process is an implicit form of stochastic integrals due to the nonlinear structure of likelihood function of the threshold diffusion process, there might arise the question: how to ”approximate” the MLEs via using the discrete data without the analytic form of the estimator? Therefore, we propose an approximate maximum likelihood method for estimating MLEs of a threshold diffusion process, and the estimator we obtain is called approximate maximum likelihood estimator(AMLE). Moreover, from the simulation results, we give some conjectures about the large sample properties of the AMLE. Finally, we apply our method to study the term structure of a long time series of US interest rates (three-month US treasury rate and 10-year treasury constant maturity rate-3-month treasury bill: secondary market rate, which are based on the Federal Reserve Bank’s H15 data set).

Keywords:irregularly-spaced data, threshold diffusion process, nonlinear continuous time series, stochastic differential equation, maximum likelihood estimator.

doi:10.6342/NTU201701414

doi:10.6342/NTU201701414

Contents

doi:10.6342/NTU201701414

doi:10.6342/NTU201701414

List of Figures

doi:10.6342/NTU201701414

doi:10.6342/NTU201701414

List of Tables

doi:10.6342/NTU201701414

doi:10.6342/NTU201701414

Chapter 1

Introduction

In recent years, the diffusion process has been a standard tool for modeling the uncertainty of risky assets in financial markets, and the results are usually considered to be the benchmarks when people trade or hedge these risky assets in practice. Specifically, for the purpose of modeling the market yield, e.g., interest rate, there arises a number of diffusion models in the literature, such as the famous CIR model(Cox et al. 1985, [11]) which is used to capture the pattern of an interest rate process. Though the CIR model is equipped a square root diffusion term, it is still fail to catch asymmetric volatility when there exists nonlinear characteristic and conditional heteroscedasticity in real data.

Since we face aforementioned problems, it is natural to extend a linear model to a nonlinear model, and the common one is the famous continuous-time threshold autoregressive(CTAR) model, which is introduced by Tong(1990) [24]. By the definition(see section 2), the first-order CTAR model turns out to be the threshold diffusion(TD) process automatically. In this case, a TD process is a solution to a stochastic differential equation(SDE) with a piecewise linear drift term and a piecewise smooth diffusion term, e.g., a piecewise constant function or a piecewise power function.

Since the underlying processes in continuous time models are usually given, the maximal likelihood(ML) theory is a natural candidate to be considered when we want to estimate the parameters in those models. In this case, Feigin(1976) [12] derives the ML theory for continuous time processes via using the martingale limit theorem and some properties of Markov processes under some regular conditions. However, Feigin's setting focused on asymptotic properties of ML estimates for continuous models with differentiable drift term and a constant diffusion term though it can be used in a wide-range regular

doi:10.6342/NTU201701414

distribution class.

After that, Tong and Yeung(1991) [26] extended the idea from threshold autoregressive(TAR) time series to CTAR. In the view of nonlinear diffusion models, Brockwell and Hyndman(1992) [4] developed a recursive method to calculate the estimates of CTAR and use it to do prediction. Moreover, Brockwell(1994) [6] used the same idea to extend CTAR models to continuous-time threshold ARMA(CTARMA) processes with a positive boundary on threshold parameters, and Brockwell and Stramer(1995) [5] emphasized on how to approximate the boundary to the threshold parameter.

It also needs to mention the brilliant works that Brockwell et al(2007) [3] did. They proposed conditional maximal likelihood estimation(CMLE) for CAR(p) model via using closely-spaced discrete data and derive its asymptotic normality. Moreover, they also used the same method to give the formula to calculate and to simulate estimates for CTAR(p) models with a constant diffusion term. To sum up, their work gives a simple but efficient approach to calculate a good estimate for CTAR(p) with constant diffusion at a heuristic stage.

On the other hand, to focus on the MLE for the differentiable nonlinear diffusion process ,Aït-Sahalia(2002) [1], Aït-Sahalia and Mykland(2004) [2] ,and Chang et al.(2011) [10] did some exciting works. The first one started to use a series-expansion approach to implement regular-spaced data to approximate the transition density of the multivariate differentiable nonlinear diffusion process, but it did not analyze the asymptotic behavior of the approximate maximum likelihood estimation(AMLE) which he named. Thus, the third one gave the details of the asymptotic results in some practical scenarios for the first one. The second one follow the first one as well. They extended such approximate approach with some modifications to the time separating successive observation which may possibly be random. Although what they have done extend our knowledge to use the discrete data to estimate the parameters on the multivariate differentiable nonlinear diffusion process, yet it is still mysterious on the field of estimating the threshold diffusion process.

Surprisingly, Su and Chan(2015) [23] proposed quasi-likelihood estimation(QLE) method to solve the problem of estimating the drift parameters indexing the drift term of a threshold diffusion process without the prior knowledge of the functional form of the diffusion term. Their work, which focused on the piecewise linear drift and allows the misspecification on the diffusion form, extended the case of CTAR(1) in Brockwell et al(2007) [3].

doi:10.6342/NTU201701414

Since the underlying process is the standard Brownian motion, the estimation framework and asymptotic theorems of QLE parallel to the those of the least squared method for TAR models, which was proposed by Chan(1993) [8], Chan and Tsay(1988) [9].

Here, we found a problem that the QL approach would fail when threshold phenomenon present in the diffusion term only. That is, the process is equipped with an affine drift and a piecewise smooth diffusion term, e.g., a piecewise constant function. On the other hand, while we want to specify the process with a piecewise smooth diffusion term, the exact MLE of the TD process is the nonlinear functional form of the stochastic integrals. Under such intractable problem, how could we "approximate" the ML estimates by using the discrete data?

To solve the problems we concerned, we propose an approximate ML approach for estimating both the drift and the diffusion parameters of a TD process simultaneously at two stages. First, we assume the threshold parameter is known and differentiate the log-likelihood function, which is constructed by the irregular-spaced discrete data and the famous Cameron-Martin-Girsanov formula, with respect to parameters, and then set the derivatives equal to zero to obtain an iterative formula.

In practice, however, the threshold parameter, say r_0, is unknown. So, in the second stage, we obtain the MLE of r_0 by optimizing log-likelihood function via using the grid search on a prespecified data interval. Moreover, the simulation result shows that the asymptotic behavior of the MLE of the drift and the diffusion parameters of a TD process assuming the threshold parameter known is the same as the results assuming the threshold parameter is unknown. For simplicity, we mainly treat the TD process as a two-regime TD process in the whole thesis, and the results could be generalized to the multiple-regime TD process.

In addition to the theoretical discussion, we demonstrate some simulation results for our AMLE. In that section, we perform the Ornstein-Uhlenbeck process with two cases:only diffusion is threshold and both drift and diffusion are threshold; as a result, our AMLE converges to the true value as the observed interval getting larger. Besides, we also demonstrate the Cox-Ingersoll-Ross model with threshold in both drift and diffusion. The simulation results perform.

In the view of application , we use our AMLE and the two-regime TD model with/without square-root diffusion term to get a simulation study and to estimate the term structure of

doi:10.6342/NTU201701414

a long time series of US interest rates (three-month US treasury rate based on the Federal Reserve Bank's H15), which was used in Su and Chan(2015), and 10-Year Treasury Constant Maturity Rate-3-Month Treasury Bill: Secondary Market Rate . The conclusion will be given in the last.

doi:10.6342/NTU201701414

Chapter 2

Introduction to Quasi-Likelihood Estimation Method

2.1 Introduction to the Threshold Diffusion Process

For the purpose of making the results clearly, it is necessary to introduce the threshold diffusion process. We start from the general nonlinear diffusion process:

$$dX(t) \quad = \quad \mu(X(t), t)dt + \sigma(X(t), t)dW(t) \tag{2.1}$$

where the function $\mu(x, t)$ is called the drift term , the function $\sigma(x, t)$ is the diffusion term, and $W(t)$ is the standard Brownian process.

In this model, the drift term,$\mu(x, t)$, represents the instantaneous mean function, and the diffusion term, $\sigma(x, t)$, is the instantaneous variance function. Here, we put our attention on the case that the drift term and diffusion term are time-homogeneous, that is, $\mu(x, t) \equiv \mu(x)$, $\sigma(x, t) \equiv \sigma(x)$. Usually, those functions are equipped some parameters, and we write μ_β for μ and σ_ν for σ, where the drift parametes β and diffusion parameters ν are vectors that may have some common values.

Among all nonlinear diffusion processes, the first-order q-regime threshold diffusion(TD) process, which is the first-order continuous-time threshold autoregressive(CTAR(1)) time series(see [26]) with the threshold diffusion, has received much attention in the literature, and it is defined to be the solution of the following stochastic differential equation(SDE):

doi:10.6342/NTU201701414

$$dX(t) = \sum_{j=1}^{q}\{\boldsymbol{\beta}_j^{\top}\begin{pmatrix}1\\X(t)\end{pmatrix}dt + \sigma_j dW(t)\}I(r_{j-1} < X(t) \le r_j). \quad (2.2)$$

where $-\infty = r_0 < r_1 < \cdots < r_q = \infty$ are the threshold parameters, $\boldsymbol{\beta}_j^{\top} = (\beta_{j0}, \beta_{j1}), j = 1, \cdots, q$, are the autoregressive parameters, $\sigma_j, j = 1, \cdots, q$, are the diffusion parameters, and $W(t)$ is the standard Brownian process.

In this case, the drift term is a piecewise function while the diffusion term is a piecewise constant function, and the two functions have the same threshold points. Specifically, $\mu(x) = \sum_{j=1}^{q}(\beta_{j0} + \beta_{j1}x)I(r_{j-1} < x \le r_j)$, and $\sigma(x) = \sum_{j=1}^{q}\sigma_j I(r_{j-1} < x \le r_j)$. When the process is in the j^{th} regime, i.e. $X(t) \in (r_{j-1}, r_j]$, it is an Ornstein-Uhlenbeck(OU) process, which is widely used in the application of financial pricing and economic forecast(see [25], [14], [22]). Thus, the q-regime TD process above can be used to model the situation that a process is doubt to be governed by the different OU mechanism in each interval. Also, due to the properties of a Brownian motion, the TD process would switch regime to regime infinitely many times in an arbitrary small time interval.

Since we have the TD process now, it might be natural to ask when the stationary solution of a TD process will exist. The following theorem give us the necessary and sufficient conditions to conclude the existence of the stationary solution of a TD process, and it gives the form of the solution as well.(see [5], [4], [21])

Theorem 2.1.1. *Suppose* $\sigma_j > 0, j = 1, \cdots, q$. *Then the process defined in (2.2) has a stationary distribution if, and only if*

$$\lim_{x\to-\infty}\mu(x) > 0 \qquad and \qquad \lim_{x\to\infty}\mu(x) < 0 \quad (2.3)$$

i.e., $\beta_{11} < 0$ *and* $\beta_{q1} > 0$, *or in the case that* $\beta_{11} = 0$ $(\beta_{q1} = 0)$ *and* $\beta_{10} < 0$ $(\beta_{q0} > 0)$. *Further, if (2.3) is hold, the stationay density of (2.2) is given by*

$$\pi(x) = \sum_{j=1}^{q} k_j exp(\frac{2\beta_{j0}x + \beta_{j1}x^2}{\sigma_j^2})I(r_{j-1} < X(t) \le r_j) \quad (2.4)$$

where the constants k_j *are determined by the following conditions:*

doi:10.6342/NTU201701414

(i)

$$\int_{-\infty}^{\infty} \pi(x) = 1 \tag{2.5}$$

(ii)

$$\sigma_j^2 \pi(r_j -) = \sigma_{j+1}^2 \pi(r_j +), \qquad j = 1, \cdots, q \tag{2.6}$$

where $\pi(r_j-)$ and $\pi(r_j+)$ are the left and right limits of π at r_j, respectively.

That is, the function $\sigma^2(x)\pi(x)$ is continuous at all threshold points, and the stationary density function, $\pi(x)$, is continuous only if the variance function, $\sigma^2(x)$, is continuous at the threshold points.

$Proof$*: see reference [4] or [7] for a clear discussion.*

$Remark$ 2.1 It is obviously that the stationary density above is generally non-Gaussian, asymmetric, and multi-modal for a TD process. Also, the form of the stationary density implies that it has finite moment of all orders and is geometrically ergodic(see [18]).

To accommodate a more general model class, we can relax the piecewise constant diffusion term to a piecewise smooth function. For example, by using power transformation, i.e., replacing σ_j by $\sigma_j x^{\gamma_j}$ where γ_j are known parameters, we can still get the result as the same as $Theorem 2.1$ via using the famous Ito formula and rewriting the stationary conditions.(see [21], [23])

$Example$ 2.1 To summarize this section, we give an example to demonstrate the results above. The threshold diffusion process we consider is $dX_t = ((-2 - 5X_t)dt + 4dW_t)I(X_t \leq 0) + ((3 - 3X_t)dt + 8dW_t)I(X_t > 0)$ where W_t stands for the standard Brownian process.

From the (2.1.1), since the process satisfies (2.3), we have the stationary distribution of this process:

$\pi(x) = k_1 exp(\frac{-5(x+2/5)^2}{4^2})I(x \leq 0) + k_2 exp(\frac{-3(x-1)^2}{8^2})I(x > 0)$

where

$k_1 = (\sqrt{\frac{16\pi}{5}}\Phi(\frac{1}{\sqrt{10}}) + \sqrt{\frac{64\pi}{5}}(1 - \Phi(-\sqrt{\frac{32}{5}})\frac{exp(\frac{-1}{320})}{4}))^{-1}$ and $k_2 = (\frac{exp(\frac{-1}{320})}{4})k_1$

Noting that $\Phi(\cdot)$ is the cumulative density function of the standard normal distribution.

doi:10.6342/NTU201701414

2.2 Estimation Framework of Quasi-Likelihood

Since our work is based on the results in Su and Chan(2015) [23], it is important to introduce the QLE. We, here, mainly focus on discussing the estimation framework and the asymptotic results, which are the factors that inspire the idea for AMLE. We start with the following model:

$$dX(t) = \sum_{j=1}^{q}\{\boldsymbol{\beta}_j^\top \begin{pmatrix} 1 \\ X(t) \end{pmatrix}\} I(r_{j-1} < X(t) \leq r_j)dt + \sigma(X(t))dW(t). \quad (2.7)$$

where $-\infty = r_0 < r_1 < \cdots < r_q = \infty$ are the threshold parameters, $\boldsymbol{\beta}_j^\top = (\beta_{j0}, \beta_{j1}), j = 1, \cdots, q$, are the autoregressive parameters, $\sigma(X(t)$,is an unspecified diffusion function, and $W(t)$ is a standard Brownian process.

Let P_θ be the probability measure of a general diffusion process, $\{ X(t), 0 \leq t \leq T\}$, with drift term indexed by θ, say μ_θ and P be a probability measure for $W(t)$, Su and Chan use the celebrated Girsanov's formula for semimartingale to derive the log-likelihood function:

$$\begin{aligned} log(\Lambda) &= log(\frac{dP_\theta}{dP}) \\ &= \int_0^T \frac{\mu_\theta(X(t))}{\sigma(X(t))}dW(t) + \frac{1}{2}\int_{0)}^T \frac{\mu_\theta^2(X(t))}{\sigma^2(X(t))}dt \\ &= \int_0^T \frac{\mu_\theta^2(X(t))}{\sigma^2(X(t))}dX(t) - \frac{1}{2}\int_0^T \frac{\mu_\theta^2(X(t))}{\sigma^2(X(t))}dt \end{aligned} \quad (2.8)$$

In the case of a constant diffusion term, i.e. $\sigma(X(t)) = c,\ for\ some\ c > 0$, the log likelihood $log(\Lambda)$ is proportional to

$$l(\theta) = \int_0^T \mu_\theta(X(t))dW(t) + \frac{1}{2}\int_0^T \mu_\theta^2(X(t))dt \quad (2.9)$$

For simplicity, we assume q=2, ie. we only have two regimes, and hence, $\mu_\theta(X(t)) = (\beta_{10} + \beta_{11}X(t))I(X(t) \leq r_0) + (\beta_{20} + \beta_{21}X(t))I(X(t) > r_0)$. To obtain the estimates, first, they fix r_0, differentiate $l(\theta)$ with respect to $\delta = (\beta_{10}, \beta_{11}, \beta_{20}, \beta_{21})$, and set the derivatives equal to zero to solve the $\tilde{\delta_r}$. Second, they put $\tilde{\delta_r}$ into $l(\theta)$ and get the $l(\tilde{\delta_r}, r)$. Since $l(\tilde{\delta_r}, r)$ is non-differentiable w.r.t r, they use grid search on the prespecified interval,

doi:10.6342/NTU201701414

say $[a, b]$, to search for the $\tilde{r}$, which leads us to get the maximum , $l(\tilde{\theta})$.

To elucidate the description above, we organize the QL estimation procedure as follow:

Denote $\tilde{\theta} = (\tilde{\delta}_r, \tilde{r})$ as the quasi-likelihood estimator of $\theta = (\beta_{10}, \beta_{11}, \beta_{20}, \beta_{21}, \sigma_1, \sigma_2, r)$, $I_1(t; r) = I(X(t) \leq r)$, and $I_2(t; r) = I(X(t) > r)$

Step 1. Given the time series data $\{X_0, X_1, ..., X_q\}$, compute the order statistics $\{X_{(0)}, ..., X_{(q)}\}$, where $X_{(0)} \leq X_{(1)} \leq \cdots X_{(q)}$. Let $a = X_{(\lfloor q*p \rfloor)}$, and $b = X_{(\lfloor q*(1-p) \rfloor)}$, where $\lfloor y \rfloor$ denotes the largest integer less than or equal to y, and $p \in [0, 1/2)$. Let $\tilde{r}_i = a + i/100$, for $i = 0, 1, 2, ..., \lambda$, where $b = \tilde{r}_\lambda = a + \lambda/100$.

Step 2. Fix $\tilde{r}_i$, for $i = 0, ..., \lambda$.

(i)For $i = 0, ..., \lambda$, compute $\tilde{\delta}_{\tilde{r}_i} = \tilde{\delta}(\tilde{r}_i)$, where

$$\tilde{\delta}_r = \begin{pmatrix} \int_0^T \frac{I_1(t;r)}{T} dt & \int_0^T \frac{X(t)I_1(t;r))}{T} dt & 0 & 0 \\ \int_0^T \frac{X(t)I_1(t;r))}{T} dt & \int_0^T \frac{X^2(t)I_1(t;r))}{T} dt & 0 & 0 \\ 0 & 0 & \int_0^T \frac{I_2(t;r)}{T} dt & \int_0^T \frac{X(t)I_2(t;r))}{T} dt \\ 0 & 0 & \int_0^T \frac{I_2(t;r)}{T} dt & \int_0^T \frac{X(t)^2 I_2(t;r))}{T} dt \end{pmatrix}^{-1} \begin{pmatrix} \int_0^T \frac{I_1(t;r)}{T} dt \\ \int_0^T \frac{X(t)I_1(t;r))}{T} dt \\ \int_0^T \frac{I_2(t;r)}{T} dt \\ \int_0^T \frac{X(t)I_2(t;r))}{T} dt \end{pmatrix}$$

(ii)For $i = 0, ..., \lambda$, compute $\tilde{\theta}_i = (\tilde{\delta}(\tilde{r}_i), \tilde{r}_i)$.

Step 3. For $i = 0, ..., \lambda$, compute (i) by plugging $\tilde{\theta}_i$ into (2.9).

Suppose (τ) is the maximum among $\{(0), ..., (\lambda)\}$, then the quasi-likelihood estimator is $\tilde{\theta}_\tau$.

Remark 2.2 The estimation procedure is the same as the section 6 in Brockwell(2007)(see [1]), yet the original model between these two papers are different at the assumption on the diffusion term. The former allows that the diffusion function can be finitely discontinuous on the threshold parameters as in drift term with left and right limits(see later) whereas the later assumes the diffusion term is a positive constant. Although the estimation procedure seems complex, it still remains the same spirit as in the discrete threshold time series(see [4]).

2.3 Asymptotic theory of Quasi-Likelihood Estimation

Since we obtain the QL estimator, we will ask whether it possesses any good property for the (2.7) with q=2? The answer is yes under some regular conditions. We demonstrate the

doi:10.6342/NTU201701414

conditions and the main results which are proposed by Chan and Su (2015) as follow.

(A1). (i) $\boldsymbol{\beta}_{1,0} \neq \boldsymbol{\beta}_{2,0}$, where $\boldsymbol{\beta}_{i,0} = (\beta_{i1}, \beta_{i2}), i = 1, 2$; (ii) $(\boldsymbol{\beta}_{1,0} - \boldsymbol{\beta}_{2,0}) \cdot (1, r_0)^\top \neq 0$

(A2). The process $\{X(t), 0 \leq t \leq T\}$ in (2.7) is stationary, geometric ergodicity, and of 4^{th} finite moment. More precisely, we give the details.

(i)(stationarity) Stationarity of (2.7) holds if (2.3) holds in $Theorem 2.1$, and the invariant density π exists uniquely;

(ii)(Geometric Ergodicity) $\exists$ a π-integrable positive function $M(x)$ and a constant $\rho \in (0, 1)$ such that $\int_{-\infty}^{\infty} |P^{t-s}(x, y) - \pi(y)| dy < \rho^{t-s} M(x)$, where $P^{t-s}(x, y) = P(X(t) = y | X(s) = x)$

(A3). Assumptions on the $\sigma(X_t)$

(i) $\sigma(\cdot)$ is time-homogeneous and positive. i.e. $\sigma(X_t, t) = \sigma(X_t) > 0$.

(ii) $\sigma(x)$ is linear growth. i.e. $\exists\, c_0, c_1$ such that $\sigma(x) \leq c_0 + c_1 x, \forall x$

(iii) $\sigma(x)$ admits finitely many discontinuous points, say$\{ r_1, ..., r_q\}$, with max$\{\sigma(r_i-), \sigma(r_i+)\} < \infty, i = 1, ..., q$

(A4). Assumptions on the threshold parameter.

(i)The true threshold parameter r_0 lies in the prespecified interval $[a, b]$.

(ii)The marginal density and invariant density of the process $\{X(t), 0 \leq t \leq T\}$ are discontinuous only if $\sigma(X_t)$ is discontinuous at r_0.

Remark 2.3 (A1) preserves the model's identifiability so that the theory can work. (A2) restricts the model class as what we do in any stationary time series analysis. (A3) is also to preserve the stationarity of the process and gives us the idea of relaxing the likelihood theory for diffusion process as in Fegin(1976)(see [12]) as well. (A4) is a natural assumption to guarantee that we can find the threshold parameter. Without loss of generality, they assume r_0=0 and $-\infty < a < 0 < b < \infty$ to simplify the asymptotic analysis of QLE.

Now, under the above conditions, Su and Chan(2015) [23] establish the large sample properties for their QLE. Noting that $\theta_T = (\tilde{\delta}_T, \tilde{r}_T)$ is the QLE on the interval $[0, T]$, and the large sample properties are based on the T goes to infinity.

doi:10.6342/NTU201701414

Theorem 2.3.1. *Under (A1)-(A4), the quasi-likelihood estimator of the threshold parameter is T-consistent:*

$$\tilde{r} = r_0 + O_p(1/T).$$

Theorem 2.3.2. *Under (A1)-(A4), and let* $\tilde{l}_T(\kappa) = \tilde{l}(\tilde{\delta}(r_0+\kappa/T), r_0+\kappa/T) - \tilde{l}(\tilde{\delta}(r_0), r_0)$, $T(\tilde{r}_T - r_0)$ *has a weakly convergence, i.e.*

$$T(\tilde{r}_T - r_0) = \arg\max_{r\in[a,b]} \tilde{l}_T(r) \Rightarrow \tilde{r} = \arg\max_{r\in[a,b]} \{\tilde{l}_1(r)I(r>0) + \tilde{l}_2(r)I(r<0)\},$$

where

$$\begin{aligned}\tilde{l}_1(\kappa) &= \frac{-1}{2} f^2(r_0)\pi(r_0+) + f(r_0)\sqrt{\pi(r_0+)}\sigma(r_0+)W(\kappa),\\ \tilde{l}_2(\kappa) &= \frac{-1}{2} f^2(r_0)\pi(r_0-) + f(r_0)\sqrt{\pi(r_0-)}\sigma(r_0-)W(-\kappa),\end{aligned}$$

$f(r_0) = (\boldsymbol{\beta}_{1,0} - \boldsymbol{\beta}_{2,0}) \cdot (1, r_0)^\top$*, and W(κ) is a standard Brownian motion on* $\kappa \in \mathbb{R}$*. Moreover,* $\tilde{r}$ *has density*

$$\begin{aligned} g_{\tilde{r}}(s) = I(s<0)&\frac{f^2(r_0)\pi(r_0-)}{2}[\frac{1}{\sqrt{-ms}}\phi(\frac{-\sqrt{-ms}}{\sigma^2(r_0-)}) - \frac{1}{\sigma^2(r_0-)}\Phi(\frac{-\sqrt{-ms}}{\sigma^2(r_0-)})]\\ +I(s>0)&\frac{f^2(r_0)\pi(r_0+)}{2}[\frac{1}{\sqrt{ms}}\phi(\frac{-\sqrt{ms}}{\sigma^2(r_0+)}) - \frac{1}{\sigma^2(r_0+)}\Phi(\frac{-\sqrt{ms}}{\sigma^2(r_0+)})], s \neq 0\end{aligned}$$

where $m = \sigma^2(r_0+)f^2(r_0)\pi(r_0+) = \sigma^2(r_0-)f^2(r_0)\pi(r_0-) > 0$*, f(·) is as above,* $\pi(\cdot)$ *is stationary density of process* $\{X(t)|t \in [0,T]\}$*,* $\phi(\cdot)$ *is pdf of* $N(0,1)$*, and* $\Phi(\cdot)$ *is cdf of* $N(0,1)$.

Remark 2.4 In the case of discrete time threshold autoregressive(TAR) model, the limiting distribution of threshold parameter is related to the compound Poisson process(see [8], [19]), and Hansen(1997)(see [13]) shows that up to scale the proceeding density, $g_{\tilde{r}}(s)$, is the same as the limiting distribution of the threshold parameter in discrete time self-exciting threshold autoregressive (SETAR) when the autoregresive coefficients in two regimes are asymptotically equal.

Theorem 2.3.3. *Under (A1)-(A4), the quasi-likelihood estimator of coefficients,* $\tilde{\delta}_r$*, is*

$\sqrt{T}$*-consistent.*

$$\tilde{\delta}_r - \delta_0 = O_p(1/\sqrt{T}).$$

Moreover, $\sqrt{T}(\tilde{\delta}_r - \delta_0)$ *is asymptotically normally distributed with the same distribution as for the case of known threshold, i.e.,* $N(0, \Sigma)$*, where* $\Sigma = \text{plim}_{T\to\infty} \frac{\partial^2(\theta_0)}{\partial\delta\partial\delta^\top}^{-1} \frac{\partial(\theta_0)}{\partial\delta} \frac{\partial(\theta_0)}{\partial\delta}^\top \frac{\partial^2(\theta_0)}{\partial\delta\partial\delta^\top}^{-1}$.

doi:10.6342/NTU201701414

Chapter 3

Approximate Maximum Likelihood Estimation

3.1 The Idea of Approximate Maximal Likelihood Estimation

Since we have some basic knowledge of the TD process and QLE, a real data set contains only finite observations and we aim to estimate the drift term and diffusion term simultaneously. It still makes some difficulty to estimate the mle for parameters in a TD model while we can not collect the whole time path of a realization.

To solve this problem, we will show how to derive the joint density for the finite observations of a TD process(see [15]) and the we can use the joint density to illustrate our idea to obtain the approximate maximal likelihood estimation(AMLE).

Let $\boldsymbol{X}$=$\{X_0, X_1, \cdots, X_q\}$ be the observed data, which are observed at $\{0 = t_0 < t_1 < \cdots < t_q = T\}$, $\boldsymbol{\beta}$ be the drift parameters, and $\boldsymbol{\sigma}$ be the diffusion parameters. Under some regularity conditions, the Gameron-Martin-Girsanov formula (see, e.g., Theorem 3.5.1 of Karatzas and Shreve, 1991, [16], or ch10 in [17]) can be applied to show that the pdf of $\boldsymbol{X}$ with respect to the Lebesgue measure exists. First, note that

doi:10.6342/NTU201701414

$$
\begin{aligned}
& P_{\boldsymbol{\beta},\boldsymbol{\sigma}}(X_0 \le x_0, X_1 \le x_1, \cdots, X_q \le x_q) \\
= \; & E_{\boldsymbol{\beta},\boldsymbol{\sigma}}[I(X_0 \le x_0, X_1 \le x_1, \cdots, X_q \le x_q)] \\
= \; & E_{\mathbf{0},\boldsymbol{\sigma}}\left[I(X_0 \le x_0, X_1 \le x_1, \cdots, X_q \le x_q)\frac{dP_{\boldsymbol{\beta},\boldsymbol{\sigma}}}{dP_{0,\boldsymbol{\sigma}}}\right] \\
= \; & E_{\mathbf{0},\boldsymbol{\sigma}}\left[E_{\mathbf{0},\boldsymbol{\sigma}}\left[I(X_0 \le x_0, X_1 \le x_1, \cdots, X_q \le x_q)\frac{dP_{\boldsymbol{\beta},\boldsymbol{\sigma}}}{dP_{0,\boldsymbol{\sigma}}}\right]\middle| \boldsymbol{X}\right] \\
= \; & E_{\mathbf{0},\boldsymbol{\sigma}}\left[I(X_0 \le x_0, X_1 \le x_1, \cdots, X_q \le x_q)E_{\mathbf{0},\boldsymbol{\sigma}}\left[\frac{dP_{\boldsymbol{\beta},\boldsymbol{\sigma}}}{dP_{0,\boldsymbol{\sigma}}}\middle| \boldsymbol{X}\right]\right] \\
= \; & \int_{-\infty}^{x_q}\cdots\int_{-\infty}^{x_0} E_{0,\boldsymbol{\sigma}}\left[\frac{dP_{\boldsymbol{\beta},\boldsymbol{\sigma}}}{dP_{\mathbf{0},\boldsymbol{\sigma}}}\middle| \boldsymbol{x}\right] f_{\boldsymbol{x};(\mathbf{0},\boldsymbol{\sigma})}(\boldsymbol{x})dx_0\cdots dx_q,
\end{aligned}
$$

where $dP_{\mathbf{0},\boldsymbol{\sigma}}$ is the measure induced by the Brownian process $W = \{W_t\}$ and $dP_{\boldsymbol{\beta},\boldsymbol{\sigma}}$ is the measure induced by the original process $X = \{X_t, t \in [0, T]\ \}$. Thus,

$$
f_{\boldsymbol{x};(\boldsymbol{\beta},\boldsymbol{\sigma})}(\boldsymbol{x}) = E_{\mathbf{0},\boldsymbol{\sigma}}\left[\frac{dP_{\boldsymbol{\beta},\boldsymbol{\sigma}}}{dP_{\mathbf{0},\boldsymbol{\sigma}}}\middle| \boldsymbol{x}\right] f_{\boldsymbol{x};(\mathbf{0},\boldsymbol{\sigma})}(\boldsymbol{x}).
$$

Therefore, $l_{\boldsymbol{X}}(\boldsymbol{\beta},\boldsymbol{\sigma})$, the log-likelihood function of $\boldsymbol{X}$ is

$$
\begin{aligned}
& l_{\boldsymbol{x}}(\boldsymbol{\beta},\boldsymbol{\sigma}) \\
= \; & log\ E_{\mathbf{0},\boldsymbol{\sigma}}\left[\frac{dP_{\boldsymbol{\beta},\boldsymbol{\sigma}}}{dP_{\mathbf{0},\boldsymbol{\sigma}}}\middle| \boldsymbol{x}\right] + l_{\boldsymbol{x}}(\mathbf{0},\boldsymbol{\sigma}) \\
= \; & log\ E_{\mathbf{0},\boldsymbol{\sigma}}\left[\exp\left\{\int_0^T \frac{\mu_{\boldsymbol{\beta}}(X_t)}{\sigma^2_{\boldsymbol{\sigma}}(X_t)}dX_t \ -\frac{1}{2}\int_0^T \frac{\mu^2_{\boldsymbol{\beta}}(X_t)}{\sigma^2_{\boldsymbol{\sigma}}(X_t)}dt\right\}\middle| \boldsymbol{x}\right] + l_{\boldsymbol{x}}(\mathbf{0},\boldsymbol{\sigma}) \quad (3.1)
\end{aligned}
$$

3.2 Estimation Procedure for AMLE

The threshold diffusion process we consider is

$$
\begin{aligned}
dX_t \;=\; & \{(\beta_{10} + \beta_{11}X_t)I(X_t \le t) + (\beta_{20} + \beta_{21}X_t)I(X_t > t)\}\,dt \\
& + \{\sigma_1 I(X_t \le t) + \sigma_2 I(X_t > r)\}X_t^{\gamma}dW_t, \qquad (3.2)
\end{aligned}
$$

where $W = \{W_t\}$ stands for the standard Brownian process.

doi:10.6342/NTU201701414

Let $\{X_0, X_1, \cdots, X_q\}$ be the observed data at observing times $\{t_0, \cdots, t_q\}$. Let $\Delta_j = t_j - t_{j-1}$, then Model (2.2) can be approximated by

$$\begin{aligned} X_j - X_{j-1} &= \Delta_j \left\{(\beta_{10} + \beta_{11}X_{j-1})I(X_{j-1} \le r) + (\beta_{20} + \beta_{21}X_{j-1})I(X_{j-1} > r)\right\} \\ &\quad + \{\sigma_1 I(X_{j-1} \le r) + \sigma_2 I(X_{j-1} > r)\}X_{j-1}^{\gamma}(W_j - W_{j-1}), \end{aligned}$$

where $W_j - W_{j-1} \sim \mathrm{N}(0, \Delta_j)$. Therefore, the log-likelihood function of $\{X_0, X_1, \cdots, X_q\}$ is

$$\begin{aligned} -2l &= C + \sum_{j=1}^{q} \log\{\sigma_1^2 I(X_{j-1} \le r) + \sigma_2^2 I(X_{j-1} > r)\} \\ &\quad + \sum_{j=1}^{q} \frac{\left[X_j - X_{j-1} - \Delta_j \left\{(\beta_{10} + \beta_{11}X_{j-1})I(X_{j-1} \le r) + (\beta_{20} + \beta_{21}X_{j-1})I(X_{j-1} > r)\right\}\right]^2}{\{\sigma_1^2 I(X_{j-1} \le r) + \sigma_2^2 I(X_{j-1} > r)\}\Delta_j X_{j-1}^{2\gamma}}. \end{aligned} \tag{3.3}$$

Note that (3.1) is equivalent to (3.3) if we approximate these two integrations of (3.1) by Euler's method.

Differentiating (3.3) with respect to β_{10}, β_{11}, β_{20}, β_{21}, σ_1^2, and σ_2^2 give

$$-2\frac{\partial l}{\partial \beta_{10}} = -2\sum_{j=1}^{q} \frac{\{X_j - X_{j-1} - \Delta_j(\beta_{10} + \beta_{11}X_{j-1})\}\, I(X_{j-1} \le r)}{\{\sigma_1^2 I(X_{j-1} \le r) + \sigma_2^2 I(X_{j-1} > r)\}\Delta_j X_{j-1}^{2\gamma}}, \tag{3.4}$$

$$-2\frac{\partial l}{\partial \beta_{11}} = -2\sum_{j=1}^{q} \frac{\{X_j - X_{j-1} - \Delta_j(\beta_{10} + \beta_{11}X_{j-1})\}\, X_{j-1} I(X_{j-1} \le r)}{\{\sigma_1^2 I(X_{j-1} \le r) + \sigma_2^2 I(X_{j-1} > r)\}\Delta_j X_{j-1}^{2\gamma}}, \tag{3.5}$$

$$-2\frac{\partial l}{\partial \beta_{20}} = -2\sum_{j=1}^{q} \frac{\{X_j - X_{j-1} - \Delta_j(\beta_{20} + \beta_{21}X_{j-1})\}\, I(X_{j-1} > r)}{\{\sigma_1^2 I(X_{j-1} \le r) + \sigma_2^2 I(X_{j-1} > r)\}\Delta_j X_{j-1}^{2\gamma}}, \tag{3.6}$$

$$-2\frac{\partial l}{\partial \beta_{21}} = -2\sum_{j=1}^{q} \frac{\{X_j - X_{j-1} - \Delta_j(\beta_{20} + \beta_{21}X_{j-1})\}\, X_{j-1} I(X_{j-1} > r)}{\{\sigma_1^2 I(X_{j-1} \le r) + \sigma_2^2 I(X_{j-1} > r)\}\Delta_j X_{j-1}^{2\gamma}}, \tag{3.7}$$

$$\begin{aligned} -2\frac{\partial l}{\partial \sigma_1^2} &= \sum_{j=1}^{q} \frac{I(X_{j-1} \le r)}{\sigma_1^2 I(X_{j-1} \le r) + \sigma_2^2 I(X_{j-1} > r)} \\ &\quad - \sum_{j=1}^{q} \frac{\{X_j - X_{j-1} - \Delta_j(\beta_{10} + \beta_{11}X_{j-1})\}^2\, I(X_{j-1} \le r)}{\{\sigma_1^4 I(X_{j-1} \le r) + \sigma_2^4 I(X_{j-1} > r)\}\Delta_j X_{j-1}^{2\gamma}}, \end{aligned} \tag{3.8}$$

doi:10.6342/NTU201701414

and

$$-2\frac{\partial l}{\partial \sigma_2^2} = \sum_{j=1}^{q} \frac{I(X_{j-1} > r)}{\{\sigma_1^2 I(X_{j-1} \le r) + \sigma_2^2 I(X_{j-1} > r)\}} - \sum_{j=1}^{q} \frac{\{X_j - X_{j-1} - \Delta_j(\beta_{20} + \beta_{21}X_{j-1})\}^2 I(X_{j-1} > r)}{\{\sigma_1^4 I(X_{j-1} \le r) + \sigma_2^4 I(X_{j-1} > r)\}\Delta_j X_{j-1}^{2\gamma}}. \quad (3.9)$$

Equating (3.4) and (3.5) to zero gives

$$\begin{bmatrix} \hat\beta_{10} \\ \hat\beta_{11} \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^q \frac{I(X_{j-1}\le r)}{\{\sigma_1^2 I(X_{j-1}\le r)+\sigma_2^2 I(X_{j-1}>r)\}X_{j-1}^{2\gamma}} & \sum_{j=1}^q \frac{X_{j-1}I(X_{j-1}\le r)}{\{\sigma_1^2 I(X_{j-1}\le r)+\sigma_2^2 I(X_{j-1}>r)\}X_{j-1}^{2\gamma}} \\ \sum_{j=1}^q \frac{X_{j-1}I(X_{j-1}\le r)}{\{\sigma_1^2 I(X_{j-1}\le r)+\sigma_2^2 I(X_{j-1}>r)\}X_{j-1}^{2\gamma}} & \sum_{j=1}^q \frac{X_{j-1}^2 I(X_{j-1}\le r)}{\{\sigma_1^2 I(X_{j-1}\le r)+\sigma_2^2 I(X_{j-1}>r)\}X_{j-1}^{2\gamma}} \end{bmatrix}^{-1} \begin{bmatrix} \sum_{j=1}^q \frac{(X_j - X_{j-1})I(X_{j-1}\le r)}{\{\sigma_1^2 I(X_{j-1}\le r)+\sigma_2^2 I(X_{j-1}>r)\}\Delta_j X_{j-1}^{2\gamma}} \\ \sum_{j=1}^q \frac{(X_j - X_{j-1})X_{j-1}I(X_{j-1}\le r)}{\{\sigma_1^2 I(X_{j-1}\le r)+\sigma_2^2 I(X_{j-1}>r)\}\Delta_j X_{j-1}^{2\gamma}} \end{bmatrix}. \quad (3.10)$$

Equating (3.6) and (3.7) to zero gives

$$\begin{bmatrix} \hat\beta_{20} \\ \hat\beta_{21} \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^q \frac{I(X_{j-1}>r)}{\{\sigma_1^2 I(X_{j-1}\le r)+\sigma_2^2 I(X_{j-1}>r)\}X_{j-1}^{2\gamma}} & \sum_{j=1}^q \frac{X_{j-1}I(X_{j-1}>r)}{\{\sigma_1^2 I(X_{j-1}\le r)+\sigma_2^2 I(X_{j-1}>r)\}X_{j-1}^{2\gamma}} \\ \sum_{j=1}^q \frac{X_{j-1}I(X_{j-1}>r)}{\{\sigma_1^2 I(X_{j-1}\le r)+\sigma_2^2 I(X_{j-1}>r)\}X_{j-1}^{2\gamma}} & \sum_{j=1}^q \frac{X_{j-1}^2 I(X_{j-1}>r)}{\{\sigma_1^2 I(X_{j-1}\le r)+\sigma_2^2 I(X_{j-1}>r)\}X_{j-1}^{2\gamma}} \end{bmatrix}^{-1} \begin{bmatrix} \sum_{j=1}^q \frac{(X_j - X_{j-1})I(X_{j-1}>r)}{\{\sigma_1^2 I(X_{j-1}\le r)+\sigma_2^2 I(X_{j-1}>r)\}\Delta_j X_{j-1}^{2\gamma}} \\ \sum_{j=1}^q \frac{(X_j - X_{j-1})X_{j-1}I(X_{j-1}>r)}{\{\sigma_1^2 I(X_{j-1}\le r)+\sigma_2^2 I(X_{j-1}>r)\}\Delta_j X_{j-1}^{2\gamma}} \end{bmatrix}. \quad (3.11)$$

Similarly, equating (3.8) and (3.9) to zero gives

$$\hat\sigma_1^2 = \frac{\sum_{j=1}^q \frac{\{X_j - X_{j-1} - \Delta_j(\beta_{10}+\beta_{11}X_{j-1})\}^2 I(X_{j-1}\le r)}{\{\sigma_1^4 I(X_{j-1}\le r)+\sigma_2^4 I(X_{j-1}>r)\}\Delta_j X_{j-1}^{2\gamma}}}{\sum_{j=1}^q \frac{I(X_{j-1}\le r)}{\sigma_1^4 I(X_{j-1}\le r)+\sigma_1^2\sigma_2^2 I(X_{j-1}>r)}}, \quad (3.12)$$

and

$$\hat\sigma_2^2 = \frac{\sum_{j=1}^q \frac{\{X_j - X_{j-1} - \Delta_j(\beta_{20}+\beta_{21}X_{j-1})\}^2 I(X_{j-1}>r)}{\{\sigma_1^4 I(X_{j-1}\le r)+\sigma_2^4 I(X_{j-1}>r)\}\Delta_j X_{j-1}^{2\gamma}}}{\sum_{j=1}^q \frac{I(X_{j-1}>r)}{\sigma_1^2\sigma_2^2 I(X_{j-1}\le r)+\sigma_2^4 I(X_{j-1}>r)}}. \quad (3.13)$$

So, we propose the following procedure to compute the approximate maximum likelihood estimator of $\theta = (\beta_{10}, \beta_{11}, \beta_{20}, \beta_{21}, \sigma_1, \sigma_2)$.

doi:10.6342/NTU201701414

Step 1. Given the time series data $\{X_0, X_1, ..., X_q\}$, compute the order statistics $\{X_{(0)}, ..., X_{(q)}\}$, where $X_{(0)} \le X_{(1)} \le \cdots X_{(q)}$. Let $a = X_{(\lfloor q/5 \rfloor)}$, and $b = X_{(\lfloor 4q/5 \rfloor)}$, where $\lfloor y \rfloor$ deontes the largest integer less than or equal to y. Let $\hat{r}_i = a + i/100$, for $i = 0, 1, 2, ..., \lambda$, where $b = \hat{r}_\lambda = a + \lambda/100$.

Step 2. Fix $\hat{r}_i$, for $i = 0, ..., \lambda$.

(i) Compute $\hat{\sigma}_1^2(0)$ and $\hat{\sigma}_2^2(0)$, the initial estimate of σ_1^2 and σ_2^2, by Equation (8) of Su and Chan (2015). More specifically,

$$\hat{\sigma}_1^2(0) = \frac{1}{T}\sum_{j=1}^{q}\{X_i I(X_i \le r) - X_{i-1} I(X_{i-1} \le r)\}^2,$$
$$\hat{\sigma}_2^2(0) = \frac{1}{T}\sum_{j=1}^{q}\{X_i I(X_i > r) - X_{i-1} I(X_{i-1} > r)\}^2.$$

(ii) Given the estimate $\hat{\theta}(k-1) = (\hat{\beta}_{10}(k-1), \hat{\beta}_{11}(k-1), \hat{\beta}_{20}(k-1), \hat{\beta}_{21}(k-1), \hat{\sigma}_1^2(k-1), \hat{\sigma}_2^2(k-1))$, compute $\hat{\theta}(k)$ as follows:

$$\begin{bmatrix} \hat{\beta}_{10}(k) \\ \hat{\beta}_{11}(k) \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^{q} \frac{I(X_{j-1} \le \hat{r}_i)}{\{\hat{\sigma}_1^2(k-1)I(X_{j-1}\le\hat{r}_i)+\hat{\sigma}_2^2(k-1)I(X_{j-1}>\hat{r}_i)\}X_{j-1}^{2\gamma}} & \sum_{j=1}^{q} \frac{X_{j-1}I(X_{j-1} \le \hat{r}_i)}{\{\hat{\sigma}_1^2(k-1)I(X_{j-1}\le\hat{r}_i)+\hat{\sigma}_2^2(k-1)I(X_{j-1}>\hat{r}_i)\}X_{j-1}^{2\gamma}} \\ \sum_{j=1}^{q} \frac{X_{j-1}I(X_{j-1} \le \hat{r}_i))}{\{\hat{\sigma}_1^2(k-1)I(X_{j-1}\le\hat{r}_i)+\hat{\sigma}_2^2(k-1)I(X_{j-1}>\hat{r}_i)\}X_{j-1}^{2\gamma}} & \sum_{j=1}^{q} \frac{\Delta_j X_{j-1}^2 I(X_{j-1} \le \hat{r}_i)}{\{\hat{\sigma}_1^2(k-1)I(X_{j-1}\le\hat{r}_i)+\hat{\sigma}_2^2(k-1)I(X_{j-1}>\hat{r}_i)\}X_{j-1}^{2\gamma}} \end{bmatrix}^{-1} \begin{bmatrix} \sum_{j=1}^{q} \frac{(X_j - X_{j-1})I(X_{j-1} \le \hat{r}_i)}{\{\hat{\sigma}_1^2(k-1)I(X_{j-1}\le\hat{r}_i)+\hat{\sigma}_2^2(k-1)I(X_{j-1}>\hat{r}_i)\}X_{j-1}^{2\gamma}} \\ \sum_{j=1}^{q} \frac{(X_j - X_{j-1})X_{j-1}I(X_{j-1} \le \hat{r}_i)}{\{\hat{\sigma}_1^2(k-1)I(X_{j-1}\le\hat{r}_i)+\hat{\sigma}_2^2(k-1)I(X_{j-1}>\hat{r}_i)\}X_{j-1}^{2\gamma}} \end{bmatrix}, \quad (3.14)$$

$$\begin{bmatrix} \hat{\beta}_{20}(k) \\ \hat{\beta}_{21}(k) \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^{q} \frac{I(X_{j-1} > \hat{r}_i)}{\{\hat{\sigma}_1^2(k-1)I(X_{j-1}\le\hat{r}_i)+\hat{\sigma}_2^2(k-1)I(X_{j-1}>\hat{r}_i)\}X_{j-1}^{2\gamma}} & \sum_{j=1}^{q} \frac{X_{j-1}I(X_{j-1} > \hat{r}_i)}{\{\hat{\sigma}_1^2(k-1)I(X_{j-1}\le\hat{r}_i)+\hat{\sigma}_2^2(k-1)I(X_{j-1}>\hat{r}_i)\}X_{j-1}^{2\gamma}} \\ \sum_{j=1}^{q} \frac{X_{j-1}I(X_{j-1} < \hat{r}_i))}{\{\hat{\sigma}_1^2(k-1)I(X_{j-1}\le\hat{r}_i)+\hat{\sigma}_2^2(k-1)I(X_{j-1}>\hat{r}_i)\}X_{j-1}^{2\gamma}} & \sum_{j=1}^{q} \frac{X_{j-1}^2 I(X_{j-1} > \hat{r}_i)}{\{\hat{\sigma}_1^2(k-1)I(X_{j-1}\le\hat{r}_i)+\hat{\sigma}_2^2(k-1)I(X_{j-1}>\hat{r}_i)\}X_{j-1}^{2\gamma}} \end{bmatrix}^{-1} \begin{bmatrix} \sum_{j=1}^{q} \frac{(X_j - X_{j-1})I(X_{j-1} > \hat{r}_i)}{\{\hat{\sigma}_1^2(k-1)I(X_{j-1}\le\hat{r}_i)+\hat{\sigma}_2^2(k-1)I(X_{j-1}>\hat{r}_i)\}X_{j-1}^{2\gamma}} \\ \sum_{j=1}^{q} \frac{(X_j - X_{j-1})X_{j-1}I(X_{j-1} > \hat{r}_i)}{\{\hat{\sigma}_1^2(k-1)I(X_{j-1}\le\hat{r}_i)+\hat{\sigma}_2^2(k-1)I(X_{j-1}>\hat{r}_i)\}X_{j-1}^{2\gamma}} \end{bmatrix}, \quad (3.15)$$

and

$$\hat{\sigma}_1^2(k) = \frac{\sum_{j=1}^{q} \frac{\left\{X_j - X_{j-1} - \Delta_j(\hat{\beta}_{10}(k-1)+\hat{\beta}_{11}(k-1)X_{j-1})\right\}^2 I(X_{j-1}\le\hat{r}_i))}{\{\hat{\sigma}_1^4 I(X_{j-1}\le\hat{r}_i)+\hat{\sigma}_2^4(k-1)I(X_{j-1}>\hat{r}_i)\}\Delta_j X_{j-1}^{2\gamma}}}{\sum_{j=1}^{q} \frac{I(X_{j-1}\le\hat{r}_i))}{\hat{\sigma}_1^4(k-1)I(X_{j-1}\le\hat{r}_i))+\hat{\sigma}_1^2(k-1)\hat{\sigma}_2^2(k-1)I(X_{j-1}>\hat{r}_i))}}, \quad (3.16)$$

doi:10.6342/NTU201701414

$$\hat{\sigma}_2^2(k) = \frac{\sum_{j=1}^{q} \frac{\left\{X_j - X_{j-1} - \Delta_j(\hat{\beta}_{20}(k-1) + \hat{\beta}_{21}(k-1)X_{j-1})\right\}^2 I(X_{j-1} > \hat{r}_i))}{\{\hat{\sigma}_1^4(k-1)I(X_{j-1} \le \hat{r}_i) + \hat{\sigma}_2^4(k-1)I(X_{j-1} > \hat{r}_i)\}\Delta_j X_{j-1}^{2\gamma}}}{\sum_{j=1}^{q} \frac{I(X_{j-1} > \hat{r}_i))}{\hat{\sigma_1}^2(k-1)\hat{\sigma}_2^2(k-1)I(X_{j-1} \le \hat{r}_i)) + \hat{\sigma}_2^4(k-1)I(X_{j-1} > \hat{r}_i))}}. \quad (3.17)$$

(iii) Repeat (ii) until $\hat{\theta}(k)$ converge.

Let the converged estimator be $\hat{\theta}_i = (\hat{\beta}_{10,i}, \hat{\beta}_{11,i}, \hat{\beta}_{20.i}, \hat{\beta}_{21,i}, \hat{\sigma}_{1,i}, \hat{\sigma}_{2,i}, \hat{r}_i)$.

Step 3. For $i = 0, ..., \lambda$, compute $-2(i)$ by plugging $\hat{\theta}_i$ into (3.3).

Suppose $-2(\tau)$ is the minimum among $\{-2(0), ..., 2(\lambda)\}$, then the approximate maximum likelihood estimator is $\hat{\theta}_\tau$.

3.3 Some Large Sample Conjectures about AMLE

The approximate maximum likelihood estimator is obtained by a two-step procedure. First, for given r, obtain the likelihood estimator for $\theta = (\beta_{10}, \beta_{11}, \beta_{20}, \beta_{21}, \sigma_1, \sigma_2)$, and denote this estimator by $\hat{\theta}_r$. This can be done by minimizing (3.3) with respect to θ. Second, we perform a grid search over the region of $r \in [a, b]$, where $[a, b]$ are often chosen to be some percentiles of the observed data in order to guarantee data abundance for estimation in each regime. Let $\hat{r}$ denote the threshold estimator which maximizes $(\hat{\delta}_k, k)$. More specifically, $\hat{\delta}_r = (\hat{\beta}_{10,r}, \hat{\beta}_{11,r}, \hat{\beta}_{20,r}, \hat{\beta}_{21,r}, \hat{\sigma}_{1,r}, \hat{\sigma}_{2,r})$, and $\hat{r} = \arg\max_{k \in [a,b]}(\hat{\delta}_k, k)$. Let $\hat{\theta}_T = (\hat{\beta}_{10}, \hat{\beta}_{11}, \hat{\beta}_{20}, \hat{\beta}_{21}, \hat{\sigma}_1, \hat{\sigma}_2, \hat{r})$ be the approximate maximum likelihood estimator of θ, and $\theta_0 = (\beta_{10,0}, \beta_{11,0}, \beta_{20,0}, \beta_{21,0}, \sigma_{1,0}, \sigma_{2,0}, r_0)$ be the true parameter of θ.

Since we have the AMLE, one will ask whether the AMLE possesses any good statistical properties? According to our model formulation, which satisfies the regular conditions (A1)-(A4) as in QLE asymptotic theory, we aim at establishing the following conjectures, which are still in progress.

Conjecture 3.3.1. *Under the regular conditions, the approximate maximum likelihood estimator of the threshold parameter is T-consistent:*

$$\hat{r} = r_0 + O_p(1/T).$$

Conjecture 3.3.2. *Under the regular conditions, the approximate maximum likelihood*

doi:10.6342/NTU201701414

estimator of the threshold parameter converges weakly to the random variable , r_∞,:

$$T(\hat{r} - r_0) \Rightarrow r_\infty.$$

Conjecture 3.3.3. *Under the regular conditions, we have：*

$$\hat{\delta}_r - \delta_0 = O_p(1/\sqrt{T}).$$

Conjecture 3.3.4. *Under the regular conditions, $\sqrt{T}(\hat{\delta}_{\hat{r}} - \delta_0)$ and $T(\hat{r} - r_0)$ are asymptotically independent, and $\sqrt{T}(\hat{\delta}_{\hat{r}} - \delta_0)$ is asymptotically normally distributed with the same distribution as for the case of known threshold, i.e.,*

$$\sqrt{T}(\hat{\delta}_{\hat{r}} - \delta_0) \xrightarrow{\mathcal{D}} N(0, \Sigma(\theta_0))$$

where

$$\Sigma(\theta_0) = \operatorname*{plim}_{T\to\infty} - \frac{\partial^2{}_T(\theta_0)}{\partial\delta\partial\delta^\top}^{-1}$$

$$= \operatorname*{plim}_{T\to\infty} \begin{bmatrix} \int_0^T \frac{X_t^{-\gamma} I(X_t \le r_0)}{T\sigma_1^2} dt & \int_0^T \frac{X_t^{1-\gamma} I(X_t \le r_0)}{T\sigma_1^2} dt & 0 & 0 & A & F \\ \int_0^T \frac{X_t^{1-\gamma} I(X_t \le r_0)}{T\sigma_1^2} dt & \int_0^T \frac{X_t^{2-\gamma} I(X_t \le r_0)}{T\sigma_1^2} dt & 0 & 0 & B & I \\ 0 & 0 & \int_0^T \frac{X_t^{-\gamma} I(X_t > r_0)}{T\sigma_2^2} dt & \int_0^T \frac{X_t^{1-\gamma} I(X_t > r_0)}{T\sigma_2^2} dt & C & J \\ 0 & 0 & \int_0^T \frac{X_t^{1-\gamma} I(X_t > r_0)}{T\sigma_2^2} dt & \int_0^T \frac{X_t^{2-\gamma} I(X_t > r_0)}{T\sigma_2^2} dt & D & K \\ A & B & C & D & E & 0 \\ F & I & J & K & 0 & L \end{bmatrix}$$

doi:10.6342/NTU201701414

$$A = \frac{1}{T\sigma_1^4}(\int_0^T (\boldsymbol{\beta}_1^\top \begin{pmatrix} 1 \\ X_t \end{pmatrix})X_t^{-\gamma}I(X_t \leq r_0)dt - \int_0^T X_t^{-\gamma}I(X_t \leq r_0)dX_t)$$

$$B = \frac{1}{T\sigma_1^4}(\int_0^T (\boldsymbol{\beta}_1^\top \begin{pmatrix} 1 \\ X_t \end{pmatrix})X_t^{1-\gamma}I(X_t \leq r_0)dt - \int_0^T X_t^{1-\gamma}I(X_t \leq r_0)dX_t)$$

$$C = \frac{1}{T\sigma_1^4}(\int_0^T (\boldsymbol{\beta}_1^\top \begin{pmatrix} 1 \\ X_t \end{pmatrix})X_t^{-\gamma}I(X_t > r_0)dt - \int_0^T X_t^{-\gamma}I(X_t > r_0)dX_t)$$

$$D = \frac{1}{T\sigma_1^4}(\int_0^T (\boldsymbol{\beta}_1^\top \begin{pmatrix} 1 \\ X_t \end{pmatrix})X_t^{1-\gamma}I(X_t > r_0)dt - \int_0^T X_t^{1-\gamma}I(X_t > r_0)dX_t)$$

$$E = \frac{1}{T\sigma_1^6}(2\int_0^T (\boldsymbol{\beta}_1^\top \begin{pmatrix} 1 \\ X_t \end{pmatrix})X_t^{-\gamma}I(X_t \leq r_0)dX_t - \int_0^T (\boldsymbol{\beta}_1^\top \begin{pmatrix} 1 \\ X_t \end{pmatrix})^2X_t^{-\gamma}I(X_t \leq r_0)dt)$$

$$F = \frac{1}{T\sigma_2^4}(\int_0^T (\boldsymbol{\beta}_2^\top \begin{pmatrix} 1 \\ X_t \end{pmatrix})X_t^{-\gamma}I(X_t \leq r_0)dt - \int_0^T X_t^{-\gamma}I(X_t \leq r_0)dX_t)$$

$$I = \frac{1}{T\sigma_2^4}(\int_0^T (\boldsymbol{\beta}_2^\top \begin{pmatrix} 1 \\ X_t \end{pmatrix})X_t^{1-\gamma}I(X_t \leq r_0)dt - \int_0^T X_t^{1-\gamma}I(X_t \leq r_0)dX_t)$$

$$J = \frac{1}{T\sigma_2^4}(\int_0^T (\boldsymbol{\beta}_2^\top \begin{pmatrix} 1 \\ X_t \end{pmatrix})X_t^{-\gamma}I(X_t > r_0)dt - \int_0^T X_t^{-\gamma}I(X_t > r_0)dX_t)$$

$$K = \frac{1}{T\sigma_2^4}(\int_0^T (\boldsymbol{\beta}_2^\top \begin{pmatrix} 1 \\ X_t \end{pmatrix})X_t^{1-\gamma}I(X_t > r_0)dt - \int_0^T X_t^{1-\gamma}I(X_t > r_0)dX_t)$$

$$L = \frac{1}{T\sigma_2^6}(2\int_0^T (\boldsymbol{\beta}_2^\top \begin{pmatrix} 1 \\ X_t \end{pmatrix})X_t^{-\gamma}I(X_t > r_0)dX_t - \int_0^T (\boldsymbol{\beta}_2^\top \begin{pmatrix} 1 \\ X_t \end{pmatrix})^2X_t^{-\gamma}I(X_t > r_0)dt).$$

Remark 3.1 Since our method is based on the ML approach, it has no too much surprise to have such results. The most different part is the convergence rate for the threshold parameter due to violation of the regularity of MLE theory in continuous-time stochastic process([12]). However, there are no standard ML theory in threshold diffusion process in the literature. Furthermore, the asymptotic results we guess in AMLE are based on the results in QLE but the concrete proofs are in progress. Last, the probability limits above will exist by the lemma 1 in [23] and the (A2) in QLE.

doi:10.6342/NTU201701414

Chapter 4

Simulation

We now report some simulation results about the finite sample performance of the maximum likelihood estimator. We have experimented with both regularly and irregularly spaced data from

$$
\begin{aligned}
dX_t &= \{(\beta_{10} + \beta_{11}X_t)I(X_t \le r) + (\beta_{20} + \beta_{21}X_t)I(X_t > r)\}\, dt \\
&\quad + \{\sigma_1 I(X_t \le r) + \sigma_2 I(X_t > r)\} X_t^{\gamma} dW_t, \qquad (4.1)
\end{aligned}
$$

where $W = \{W_t\}$ stands for the standard Brownian process.

(A)Regularly spaced data were sampled as follows.
First, we specify the time interval unit and split the unit into m pieces. That is, let [0,1] be the time interval unit and $dt \approx \Delta t = \frac{1}{m}$, we get $\{0, \frac{1}{m}, \cdots, \frac{m-1}{m}, 1\}$. Since the time interval is $[0, T]$, we get $\{t_k = \frac{k*T}{m} : k = 0, \cdots, m*T\}$. Second, we simulated the regularly spaced time series data from Model (4.5) via using Euler approximation.
(B)Irregularly spaced data were sampled as follows.
First, we simulated s_i, $i = 1, ..., q$, independently from the exponential distribution with mean equal to 0.005, and recursively set the sampling epochs with $t_0 = 0$, and $t_i = t_{i-1} + s_i + 0.5$, $i = 1, ..., q$; hence the sampling intervals are independent, identically distributed and of unit mean. Second, we simulated the irregularly spaced time series data from Model (4.5) via using Euler approximation.

doi:10.6342/NTU201701414

4.1 Ornstein-Uhlenbeck process with Both the Drift and the Diffusion Are Nonlinear

Model:

$$\begin{aligned} dX_t &= \{(\beta_{10}+\beta_{11}X_t)I(X_t \le r) + (\beta_{20}+\beta_{21}X_t)I(X_t > r)\}\,dt \\ &\quad + \{\sigma_1 I(X_t \le r) + \sigma_2 I(X_t > r)\}dW_t, \end{aligned} \tag{4.2}$$

where $W = \{W_t\}$ stands for the standard Brownian process.

Table 4.1: Averages (standard deviations) of 1,000 simulations of the approximate maximum likelihood estimators of the parameters r, β_{10}, β_{11}, β_{20}, β_{21}, σ_1, and σ_2. (Ornstein - Uhlenbeck process with Both the Drift and the Diffusion Are Nonlinear)

	parameter	true value \ T	10	100	200
	r	0.75	0.8227(0.1563)	0.7529(0.0134)	0.7512(0.0061)
	β_{10}	1	1.8348(1.8773)	1.0231(0.2259)	1.0138(0.1504)
regularly	β_{11}	-2	-2.8387(2.3957)	-2.0059(0.4930)	-2.0121(0.3116)
-spaced	β_{20}	1.5	2.5165(2.9872)	1.5348(0.2637)	1.5159(0.1904)
	β_{21}	-1.5	-2.4749(2.5171)	-1.5382(0.2548)	-1.5183(0.1836)
	σ_1	0.4	0.3740(0.0415)	0.3990(0.0062)	0.3996(0.0036)
	σ_2	0.3	0.3013(0.0189)	0.3000(0.0026)	0.3000(0.0181)
	r	0.75	0.8143(0.1570)	0.7525(0.0122)	0.7509(0.0043)
	β_{10}	1	1.9375(2.1259)	1.0327(0.2443)	1.0184(0.1626)
irregularly	β_{11}	-2	-2.9336(2.7337)	-2.0330(0.5229)	-2.0195(0.3292)
-spaced	β_{20}	1.5	2.4076(2.8209)	1.5343(0.2944)	1.5136(0.1968)
	β_{21}	-1.5	-2.4077(2.4245)	-1.5385(0.2824)	-1.5169(0.1904)
	σ_1	0.4	0.3748(0.0411)	0.3993(0.0060)	0.3997(0.0035)
	σ_2	0.3	0.3022(0.0249)	0.3000(0.0026)	0.3001(0.0018)

doi:10.6342/NTU201701414

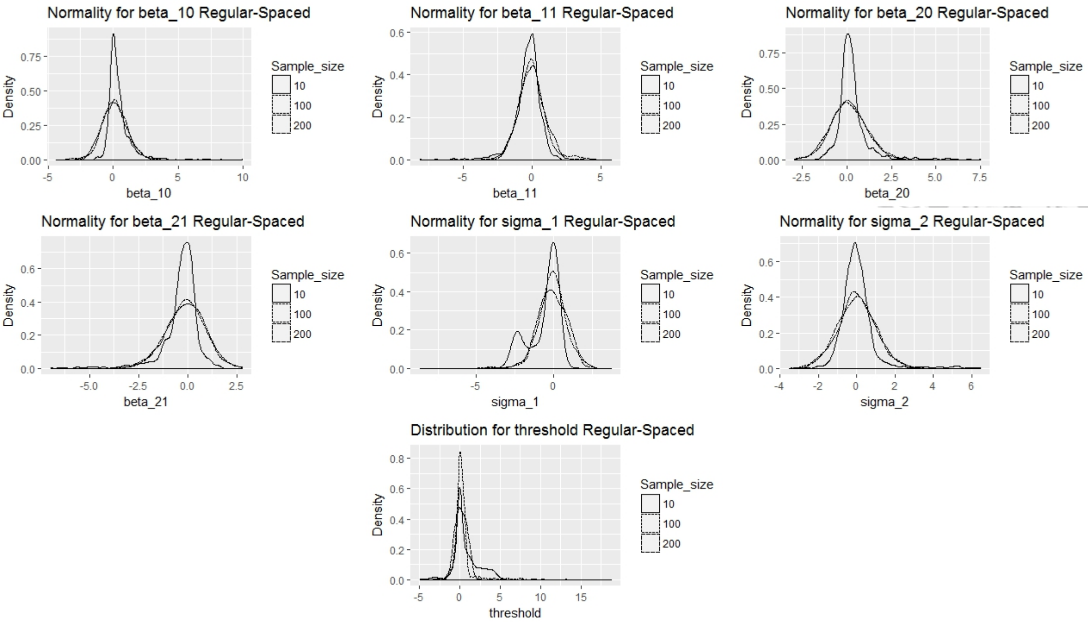

Figure 4.1.1: Asymptotic Density for the AMLEs - OU Process with Both the Drift and the Diffusion Are Nonlinear - Regular Data

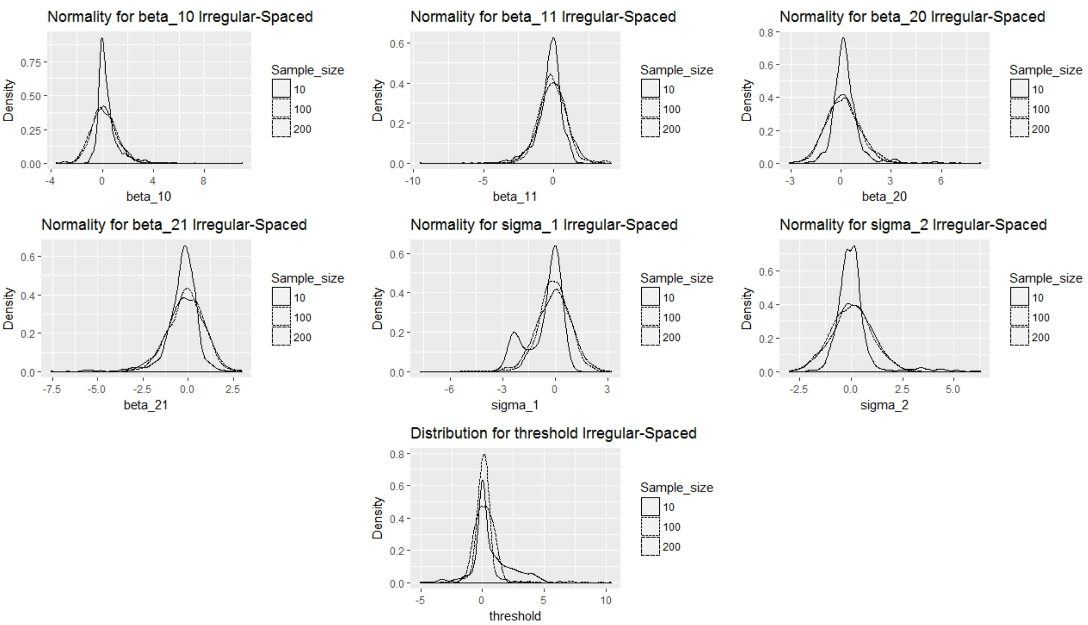

Figure 4.1.2: Asymptotic Density for the AMLEs - OU Process with Both the Drift and the Diffusion Are Nonlinear - Irregular Data

doi:10.6342/NTU201701414

4.2 Ornstein-Uhlenbeck process with Only the Diffusion Is Nonlinear

Model:

$$dX_t \;=\; (\beta_0 + \beta_1 X_t)dt + \{\sigma_1 I(X_t \leq r) + \sigma_2 I(X_t > r)\}dW_t, \qquad (4.3)$$

where $W = \{W_t\}$ stands for the standard Brownian process.

Table 4.2: Averages (standard deviations) of 1,000 simulations of the approximate maximum likelihood estimators of the parameters r, β_0, β_1, σ_1, and σ_2. (Ornstein - Uhlenbeck process with Only the Diffusion Is Nonlinear)

	parameter	true value \ T	10	100	200
	r	0.75	0.7492(0.0226)	0.7507(0.0033)	0.7508(0.0031)
	β_0	1.5	1.8456(0.6423)	1.5243(0.1621)	1.5145(0.1122)
regularly	β_1	-2	-2.4438(0.7867)	-2.0315(0.2039)	-2.0194(0.1411)
-spaced	σ_1	0.4	0.4002(0.0153)	0.3995(0.0044)	0.3993(0.0031)
	σ_2	0.3	0.2992(0.0104)	0.3002(0.0028)	0.3004(0.0020)
	r	0.75	0.7492(0.0186)	0.7510(0.0033)	0.7507(0.0032)
	β_0	1.5	1.8418(0.6977)	1.5251(0.1718)	1.5143(0.1224)
irregularly	β_1	-2	-2.4392(0.8582)	-2.0329(0.2156)	-2.0194(0.1534)
-spaced	σ_1	0.4	0.4004(0.0154)	0.3994(0.0043)	0.3993(0.0031)
	σ_2	0.3	0.2993(0.0101)	0.3002(0.0028)	0.3004(0.0020)

doi:10.6342/NTU201701414

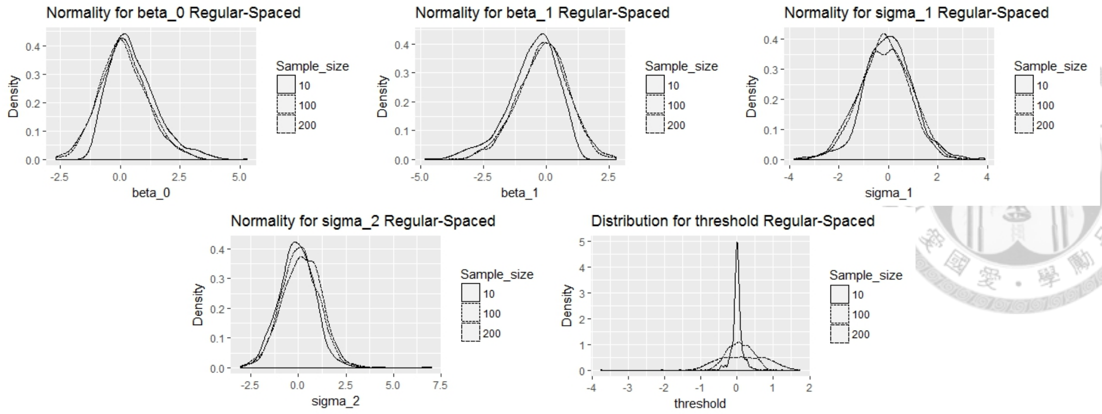

Figure 4.2.3: Asymptotic Density for the AMLEs - OU Process with Only the Diffusion Is Nonlinear - Regular Data

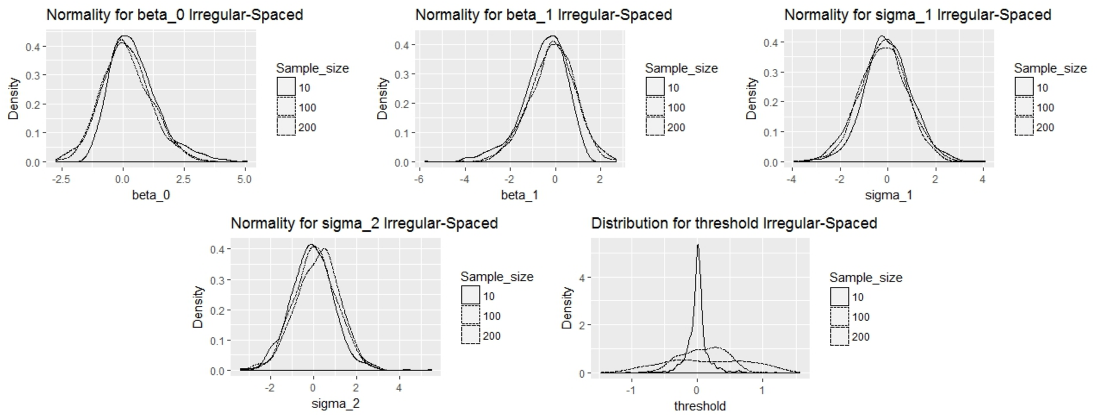

Figure 4.2.4: Asymptotic Density for the AMLEs - OU Process with Only the Diffusion Is Nonlinear - Irregular Data

4.3 Cox-Ingersoll-Ross model with Both the Drift and the Diffusion Are Nonlinear

$$\begin{aligned} dX_t &= \{(\beta_{10} + \beta_{11}X_t)I(X_t \leq r) + (\beta_{20} + \beta_{21}X_t)I(X_t > r)\}\, dt \\ &\quad + \{\sigma_1 I(X_t \leq r) + \sigma_2 I(X_t > r)\}\sqrt{X_t} dW_t, \end{aligned} \tag{4.4}$$

where $W = \{W_t\}$ stands for the standard Brownian process.

doi:10.6342/NTU201701414

Table 4.3: Averages (standard deviations) of 1,000 simulations of the approximate maximum likelihood estimators of the parameters r, β_{10}, β_{11}, β_{20}, β_{21}, σ_1, and σ_2. (Cox-Ingersoll-Ross model with Both the Drift and the Diffusion Are Nonlinear)

	parameter	true value \ T	10	100	200
	r	0.75	0.7622(0.1764)	0.7509(0.0084)	0.7507(0.0039)
	β_{10}	1	1.6374(1.7315)	1.0101(0.1592)	1.0029(0.1060)
regularly	β_{11}	-2	-2.7261(2.4105)	-2.0025(0.3507)	-1.9986(0.2288)
-spaced	β_{20}	1.5	2.3057(2.6327)	1.5346(0.2979)	1.5151(0.2102)
	β_{21}	-1.5	-2.5508(2.7468)	-1.5420(0.2948)	-1.5195(0.2079)
	σ_1	0.4	0.3812(0.0381)	0.3998(0.0045)	0.3998(0.0029)
	σ_2	0.3	0.3127(0.0337)	0.3000(0.0031)	0.3000(0.0021)
	r	0.75	0.7548(0.1764)	0.7506(0.0065)	0.7505(0.0038)
	β_{10}	1	1.6438(1.7125)	1.0226(0.1662)	1.0090(0.1164)
irregularly	β_{11}	-2	-2.7315(2.4472)	-2.0334(0.3568)	-2.0111(0.2487)
-spaced	β_{20}	1.5	2.2186(2.5855)	1.5465(0.3374)	1.5226(0.2224)
	β_{21}	-1.5	-2.4400(2.5341)	-1.5543(0.3320)	-1.5277(0.2206)
	σ_1	0.4	0.3827(0.0379)	0.4000(0.0041)	0.3999(0.0028)
	σ_2	0.3	0.3133(0.0342)	0.3000(0.0033)	0.3001(0.0021)

doi:10.6342/NTU201701414

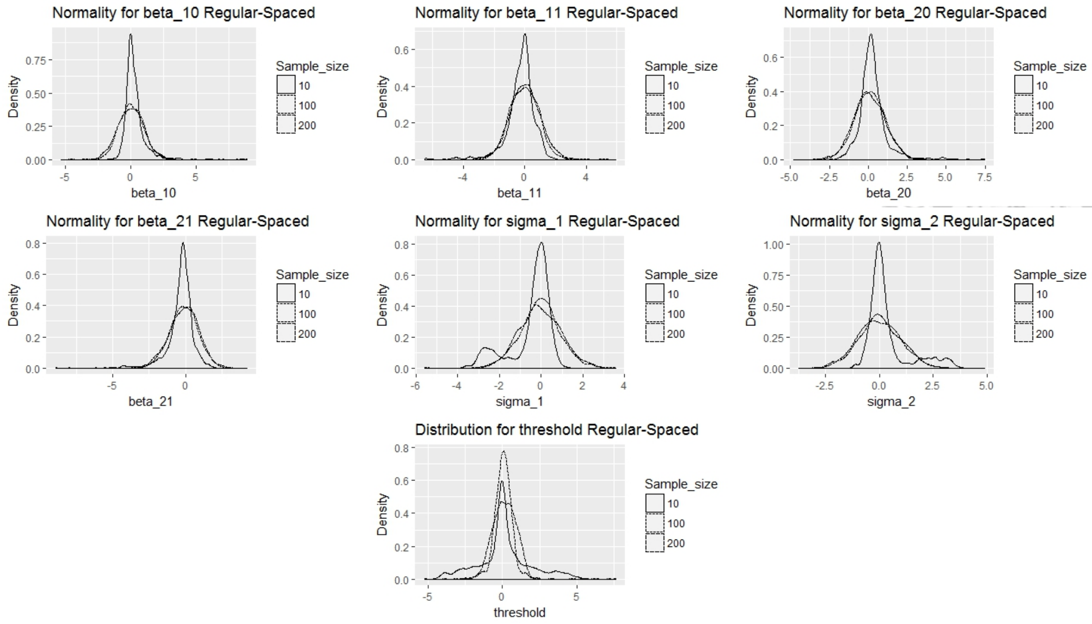

Figure 4.3.5: Asymptotic Density for the AMLEs - CIR Process with Both the Drift and the Diffusion Are Nonlinear - Regular Data

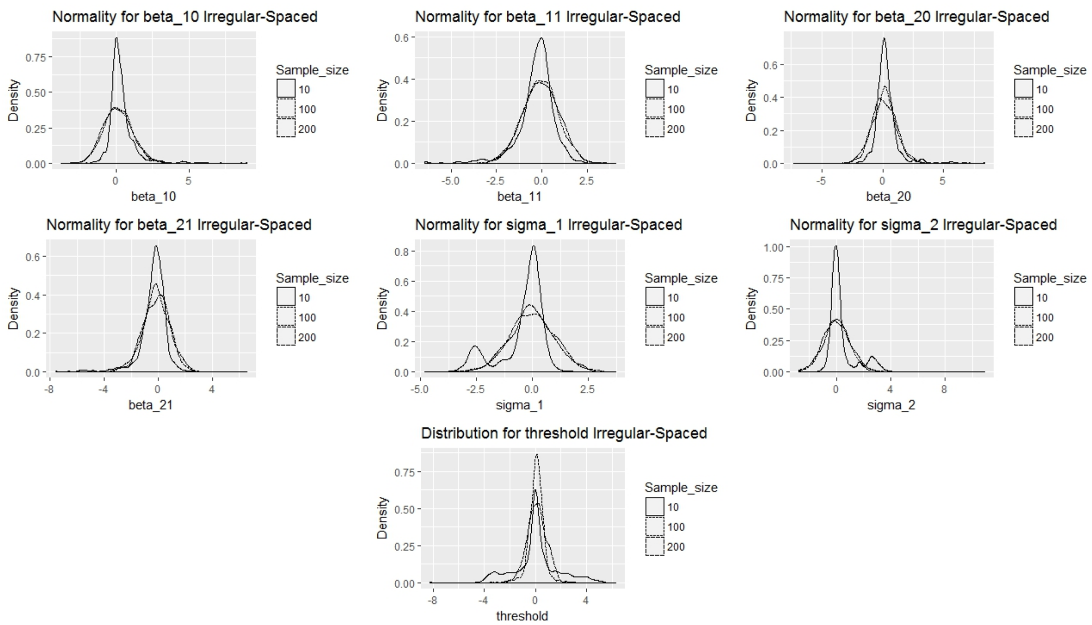

Figure 4.3.6: Asymptotic Density for the AMLEs - CIR Process with Both the Drift and the Diffusion Are Nonlinear - Irregular Data

doi:10.6342/NTU201701414

4.4 Cox-Ingersoll-Ross model with Only the Diffusion Is Nonlinear

Model:

$$dX_t = (\beta_0 + \beta_1 X_t)dt + \{\sigma_1 I(X_t \leq r) + \sigma_2 I(X_t > r)\}\sqrt{X_t}dW_t, \qquad (4.5)$$

where $W = \{W_t\}$ stands for the standard Brownian process.

Table 4.4: Averages (standard deviations) of 1,000 simulations of the approximate maximum likelihood estimators of the parameters r, β_0, β_1, σ_1, and σ_2. (Cox-Ingersoll-Ross model with Only the Diffusion Is Nonlinear)

	parameter	true value \ T	10	100	200
	r	0.75	0.7498(0.0189)	0.7506(0.0034)	0.7508(0.0030)
	β_0	1.5	1.8303(0.5892)	1.5237(0.1484)	1.5138(0.1030)
regularly	β_1	-2	-2.4335(0.7529)	-2.0314(0.1953)	-2.0189(0.1355)
-spaced	σ_1	0.4	0.3998(0.0147)	0.3994(0.0044)	0.3993(0.0030)
	σ_2	0.3	0.2994(0.0108)	0.3003(0.0029)	0.3005(0.0021)
	r	0.75	0.7499(0.0181)	0.7508(0.0034)	0.7507(0.0031)
	β_0	1.5	1.8265(0.6403)	1.5246(0.1571)	1.5144(0.1121)
irregularly	β_1	-2	-2.4290(0.8218)	-2.0330(0.2065)	-2.0198(0.1473)
-spaced	σ_1	0.4	0.4000(0.0148)	0.3993(0.0042)	0.3992(0.0030)
	σ_2	0.3	0.2994(0.0108)	0.3003(0.0030)	0.3005(0.0021)

doi:10.6342/NTU201701414

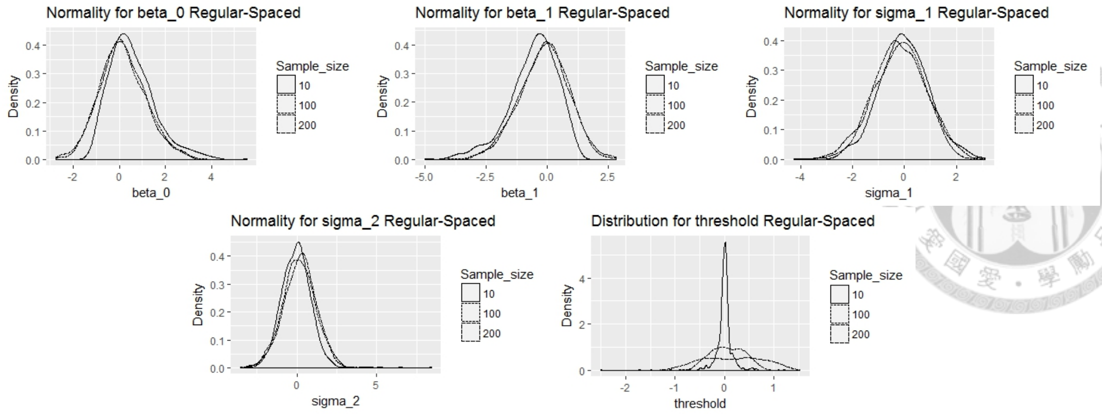

Figure 4.4.7: Asymptotic Density for the AMLEs - CIR Process with Only the Diffusion Is Nonlinear - Regular Data

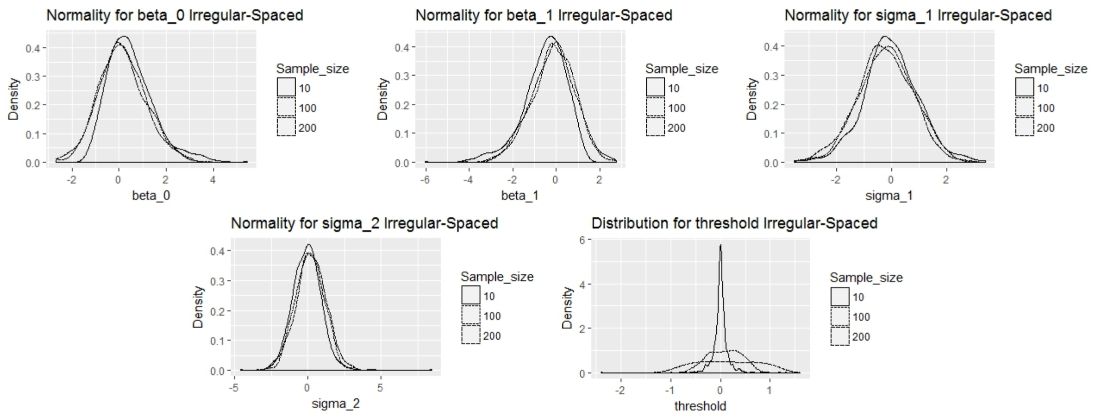

Figure 4.4.8: Asymptotic Density for the AMLEs - CIR Process with Only the Diffusion Is Nonlinear - Irregular Data

$Remark$ 4.1 From the simulation results in this chapter, we have some evidence, which is still lack of mathematical proof, to conclude our conjectures in chapter 3 for the large sample properties for AMLE.

doi:10.6342/NTU201701414

doi:10.6342/NTU201701414

Chapter 5

Application

We consider two series, namely the 3-Month Treasury Bill: Secondary Market Rate (from 1934.1.1 to 2017.4.1), which are downloable at https://fred.stlouisfed.org/series/TB3MS, and the 10-Year Treasury Constant Maturity Rate-3-Month Treasury Bill: Secondary Market Rate (from 1962.1.2 to 2017.5.11), which are available at https://fred.stlouisfed.org/graph/?g=oGg.
Note that the first series are regularly-spaced and count on every first day in month(1001 data points), whereas the second one is irregulary-spaced count on every day(14444 data points). The values of the first series are always positive, whereas the values of the second one could be negative.

doi:10.6342/NTU201701414

5.1 3-Month Treasury Bill:from 1934.1.1 to 2017.4.1

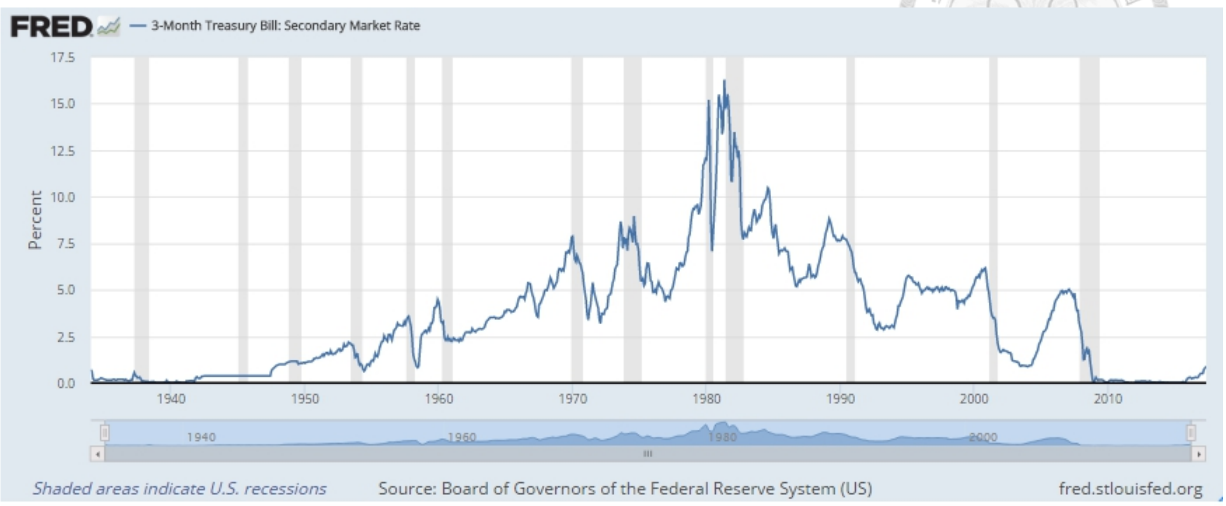

Figure 5.1.1: 3-Month Treasury Bill: Secondary Market Rate (from 1934.1.1 to 2017.4.1)

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.010	0.380	3.060	3.527	5.360	16.300

Table 5.1: Summary Statistics:3-Month Treasury Bill

For the 3-Month Treasury Bill, the fitted models are

$$
\begin{aligned}
dX_t &= \{(0.4518 - 0.5520X_t)I(X_t \leq 0.510) + (4.6530 - 1.0174X_t)I(X_t > 0.510)\}\,dt \\
&\quad + \{0.4288I(X_t \leq 0.510) + 4.2345I(X_t > 0.510)\}dW_t, \\
dX_t &= \{(0.4554 - 0.3803X_t)\}\,dt + \{0.4288I(X_t \leq 0.510) + 4.2403I(X_t > 0.510)\}dW_t, \\
dX_t &= \{(0.6846 - 0.1164t)I(X_t \leq 5.780) + (21.649 - 2.7826X_t)I(X_t > 0.578)\}\,dt \\
&\quad + \{1.2943I(X_t \leq 5.780) + 2.1984I(X_t > 5.780)\}\sqrt{X_t}dW_t, \\
dX_t &= \{(0.6940 - 0.1579X_t)\}\,dt + \{1.2943I(X_t \leq 5.780) + 2.2063I(X_t > 5.780)\}\sqrt{X_t}dW_t,
\end{aligned}
$$

Remark 5.1 For those working model above, the threshold values could be viewed as the level that the whole investor in the financial market will change their attitude toward the 3-Month Treasury Bill under the assumptions to those specific models.

doi:10.6342/NTU201701414

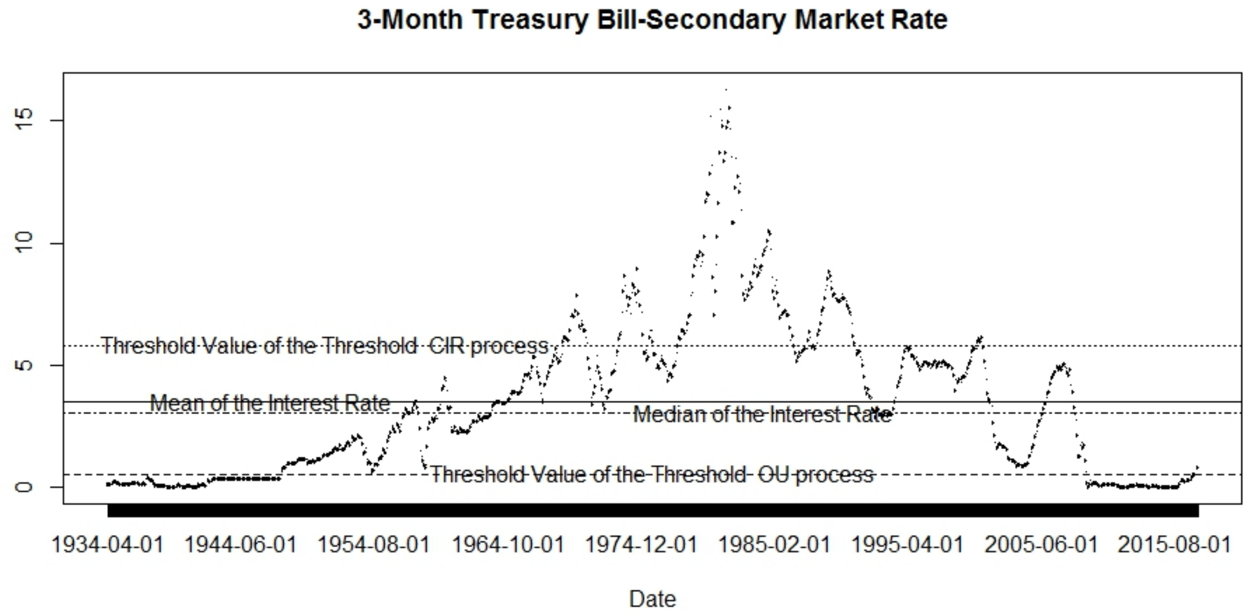

Figure 5.1.2: 3-Month Treasury Bill: Secondary Market Rate with Threshold Values

5.2 10-Year Treasury Constant Maturity Rate-3-Month Treasury Bill:from 1962.1.2 to 2017.5.11

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-3.73	0.62	1.61	1.56	2.58	5.41

Table 5.2: Summary Statistics:10-Year Treasury Constant Maturity Rate-3-Month Treasury Bill

For the 10-Year Treasury Constant Maturity Rate-3-Month Treasury Bill, the fitted models are

$$\begin{aligned} dX_t &= \{(0.2816 - 1.0308X_t)I(X_t \leq 0.390) + (0.2889 - 0.1330X_t)I(X_t > 0.390)\}\, dt \\ &\quad + \{1.0458I(X_t \leq 0.390) + 0.6834I(X_t > 0.390)\}dW_t, \\ dX_t &= \{(0.0164 - 0.1865X_t)\}\, dt + \{0.1577I(X_t \leq 1.110) + 0.2367I(X_t > 1.110)\}dW_t, \end{aligned}$$

Remark 5.2 For those working model above, again, the threshold values could be

doi:10.6342/NTU201701414

viewed as the level that the whole investor in the financial market will change their attitude toward the 10-Year Treasury Constant Maturity Rate-3-Month Treasury Bill under the assumptions to those specific models. Moreover, we can use this data to infer the arbitrage direction if we should hold 10-Year Treasury Constant Maturity Rate or short 3-Month Treasury Bill.

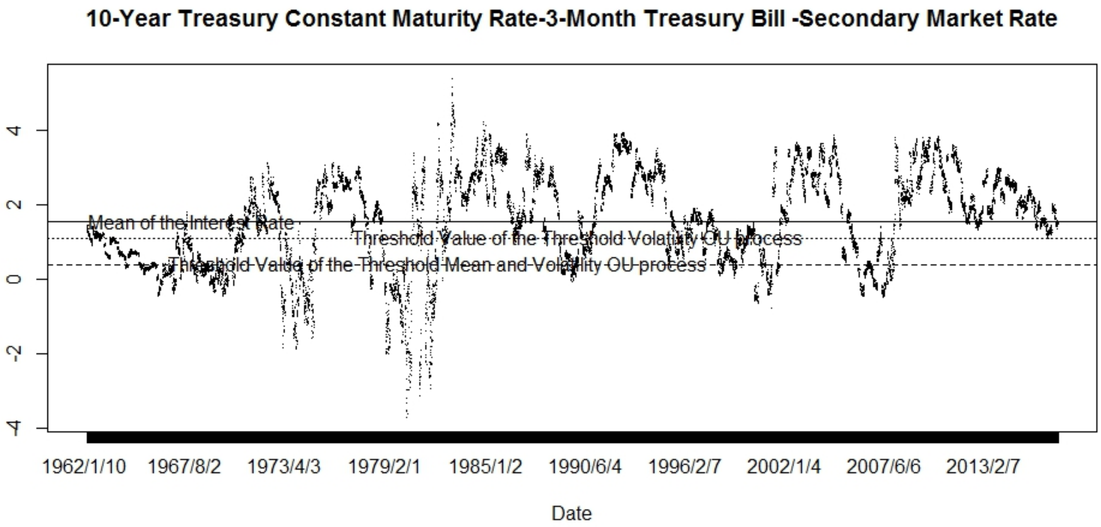

Figure 5.2.3: 10-Year Treasury Constant Maturity Rate-3-Month Treasury Bill with Threshold Values

doi:10.6342/NTU201701414

Chapter 6

Conclusion

In this thesis, we introduce the threshold diffusion process and the quasi-likelihood estimation(QLE), and we point out that searching maximum likelihood estimator via irregular-spaced data lead to an estimating problem. Thus, we propose approximate maximum likelihood estimation(AMLE) to estimate the parameters in drift term and diffusion term.

Moreover, we use simulation study to confirm our conjecture about some large sample properties of AMLE, such as consistency and asymptotic normality. Although we do not give more specifically sufficient conditions to prove these properties transparently, we still formalize these asymptotic results and derive the asymptotic variance for the asymptotic normality. In simulation, we can see that AMLE converges to the true value in the sense of larger time interval when the specification is right.

In addition to the theoretical discussion, we also apply the model and our AMLE to real data, the application of real data gives us a idea to diagnostic whether the diffusion is a threshold constant function , a threshold smooth power function. or even a nonstationary process.

The advantage is that one can do the hypothesis test for all parameters jointly except the threshold parameters. However, it remains some technical parts which may left as future work. Also, a test procedure for the existence of threshold parameters and the pricing formula of a threshold diffusion process might follow the results in Su's PhD thesis(2011) [22].

As what we might ask in discrete time series, procedures of prediction and model selection for a threshold diffusion process based on AMLE are also in our interest for the future work.

doi:10.6342/NTU201701414

Furthermore, the results in the discrete time threshold time series such as [9], [20], and [25] are potential to extended to the CTAR model or the CTARMA model.

doi:10.6342/NTU201701414

Bibliography

[1] Y. Aït-Sahalia. Maximum likelihood estimation of discretely sampled diffusions: A closed-form approximation approach. Econometrica, 70(1):223–262, 2002.

[2] Y. Aït-Sahalia and P. A. Mykland. Estimators of diffusions with randomly spaced discrete observations: a general theory. Annals of Statistics, pages 2186–2222, 2004.

[3] P. Brockwell, R. Davis, and Y. Yang. Continuous-time gaussian autoregression. Statistica Sinica, 17(1):63, 2007.

[4] P. Brockwell and R. Hyndman. On continuous-time threshold autoregression. International Journal of Forecasting, 8(2):157–173, 1992.

[5] P. Brockwell and O. Stramer. On the approximation of continuous time threshold arma processes. Annals of the Institute of Statistical Mathematics, 47(1):1–20, 1995.

[6] P. J. Brockwell. On continuous-time threshold arma processes. Journal of Statistical Planning and Inference, 39(2):291–303, 1994.

[7] P. J. Brockwell, R. J. Hyndman, and G. K. Grunwald. Continuous time threshold autoregressive models. Statistica Sinica, pages 401–410, 1991.

[8] K.-S. Chan. Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model. The annals of statistics, pages 520–533, 1993.

[9] K.-S. Chan and R. S. Tsay. Limiting properties of the least squares estimator of a continuous threshold autoregressive model. Biometrika, 85(2):413–426, 1998.

[10] J. Chang, S. X. Chen, et al. On the approximate maximum likelihood estimation for diffusion processes. The Annals of Statistics, 39(6):2820–2851, 2011.

doi:10.6342/NTU201701414

[11] J. C. Cox, J. E. Ingersoll Jr, and S. A. Ross. A theory of the term structure of interest rates. Econometrica: Journal of the Econometric Society, pages 385–407, 1985.

[12] P. D. Feigin. Maximum likelihood estimation for continuous-time stochastic processes. Advances in Applied Probability, 8(4):712–736, 1976.

[13] B. E. Hansen. Inference in tar models. Studies in nonlinear dynamics & econometrics, 2(1), 1997.

[14] B. E. Hansen. Threshold autoregression in economics. Statistics and its Interface, 4(2):123–127, 2011.

[15] T. Henghsiu. Topics in continuous-time modeling, 1998.

[16] I. Karatzas and S. Shreve. Brownian motion and stochastic calculus, volume 113. Springer Science & Business Media, 2012.

[17] F. C. Klebaner. Introduction to stochastic calculus with applications. World Scientific Publishing Co Inc, 2005.

[18] Y. A. Kutoyants. On identification of the threshold diffusion processes. Annals of the Institute of Statistical Mathematics, 64(2):383–413, 2012.

[19] D. Li and S. Ling. On the least squares estimation of multiple-regime threshold autoregressive models. Journal of Econometrics, 167(1):240–253, 2012.

[20] D. Li, S. Ling, and J.-M. Zakoïan. Asymptotic inference in multiple-threshold double autoregressive models. Journal of Econometrics, 189(2):415–427, 2015.

[21] O. Stramer, R. Tweedie, and P. Brockwell. Existence and stability of continuous time threshold arma processes. Statistica Sinica, pages 715–732, 1996.

[22] F. Su. Statistical analysis of non-linear diffusion process, 2011.

[23] F. Su and K.-S. Chan. Quasi-likelihood estimation of a threshold diffusion process. Journal of Econometrics, 189(2):473–484, 2015.

[24] H. Tong. Non-linear time series. A Dynamical System Approach, 1990.

[25] H. Tong. Nonlinear time series analysis since 1990: Some personal reflections. Acta Mathematicae Applicatae Sinica (English Series), 18(2):177–184, 2002.

doi:10.6342/NTU201701414

[26] H. Tong and I. Yeung. Threshold autoregressive modelling in continuous time. Statistica Sinica, pages 411–430, 1991.

doi:10.6342/NTU201701414