Journal of Computational and Applied Mathematics 348 (2019) 282–297

Contents lists available at ScienceDirect

Journal of Computational and Applied Mathematics

journal homepage: www.elsevier.com/locate/cam

Data driven confidence intervals for diffusion process using double smoothing empirical likelihood

Qi Yang [a], Yuping Song [b,*]

[a] *Department of Statistics, The Chinese University of Hong Kong, Hong Kong, PR China*
[b] *School of Finance and Business, Shanghai Normal University, Shanghai, PR China*

A R T I C L E I N F O

Article history:
Received 13 April 2018
Received in revised form 16 August 2018

Keywords:
Empirical likelihood
Diffusion process
Volatility function
Confidence interval
Double smoothing

A B S T R A C T

In this paper, we propose an innovative estimation method for the volatility function in diffusion process and construct the empirical likelihood confidence interval for it. Compared with double smoothing local constant estimator, double smoothing local linear estimator proposed in this paper is an inventive estimation method. Moreover, we find that the empirical likelihood confidence interval constructed with the approximate estimating equation is much better than the one based on the estimating equation, since the latter undermines the predictability of the estimator. Accordingly, new algorithm for simulation is proposed. Through Monte Carlo simulation and empirical analysis, both these approaches are superior to traditional asymptotic normality confidence interval in terms of coverage rate, etc.

1. Introduction

This paper is concerned mainly about the construction of point-wise confidence interval for the diffusion coefficient of the following diffusion process via empirical likelihood:

$$dX_t = \mu(X_t)dt + \sigma(X_t)dB_t. \tag{1.1}$$

where B_t stands for the standard Brownian motion. These coefficients of model (1.1) economically represent the time trend referred as expected risk return, which reports a complete and scale-free summary of the investment opportunity, and the conditional variance of the return for an underlying asset, respectively. Supposing that price process of the risky stock follows a specific form of model (1.1), that is geometric Brownian motion, Gao and Yin [1] discussed a perturbed risk process compounded by it with a dividend barrier strategy, Zhu [2] tested the expected return and market price of risk in Chinese A–B share market, Liu and Xu [3] studied the capped stock loans for the stock price, Abidin and Jaffar [4] forecasted the share prices of small size companies in Bursa Malaysia, Vajargah and Shoghi [5] predicted the total index of stock market and value at risk. Furthermore, Yuan and Lai [6] surveyed financial markets with volatility uncertainty based on the price process governed by the constant elasticity of variance, which is a natural extension of the geometric Brownian motion.

As we all know, financial phenomena can be explained and modeled by Eq. (1.1). Statisticians and economists spared no efforts to research on (1.1) and made much progress since 1970s to characterize the applications of the above equation. For model (1.1), the statistical inference of continuous-path diffusion process is usually focused on the estimation of $\mu(x)$ and

* Corresponding author.
E-mail address: songyuping@shnu.edu.cn (Y. Song).

https://doi.org/10.1016/j.cam.2018.08.027
0377-0427/

$\sigma(x)$. In general, the nonparametric estimation usually depends on the following two relations:

$$E\left(\frac{X_{t+\Delta}-X_t}{\Delta}\middle| X_t\right)=\mu(X_t)+o(1), \tag{1.2}$$

$$E\left(\frac{(X_{t+\Delta}-X_t)^2}{\Delta}\middle| X_t\right)=\sigma^2(X_t)+o(1). \tag{1.3}$$

which can be calculated through the infinitesimal generator of X_t. Based on the infinitesimal generator, Ramsden and Papaioannou [7] derived integro-differential equations for the ruin probabilities, Hozman and Tichý [8] studied the pricing of European options under a general one-factor stochastic volatility model. Volatility is an important indicator of the quality and efficiency of financial markets and is also a key variable in the field of price of derivatives, portfolio construction and financial risk management, one can refer to Noh and Kim [9], Chiu and Wong [10], Zhang and Ge [11] and Li [12]. Therefore, how to effectively describe the dynamic behavior of financial market volatility has been a hot topic in applied mathematics, such as statistics and econometrics. The past several decades saw many estimation methods for the diffusion coefficient for its essential role played in the financial application. The pioneers of this field are Florens-Zmirou [13] and Aït-Sahalia [14]. Among them, Aït-Sahalia [14] firstly introduced the following estimator (1.4) as a representation of their idea for the diffusion coefficient of the instant interest model in a semiparametric context based on the data:

$$\hat{\sigma}^2(x)=\frac{2\sum_{i=1}^{T}\int_{-\infty}^{x}\mu(s)K\left(\frac{s-\hat{x}_i}{h_s}\right)ds}{\sum_{i=1}^{T}K\left(\frac{x-\hat{x}_i}{h_s}\right)}. \tag{1.4}$$

Stanton [15] estimated both the drift and diffusion coefficients in a fully nonparametric context and the estimator for the diffusion coefficient was

$$\hat{\sigma}^2(x,h)=\frac{\sum_{i=1}^{n}K_i(x,h)(X_{(i+1)\Delta}-X_{i\Delta})^2}{\Delta\sum_{i=1}^{n}K_i(x,h)}. \tag{1.5}$$

Bandi and Phillips [16] proposed the double smoothing inference method for the drift and diffusion coefficients of the continuous-time diffusion process, that is,

$$\hat{\sigma}^2(x)=\frac{\sum_{i=1}^{n}K\left(\frac{X_{i\Delta}-x}{h}\right)\left(\frac{1}{m_i\Delta}\sum_{j=0}^{m_i-1}(X_{t_j+\Delta}-X_{t_j})^2\right)}{\sum_{i=1}^{n}K\left(\frac{X_{i\Delta}-x}{h}\right)}. \tag{1.6}$$

They just did some research on theoretical results, however, did not show the goodness of their proposed estimator through simulations.

With development of computer technology and wide application of statistical software, Monte Carlo simulation became increasingly popular. Through simulation, Chapman and Pearson [17] did some research on the properties of the estimators proposed in [14,15]. Renò et al. [18] compared these three estimators (1.4)–(1.6) through Monte Carlo simulation with the data generated by CIR and Vasicek model. They characterized the estimator in [14] as problematic one, however, the estimators in [15,16], performed better despite of the bias in the tail distribution. In general, owing to the possibility of the nonlinearity between the conditional variation and error rate in the three estimators above, authors held the view that the nonparametric estimation for diffusion coefficient of short-term interest rate was not reliable. As an improvement in the estimation method, Xu [19] estimated both the drift and diffusion coefficient in model (1) based on local linear method, and constructed the empirical likelihood (EL) confidence interval for them. The asymptotic normality (AN) confidence interval was also constructed for the sake of comparison. Simulation results in [19] showed that EL confidence interval outperforms the AN confidence interval in terms of many important statistical properties, especially in the coverage rate.

Empirical likelihood is an important nonparametric estimation method in statistics, which has been enjoying great popularity for the past two decades within the statistical academic community. For more details, one can refer to Owen [20–22], Qin and Lawless [23] and Zhao et al. [24]. Empirical likelihood possesses many advantages as follows. Compared with the maximum likelihood estimation, this method requires less assumptions on the error distribution, which also performs pretty well even when the data is asymmetric or censored. Besides, the shape of EL confidence interval is data-driven, which makes it better than the symmetric confidence interval constructed based on the traditional asymptotic normality method. Additionally, the EL confidence interval can be Bartlett corrected and keeps unchanged under transformation. Furthermore, empirical likelihood estimation method does not have to construct moment statistic or estimate the asymptotic square deviation, which makes it outstanding when it is difficult to determine the standard deviation of the data. Lin and Wang [25] considered empirical likelihood inference for diffusion processes with jumps, Hanif [26] employed empirical likelihood to construct re-weighted Nadaraya–Watson estimator for scalar diffusion models using asymmetric kernels, Wang et al. [27] proposed empirical likelihood based inference for second-order diffusion models. Xu [28] provides a generic framework of obtaining empirical likelihood using the properly constructed local (kernel-smoothed) estimating equations (allowing the existence of nuisance parameters) in the discrete-time setting.