Journal of Econometrics 189 (2015) 473–484

Contents lists available at ScienceDirect

Journal of Econometrics

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

Quasi-likelihood estimation of a threshold diffusion process

Fei Su [a], Kung-Sik Chan [b,*]

[a] ISO Innovative Analytics, San Francisco, CA, United States

[b] Department of Statistics and Actuarial Science, The University of Iowa, Iowa City, IA 52242, United States

ARTICLE INFO

Article history:
Available online 20 March 2015

JEL classification:
C22
E43

Keywords:
Girsanov's theorem
Interest rates
Nonlinear time series
Stochastic differential equation
Term structure

ABSTRACT

The threshold diffusion process, first introduced by Tong (1990), is a continuous-time process satisfying a stochastic differential equation with a piecewise linear drift term and a piecewise smooth diffusion term, e.g., a piecewise constant function or a piecewise power function. We consider the problem of estimating the (drift) parameters indexing the drift term of a threshold diffusion process with continuous-time observations. Maximum likelihood estimation of the drift parameters requires prior knowledge of the functional form of the diffusion term, which is, however, often unavailable. We propose a quasi-likelihood approach for estimating the drift parameters of a two-regime threshold diffusion process that does not require prior knowledge about the functional form of the diffusion term. We show that, under mild regularity conditions, the quasi-likelihood estimators of the drift parameters are consistent. Moreover, the estimator of the threshold parameter is super consistent and weakly converges to some non-Gaussian continuous distribution. Also, the estimators of the autoregressive parameters in the drift term are jointly asymptotically normal with distribution the same as that when the threshold parameter is known. The empirical properties of the quasi-likelihood estimator are studied by simulation. We apply the threshold model to estimate the term structure of a long time series of US interest rates. The proposed approach and asymptotic results can be readily lifted to the case of a multi-regime threshold diffusion process.

1. Introduction

In financial and insurance markets, diffusion processes have become the standard tool for modeling returns and values for risk management purposes. For example, a number of diffusion processes have been used to model the term structure of market yields such as interest rate (Vasicek, 1977; Cox et al., 1985; Black and Karasinski, 1991), some of which include time-dependent covariates in the mean function (Hull, 2010; Black et al., 1990). While the functional form of the diffusion term differs in these models, their drift terms stay affine (or can be transformed to linear functions). Despite their relative computational convenience, linear diffusion processes fail to capture nonlinear characteristics such as multimodality, asymmetric periodic behavior, time-irreversibility, and the occurrence of occasional extreme events that are commonly found in real data.

Continuous-time nonlinear models have proved increasingly useful over the past decade for capturing the aforementioned nonlinear properties (Tong, 1990; Decamps et al., 2006). Although continuous-time nonlinear diffusion processes form a relatively large model class, the field of empirical nonlinear time series modeling is relatively under-explored, except for the first-order continuous-time threshold autoregressive (CTAR) model proposed by Tong (1990); see Section 2 for the definition of the CTAR model, and some of its properties. The first order CTAR model will be simply referred to as the threshold diffusion (TD) process below. Several approaches on the inference of TD processes with discrete-time data have been proposed in the literature, e.g., Gaussian likelihood estimation (Tong and Yeung, 1991; Brockwell and Hyndman, 1992; Brockwell, 1994; Brockwell et al., 2007), moment-based estimators (Chan et al., 1992; Coakley et al., 2003), and Bayesian approach (Pai and Pedersen, 1999). If sufficiently fine data are available, the likelihood function can be approximated by Girsanov's formula (at least for the case of known diffusion term). An advantage of Bayesian estimation is that even when the data are not sufficiently fine, Bayesian data augmentation techniques could be used; (see Elerian et al. (2001), Eraker (2001, 2004), Roberts and Stramer (2001), Stramer and Roberts (2007)).

Within the under-developed literature on the inference of the TD processes with continuous-time data, maximum likelihood is preferable for efficiency consideration. Recently, Kutoyants (2012) derived the asymptotic distribution for maximum-likelihood estimation of a TD model under restrictive conditions including

* Corresponding author. Tel.: +1 319 335 2849; fax: +1 319 335 3017.
E-mail addresses: sophysu@gmail.com (F. Su), kung-sik-chan@uiowa.edu (K.-S. Chan).

http://dx.doi.org/10.1016/j.jeconom.2015.03.038
0304-4076/

bounded parameter space, known ordering among some parameters, and known functional form of the diffusion term. In practice, the functional form of the diffusion term is generally unknown. Thus, it is desirable to develop an estimation method that does not require knowing the functional form of the diffusion term.

Here, we introduce a quasi-likelihood approach to estimate the drift parameters of a TD model, without requiring prior knowledge of the functional form of the diffusion term. The quasi-likelihood is obtained by applying Girsanov's theorem to the TD model with constant diffusion coefficient even though the true diffusion term may be non-constant and even nonlinear. The consistency and the limiting distribution of the quasi-likelihood drift estimator of a 2-regime TD model are derived in Section 4, under some regularity conditions. Given data over T units of time, we show that the threshold parameter is T-consistent and its limiting distribution admits a closed-form pdf. Moreover, the autoregressive parameter estimators are $\sqrt{T}$-consistent, and asymptotically independent of the threshold estimator, with a limiting normal distribution which is the same as that assuming known threshold. A simulation study is conducted in Section 5 to illustrate the asymptotic results. In Section 6, we apply the proposed method to study the term structure of the US interest rate. We conclude briefly in Section 7. All proofs are collected in Appendix A.

2. Nonlinear diffusion processes

We begin with the general nonlinear diffusion process:

$$dX(t) = \mu(X(t), t)dt + \sigma(X(t), t)dW(t) \tag{1}$$

where the function $\mu(x, t)$ is the drift term (instantaneous mean function), $\sigma(x, t)$ is the diffusion term ($\sigma^2(x, t)$ instantaneous variance function) and $W = \{W(t)\}$ stands for the standard Brownian process. Here, we focus on the case that both the drift and diffusion terms are time-homogeneous, i.e., $\mu(x, t) \equiv \mu(x)$ and $\sigma(x, t) \equiv \sigma(x)$. The drift and the diffusion terms are generally known up to some parameters, in which case we write $\mu_{\boldsymbol{\theta}}$ for μ and $\sigma_{\boldsymbol{\gamma}}$ for σ where the drift parameter $\boldsymbol{\theta}$ and the diffusion parameter $\boldsymbol{\gamma}$ are vectors that may share some common parameters. For conciseness, these parameters are often suppressed.

Among all nonlinear diffusion processes, the first-order m-regime threshold diffusion (TD) model, which is the first-order case of the continuous-time threshold autoregressive process (Tong, 1990; Tong and Yeung, 1991), has received much attention in the literature, and it is defined to be the solution of the following stochastic differential equation

$$dX(t) = \sum_{1}^{m} \left\{ \boldsymbol{\beta}_i^{\top} \begin{pmatrix} 1 \\ X(t) \end{pmatrix} dt + \sigma_i dW(t) \right\} I(r_{i-1} < X(t) \le r_i) \tag{2}$$

where $-\infty = r_0 < r_1 < \cdots < r_m = \infty$ are the threshold parameters, $\boldsymbol{\beta}_i^{\top} = (\beta_{i0}, \beta_{i1})$ are the autoregressive parameters and σ_i's are the diffusion parameters. In other words, the drift term is piecewise linear while the diffusion term is piecewise constant, and the two functions have identical break points. Specifically, $\mu(x) = \sum_{i=1}^{m}(\beta_{i0} + \beta_{i1}x)I(r_{i-1} < x \le r_i)$ and $\sigma(x) = \sum_{i=1}^{m}\sigma_i I(r_{i-1} < x \le r_i)$. Thus, the TD process models the situation that the underlying process is governed by m Ornstein–Uhlenbeck (OU) sub-processes, with the ith OU governing mechanism in effect whenever the process $X(t)$ is in the ith regime, i.e., $X(t) \in (r_{i-1}, r_i]$. The TD process may switch regimes infinitely many times within an arbitrary small interval of time due to the properties of the Brownian motion.

Similar to Chan and Tong (1986), the hard-thresholding regime switching mechanism may be smoothed by employing a soft-thresholding rule. A smooth threshold diffusion (STD) model can be obtained by replacing $I(r_{i-1} < x \le r_i)$ by $F(x; r_i, s_i) - F(x; r_{i-1}, s_{i-1})$ where $F(\cdot; r, s)$ denotes the cumulative distribution function of some location-scale family with location parameter r and scale parameter s, for instance the family of normal or logistic distributions. The proposed estimation method and much of the theory developed below can be lifted to the STD model, with details to be reported elsewhere.

For the stationary solution of a TD model to exist, the sub-models of the two outermost regimes must be "stationary". The following theorem is due to Brockwell and Hyndman (1992) (see also Brockwell et al. (1991)).

Theorem 1. *Suppose that $\sigma_i > 0, i = 1, \ldots, m$. Then the process defined by (2) has a stationary distribution if and only if*

$$\lim_{x\to-\infty} \mu(x) > 0; \qquad \lim_{x\to\infty} \mu(x) < 0,$$

i.e., $\beta_{1,1} < 0$ and $\beta_{m,1} < 0$, or in the case that $\beta_{1,1} = 0$ ($\beta_{m,1} = 0$), then $\beta_{1,0} > 0$ ($\beta_{m,0} < 0$). Further, if the stationarity condition is satisfied, the stationary density is given by

$$\pi(x) = \sum_{i=1}^{m} k_i \exp\{(\beta_{i1}x^2 + 2\beta_{i0}x)/\sigma_i^2\} I(r_{i-1} < x \le r_i),$$

where the constants $\{k_i\}$ are determined by the conditions that (i) $\int_{-\infty}^{\infty} \pi(x)dx = 1$ *and* (ii) $\sigma_i^2\pi(r_i-) = \sigma_{i+1}^2\pi(r_i+), i = 1, \ldots, m-1$, *where $\pi(r_i-)$ and $\pi(r_i+)$ are the left and right hand limits of π at r_i. 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 instantaneous variance function $\sigma^2(x)$ is continuous at the threshold points.*

Note that the stationary density is generally non-Gaussian, asymmetric and often multi-modal for a TD process. For instance, Fig. 1 displays the stationary density function of the process $dX(t) = \{(-2 - 4X(t))I(X(t) \le 0) + (3 - 3X(t))I(X(t) > 0)\}dt + 4dW(t)$, which is non-Gaussian and bimodal. The form of the stationary density implies that it has finite moments of all orders. Also, a stationary TD model is geometrically ergodic (Stramer et al., 1996).

A more general TD model may be obtained by relaxing the piecewise constant diffusion term to a piecewise smooth function, for instance, a piecewise power diffusion term obtained by replacing σ_i by $\sigma_i X^{\gamma_i}(t)$ where γ_i are parameters. The preceding more general formulation enables us to model positive data without the need for data transformation. The stationarity results stated in Theorem 1 can be extended to the more general TD model. As an illustration, consider the stationarity condition for the square-root case when X is a positive process a.s., and $\sigma(x) = \sum_{i=1}^{m}\sigma_i\sqrt{x}I(r_{i-1} < x \le r_i)$, where $0 = r_0 < r_1 < \cdots < r_m = \infty$. We shall assume that $\sigma_i > 0, \forall i$. Let $Y(t) = \sqrt{X(t)}$. Then the stationary condition for $\{X(t)\}$ and $\{Y(t)\}$ should be the same. By Ito's formula,

$$dY(t) = \sum_{i=1}^{m} \left\{ \left(\frac{4\beta_{i0} - \sigma_i^2}{8Y(t)} + \frac{\beta_{i1}}{2}Y(t) \right) dt + \frac{\sigma_i}{2}dW(t) \right\} I(\sqrt{r_{i-1}} < Y(t) \le \sqrt{r_i}).$$

Thus, $\{X(t)\}$ is stationary if $4\beta_{10} - \sigma_1^2 > 0$ and $\beta_{m1} < 0$. Following an argument in Karlin and Taylor (1981, p. 221), the stationary density function can be shown to be

$$\pi(x) = \sum_{i=1}^{m} k_i x^{2\beta_{i0}/\sigma_i^2 - 1} \exp(2\beta_{i1}x/\sigma_i^2) I(r_{i-1} < x \le r_i) \tag{3}$$

where the constants k_i satisfy condition (i) of Theorem 1 and (ii') $\sigma_i^2 r_i \pi(r_i-) = \sigma_{i+1}^2 r_i \pi(r_i+), i = 1, \ldots, m-1$. Thus, the stationary density is piecewise "Gamma"-distributed. In summary, the TD