Journal of Econometrics 150 (2009) 139–150

Contents lists available at ScienceDirect

Journal of Econometrics

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

A two-stage realized volatility approach to estimation of diffusion processes with discrete data[☆]

Peter C.B. Phillips [a,b,c,d], Jun Yu [e,*]

[a] Cowles Foundation, Yale University, United States
[b] Department of Economics, University of Auckland, New Zealand
[c] Department of Economics, University of York, United Kingdom
[d] School of Economics, Singapore Management University, Singapore
[e] School of Economics, Singapore Management University, 90 Stamford Road, Singapore 178903, Singapore

ARTICLE INFO

Article history:
Available online 25 December 2008

JEL classification:
C13
C22
E43
G13

Keywords:
Maximum likelihood
Girsanov theorem
Discrete sampling
Continuous record
Realized volatility

ABSTRACT

This paper motivates and introduces a two-stage method of estimating diffusion processes based on discretely sampled observations. In the first stage we make use of the feasible central limit theory for realized volatility, as developed in [Jacod, J., 1994. Limit of random measures associated with the increments of a Brownian semiartingal. Working paper, Laboratoire de Probabilities, Universite Pierre et Marie Curie, Paris] and [Barndorff-Nielsen, O., Shephard, N., 2002. Econometric analysis of realized volatility and its use in estimating stochastic volatility models. Journal of the Royal Statistical Society. Series B, 64, 253–280], to provide a regression model for estimating the parameters in the diffusion function. In the second stage, the in-fill likelihood function is derived by means of the Girsanov theorem and then used to estimate the parameters in the drift function. Consistency and asymptotic distribution theory for these estimates are established in various contexts. The finite sample performance of the proposed method is compared with that of the approximate maximum likelihood method of [Aït-Sahalia, Y., 2002. Maximum likelihood estimation of discretely sampled diffusion: A closed-form approximation approach. Econometrica. 70, 223–262].

1. Introduction

For many years, continuous time models have enjoyed a great deal of success in finance (Merton, 1990) as well as wide applications in economics (e.g., Dixit (1993)). Correspondingly, there has been growing interest in estimating continuous systems using econometric methods with discrete data.

Many models used in finance for modelling asset prices can be written in terms of a diffusion process as

$$\mathrm{d}X_t = \mu(X_t;\theta_1)\mathrm{d}t + \sigma(X_t;\theta_2)\mathrm{d}B_t, \quad (1)$$

where B_t is a standard Brownian motion, $\sigma(X_t;\theta_2)$ is a known diffusion function, $\mu(X_t;\theta_1)$ is a known drift function, and $\theta = (\theta_1,\theta_2)'$ is a vector of $k_1 + k_2$ unknown parameters. Note that we isolate the vector of parameters θ_2 in the diffusion function from θ_1 for reasons which will be clear below. The attractions of the Ito calculus make it easy to work with processes generated by diffusions like (1) and as a result these processes have been used widely in finance to model asset prices, including stock prices, interest rates, and exchange rates.

From an econometric standpoint, the estimation problem is to estimate θ from observed data, which are typically recorded discretely at $(\Delta, 2\Delta, \ldots, n_\Delta\Delta(\equiv T))$ over a certain time interval $[0, T]$, where Δ is the sampling interval and T is the time span of the data. For example, if X_t is recorded as the annualized interest rate and observed monthly (weekly or daily), we have $\Delta = 1/12$ $(1/52$ or $1/250)$. Typically, T can be as large as 50 for US Treasury Bills, but is generally much smaller for data from swap

☆ We thank two anonymous referees for their constructive comments. We also thank Yacine Aït-Sahalia, Yongmiao Hong, Chung-ming Kuan, Andy Lo, and Neil Shephard, and seminar participants at the First Symposium on Econometric Theory and Applications in Taiwan, the 2006 North American Winter Meeting of Econometric Society, and the First Finance Summer Camp at Singapore Management University for helpful discussions. Phillips gratefully acknowledges visiting support from the School of Economics at Singapore Management University, and support from a Kelly Fellowship at the University of Auckland Business School and from the NSF under Grant No. SES 04-142254. Yu gratefully acknowledges financial support from the Ministry of Education AcRF Tier 2 fund under Grant No. T206B4301-RS.

* Corresponding author.
E-mail addresses: peter.phillips@yale.edu (P.C.B. Phillips), yujun@smu.edu.sg (J. Yu).

0304-4076/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.jeconom.2008.12.006

markets. Also, due to time-of-day effects and possibly other market microstructure frictions, it is commonly believed that intra-day data do not completely follow diffusion models such as (1). As a result, daily and lower frequencies are most frequently used to estimate continuous time models. However, Barndorff-Nielsen and Shephard (2002) and Bollerslev and Zhou (2002) recently showed how to use information from intra-day data to estimate continuous time stochastic volatility models.

A large class of estimation methods is based on the likelihood function derived from the transition probability density of discrete sampling and then resorts to long span asymptotic theory (i.e. $T \to \infty$). Except for a few cases, the transition probability density does not have a closed form expression and hence the exact maximum likelihood (ML) method based on the likelihood function for the discretely sampled data is not directly available. In the financial econometrics literature, interest in obtaining estimators which approximate or approach ML estimators has been growing, in view of the natural attractiveness of maximum likelihood and its asymptotic properties. Several alternative methods of this type have been developed in recent years. See Phillips and Yu (in press-a) for a survey of various alternative methods and the discussion of their advantages and drawbacks.

The main purpose of the present paper is to propose an alternative method of estimating diffusion processes of the form given by model (1) from discrete observations and to establish asymptotic properties by resorting to both the long span (i.e. $T \to \infty$) and in-fill asymptotics (i.e. $\Delta \to 0$). The estimation procedure involves two steps. In the first step, we propose to use a quadratic variation type estimator of θ_2. In the second step, an approximate in-fill likelihood function is maximized to obtain a ML estimator of θ_1. This method has several advantages over the existing method. First, it is not dependent on finding an appropriate auxiliary model. Second, it does not require simulations or polynomial expansions and hence is straightforward to implement. Third, it decomposes the optimization problem into two smaller scale optimization problems, making the approach computationally more attractive. Finally, experience with the procedure both in simulations and empirical applications indicates that the method works well in finite samples.

The paper is organized as follows: Section 2 reviews the literature on the ML estimation of diffusion processes and motivates the approach. Section 3 introduces the new method and Section 4 derives the asymptotic properties of the estimates. Section 5 presents some Monte Carlo evidence. Section 6 discusses the case of microstructure noises and Section 6 concludes. Proofs are provided in the Appendix.

2. Literature review and motivation

2.1. Literature review

2.1.1. Transition probability density based approaches

As explained above, a large class of estimation methods is based on the likelihood function derived from the transition probability density of the discretely sampled data. Suppose $p(X_{i\Delta}|X_{(i-1)\Delta}, \theta)$ is the transition probability density. The Markov property of model (1) implies the following log-likelihood function for the discrete sample

$$\ell_{TD}(\theta) = \sum_{i=2}^{n_\Delta} \log(p(X_{i\Delta}|X_{(i-1)\Delta}, \theta)). \tag{2}$$

Under regularity conditions, the resulting estimator is consistent, asymptotically normally distributed and asymptotically efficient (Billingsley, 1961). Unfortunately, except for a few cases, the transition density does not have a closed form expression and hence the exact ML method based on the likelihood function of the discrete sample is not a practical procedure. In the financial econometrics literature, interest in finding estimators that approach ML estimators in some quantifiable sense has been growing and many alternative methods have been developed in recent years. For example, Lo (1988) suggested calculating the transition probability density by solving a partial differential equation numerically. Nowman (1997) suggested an approach which assumes that the conditional volatility remains unchanged over the unit intervals so that he can approximate the transition density using a Gaussian method. Yu and Phillips (2001) used the stoping time technique to develop an exact Gaussian method. Pedersen (1995) and Brandt and Santa-Clara (2002) advocated an approach which calculates the transition probability density using simulation with some auxiliary points between each pair of consecutive observations introduced. This method is also closely related to the Bayesian MCMC method proposed by Elerian et al. (2001) and Eraker (2001). A drawback of these simulation-based approaches is that the corresponding computational cost will inevitably be high.

As an important alternative to these numerical and simulated ML methods, Aït-Sahalia (2002) proposed to approximate the transition probability density of diffusions using analytical expansions via Hermite polynomials. Before obtaining the closed-form expansions, a Lamperti transform is performed on the continuous time model so that the diffusion function becomes a constant. After that one then obtains a Hermite polynomial expansion of the transition density of the transformed variable around the normal distribution. Aït-Sahalia (1999) implemented the approximate ML method and documents its good performance. As it is typically tedious and error prone to derive the Hermite expansion by hand, Aït-Sahalia (2002) suggested using symbolic softwares, such as MATHEMATICA.

Apart from these likelihood-based approaches, numerous alternative methods are available. We simply refer readers to the book by Prakasa Rao (1999a) for a review of many alternative approaches.

2.1.2. Approaches based on realized volatility and in-fill likelihood

When the transition probability density does not have a closed form expression but X_t is observed continuously over $[0, T]$, an alternative method can be used to estimate diffusion models. We now introduce and motivate the approach.

When the diffusion term is known (i.e. $\sigma(X_t; \theta_2) = \sigma(X_t)$) and so does not depend on any unknown parameters, one can construct the exact continuous record log-likelihood via the Girsanov theorem (e.g., (Liptser and Shiryaev, 2000) as follows.

$$\ell_{IF}(\theta_1) = \int_0^T \frac{\mu(X_t; \theta_1)}{\sigma^2(X_t)} \mathrm{d}X_t - \frac{1}{2}\int_0^T \frac{\mu^2(X_t; \theta_1)}{\sigma^2(X_t)} \mathrm{d}t.$$

Lánska (1979) established the consistency and asymptotic normality of the continuous record ML estimator of θ_1 when $T \to \infty$ under a certain set of regularity conditions.

The assumptions of a known diffusion function and the availability of a continuous time record are not realistic in financial and other applications. Motivated by the fact that the drift and diffusion functions are of different orders (Bandi and Phillips, 2003, 2007), we argue that there can be advantages to estimating the diffusion parameters separately from the drift parameters. For example, when $\sigma(X_t; \theta_2) = \theta_2$, i.e., the diffusion function is an unknown constant, a two-stage approach can be used to estimate the model. First, θ_2 can be estimated directly by the realized volatility function, i.e.,

$$\hat{\theta}_2 = \sqrt{\frac{[X_\Delta]_T}{T}}, \tag{3}$$

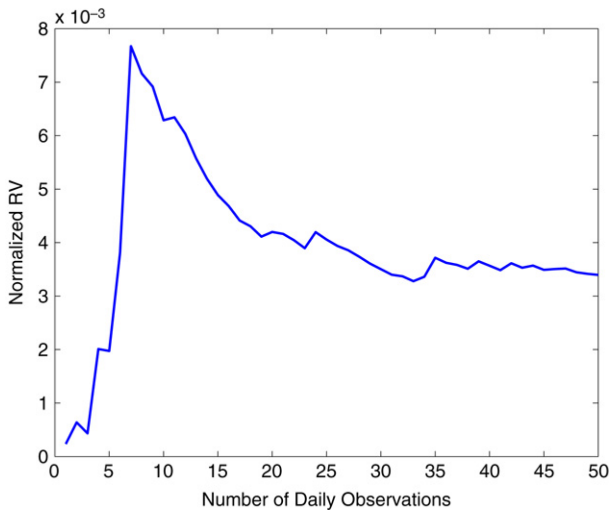

Fig. 1. Standardized realized volatility against the number of daily observations used to calculate the realized volatility. The daily data are simulated from the Vasicek model $\mathrm{d}X_t = 0.6(0.09 - X_t)\mathrm{d}t + 0.06\mathrm{d}B_t$.

where $[X_\Delta]_T = \sum_{i=2}^{n_\Delta}(X_{i\Delta} - X_{(i-1)\Delta})^2$. This is because model (1) implies that

$$(\mathrm{d}X_t)^2 = \theta_2^2 \mathrm{d}t, \quad \forall t,$$

and hence

$$[X]_T = \int_0^T (\mathrm{d}X_t)^2 = \int_0^T \theta_2^2 \mathrm{d}t = T\theta_2^2,$$

where $[X]_T$ is the quadratic variation of X, which can be consistently estimated by $[X_\Delta]_T$ as $\Delta \to 0$. As a result, $\hat{\theta}_2$ should be a very good estimate of θ_2 when Δ is small, which is typically the case for interest rate data. Indeed, in the special case where the drift term is zero and the diffusion term is an unknown constant, the exact discrete model (Phillips, 1972) for the data is $X_{i\Delta} - X_{(i-1)\Delta} = \theta_2\left(B_{i\Delta} - B_{(i-1)\Delta}\right)$ and so the maximum likelihood estimator is trivially $\hat{\theta}_2$ (see also Aït-Sahalia et al. (2005)). Although this correspondence clearly does not apply to more general specifications, it seems likely that when Δ is small the two estimators will be close to each other. Second, the following logarithmic continuous record likelihood function of model (1),

$$\ell_{IF}(\theta_1) = \int_0^T \frac{\mu(X_t;\theta_1)}{\sigma^2(X_t;\hat{\theta}_2)}\mathrm{d}X_t - \frac{1}{2}\int_0^T \frac{\mu^2(X_t;\theta_1)}{\sigma^2(X_t;\hat{\theta}_2)}\mathrm{d}t,$$

may be approximated by the in-fill likelihood function

$$\ell_{AIF}(\theta_1) = \sum_{i=2}^{n_\Delta} \frac{\mu(X_{(i-1)\Delta};\theta_1)}{\hat{\theta}_2^2}(X_{i\Delta} - X_{(i-1)\Delta}) - \frac{\Delta}{2}\sum_{i=2}^{n_\Delta}\frac{\mu^2(X_{(i-1)\Delta};\theta_1)}{\hat{\theta}_2^2}, \tag{4}$$

which is in turn maximized with respect to θ_1. Because θ_1 is estimated by ML and $\widehat{\theta_2}$ is close to an MLE, the two-stage procedure may be interpreted as a form of profile ML estimation. This two-stage approach is closely related to the method proposed by Florens-Zmirou (1989), where a contrast function instead of the logarithmic in-fill likelihood function was used in the second step.

To better appreciate the quality of approximation of $[X_\Delta]_T$ to $[X]_T$, we simulate daily data from the Vasicek model

$$\mathrm{d}X_t = 0.6(0.09 - X_t)\mathrm{d}t + 0.06\mathrm{d}B_t, \tag{5}$$

Fig. 2. Standardized realized volatility against the number of daily observations used to calculate the realized volatility. The daily data are simulated from the Vasicek model $\mathrm{d}X_t = 0.6(0.09 - X_t)\mathrm{d}t + 0.12\mathrm{d}B_t$.

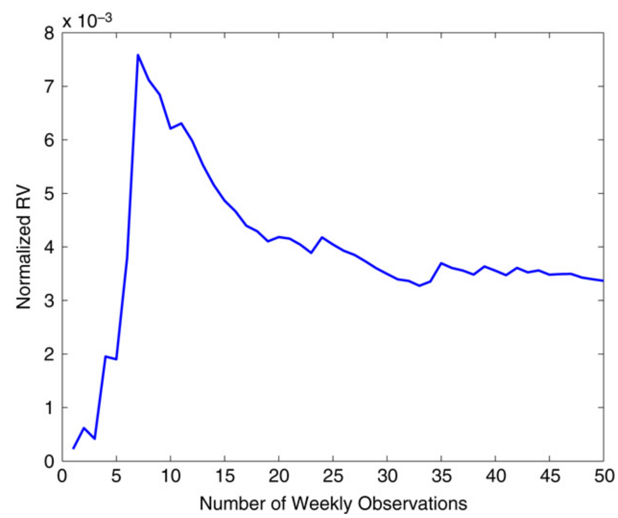

Fig. 3. Standardized realized volatility against the number of weekly observations used to calculate the realized volatility. The weekly data are simulated from the Vasicek model $\mathrm{d}X_t = 0.6(0.09 - X_t)\mathrm{d}t + 0.06\mathrm{d}B_t$.

and plot the standardized realized volatility, $[X_\Delta]_T/T$, against the number of daily observations used to calculate $[X_\Delta]_T$. Quadratic variation theory implies that as $\Delta \to 0$, $[X_\Delta]_T/T \overset{a.s.}{\to} \sigma^2$. It is clear from Fig. 1 that although $[X_\Delta]_T/T$ is quite erratic initially but quickly settles down around σ^2. It seems that with only 30–35 observations one can get a good estimate of σ^2. We then increase the volatility rate from 0.06 to 0.12 and decrease the sampling frequency from daily to weekly. Results are plotted in Figs. 2 and 3, respectively. It clear that the conclusion about the rapid convergence of $[X_\Delta]_T/T$ is quite robust to these changes.

When the diffusion term is only known up to a scalar factor, that is when

$$\mathrm{d}X_t = \mu(X_t;\theta_1)\mathrm{d}t + \theta_2 f(X_t)\mathrm{d}B_t, \tag{6}$$

the above two-stage method is easily modified. First, θ_2^2 can be estimated by

$$\hat{\theta}_2^2 = \frac{[X_\Delta]_T}{\Delta\sum_{i=2}^{n_\Delta} f^2(X_{(i-1)\Delta})}. \tag{7}$$