ARTICLE IN PRESS

Journal of Econometrics ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Journal of Econometrics

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

Estimating spot volatility with high-frequency financial data

Yang Zu [a,*], H. Peter Boswijk [b,1]

[a] Department of Economics, City University London, Northampton Square, EC1V 0HB, London, United Kingdom

[b] Amsterdam School of Economics, University of Amsterdam, Valckenierstraat 65-67, 1018 XE Amsterdam, The Netherlands

ARTICLE INFO

Article history:
Received 14 January 2014
Received in revised form
14 January 2014
Accepted 19 January 2014
Available online xxxx

JEL classification:
C58
C13
C14

Keywords:
Spot volatility
Market microstructure noise
Subsampling
Scale selection
Bandwidth selection

ABSTRACT

We construct a spot volatility estimator for high-frequency financial data which contain market microstructure noise. We prove consistency and derive the asymptotic distribution of the estimator. A data-driven method is proposed to select the scale parameter and the bandwidth parameter in the estimator. In Monte Carlo simulations, we compare the finite sample performance of our estimator with some existing estimators. Empirical examples are given to illustrate the potential applications of the estimator.

1. Introduction

Spot volatility, also known as *instantaneous volatility*, measures the strength of return variations at a certain time point, expressed per unit of time (Andersen et al. (2010)). Spot volatility has important applications in studying the intraday patterns of the volatility process, testing price jumps (Lee and Mykland (2007), Veraart (2010)), and estimating parametric stochastic volatility models (Bandi and Reno (2009), Kanaya and Kristensen (2010)). In this paper, we are interested in the nonparametric estimation of *spot volatility* with high-frequency financial data.

Spot volatility estimation in the literature dates back to Merton (1980), who considered a constant volatility model. Later on, researchers tended to estimate volatility in the context of the ARCH model (Engle (1982)), the GARCH model (Bollerslev (1986)), and their numerous variations. Nonparametric estimation of spot volatility in the context of diffusion models was firstly considered by Foster and Nelson (1996). Andreou and Ghysels (2002) conducted simulation studies using Foster and Nelson's estimator and some related estimators. Recent contributions include Mykland and Zhang (2008), Fan and Wang (2008) and Kristensen (2010).

High-frequency financial data have become more accessible for academic research in recent years. In contrast to low frequency (daily, weekly or longer sampling frequency) financial datasets, high-frequency datasets are characterized by the large number of observations they contain and the existence of so-called market microstructure noise. O'Hara (1998) made theoretical studies of market microstructure noise; Andersen et al. (2000) and Hansen and Lunde (2006) analyzed the empirical characteristics of the noise.

Existing research on volatility measurement for high-frequency data focuses mainly on the *ex post* nonparametric measurement of the *integrated volatility* of the underlying efficient price process. Andersen et al. (2001) and Barndorff-Nielsen and Shephard (2002) made important early contributions to the use of *realized variance* to estimate the integrated volatility. However, they did not consider the effects of market microstructure noise, and the realized variance estimator can only be applied to sparsely sampled data, where the effects of noise are small. The problem of estimating integrated volatility under noise was first studied by Zhou (1996), who gave an unbiased but inconsistent estimator for integrated volatility. Aït-Sahalia et al. (2005) considered a constant variance model and gave a Maximum Likelihood Estimator for the constant variance. Later, four types of estimators were proposed

* Corresponding author. Tel.: +44 0 20 7040 8619; fax: +44 0 20 7040 8580.
E-mail addresses: Yang.Zu@city.ac.uk (Y. Zu), H.P.Boswijk@uva.nl (H. Peter Boswijk).
[1] Tel.: +31 0 20 525 4316.

http://dx.doi.org/10.1016/j.jeconom.2014.04.001
0304-4076/

for estimating integrated volatility in the presence of noise. These are the subsampling-based Two Scale Realized Variance (TSRV) estimator by Zhang et al. (2005) and the Multiscale Realized Variance (MSRV) estimator by Zhang (2006); the Realized Kernel (RK) estimator by Barndorff-Nielsen et al. (2008), which is based on Zhou (1996)'s first order moving average correction; the pre-averaging method by Podolskij and Vetter (2009), and Jacod et al. (2009); and the Quasi-Maximum Likelihood Estimator (QMLE) by Xiu (2010), which is based on the estimator in Aït-Sahalia et al. (2005).

In this paper, we study the problem of estimating spot volatility with high-frequency data and we explicitly consider the effects of market microstructure noise. Our approach is closely related to the literature on integrated volatility measurement with noise. We construct our estimator based on the Two Scale Realized Variance estimator by Zhang et al. (2005) — our estimator calculates the increment of the Two Scale Realized Variance estimator over a small interval and applies an appropriate normalization. Under appropriate conditions, we prove consistency and derive the asymptotic distribution of our estimator and propose a data-driven procedure to select tuning parameters. In practically meaningful Monte Carlo simulations, we compare our estimator with existing methods in terms of several error measures and we demonstrate the improved accuracy in using our estimator.

Some recent research is closely related to this paper. Mykland and Zhang (2008) independently proposed the same estimator as in our paper, but did not provide a complete asymptotic theory. Bandi and Reno (2009), Ogawa and Sanfelici (2011), Bos et al. (2012), among others, have considered spot volatility estimators based on the Realized Kernel estimator and the Pre-Averaging estimator. In a concurrent paper, Mancini et al. (2012) (Section 3.1) have proposed a two-scale estimator for spot volatility weighted by the so-called delta sequence and have provided theoretical analysis; our estimator is a special case of their estimator with equal weights. We provide more comprehensive asymptotic and finite sample studies for our estimator; we also study the problem of bandwidth and scale parameters selection, which is important for practical implementation. In the presence of jumps but without noise, spot volatility has been studied by Aït-Sahalia and Jacod (2009), Ngo and Ogawa (2009) and Andersen et al. (2009). Munk and Schmidt-Hieber (2010b,a) studied the best possible convergence rate of any spot volatility estimator in a volatility model observed with noise, where the volatility process is assumed to be a deterministic function. Hoffmann et al. (2010) derived a minimax bound for the same problem in a genuine stochastic volatility model observed with noise, they showed that this bound is "nearly optimal" in their definition and they proposed a wavelet estimator that achieves this rate. Their rate is $n^{-1/8}$ if translated to the present context, up to some logarithmic corrections. Our estimator does not have the best rate of convergence in their sense, we discuss possible extensions to improve the convergence rate in Section 6.

The structure of this paper is as follows. Section 2 introduces the setup of the problem. Section 3 defines the estimator, studies its asymptotic properties and the problem of bandwidth and scale selection. Section 4 conducts Monte Carlo studies on the finite-sample properties of the estimator and Section 5 contains two empirical applications to Euro FX futures data. Section 6 discusses possible extensions to our model. Section 7 concludes the paper. Proofs are collected in Appendix A and technical lemmas are collected in Appendix B.

Throughout the paper, $\langle X, Y\rangle$ denotes the quadratic covariation of two processes X and Y; $\xrightarrow{d}$ denotes converge in distribution; $\xrightarrow{st}$ denotes stable convergence in distribution; $\xrightarrow{p}$ denotes converge in probability; for a real number x, $\lfloor x\rfloor$ denotes its integer part. We call σ_t^2 the *spot variance* at time t, and we call σ_t the *spot volatility* at time t. However, as in the financial econometrics literature, when we use the term *spot volatility* in general discussions, it could refer to either σ_t^2 or σ_t, depending on the context.

2. The model

Let $\{X_t\}$ be a univariate log price process, assumed to be a Brownian semimartingale, satisfying

$$\mathrm{d}X_t = \mu_t \mathrm{d}t + \sigma_t \mathrm{d}W_t, \quad t \in [0, 1],$$

where $\{W_t\}$ is a standard Brownian motion; $\{\mu_t\}$ is the spot drift process, and $\{\sigma_t\}$ is the spot volatility process; both are predictable. We further assume:

A1 the processes $\{\mu_t\}$ and $\{\sigma_t\}$ have continuous sample paths.
A2 the process $\{\sigma_t\}$ is positive.

Since X is a Brownian semimartingale, it has continuous sample paths, and its quadratic variation process satisfies

$$\langle X, X\rangle_t = \int_0^t \sigma_s^2 \mathrm{d}s, \quad t \in [0, 1],$$

such that the spot volatility satisfies

$$\sigma_t^2 = \frac{\mathrm{d}\,\langle X, X\rangle_t}{\mathrm{d}t}. \tag{1}$$

Taking into account the market microstructure noise existing in high-frequency financial data, we further assume

A3 X is not observable, but

$$Y_t = X_t + \varepsilon_t$$

is observed over the interval $[0, 1]$ in discrete time over a grid $t_i = i/n$ for $i = 0, 1, \ldots, n$ with equal distance $\Delta_n = 1/n$.

A4 $\{\varepsilon_{t_i}\}_{i=1}^n$ are independent and identically distributed (i.i.d.) with mean 0, variance ω^2, and with finite fourth moment. Furthermore, $\{\varepsilon_{t_i}\}_{i=1}^n$ are independent of the $\{X_t\}$ process.

The continuity assumption on the volatility sample paths accommodates a large class of spot variance processes such as diffusion processes, long memory, deterministic patterns as well as nonstationarity. The model allows for possible dependence between $\{W_t\}$ and $\{\sigma_t\}$, so leverage effects are allowed in this model.

The model specification and Assumption A1 exclude the possibility of jumps in both the price process and the volatility process. Assumption A4 excludes the possibility that the noise is dependent over time (so called dependent noise) and that the noise is dependent of the efficient price process (so-called endogenous noise). We discuss possible extensions to these cases in Section 6.

3. The estimator and its properties

3.1. The estimator

We are interested in estimating the realization of the spot variance process $\left\{\sigma_t^2\right\}$ at any time $t \in (0, 1)$. Our estimator is based on the Two Scale Realized Variance estimator (TSRV) by Zhang et al. (2005).

The TSRV estimator uses a subsampled and averaged Realized Variance (RV) estimator over a scale K, together with a usual Realized Variance estimator to correct the effects of noise. It is defined as

$$\mathrm{TSRV} = [Y, Y]^K - \frac{\bar{n}}{n}[Y, Y],$$