International Review of Financial Analysis 28 (2013) 34–45

Contents lists available at SciVerse ScienceDirect

International Review of Financial Analysis

Nonparametric realized volatility estimation in the international equity markets

Dimitrios I. Vortelinos [a,*], Dimitrios D. Thomakos [b,1]

[a] Lincoln Business School, University of Lincoln, LN6 7TS, UK
[b] School of Management and Economics, University of Peloponnese, 22 100, Greece

ARTICLE INFO

Article history:
Received 26 June 2012
Accepted 10 February 2013
Available online 26 February 2013

Keywords:
Realized volatility
Optimal sampling
Kernels
Two scales
Moving average
Long memory
Jumps
Heterogeneous autoregressive models

ABSTRACT

Using high-frequency intraday data, we construct, test and model seven new realized volatility estimators for six international equity indices. We detect jumps in these estimators, construct the jump components of volatility and perform various tests on their properties. Then we use the class of heterogeneous autoregressive (HAR) models for assessing the relevant effects of jumps on volatility. Our results expand and complement the previous literature on the nonparametric realized volatility estimation in terms of volatility jumps being examined and modeled for the international equity market, using such a variety of new realized volatility estimators. The selection of realized volatility estimator greatly affects jump detection, magnitude and modeling. The properties each volatility estimator tries to incorporate affect the detection, magnitude and properties of jumps. These volatility-estimation and jump properties are also evident in jump modeling based on statistical and economic terms.

1. Introduction

Quadratic variation is aimed by the volatility estimators, because it is the type of variation that is most close to the “true”-latent volatility. References for quadratic variation go back to Wiener (1924) and Levy (1937), as a development in semimartingales analysis. Rosenberg (1972b), Officer (1973), as well as the latest Merton (1980), and Schwert (1998) define and refine the elements of realized quadratic volatility. Consistently estimating the quadratic variation and integrated covariance, Back (1991) and Protter (2004) are the right references for the economics and the mathematics of semimartingales respectively. The stochastic integral of the adapted caglad process, with respect to the cadlag semimartingale, in the unique quadratic variation process of any semimartingale, is well defined by Protter (2004). Also, Barndorff-Nielsen and Shephard (2001) and Andersen, Bollerslev, Diebold, and Labys (2001a) are the ones that first connect the quadratic variation with realized quadratic volatility from the financial perspective. The notional (i.e. expected in a continuous time frame) volatility concept, for diffusions with and without jumps, has recently been highlighted in a series of papers by Andersen et al. (2001a), Andersen, Bollerslev, Diebold, and Labys (2001b) and Barndorff-Nielsen and Shephard (2002a, 2002b, 2002c).

Realized volatility has a long history, and is the mostly used nonparametric volatility estimator, after the introduction of high-frequency data in academics. It appears in Rosenberg (1972b), Merton (1980) and French, Schwert, and Stambaugh (1987), with Merton (1980) making the implicit connection of realized volatility with the pure scaled Brownian motion plus drift case. Poterba and Summers (1986), French et al. (1987) and Schwert (1989) were the first relying on monthly sample variances computed from daily returns and Hsieh (1989), Schwert (1990a, 1990b), Hsieh (1991) and Taylor and Xu (1997) exploit intraday data to compute daily sample return variance measures, with high-frequency (intraday) data. In order to estimate volatility, what we have to estimate is integrated volatility. Under the natural estimate of integrated volatility (the sum of squared returns), assuming a standard Brownian motion as price process, realized volatility estimate comes more close to the true IV, as the sampling frequency increases. Though, as sampling frequency increases, the microstructure noise also increases.

Realized variation converges uniformly in probability to the quadratic variation process as the sampling interval shrinks —as shown in Andersen and Bollerslev (1998) and Barndorff-Nielsen and Shephard (2002b, 2002c). When the price path displays discontinuities, due to jumps, in the general semimartingale setting, the quadratic variation is no longer identical to the integrated volatility, because it should also include the cumulative squared jumps. The main explanation for this phenomenon is a vast array of issues collectively known as market microstructure noise, including – but not limited to – the existence of the bid-ask spread. Market microstructure noise induces autocorrelation in the intraday returns, imposing bias in the realized volatility, because of its assumption of no autocorrelation among intraday returns.

In this paper we empirically investigate the presence, properties and effects of jumps on the volatility of two most important European

* Corresponding author. Tel.: +44 1522 835634.
E-mail addresses: dvortelinos@lincoln.ac.uk (D.I. Vortelinos), thomakos@uop.gr (D.D. Thomakos).
[1] Tel.: +30 2710 230128.

1057-5219/$ – see front matter
http://dx.doi.org/10.1016/j.irfa.2013.02.010

indices and four most important U.S. American indices. Using intraday data we first construct several realized volatility estimators and then use these estimators as inputs in further analysis: in testing for jumps, in examining the properties of the resulting jump series and in modeling the jump component of volatility. We employ the methodology of Andersen, Bollerslev, and Diebold (2007), ABD hereafter, in testing for and modeling jumps. Their work, in turn, builds on earlier results of Barndorff-Nielsen and Shephard (2004, 2006, 2007). Of particular interest here are certain statistics of the jump component of volatility, such as the temporal dependence and duration of jumps. Then, we employ the class of heterogeneous autoregressive (HAR) models for assessing the relevant effects of jumps on volatility. In this type of models we can also disentangle the differential effect that the jump and continuous components have on volatility. The HAR class of models was introduced by Corsi (2009 — the working paper was available from 2004) and see also Corsi, Pirino, and Reno (2008) and Andersen et al. (2007). The analysis is performed for six important international equity market indices, the CAC40, DAX, COMPX, NDX, SPX and DJI indices, that differ on capitalization and other characteristics, in an attempt to see what type of effects do volatility jumps have on different markets. This is the first time, to the best of our knowledge, that this type of empirical analysis is performed for the realized volatility estimators studied.

The first trial to correct realized volatility for the market microstructure noise (i.e. jumps) was from Ebens (1999) and Bollen and Inder (2002), who have used either the moving average or the autoregressive model to whiten the intraday returns, before computing the realized volatility. The realized volatility estimators – studied in the present paper – comprehend various corrections for microstructure noise. Hansen, Large, and Lunde (2005) denote a univariate and a multivariate version of a moving average based realized volatility estimator. This estimator uses a first-order moving average regression — symbolized as $RV_t^{(ma.\ adj1)}$. Another moving average based realized volatility estimator is also used; in specific, this estimator uses the residuals of a second-order moving average regression and is symbolized as $RV_t^{(ma.\ adj2)}$. They also propose an adjustment for all the cases, in order to capture the long-run variance of the "whiten" returns. After finding this residual series, they sum the squared residual series and multiply this sum with a numerical function of the MA coefficient. Bandi, Russell, and Yang (2008) propose different types of realized kernels. They define the Bartlett kernel estimator ($RV_t^{(bar)}$) in Barndorff-Nielsen, Hansen, Lunde, and Shephard (2006, 2008) with a number of the optimal number of return autocovariances defined by Bandi and Russell (2005). They also define the two scale realized volatility estimator of Zhang, Mykland, and Ait-Sahalia (2005) with an optimal sampling frequency ($RV_t^{(m_{opt})}$), estimating this frequency as the signal-to-noise ratio, like in Bandi and Russell (2005). They also define the two-scale realized volatility estimator of Zhang et al. (2005) with the optimal number of subsamples ($RV_t^{(TS,fs)}$), coming from Zhang et al. (2005). Bandi et al. (2008) also define the two-scale realized volatility estimator of Zhang et al. (2005) with the number of subsamples estimated by the Bartlett-type kernel weights ($RV_t^{(TS)}$), like the optimal number of return autocovariances in Bandi and Russell (2005). Bandi et al. (2008) also define an adjustment of the above two-scale estimator as its downward bias corrected version ($RV_t^{(TS,BC)}$). The number of subsamples is defined like in estimator $RV_t^{(TS)}$.

The rest of the paper is organized as follows. In Section 2 we have an overview of our empirical methodology; in Section 3 we present the data used in the paper; in Section 4 we discuss our results and in Section 5 we offer some concluding remarks.

2. Empirical methodology

2.1. Volatility estimators

There is an active literature on the many available realized volatility estimators, from the original "naive" estimator to advanced estimators that correct for the presence of market microstructure effects, like microstructure noise which can be correlated with the underlying high-frequency prices. We cannot possibly review this line of the literature here so we next present the estimators that we use in this paper along with their main references from the literature.

To begin with assume that for each day, with time confined in the unit interval [0,1], the observed intraday logarithmic asset prices follow the noise contaminated process:

$$p_{t_i,m} = p^*_{t_i,m} + u_{t_i} \tag{1}$$

where p denotes the observed logarithmic price, p^* denotes the unobservable equilibrium logarithmic price and u denotes the unobservable market microstructure noise.[2] The time index t_i represents the i^{th} observation in the $m+1$ intraday observations with a sampling frequency equal to $1/m$. Denote by $r_{i,m} = p_{t_i,m} - p_{t_i-1,m}$ and $r^*_{i,m} = p^*_{t_i,m} - p^*_{t_i-1,m}$ the corresponding intraday returns (at the highest frequency of observation) and by $e_{i,m} = u_{t_i} - u_{t_i-1}$ the difference of the noise component and assume that the equilibrium price evolves as a function of a stochastic volatility process as:

$$p^*_{t_i} \underline{\underline{\mathrm{def}}} \int_0^{t_i} \sigma_s dW_s + j_{t_i} \tag{2}$$

where σ_t is a stochastic volatility process and W_t is standard Brownian motion and where j_{t_i} denotes the component that will appear in the price process in the case there are discrete jumps. The integrated volatility over the whole day is then given by:

$$V_t \underline{\underline{\mathrm{def}}} \int_0^1 \sigma_s^2 ds + \lambda_t \tag{3}$$

where $\lambda_t = \sum_{0<s\leq 1} \kappa_s^2$ is the contribution of the jumps into the volatility, with κ_s denoting the size of the discrete jumps.

To present our realized volatility estimators consider first the case where $j_{t_i} = q_t = 0$ so that there are no jumps present. Then, a consistent estimator for volatility, as $m \to \infty$, is given by the sum of intraday squared equilibrium returns as:

$$RV_{t^*}^{(m)} \underline{\underline{\mathrm{def}}} \sum_{i=1}^{m} r^{*2}_{i,m} \to V_t. \tag{4}$$

However, $r^{*2}_{i,m}$ is latent and thus the above estimator cannot be implemented. The obvious alternative is to use the sum of intraday squared observable returns but this alternative is not robust to the presence of microstructure noise (leading to various inconsistencies). One has therefore to consider various other estimators. We begin with the naive benchmark, the realized 5-minute estimator.

Hsieh (1989, 1991) and Schwert (1990a, 1990b) were the first to consider the use of high-frequency data for computing an early version of realized volatility estimators. However, the currently accepted prototype of a realized volatility estimator comes from the work of Andersen et al. (2001a), ABDL hereafter, and is simply the sum of the observable intraday squared returns:

$$RV_t^{(m)} \underline{\underline{\mathrm{def}}} \sum_{i=1}^{m} r^2_{i,m}. \tag{5}$$

In the absence of noise, this estimator is a consistent estimator of V_t as the sampling frequency increases. However, given the existence

[2] The microstructure noise variable is treated under various probabilistic assumptions in the literature, the easiest of which is that it is an i.i.d. sequence.