Journal of Banking & Finance 35 (2011) 3319–3334

Contents lists available at ScienceDirect

Journal of Banking & Finance

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

Volatility and covariation of financial assets: A high-frequency analysis

Álvaro Cartea [a], Dimitrios Karyampas [b,*]

[a] Department of Business Administration, Universidad Carlos III de Madrid, Spain
[b] Department of Economics, Mathematics and Statistics, Birkbeck, University of London, UK

ARTICLE INFO

Article history:
Received 7 October 2010
Accepted 15 May 2011
Available online 20 May 2011

JEL classification:
G12
G14
C22

Keywords:
Volatility estimation
High-frequency data
Market microstructure noise
Covariation of assets
Kalman filter

ABSTRACT

Using high frequency data for the price dynamics of equities we measure the impact that market microstructure noise has on estimates of the: (i) volatility of returns; and (ii) variance–covariance matrix of n assets. We propose a Kalman-filter-based methodology that allows us to deconstruct price series into the true efficient price and the microstructure noise. This approach allows us to employ volatility estimators that achieve very low Root Mean Squared Errors (RMSEs) compared to other estimators that have been proposed to deal with market microstructure noise at high frequencies. Furthermore, this price series decomposition allows us to estimate the variance covariance matrix of n assets in a more efficient way than the methods so far proposed in the literature. We illustrate our results by calculating how microstructure noise affects portfolio decisions and calculations of the equity beta in a CAPM setting.

1. Introduction

Volatility of asset returns is one of the most important variables in finance. It is an important "building block" in many areas including portfolio and risk management, investment appraisal and derivatives pricing. In general, measuring volatility is not a straightforward task because it is not directly observable from the data. Focusing on financially traded assets, the answer to the question of how to estimate volatility, or quantities that use volatility as an input, will inevitably depend on modeling assumptions and compromises that will have an effect on both its estimate and estimator. For example, some of the difficulties that arise when measuring volatility stem from assumptions such as the model driving the asset price dynamics or practical issues such as the frequency or amount of data that should be employed in the estimation.

Model assumptions for asset price dynamics and the choice of data employed in the estimation of volatility are generally not independent modeling decisions. Until now, the literature has developed a large number and diverse range of volatility models that are applied in different contexts and to various financial applications. However, most of these models have been developed and tested by employing data sets that make use of a very small subset of the complete sample of available trades. In fact, the common approach has been to employ low frequency data, say one data point per trading day, when there could be thousands of intra-day observations.

Relying only on daily observations results in the discarding of a significant amount of information which in some asset classes, such as equity, can account for more than 99% of the available data. On the other hand, it is not clear whether employing as much data as possible will unequivocally improve the accuracy of the volatility estimates. The answer to the question of whether more data is preferred to less depends not only on quantity, but also on quality.

The highest resolution of stock price data is tick-by-tick data. It could be either a record of every trade or every trade and quote (including bid and ask). For a long time, the market microstructure literature has highlighted the difficulties arising from such high frequency data (see for instance Black (1986)). One of the key problems is a 'quality' issue since tick-by-tick data contains microstructure noise. In other words, tick-by-tick prices consist of the true or efficient price plus noise. Therefore, the approach of using all observations may lead to entirely different, and possibly misleading, results to those obtained if the high frequency data were to only contain the true price, i.e. no microstructure noise. Zhang et al. (2005) look into the question of how often it is optimal to

* Corresponding author.
E-mail addresses: alvaro.cartea@uc3m.es (Á. Cartea), d.karyampas@ems.bbk.ac.uk (D. Karyampas).

0378-4266/$ - see front matter
doi:10.1016/j.jbankfin.2011.05.012

sample a continuous-time diffusion process in the presence of microstructure noise. One of the objectives of their work is to look at the delicate balance between quality and quantity of the data. They show, under a set of assumptions governing the dynamics of share prices, that estimating the volatility of the true price process uses neither all available data nor a limited subset of trades such as the daily observations previously described.

In this paper we aim to provide estimators for the variance and covariance matrices of stock returns. The main obstacle we face is to identify estimators that are not tainted by market microstructure noise. To this end, we propose a Kalman filter-based approach where we not only obtain efficient volatility estimators, but also generate efficient covariances between several assets and high frequency returns that are not affected by the microstructure noise. The Kalman-based volatility estimator we propose is a best estimator in the Cramér–Rao criterion sense, with a RMSE as small as that of the Maximum Likelihood Estimator proposed by Aït-Sahalia et al. (2005). Moreover, our proposed estimator is better, according to the RMSE metric, than the non-parametric estimators that the current literature has proposed to deal with the presence of market microstructure noise. The Kalman filter approach also offers us the ability to generate de-noised series for each asset price, and by applying common estimators of covariance to the filtered series we can build a best unbiased estimator of the covariation matrix.[1]

We illustrate the significance of our results with two examples. First, we show that the efficient frontier and the optimal weights of a portfolio differ significantly when high frequency data, rather than daily data, are used. More precisely, we show that an efficient frontier will vary in location and shape according to whether it is calculated using daily data, high frequency data with microstructure noise, or high frequency data without microstructure noise. Therefore, for a given risk target, a portfolio manager would have three different portfolio mixes to choose from depending on the data set employed, when in theory only one of these choices is correct. For example, an efficient frontier calculated with high frequency data, with microstructure noise, exhibits greater levels of risk, per level of return, than those exhibited by a frontier calculated using daily observations or a frontier calculated using high frequency data where the microstructure noise has been filtered out. Moreover, we find that the filtered high frequency efficient frontier also differs from the daily one. Finally, our results also show that measurements of log-returns will be different depending on the frequency and quality of the data used. As an example, we calculate the log-returns of the Dow Jones Industrial Average (DJIA) index constituents and find that for the majority of stocks, daily log-returns are lower than high-frequency (filtered and unfiltered) log-returns. We find that the difference in returns is not negligible; on average daily log returns are around 2.5% lower than log-returns calculated with a high frequency data base where the microstructure noise has been filtered out.

In the second example, we compute the equity beta in the Capital Asset Pricing Model (CAPM). We show that assets' systematic and unsystematic risk are different when we take into account the additional information provided by high frequency data.[2] We calculate equity betas for the DJIA constituents over the period January 2005–December 2006. On average the equity betas are approximately the same whether we use daily or filtered high frequency data. However, when we look at every individual stock the differences in the equity beta estimates are frequently significant. For example, the equity beta for Eastman Kodak using daily data is 1.071, but when filtered high frequency is employed the equity beta becomes 0.61.

The rest of the paper is organized as follows. Section 2 describes the framework for stock dynamics which we use in our volatility and covariation analysis. It also summarizes the existing literature on estimating the volatility of assets and highlights the difficulties that arise due to the existence of market microstructure noise. Section 3 proposes a Kalman filter-based approach to: estimate the volatility of log-returns; construct the de-noised true path of the assets; and finally, estimate the variance–covariance matrix (VCM) between multiple assets. Section 4 considers how to consistently estimate the total covariation between a large number of financial assets. Section 5 discusses how decision making in portfolio theory and measurements of cost of capital components, such as the equity beta in the CAPM framework, are affected by the frequency of the data set employed in the calculations. Section 6 concludes.

2. Background

Engle (2000) mentions that "one measure of progress in empirical econometrics is the frequency of data use". Initially one would expect that the consistency of estimators is improved by the resolution of data employed, that is, the higher the frequency the better the consistency. This is one of the main reasons why the current literature has focused on the use of financial data at higher frequencies, for instance tick-by-tick data. The study of volatility of stock returns in a high frequency setting is perhaps the most active exponent of this line of research. Below, we summarize the existing methods currently proposed in the literature to measure realized volatility based on high frequency data for a diffusion model, which is the framework we will use in this paper.

Following Zhang et al. (2005), we assume that the log-price of a security follows a semimartingale process defined in $(\Omega, \mathcal{F}, \mathbb{P})$ given by

$$dX_t = \underbrace{\mu(X_t;\theta)dt}_{\text{drift component}} + \underbrace{\sigma dW_t}_{\text{diffusion component}}, \tag{1}$$

where $X_t = \ln S_t$ and S_t is the price of the security, $X_0 = 0$, $\mu(X_t;\theta)$ is the drift function with θ a drift parameter, $\sigma > 0$ is the diffusion coefficient and W_t is a standard Brownian motion under the probability measure $\mathbb{P}$. For our analysis, we assume that the object of interest σ is constant through each day and estimate it using high frequency data where it is possible to sample at different resolutions; the highest possible being tick-by-tick data. In our study the resolution is such that time-intervals between observations are short enough to make the drift component in (1) negligible. This is because the drift component $\mu(X_t;\theta)dt$ is of order dt, while the order of the diffusive component σdW_t is $dt^{1/2}$. Therefore, as $dt \to 0$ the drift term is much smaller than the diffusion. Hence we can assume that

$$X_t = \sigma W_t, \qquad t \in [0, T]. \tag{2}$$

Furthermore, we assume that the observations are equally spaced, so the time interval between them is constant and equal to Δ. The observations are recorded at times $t_i = i\Delta$ with $t_N = N\Delta = T$ for $i = 0,\ldots,N$.

The assumption that data are equally spaced is arguably not optimal if we consider actual tick-by-tick data. Generally, with financial data the time intervals between consecutive trades is

[1] There are other recent studies that also employ state-space models to extract the true efficient price from prices with microstructure noise, see for instance Menkveld et al. (2007), or to model the dynamics of the volatility skew Bedendo and Hodges (2009). For other studies that employ intraday data to calculate realized volatility see Chan and Fong (2006) and Ferland and Lalancette (2006).

[2] See Guo and Savickas (2010) for a study on cross-sectional effects of idiosyncratic variance on stock returns.