ABSTRACT

The availability of high frequency data in finance may lead to the project of getting almost direct observations of continuous time dynamics popular in mathematical finance. The reason why such a project is utopia is twofold. First, as pointed out by Merton (1980), high frequency data are of little use to identify the drift of a diffusion process. This is the reason why the main statistical use of high frequency data in finance is estimation of volatility. The second utopia is then to believe that we have a direct observation of the diffusion process (or more generally of the semimartingale) of interest for continuous time asset pricing theory based on the absence of arbitrage. Due to market frictions (bid-ask spreads, transactions costs, rounding errors, etc.) each observed price is a noisy observation of the underlying arbitrage-free semi-
martingale. While the so-called market microstructure noise has been to a large extent neglected as long as people were working with daily data or even intraday data with moderate frequency, it cannot be overlooked when working with high-frequency data.

The purpose of this talk is to review some approaches and some outputs of recent research to address this issue, while not throwing the baby with the bath water by neglecting the information content of high frequency data. It is worth reminding in this respect that there is always a possible naive approach to accommodate microstructure noise: just performing a sufficiently sparse sampling to be sure that noise can be neglected. Bandi and Russell (2007) go even further by looking for the “optimal sampling frequency” where the bias introduced by microstructure noise in the estimation of volatility is well balanced with the reduction in variance allowed by more frequent data in order to minimize the overall mean squared error.

There will be no such thing as an optimal sampling frequency in the three approaches presented in this talk. The golden rule is to use all the available data that is the highest possible frequency. It turns out that there are several ways to kill the noise, while computing objects similar to realized variance. While the three approaches presented here are all akin to some averaging allowing us to use the data at the highest frequency available in spite of the microstructure noise, the price to pay is twofold. First the effective number of data is infinitely smaller than what it would be without the need of averaging. Second, averaging raises the issue of edge effects that are the main reason why the three approaches may deliver different results. The fact that the effective number of observations may be smaller than the nominal one is a key motivation for considering not only volatility estimation on a given day from intraday data of this particular day but also taking advantage of volatility persistence between days. The fact that edge effects may matter is the key reason why microstructure noise makes even more challenging the estimation of instantaneous volatility than estimation of daily volatility.

To summarize, there is a tight connection between the fact that three approaches are currently available for estimating daily volatility and three horizons may be of interest: not only the daily one, but also taking advantage of across days information to improve estimators and also focusing on estimation of spot volatility. The survey part of this talk will be the recapitulation of the available approaches for estimating daily volatility with high frequency data. While focusing on the so-called three approaches all based on some kind of averaging, we will overlook some alternative approaches that will be only briefly quoted. The innovative part of the talk is about alternative horizons. We will first discuss the efficiency gains provided by a time series of daily measurements. Second, we will propose a general framework for comparing standard integrated variance estimates with estimates of instantaneous volatility. We will provide a representation which yields asymptotic distributions of estimators of instantaneous volatility from those of quadratic variation. We shall see that while the two types of estimators are closely connected, their asymptotic properties emphasize different aspects of their construction.