High-Frequency Financial Econometrics

Harley Thompson

Global Head of Market Risk Analytics, Zurich Investment Management, 8002 Zurich, Switzerland;
harley_thompson@hotmail.com

This book is fundamentally about the estimation of risk. At an intuitive level, risk is easy to understand: given an asset with a current price of say $100, what is the likelihood that at some future time—the risk horizon—its price will be less than $90? Or more than $120? And how does the probability of observing a price below $90 change as a function of the risk horizon?

Such heuristic notions can be formalized in several different ways. Suppose that the (log) price of the asset follows $X_t = X_0 + \sigma W_t$ where W_t is a standard Brownian motion and $\sigma > 0$ is a parameter quantifying the risk of X, which needs to be estimated. Suppose also that the price is observed at a sequence of fixed points in time, say at the end of each trading day. Then a simple measure of risk is the squared change in the observed price from one day to the next, possibly averaged over several adjacent intervals to obtain a smoothed estimate—this is, of course, the (daily) variance. As the number of intervals increases, and assuming the statistical properties of the underlying price process remain unchanged, this estimator eventually yields the true value of σ (or in fact its square). But asymptotic results can be considered in another sense: suppose that the total length of the observation interval is fixed (say one day) but the distance between observations within the interval becomes increasingly small—that is, one is in the realm of high-frequency data. Then another way to estimate risk on the interval is to sum the squared incremental price changes. This estimator is called the realized volatility, and it is this estimator, along with its variants and applications, that form the main focus of the book.

Formally, if the observation interval has length T and the distance between observations is Δ then the realized volatility is:

$$RV(\Delta, T) = \sum_{i=1}^{|T/\Delta|} \left(X_{i\Delta} - X_{(i-1)\Delta} \right)^2$$

with asymptotic behavior defined in terms of Δ tending to zero for fixed T.

In common with much time-series analysis, the maintained assumption throughout the book is that the observations (log prices) come from an underlying continuous-time stochastic process, in this case an Itô semi-martingale of the form:

$$X_t = X_0 + \int_0^t b_s ds + \int_0^t \sigma_s dW_s + \text{"jumps"}$$

(in several instances simpler cases are considered, for instance with volatilities held constant or jumps removed, in order to illustrate specific features of the problem). In this setting, the limiting process of realized volatility, called the integrated (squared) volatility, is in general random. The goal is to use the estimator, and extensions and generalizations thereof, to deduce properties of the underlying process. The kinds of questions that arise include:

- How can variation be decomposed to continuous ("volatility") and discontinuous ("jump") components?
- How can the estimator and generalizations of it be applied when the object of interest is solely the integrated volatility (the continuous part)?

- How can the estimators be adapted to ensure robustness to microstructure noise and irregular sampling?
- How can one test for the presence of jumps and their activity rate? And if the activity rate is infinite how can one decide if a continuous component required at all?

The material is presented in four parts. Part 1 reviews semi-martingale theory and describes related data considerations, for instance market data generating mechanisms and characteristics of high-frequency data. The second part introduces the basic high-frequency estimator—the realized volatility given above—and presents the relevant first and second-order asymptotic theory, which as noted above is based on the sampling frequency tending to zero on fixed observation intervals. Results are presented in a series of increasingly complex model settings, starting with continuous processes (with and without stochastic volatility) then considering processes with jumps. Particular issues raised by the introduction of jumps are demonstrated in the simple setting of a compound Poisson process plus Brownian motion, which then serves as an introduction to the notion of power variations. These important generalization of realized volatilities enable to obtain in certain circumstances non-trivial asymptotic results when the underlying process contains jumps. A discussion on identifiability and asymptotic efficiency then follows, the purpose of which is to detail econometric tools available for parameter estimation and hypothesis testing when quantities are fundamentally random, as is the case here. Part 3 focuses solely on the estimation of volatility in (possibly multidimensional) semi-martingale models, principally integrated volatility, but also its derivative, spot volatility. While this is generally a well-understood problem it is rendered non-trivial by the presence of jumps, and the authors discuss several estimation approaches that allow the volatility component to be isolated and consistently estimated; methods include truncation, multi-power variation, range-based estimators and Fourier-based approaches. Finite sample refinements for addressing issues around skews and heavy tails in the estimation error distributions are also covered. The discussion begins with the simplest case when observations are equidistant and observed without noise then progressively relaxes these assumptions. Again, asymptotic results are presented wherever possible. The final part of the book focuses specifically on jumps, which up to that point were present but treated as essentially a nuisance. A compendium of topics is covered: testing for the presence of jumps, determining the proportion of total variation due to components (continuous versus discontinuous), detecting the size and location of large jumps, testing the degree of jump activity (under the assumption that jumps exist at all), testing whether the activity rate is finite or infinite, and testing for the necessity of a continuous component.

Throughout, each chapter closes with a bibliographical guide and proofs for the most part are relegated to the appendices, a strategy which contributes significantly to readability.

Many of the results are recently-developed, especially those on microstructure noise and irregular sampling, and the authors borrow heavily (but not unreasonably) from their own published work, which has been substantial. Reflecting the bulk of known results, the presentation tends to focus on the univariate case, although multivariate extensions are presented where possible. The relative lack of multivariate results may be a minor disappointment to practitioners, whose interest is as often in multivariate as univariate results, although it is largely unavoidable given the complexity of the material and the developing nature of the field. It would also have been interesting to see more material on finite-sample results given that this is ultimately the situation that applies in practice, although again this is difficult given the case-by-case nature of such results, which essentially necessitate a topic of their own. Also interesting would have been a treatment of intra-day features, such as the fact that markets exhibit varying degrees of trading activity and liquidity during the day, which at the microscopic level may have consequences for the inference. It is worth noting too that the treatment is mostly non-parametric which means that some work is required of the reader to pin down what the "total squared size" of the realized volatility estimator actually represents; for instance, how it enables to answer simple questions such as posed at the beginning of this review.

The book is exceptionally well written and the pedagogical talent of the authors is clearly evident. It is evident too that careful thought was given to the presentation of ideas, which follow a logical progression, albeit with a minor caveat regarding the presentation in the first part, where material on market data and price formation is placed alongside mathematical material, causing a slight dissonance in the flow of ideas. While this partly reflects the challenges of a comprehensive exposition, needing in this case to bridge mathematics, econometrics and applied finance, the relative brevity of the non-technical sections means the material could perhaps have been placed in appendices. The presentation may also have been improved by introducing the basic estimator earlier in the book. These are however minor criticisms of what is overall a very readable and exceptionally informative book. To the reviewer's knowledge the book represents the first exposition of such high-frequency econometric techniques, collecting into a single volume an extensive literature on these fascinating topics, and in this sense the scale and comprehensiveness of the coverage is especially impressive. Overall the book was a real pleasure to read and is highly recommended.

Conflicts of Interest: The author declares no conflict of interest.