Princeton ORFE-569 (Spring 2007)

Special Topics in Statistics and Operations Research:
Statistical Analysis of Ultra-High Frequency Data
- An Overview and A New Filtering Approach

Yong Zeng

Course Description:

This graduate seminar course has two objectives. The first is to provides an overview on modelling and statistical inference of ultra-high frequency (UHF) data and their related market microstructure theory. The second is to elaborate a new nonlinear filtering approach in modelling the micromovement of asset prices and its related statistical, computational and mathematical finance topics. This includes, but not limited to filtering with counting process observations, filtering with marked point process observation, their related filtering equations and computational methods including particle filtering, the Bayesian inference (estimation and model selection) via filtering of the models, and the risk minimization hedging of the model. Students will also learn simulation programs to simulate the micromovement of stock price, and learn to do estimation and model selection using the provided Fortran programs on simulated and real UHF stock price data sets for some specific micromovement models.

This seminar course consists of two parts. The main part is lectures on statistical analysis of the filtering models for UHF data and their applications. The other part is student presentations and discussions (strengths and weaknesses, contributions and limitations, future research ideas) on some key papers in the literature of modelling and analyzing UHF data. Students can choose or are assigned one or two papers in the overview on modelling UHF data in the beginning of the semesters for presentation and they will give their presentations in the last two or three weeks.

Below two sections are more detail descriptions of the contents for the aforementioned two objectives.

An Overview on Modelling UHF Data

UHF data have two fundamental properties that do not emerge when data are aggregated into lower frequencies. First, the occurrence times are irregular and random. Second, UHF data contain the so-called microstructure (or trading) noise due to price discreteness, price clustering, ask and bid bounce and other market microstructure issues. In spite of the brief history of UHF data analysis, there have been many exciting econometric developments over the past decade. The literature is classified into *three major directions* of development. Direction One aggregates UHF data by one or five minutes (usually) high frequency data and produces "realized volatility". Direction Two separates the transaction price into *efficient price* (or intrinsic value) and *pricing error* (or trading noise, or microstructure noise). The main methods used are the time series tools of vector autoregressive impulse response analysis and the technique of state-space modelling. The tools of the first two directions are closely related to standard time series analysis for fixed time intervals. Direction Three develops new econometric tools to account for both the irregularity of arrival times and microstructure noise.

For each direction, several papers are selected to assign for students' reading and a list of papers is provided at the end.

A Nonlinear Filtering Approach for Modelling UHF Data

The filtering approach belongs to Direction Three also, but with a different view of UHF data by most econometricians, who view UHF data as an irregularly-spaced time series. From the standpoint of stochastic processes, we treat the transaction observations as an observed sample path of a collection of counting process observations, a special case of marked point process observations. In the model, the intrinsic value process of an asset, which connects to the usual models in option pricing and the empirical econometric literature of daily closing prices, is assumed not to be observed directly. But it can be partially observed through the prices, which are distorted by the microstructure noises. Then, the model is formulated as a filtering problem with counting process observations. Hence, the powerful tools of stochastic calculus for filtering are introduced.

The following topics will be covered: **Part (I)** review of Stochastic calculus including Brownian motion, Poisson process, Semimartingale and stochastic integral, Itô's Formula for semimartingale and Girsanov's Theorems, etc. **Part (II)** A Filtering Model with Counting Process Observations including (1) motivations for the model, another equivalent representation of the model; and interesting specific models; (2) joint likelihood, integrated likelihoods, posterior, likelihood ratio and Bayes factor for the model; unnormalized and normalized filtering equations, and the evolution equations for Bayes factors; Bayesian inference via filtering for the model. (3) Computation of Filters including convergence theorems on stochastic integrals and on conditional expectation; "nearly" optimal filters; Markov chain approximation method to construct "nearly" likelihoods, posteriors and Bayes factors and their applications to simulated and real stock price data sets; A branching particle filter and its convergence. (4) Local risk minimization hedging of the model including an overview of risk minimization hedging and local risk minimization hedging and its application to the filtering model. **Part (III)** Filtering with Marked Point Process Observations including random Poisson measure and repeating (1) - (3) of Part (II) for the more general model, which subsumes many existing important models.

A List of Papers on Overview (more maybe added)

Direction One:

1. Anderson, T. G., Bollerslev, T., Diebold, F. X. & Labys, P. (2001), 'The distribution of realized exchange rate volatility', *Journal of the American Statistical Association*, 96(453), 42-55.

2. Anderson, T. G., Bollerslev, T., Diebold, F. X. & Labys, P. (2003), 'Modeling and forecasting realized volatility', *Econometrica*, 71, 579-625.

3. Barndorff-Nielsen, O. E. & Shephard, N. (2004), 'Econometric analysis of realized covariation: high frequency based covariance, regression and correlation in Financial economics', *Econometrica*, 72, 885 - 925.

4. Bandi, F. M. & Russell, J. R. (2004), 'Separating microstructure noise from volatility'. AFA 2005 Philadelphia Meetings.

5. Zhang, L., Mykland, P. A. & Ait-Sahalia, Y. (2005), 'A tale of two time scales: Determining integrated volatility with noisy high frequency data', Journal of the American Statistical Association 100, (in press).

Direction Two:

1. Hasbrouck, J. (1996), Modelling market microstructure time series, in G. Maddala & C. Rao, eds, 'Handbook of Statistics', Vol. 14, North-Holland, Amsterdam, pp. 647-692.

2. Hasbrouck, J. (1999), 'The dynamics of discrete bid and ask quotes', Journal of Finance 54(6), 2109 - 2142.

3. George, T. J. & Hwang, C. Y. (2001), 'Information ow and pricing errors: A unified approach to estimation and testing', Reviews of Financial Studies 14, 979-1020.

4. Hasbrouck, J. (2002), 'Stalking the "efficient price" in market microstructure specifications: an overview', Journal of Financial Markets 5(3), 329 - 339.

Direction Three:

1. Hausman, J., Lo, A. & Mackinlay, C. (1992), 'An ordered probit analysis of stock transaction prices', Journal of Financial Economics 31, 319-379.

2. Engle, R. & Russell, J. (1998), 'Autoregressive conditional duration: A new model for irregularly spaced transaction data', Econometrica 66, 1127-1162.

3. Engle, R. (2000), 'The econometrics of ultra-high-frequency data', Econometrica 68, 1-22.

4. Dufour, A. & Engle, R. F. (2000), 'Time and the price impact of a trade', Journal of Finance 55(6), 2467-2498.

5. Rydberg, T. H. & Shephard, N. (2003), 'Dynamics of trade-by-trade price movements: decomposition and models', Journal of Financial Econometrics 1, 2-25.

6. Engle, R. & Lunde, A. (2003), 'Trades and quotes: A bivariate point process', Journal of Financial Econometrics 1, 159-188.

7. Duffie, D. & Glynn, P. (2004), 'Estimation of continuous-time Markov processes sampled at random time intervals', Econometrica 72, 1773 - 1808.

List of Papers for the Filtering Approach

1. Frey, R. & Runggaldier, W. J. (2001), 'A nonlinear filtering approach to volatility estimation with a view towards high frequency data', International Journal of Theoretical and Applied Finance 4, 199-210.

2. Cvitanic, J., Liptser, R. & Rozovskii, B. (2003), 'A filtering approach to tracking volatility from prices observed at random times'. Working Paper, Southern California University.

3. Zeng, Y. (2003), 'A partially-observed model for micro-movement of asset prices with Bayes estimation via filtering', Mathematical Finance 13, 411-444.

4. Kouritzin, M. & Zeng, Y. (2005), 'Bayesian model selection via filtering for a class of micro-movement models of asset price', International Journal of Theoretical and Applied Finance 8, 97-121.

5. Xiong, J. & Zeng, Y. (2006) 'A branching particle approximation for the filtering equations with counting process observations', Working paper.

6. Frey, R. (2000), 'Risk-minimization with incomplete information in a model for high-frequency data', Mathematical Finance 10, 215-225.

7. Lee, K. & Zeng, Y. (2005) 'Filtered Local Risk Minimization for a Partially-Observed Micro-movement Model of Asset Price', under review.

8. Other related papers, data sets, and Fortran programs are available at Yong Zeng's Webpage `http://mendota.umkc.edu/paper-tick.html`.