Princeton ORFE-569 (Spring 2007)

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

11:00-12:20, TuTh; Room 111, Friend Center

Yong Zeng

Course Description: This graduate special topic course has two objectives. The first is to provides an overview on modeling 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 modeling the micromovement of asset prices and its related statistical, computational and mathematical finance topics. This includes, but not limited to **Fi**ltering with **C**ounting **P**rocess **O**bservations (FICPO model), **Fi**ltering with **M**arked **P**oint **P**rocess **O**bservations (FIMPPO model, which further allows other observable factors such as duration, trade size and others), their related filtering equations and computational methods including particle filtering, the Bayesian inference (estimation and model selection) via filtering of the models, and related financial applications such as 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 by using the provided Fortran programs on simulated and real UHF stock price data sets for some specific micromovement models. Since there are many opportunities in this unexplored research area, furthermore, students or teams of two students are expected to repeat all of these for new micromovement models and write term papers.

This special topic 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 students' 20-minute presentation. Students are to present one or two key papers in the literature of modeling and analyzing UHF data and discuss the strengths and weaknesses, contributions and limitations, and future research ideas. Students can choose or are assigned one or two papers in the overview on modeling UHF data in the middle of the semesters for presentation and they will give their presentations in the last few classes.

Close interactions with the instructor is strongly recommended along the production of term paper.

Brief Outline: The lectures consist of four parts. Since the goal of this course is to produce a term paper, *Part I* is a quick introduction to give students just enough to start to work on data, to construct suitable noise and choose appropriate models, to do simulation (using R or Splus), and to do Bayesian inference (estimation and model selection) via filtering (using Fortran program language). *Part II* is theory. It contains (i) a brief review of stochastic calculus including Brownian motion, Poisson process, Poisson random measure and marked point process, semimartingale and stochastic integral, Itô's Formula for semimartingale and Girsanov's Theorems, etc; (ii) the derivation of the filtering equations and the evolution equations of Bayes factors for FICPO model; (iii) the 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 branching particle filtering; and (iv) repeat (i), (ii) and (iii) for FIMPPO model. *Part III* is the literature overview and selected review for UHF data and related market microstructure theory. *Part IV* is financial applications of FICPO including option pricing and hedging problem, mean-variance portfolio selection problem, and others in market

microstructure theory if time permits.
Prerequisites: Familiar with graduate level statistics and stochastic calculus. Computer skills of SAS, R or Splus, Fortran or C++ are plus, but not required. Or with the consent of the instructor.
References:
"Stochastic Integration and Differential Equations" by Philip Protter, 2nd Edition, Springer, 2003.
"Markov Processes: Characterization and Convergence", by Ethier and Kurtz, Wiley, 1986.
"Limit Theorems for Stochastic Processes" by J. Jacob and A. N. Shiryaev, 2nd Ed., Springer 2003.
"Marked Point Processes on the Real Line", by G. Last and A. Brandt, Springer 1995.
"Market Microstructure Theory" by M. O'Hara, Blackwel, 1995.
"The Microstructure Approach to Exchange Rates", by R. Lyons, MIT Press, 2001.
"An introduction to High-Frequency Finance" by M. M. Dacorogna et.al., Academic Press, 2001.
"Trading and exchanges" by Larry Harris, Oxford University Press, 2003.
"Empirical Market Microstructure" by J. Hasbrouck, Oxford University Press, 2007.

Grading: There will be 3 - 4 Lab assignments worth 35%. There will be 3 - 5 homework (theoretical) worth 15%. The Lab assignments with some homeworks are designed to build up the term paper. The 20-minute presentations will be worth 15%. The final term paper will be worth 35%.

Team works are strongly encouraged. Students are encouraged to discuss lab assignments and homework problems, but you must write up the solutions independently. No credit will be given to assignments found to be substantially similar.
How to Reach Me:

E-mail:	zeng@princeton.edu, or zeng@mendota.umkc.edu
Phone:	(609) 258 6494 (office) or (913) 907 4907 (cell)
Office Hours:	TuTh 2:30-3:30pm, Room ACE43 or by appointment gladly
Course Website:	http://mendota.umkc.edu/paper-tick.html

Appendix

Below are a more detail description of the content and partial list of related papers.

An Overview on Modeling 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 can be 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 modeling. 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 Modeling 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.

A Partial List of Papers on Overview (more to be added)

Direction One: *Realized Volatility*

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, 1394-1411.

Direction Two: *Decomposition of Time Series*

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 flow 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: *ACD type Models*

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. (2006), 'A filtering approach to tracking volatility from prices observed at random times', *Annals of Applied Probability*, 16, 1633-1652.

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. 'Estimating Stochastic Volatility via Filtering for the Micro-movement of Asset Prices', (2004) *IEEE Transactions on Automatic Control*, Vol. 49(3), pp. 338-348.

5. 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.

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', Working paper.

8. Other related papers, data sets, and Fortran programs are available the course webpage `http://mendota.umkc.edu/paper-tick.html`.