ASC2017: Asymptotic Statistics and Computations

Graduate School of Mathematical Sciences, The University of Tokyo
Room 123, January 30 ~ February 1, 2017
http://www.ms.u-tokyo.ac.jp/access/index.html

This series of workshops aims at exchanges of the state-of-the-art in a wide range of asymptotic statistics, computational statistics, and statistical modeling from theoretical, methodological, and implementation points of view.

This workshop is partially supported by CREST, JST:
Mathematical statistics and stochastic analysis for modeling and analysis of complex random systems (Research Director: Nakahiro Yoshida) http://www.sigmath.es.osaka-u.ac.jp/statmodel/

Program

January 30 (Mon.)

p.m. Chair: Hiroki Masuda (Kyushu University)

12:55 - 13:00	Opening
13:00 - 13:40	Stefano M. Iacus (University of Milan) *Some results on Lasso-model selection for dynamical systems with small noise*
13:40 - 14:20	Yasutaka Shimizu (Waseda University) *Applications of central limit theorems for equity-linked insurance*
14:30 - 15:10	Claudio Heinrich (Aarhus University) *Limit theory for Lévy semistationary processes*
15:10 - 15:50	Mark Podolskij (Aarhus University) *Statistical inference for linear fractional stable motion*
16:00 - 16:40	Masayuki Uchida (Osaka University, MMDS) *Hybrid estimators with the initial Bayes estimators for ergodic diffusion processes based on reduced data*

January 31 (Tue.)

a.m. Chair: Yuta Koike (Tokyo Metropolitan University)

9:30 - 10:10	Takaki Hayashi (Keio University) *Wavelet-based methods for high-frequency lead-lag analysis*
10:10 - 10:50	Xiaofei Lu (Ecole CentraleSupelec) *Limit order book modelling with high dimensional Hawkes processes*
11:00 - 11:40	Simon Clinet (University of Tokyo) *Estimating the integrated parameter of the time-varying parameter self exciting process*
12:00 – 12:50	Luncheon party (Lever son Verre)

January 31 (Tue.)

p.m. Chair: Yasutaka Shimizu (Waseda University)

13:30 - 14:10	Kengo Kamatani (Osaka University, MMDS) *MCMC in Yuima Package*
14:10 - 14:50	Lorenzo Mercuri (University of Milan) *New classes and methods in yuima package*
15:00 - 15:40	Hiroki Masuda (Kyushu University) *Remarks on Gaussian quasi-likelihood inference for Lévy driven SDE*
15:40 - 16:20	Niels Richard Hansen (University of Copenhagen) *Recurrent event time modeling: predictive and causal models*

February 1 (Wed.)

a.m. Chair: TBA

9:30 - 10:10	Emil S. Jørgensen (University of Copenhagen) *Prediction-based estimating functions and high-frequency data*
10:10 - 10:50	Nina Munkholt Jakobsen (University of Copenhagen) *Efficient estimation for diffusions with jumps*
11:00 - 11:40	Michael Sørensen (University of Copenhagen) *Maximum likelihood estimation for stochastic differential equations with random effects*

Lunch break

p.m. Chair: Kengo Kamatani (Osaka University)

13:00 - 13:40	Bezirgen Veliyev (Aarhus University) *Edgeworth expansion for Euler approximation of continuous diffusion processes*
13:40 - 14:20	Yuta Koike (Tokyo Metropolitan University) *Capturing heterogeneous lead-lag effects from ultra high frequency data*
14:30 - 15:10	Ioane Muni Toke (Ecole CentraleSupélec) *Some models of intensities for limit order books*
15:10 - 15:50	Frédéric Abergel (Ecole Centrale Paris) *Optimal order placement in limit order books*
16:00 - 16:40	Nakahiro Yoshida (University of Tokyo) *Applications of the quasi likelihood analysis to point processes*
16:40	Closing

Some results on Lasso-model selection for dynamical systems with small noise

Stefano, M. Iacus

University of Milan

We consider a dynamical system with small noise for which the drift is parametrized by a finite dimensional parameter. For this model we consider minimum distance estimation from continuous time observations under L^p penalty imposed on the parameters in the spirit of the Lasso approach with the aim of simultaneous estimation and model selection. We study the consistency and the asymptotic distribution of these Lasso-type estimators for different values of p. For $p = 1$ we also consider the adaptive version of the Lasso estimator.

Applications of Central Limit Theorems for Equity-linked Insurance

Yasutaka Shimizu
Department of Applied Mathematics
Waseda University
shimizu@waseda.jp

Abstract

In both the past literature and industrial practice, it was often implicitly used without any justification that the classical strong law of large numbers applies to the modeling of equity-linked insurance. However, as all policyholders' benefits are linked to common equity indices or funds, the classical assumption of independent claims is clearly inappropriate for equity-linked insurance. In other words, the strong law of large numbers fails to apply in the classical sense. In this paper, we investigate this fundamental question regarding the validity of strong laws of large numbers for equity-linked insurance. As a result, extensions of classical laws of large numbers and central limit theorem are presented, which are shown to apply to a great variety of equity-linked insurance products.

This is a joint work with R. Feng in University of Illinois at Urbana-Champaign.

Key Words. Equity-linked insurance; variable annuity guaranteed benefits; risk measures; strong law of large numbers; central limit theorem; individual model; aggregate model.

On limit theory for Lévy semistationary processes

Claudio Heinrich

Joint work with Andreas Basse-O'Connor and Mark Podolskij

Lévy semistationary (LSS) processes are continuous time stochastic processes defined as the stochastic integral

$$X_t := \int_{-\infty}^{t} g(t-s)\sigma_s dL_s,$$

where L is a Lévy process, g a deterministic kernel and σ a stochastic volatility process, not necessarily independent of L. These processes form an important subclass of *ambit fields*, which were proposed for modeling particle velocities in turbulent flows. We introduce limit theorems for the power variation of LSS processes in the high frequency setting, i.e. for the functional

$$V(p)_n := \sum_{i=1}^{n} |\Delta_i^n X|^p, \quad \Delta_i^n X := X_{i/n} - X_{(i-1)/n},$$

as $n \to \infty$, where $p > 0$. It turns out that the limiting behavior of $V(p)_n$ is divided into 3 different regimes with different limits and different rates of convergence. Which regime applies depends on the power p, the Blumenthal-Getoor index β of the driving Lévy motion L, and the behavior of the function g near 0, specified by a parameter α. Our results can be used to estimate the model parameters α and β. Moreover, for application in turbulence, the quantity $\sigma_t^{2+} := \int_0^t \sigma_s^2 ds$ is of key interest, as it models energy dissipation. This object can also be estimated from our results.

Mark Podolskij, Aarhus University, Denmark

Title of the presentation

Statistical inference for the fractional stable motion.

Abstract: In this paper we investigate the parametric inference for the linear fractional stable motion in high and low frequency setting. The symmetric linear fractional stable motion is a three-parameter family, which constitutes a natural non-Gaussian analogue of the scaled fractional Brownian motion. It is fully characterised by the scaling parameter $\sigma > 0$, the self-similarity parameter $H \in (0,1)$ and the stability index $\alpha \in (0,2)$ of the driving stable motion. The parametric estimation of the model is based upon the limit theory for stationary increments Lévy moving average processes that has been recently studied in [1]. More specifically, we combine power variation statistics and empirical characteristic functions to obtain consistent estimates of (σ, α, H). We present the law of large numbers and fully feasible central limit theorems.

This talk is based on a joint work with Stepan Mazur

References

[1] A. Basse-O'Connor, R. Lachièze-Rey and M. Podolskij (2016): Limit theorems for stationary increments Lévy driven moving averages. To appear in *Annals of Probability*.

Hybrid estimators with the initial Bayes estimators for ergodic diffusion processes based on reduced data

Masayuki Uchida
Graduate School of Engineering Science, Osaka University and Center for Mathematical Modeling and Data Science (MMDS), Osaka University, CREST, JST, Toyonaka, Osaka 560-8531, Japan

We consider efficient estimation of both drift and diffusion coefficient parameters for an ergodic diffusion process from discrete observations. First of all, four kinds of the initial Bayes type estimators based on reduced data are proposed and the asymptotic properties of the estimators are stated. Next, multi-step estimators with the initial Bayes type estimator based on reduced data are described. Furthermore, four kinds of hybrid type estimators are studied and their asymptotic properties, including convergence of moments, are shown. We also give an example and simulation studies. This is a joint work with Yuto Yoshida.

Wavelet-based methods for high-frequency lead-lag analysis

Takaki Hayashi, Keio University

We propose a novel framework to investigate lead-lag relationships between two financial as- sets. Our framework bridges a gap between continuous-time modeling based on Brownian motion and the existing wavelet methods for lead-lag analysis based on discrete-time models and enables us to analyze the multi-scale structure of lead-lag effects. We also present a statistical methodology for the scale-by-scale analysis of lead-lag effects in the proposed framework and develop an asymptotic theory applicable to a situation including stochastic volatilities and irregular sampling. We then report several empirical results. (Joint work with Yuta Koike.)

Keywords: High-frequency data; Lead-lag effect; Wavelet.

Limit order book modelling with high dimensional Hawkes processes

Xiaofei LU*[1,2] and Frédéric ABERGEL†[1]

[1]MICS laboratory, Ecole CentraleSupélec
[2]BNP Paribas Arbitrage

January 13, 2017

Abstract

With the rapid emerging of electronic exchanges, the limit order book (LOB) has become the heart of financial market microstructure research. The LOB is the aggregation of all interest of buyers and sellers in a particular financial instrument. Usually the best 5 bids (buyer prices) and asks (seller prices) as well as their corresponding volumes are displayed so that traders may react to market movements. LOB provides much more abundant information than just the price series. Thus an appropriate mathematical LOB model allow to better understand agent behaviors and make predictions.

The classical approach is to model the order flows of limit orders, market orders and cancellations which are instantaneous changes of order book states by point processes. The clustering of event arrivals is a well acknowledged stylized fact in financial markets. The branch of order book models with Hawkes processes is thus studied, which model the self-exciting effects of systems.

Previously, Bacry et al. has proposed a price model with a 2-dimensional Hawkes process [1] and an order book model with a 8-dimensional Hawkes process [2]. For the latter, the authors choose to consider limit orders, market orders and cancellations on bid and ask side that do not change prices (6 dimensions) and the increase and decrease of price (2 dimensions). However, the price changes actually result from events (aggressive limit orders, total cancellations of best limit orders and total take by market orders) that follow very different dynamics. A higher dimensional model with different event types helps to better understand the price formation.

In our work, we first show empirical evidences of structural exciting relations between different events, raising the interest and necessity of using Hawkes process for order book modelling. Then, we review the 2-dimensional Hawkes process price model, especially to compare the full model with that restricted to pure cross-excitation and the use of 1 or 2 exponential functions for the kernel. Besides, we also show via a simulation experiment that the time variant nature of intensities does not necessarily destruct the dependence structure. A 12-dimensional Hawkes process order book model is constructed to model events on bid and on ask (limit orders, market orders and cancellations that change and do not change the mid-price). Exponential kernels are used in all of our models in order to benefit from the $O(n)$ complexity ($O(n^2)$ in general) for the likelihood function computation. The likelihood function is not concave, which makes it difficult to be optimized by gradient descent algorithms. In addition calibrating such a high-dimensional Hawkes with the real data encounters severe local minimum problem. Adapted differential evolution algorithm is used to boost the convergence and efficiency. It is further found that inhibiting relations between events are non-negligible to properly model the dependences of the events. We propose a heuristic solution with non-linear Hawkes process. The resulting model conforms well with the real event dependences, and exhibits close high-frequency and low-frequency volatilities in signature plots.

References

[1] E. Bacry, S. Delattre, M. Hoffmann, and J.-F. Muzy. Modelling microstructure noise with mutually exciting point processes. *Quantitative Finance*, 13(1):65–77, 2013.

[2] E. Bacry, I. Mastromatteo, and J.-F. Muzy. Hawkes processes in finance. *Market Microstructure and Liquidity*, 1(01):1550005, 2015.

*xiaofei.lu@ecp.fr
†frederic.abergel@centralesupelec.fr

Statistical inference for the doubly stochastic self-exciting process in high-frequency financial data

Running title: Doubly stochastic self-exciting process

Simon Clinet* and Yoann Potiron**

* Graduate School of Mathematical Sciences, University of Tokyo: 3-8-1 Komaba, Meguro-ku, Tokyo 153-8914, Japan. Email: simon@ms.u-tokyo.ac.jp

** (Corresponding Author) Faculty of Business and Commerce, Keio University. 2-15-45 Mita, Minato-ku, Tokyo, 108-8345, Japan. Phone: +81-3-5418-6571. Email: potiron@fbc.keio.ac.jp, website: http://www.fbc.keio.ac.jp/~potiron

January 27, 2017

Abstract

We introduce and show the existence of a continuous time-varying-parameter-extension model to the Hawkes self-exciting point process with exponentially-decreasing kernel. The quantity of interest is defined as the integrated parameter $T^{-1}\int_0^T \theta_t^* dt$, where θ_t^* is the time-varying parameter, and we consider the financial high-frequency data asymptotics. To estimate it naïvely, we chop the data into several blocks, compute the maximum likelihood estimator (MLE) on each block, and take the average of the local estimates. The asymptotic bias explodes asymptotically, thus we provide a non-naïve estimator which is constructed as the naïve one when applying a first-order bias reduction to the local MLE. We show the associated central limit theorem.

Keywords: finance; Hawkes process; high-frequency data; integrated parameter

MCMC in Yuima Package

Kengo Kamatani, Graduate School of Engineering Science, and MMDS, Osaka University

31 Jan 2017

Abstract

In Yuima package in R, a Bayesian estimation function "adaBayes" is already available, and a new Bayesian function "lseBayes" will be added. These functions calculate the posterior inferences from discretely observed diffusion process. A new Markov chain Monte Carlo method, the MpCN algorithm will be implemented which outperforms traditional random-walk Metropolis algorithm in some scenarios. In this talk, we study some properties of MpCN algorithm as a tool for inference for stochastic processes.

The following is an example of `adaBayes` with `method = mpcn`.

```
# Yuima object construction
set.seed(123)
n <- 500
true_parameter <- list(theta2 = -2, theta1 = -1)
yuima_model <- setModel(drift = "(-1) * exp(theta2) * x",
                        diffusion =  "sqrt(exp(theta1) ^ 2 + 1)",
                        time.variable = "t", state.variable = "x",
                        solve.variable = "x")
yuima_sampling <- setSampling(Terminal = n ^ (1 / 3), n = n)
yuima <- setYuima(model = yuima_model, sampling = yuima_sampling)
yuima <- simulate(yuima, xinit = 1, true.parameter = true_parameter, sampling = yuima_sampling)
```

```
# Estimation setting
lower = list(theta1 = - 10, theta2 = - 10)
upper = list(theta1 = 10,theta2 = 10)
start = list(theta1 = runif(1, min = - 10, max = 10), theta2 = rnorm(1))
```

```
# Prior construction via 'code'
prior <- list(theta2 = list(measure.type = "code", df = "dnorm(z, 0, 1)"),
              theta1 = list(measure.type = "code", df = "dnorm(z, 0, 1)"))
bayes <- adaBayes(yuima, start = start, prior = prior,
                  mcmc = 10^5, lower = lower, upper = upper, method="mpcn")
bayes@coef
```

```
##    theta1    theta2
## -1.374964 -1.092669
```

This talk is based on (Kamatani 2017).

Reference

Kamatani, Kengo. 2017. "Ergodicity of Markov chain Monte Carlo with reversible proposal." *Journal of Applied Probability* 54 (2). http://arxiv.org/abs/1602.02889.

New Classes and Methods in YUIMA package

Lorenzo Mercuri

University of Milan.

Abstract

Recently Lévy CARMA(p,q) models, COGARCH(p,q) processes and Point Process Regression model have been introduced in YUIMA package. In this presentation, we review classes and methods for the estimation and the simulation of these models. We also discuss some useful advances with respect to the existing theoretical literature that are used in the implementation schemes.

Remarks on Gaussian quasi-likelihood inference for Lévy driven SDE

Hiroki Masuda

Kyushu University
hiroki@math.kyushu-u.ac.jp

We consider an ergodic Lévy driven stochastic differential equation (SDE)

$$dX_t = a(X_{t-}, \alpha)dZ_t + b(X_t, \beta)dt,$$

where the coefficients are assumed to be known up to the statistical parameter $\theta := (\alpha, \beta)$, and the driving noise Z is a Lévy process such that $\mathbb{E}(Z_1) = 0$ and $\mathrm{var}(Z_1) = 1$. We observe discrete-time data $X_{t_0}, X_{t_1}, \ldots, X_{t_n}$, where $t_j = t_j^n = jh$ with $h = h_n$ satisfying that $nh^2 \to 0$ and $nh \to \infty$. The coefficients $a(a, \alpha)$, $b(x, \beta)$ are correctly specified and satisfy the standard smoothness and non-degeneracy conditions.

In this talk, we are concerned with making inferences for the (semi-)parametric SDE model in a unified manner, regardless of the presence of jumps. Everything is done based on the fully explicit (logarithmic) Gaussian quasi-likelihood

$$\mathbb{H}_n(\theta) := \sum_{j=1}^{n} \log \phi\left(X_{t_j};\, X_{t_{j-1}} + hb(X_{t_{j-1}}, \beta),\, ha^2(X_{t_{j-1}}, \alpha)\right),$$

together with the Gaussian quasi-likelihood estimator $\hat{\theta}_n \in \mathrm{argmax}\, \mathbb{H}_n$, where $\phi(\cdot; \mu, \sigma^2)$ denotes the $N(\mu, \sigma^2)$-density.

The Gaussian quasi-likelihood is the best only for the diffusion case, where the driving noise Z is a standard Wiener process; the previous studies (e.g. [1]) say that the Gaussian quasi-likelihood estimator takes essentially different asymptotic behaviors according to whether or not Z has jumps. The unified inference strategies automatically and quantitatively adjust the situation.

References

[1] Masuda, H. (2011), Approximate quadratic estimating function for discretely observed Lévy driven SDEs with application to a noise normality test. RIMS Kokyuroku 1752, 113–131.

RECURRENT EVENT TIME MODELING: PREDICTIVE AND CAUSAL MODELS

NIELS RICHARD HANSEN

Abstract. Spike times for neurons constitute one main example of a multivariate system of recurrent events. The nonlinear Hawkes process is useful for modeling such a system. Conditional local independence and related Markov properties for multivariate point processes are well established, and causal interpretations are possible[1, 2]. If a causal Hawkes process is completely observed, estimation is – in principle – straight forward[3, 4], and intervention effects can be identified. If the process is only partially observed, causal search algorithms based on global Markov properties of the causal model are possible, but identification and estimation of intervention effects may not be possible.

In the talk I will give a brief introduction to the main ideas of causal models, interventions and graphical representations of dependencies, and I will discuss their relevance for the modeling of dynamic processes. I will also demonstrate how the R package ppstat[5], can be used for estimation of multivariate nonlinear Hawkes models. The package includes some newly implemented causal search algorithms. The theory is currently mostly developed for point processes, and I will conclude by describing some open problems related to global Markov properties of general stochastic processes. These are central if we want to generalize causal search algorithms to stochastic dynamic systems beyond point processes[6].

References

[1] Vanessa Didelez. Graphical models for marked point processes based on local independence. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 70(1):245–264, 2008.

[2] Vanessa Didelez. Causal reasoning for events in continuous time: A decision - theoretic approach. In *Proceedings of the UAI 2015 Workshop on Advances in Causal Inference*, 2015.

[3] Niels Richard Hansen. Nonparametric likelihood based estimation of linear filters for point processes. *Statistics and Computing*, 25(3):609–618, 2015.

[4] Niels Richard Hansen, Patricia Reynaud-Bouret, and Vincent Rivoirard. Lasso and probabilistic inequalities for multivariate point processes. *Bernoulli*, 21(1):83–143, 02 2015.

[5] Niels Richard Hansen. `https://github.com/nielsrhansen/ppstat`.

[6] Alexander Sokol and Niels Richard Hansen. Causal interpretation of stochastic differential equations. *Electron. J. Probab.*, 19(100):1–24, 2014.

Department of Mathematical Sciences, University of Copenhagen Universitetsparken 5, 2100 Copenhagen Ø, Denmark, Niels.R.Hansen@math.ku.dk

Prediction-Based Estimating Functions for Discretized Diffusions

Emil S. Jørgensen
University of Copenhagen

Prediction-based estimating functions (PBEFs) provide a general framework for parameter estimation in discretely observed diffusion-type models. This talk deals with high-frequency asymptotics for estimators obtained from PBEFs when the statistical model is that of a parametric diffusion process (X_t). The discretization $\{X_{t_i^n}\}_{i=0}^n$ is observed at equidistant time points $t_i^n = i\Delta_n$ for some $\Delta_n > 0$, and we consider the infinite horizon scenario where $\Delta_n \to 0$ and $n\Delta_n \to \infty$. Subject to standard regularity conditions on (X_t), we prove the existence of a consistent and asymptotically normal sequence of estimators for some PBEFs of applied interest. The latter requires the additional assumption that $n\Delta_n^3 \to 0$. To complement the asymptotic results, we construct an explicit estimating function for the square-root (CIR) model and a simulation-based extension for the volatility process of a 3/2 model from finance. Joint work with Michael Sørensen.

Efficient estimation for diffusions with jumps

Nina Munkholt Jakobsen
University of Copenhagen, Denmark
munkholt@math.ku.dk

This talk is based on joint work with Michael Sørensen, University of Copenhagen.

The talk concerns parametric estimation for diffusion processes with finite activity jumps. These processes are assumed to be observed at high frequency over an increasing time interval, with terminal sampling time going to infinity. We will discuss asymptotic properties of estimators based on approximate martingale estimating functions, with focus on conditions for rate optimality and efficiency.

Maximum likelihood estimation for stochastic differential equations with random effects

Michael Sørensen,
Department of Mathematical Sciences, University of Copenhagen

Simulation-based methods for maximum likelihood estimation for stochastic differential equations with random effects are presented. The focus is on a Gibbs sampler that uses the new simple method for simulation of diffusion bridges developed in Bladt and Sørensen (2014) and Finch, Bladt and Sørensen (2016). The Gibbs sampler is illustrated by an application to an Ornstein-Uhlenbeck process with random effects in the drift and in the diffusion coefficient. An EM-algorithm is briefly considered.

The bridge simulation method is explained in detail. By a novel application of classical coupling methods, approximate, but often very accurate, diffusion bridges can be simulated. These approximate bridges are used as proposal for an easily implementable MCMC algorithm that produces exact diffusion bridges. An advantage of the new bridge simulation method, compared to the methods of Roberts and coworkers, is that it works well for diffusion bridges in long time intervals. This is because the computational complexity of the method is linear in the length of the interval. In the statistical application this property implies that the method works well at low sampling frequencies.

The lecture is based on joint work with Mogens Bladt and Fernando Baltazar Larios.

References:

Bladt, M. and Sørensen, M. (2014): Simple simulation of diffusion bridges with application to likelihood inference for diffusions. *Bernoulli*, **20**, 645 – 675.

Bladt, M., Finch, S. and Sørensen, M. (2016): Simulation of multivariate diffusion bridges. *J. Roy. Statist. Soc., Ser. B*, **78**, 343 – 369.

Edgeworth expansion for Euler approximation of continuous diffusion processes

Mark Podolskij Bezirgen Veliyev Nakahiro Yoshida

January 24, 2017

Abstract

In this paper we present the Edgeworth expansion for the Euler approximation scheme of a continuous diffusion process driven by a Brownian motion. Our methodology is based upon a recent work Yoshida (2013), which establishes Edgeworth expansions associated with asymptotic mixed normality using elements of Malliavin calculus. Potential applications of our theoretical results include higher order expansions for weak and strong approximation errors associated to the Euler scheme, and for studentized version of the error process.

Capturing heterogeneous lead-lag effects from ultra high frequency data

Yuta Koike*†‡

"ASC2017: Asymptotic Statistics and Computations" Abstaract

A new framework for modeling lead-lag relationships in high frequency financial markets is proposed. The model can be accommodated to non-synchronous trading and market microstructure noise as well as the intraday heterogeneity of the lead-lag relationships, which are essential for empirical applications. A simple statistical methodology for analyzing the proposed model is presented as well. In particular, we can conduct a statistical test to detect the presence of such a lead-lag relationship. The methodology is illustrated by an empirical study to detect lead-lag relationships in multi-market trading.

Keywords: High-frequency data; Lead-lag relationship; Microstructure noise; Non-synchronous observations; Semimartingale; Stable convergence.

*Tokyo Metropolitan University, Department of Business Administration, Graduate School of Social Sciences, Marunouchi Eiraku Bldg. 18F, 1-4-1 Marunouchi, Chiyoda-ku, Tokyo 100-0005 Japan

†The Institute of Statistical Mathematics, 10-3 Midori-cho, Tachikawa, Tokyo 190-8562, Japan

‡CREST, Japan Science and Technology Agency

Intensity models for order flows in a limit order book

Ioane Muni Toke, CentraleSupélec, France
Nakahiro Yoshida, University of Tokyo, Japan

Keywords : Limit order book; Point processes; Intensity models; Cox models:

Abstract

We review some of the existing models of limit order book based on point processes. We propose a parametric model for the order intensities that is specially designed to provide sound simulations of limit order books. We assume that limit orders, market orders and cancellations are submitted according to point processes with state-dependent intensities. We propose functional forms for these intensities, as well as new models for the placement of limit orders and cancellations. For cancellations, we introduce the concept of priority index to describe the selection of orders to be cancelled in the order book. Parameters of the model are estimated using likelihood maximization. We illustrate the performance of the model by providing extensive simulation results, with a comparison to empirical data and a standard Poisson reference. Finally, we continue this study with some ongoing work based on Cox-type models for this modeling of limit order books. By estimating intensity ratios of Cox-type processes using quasi likelihood maximization, we are able to retrieve characteristics of order flows that are coherent with the financial literature.

References

[1] Abergel, F., Anane, M., Chakraborti, A., Jedidi, A. and Muni Toke, I. (2016) Limit Order Books, Cambridge University Press.
[2] Cont, R., Stoikov, S. and Talreja, R. (2010) *A stochastic model for order book dynamics*, Oper. Res. 58(3), 549–563
[3] Huang, W., Lehalle, C.-A. and Rosenbaum, M. (2015) *Simulating and analyzing order book data: The queue-reactive model*, J. Am. Stat. Assoc. 110(509), 107–122.
[4] Mike, S. and Farmer, J.D. (2008) *An empirical behavioral model of liquidity and volatility*, J. Econ. Dyn. Control. 32(1), 200–234.
[5] Muni Toke, I. and Yoshida, N. (2017) *Modelling intensities of order flows in a limit order book*, Quant.finance. To appear.
[6] Smith, E., Farmer, J.D., Gillemot, L. and Krishnamurthy, S. (2003) *Statistical theory of the continuous double auction*, Quant. Finance. 3(6), 481–514.

Optimal order placement in limit order books

Frédéric Abergel (Ecole Centrale Paris)

Applications of the quasi likelihood analysis to point processes

Nakahiro Yoshida

Graduate School of Mathematical Sciences, University of Tokyo *

CREST, Japan Science and Technology Agency

Institute of Statistical Mathematics

February 1, 2017

The quasi likelihood analysis (QLA) gives a basis of statistical inference for stochastic processes. The polynomial type large deviation (PLD) inequality featuring in the QLA provides estimates of the tail of the quasi likelihood random field and hence L^p-boundedness of the quasi maximum likelihood estimator and the quasi Bayesian estimator. The PLD inequality follows from only the local asymptotic quadratic structure of the quasi log likelihood function. Since most of differentiable experiments admit this structure, the QLA was successfully applied to various inferential problems such as ergodic diffusion process, volatility parametric estimation, non-synchronous volatility parametric estimation, model selection for ergodic diffusion processes, spot volatility information criterion, point processes, asymptotic expansion in volatility estimation, and so on. In this talk, we discuss some applications of the QLA to point processes in ergodic and non-ergodic scenarios.

We may say the QLA is one of the mathematical fundamentals in the theory of information criteria for model selection. The QLA can extend to the penalized QLA in an abstract manner, keeping PLD. Consequently, L^p-boundedness of the penalized estimator and a precise estimate of the probability of selection consistency are obtained. The penalized QLA so obtained can be applied with high universality to various dependent structures. We will discuss applications of the QLA to information criteria and penalized methods for estimation of point processes, emphasizing universality of the method with some applications to diffusion type processes as well.

This talk is in part based on joint works and discussions with F. Abergel, S. Clinet, S. Iacus, Y. Kinoshita, H. Masuda, L. Mercuri, I. Muni Toke, T. Ogihara, H. Sato, Taiji Suzuki, Takumi Suzuki, M. Uchida, and others.

*Graduate School of Mathematical Sciences, University of Tokyo: 3-8-1 Komaba, Meguro-ku, Tokyo 153-8914, Japan. e-mail: nakahiro@ms.u-tokyo.ac.jp