SDS-7

データサイエンス 松本キャンプ2018
（新しい時系列計量分析の理論と応用）

国友直人

December 2018

Statistics & Data Science Series back numbers:
http://www.mims.meiji.ac.jp/publications/datascience.html

データサイエンス
松本キャンプ2018

(新しい時系列計量分析の理論と応用) [1]

国友直人[2](編)

2018 年 8 月

[1]学術振興会・科学研究プロジェクト「新しい時系列計量分析の理論と応用」(2017 年度〜2020 年度) が主催した（2018 年 8 月 6 日, 信州大学経法学部) 研究集会 (明治大学先端数理科学インスティテュート (MIMS) 協賛）における講演集。

[2]明治大学政治経済学部

概要

金融・ファイナンスデータ、経済データなどを分析対象とした統計的時系列分析では幾つかの検討するべき基本的な問題が存在している。例えば経済において「通常の常識では起こりにくいとされる事象」についてのリスク解析や対策の重要性についての認識が高まっているが、実際に 2008 年に起きたリーマンショックや 2011 年ごろに発生したヨーロッパ諸国の金融危機などが顕著な例として挙げることができる。国際的に連動している現代の経済・金融市場では従来の計量分析ではほとんど考慮されてこなかった変動を経験しているのである。社会・経済を理解し、より良いものにしていくには、事前には予想が困難である自然災害や経済変動における稀ではあるが実際に起きると大きな影響のある不確実な事象を科学的に分析し、有効な対策を考察することが必要であり重要となっている。また近年に明らかになりつつあるように高頻度金融データやマクロ経済データの分析においては、経済変数間の関係の統計的分析が重要であるが、確率過程の統計分析や非定常多次元時系列分析にはなお解決すべき課題が少なくない。

科学研究プロジェクト「新しい時系列計量分析の理論と応用」ではこうした経済・金融・社会における新しい動向を背景として、近年の日本経済・社会の理解の方法として重要な新しい計量分析の方法を開発、応用を検討している。特に金融現象やマクロ経済の統計分析では、時々起きる大きな経済変動は重要であるにもかかわらず、なお研究の蓄積が不十分であり、様々な研究の可能性がある。また、ミクロ金融の分析ではジャンプや確率過程の一般理論を踏まえた金融時系列分析はなお十分とは言えず、マイクロマーケット分析や確率過程の計量分析の方法を確立が望まれる。こうした経済・金融におけるマクロ分野、ミクロ分野と云う二つの時系列分析において新しい分析の枠組みを構築する必要がある。またマクロ経済データが象徴的であるように、経済・金融分野で観察されるデータは統計的には非定常性が見られるとともに本来的に多次元時系列であり、次元数も必ずしも小さくない場合も研究対象である。

本年度は研究プロジェクトの二年目であり、特にファイナンス市場の分析では重要なマイクロマーケット、確率過程の理論と計測理論などを中心に活発な議論を行った。ここに収録した研究報告が計量ファイナンス、経済時系列分析の理論と応用を展開する一助になることを期待する。

2018 年 8 月
編者

研究集会・プログラム

科学研究プロジェクト「新しい時系列計量分析の理論と応用」
日程：2018 年 8 月 6 日 (月)
会場：信州大学経法学部 4F403
オーガナイザー：国友直人
プロジェクト参加者：国友直人 (明治大学)・大屋幸輔 (大阪大学)・佐藤整尚 (東京大学)・栗栖大輔 (東京工業大学)
会場責任者：椎名洋 (信州大学)
協賛：明治大学先端数理科学インスティテュート (MIMS)

< セッション I：特別講演 >
Chair: 国友直人
13:00〜13:50「Detecting mean-field in a financial network model」
Ichiba Tomoyuki (UC Santa-Barbara)
13:50〜14:40「アルゴリズム取引の実際」
足立高徳 (首都大学東京)

< 休憩 >

< セッション II：Financial Econometrics >
Chair: 都築幸宏 (信州大学)
14:50〜15:20「Local SIML Estimation of Some Brownian Functionals」
佐藤整尚・国友直人
15:20〜16:00「Estimation for affine term structure with smooth transition」
Shingo Mukunoki and Kosuke Oya

< 休憩 >

< セッション III：Econometrics of Time Series>
Chair:矢部竜太 (信州大学)
16:10〜16:50「Detecting Number of Factors in Non-stationary Errors-in-Variables Models」
国友直人
16:50〜17:30「不等間隔観測の下でのノンパラメトリック空間回帰モデルに対する統計的推測」
栗栖大輔

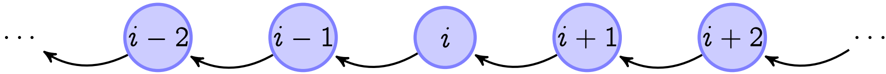
Detecting mean-field in a large financial network model

TOMOYUKI ICHIBA
Department of Statistics & Applied Probability,
Center for Financial Mathematics and Actuarial Research,
University of California, Santa Barbara

Joint work with
NILS DETERING & JEAN-PIERRE FOUQUE

Shinshu University, Japan, August 2018

Let us consider a linear directed graph (or directed network) of vertices $\{1, \dots, n\}$



On some probability space, based on this graph let us consider the simple Ornstein-Uhlenbeck system (or a Gaussian cascade)

$$\mathrm{d}X_{t,i} = (X_{t,i+1} - X_{t,i})\mathrm{d}t + \mathrm{d}\,W_{t,i}\,; \quad t \geq 0\,, \quad i = 1, \dots, n-1\,,$$
$$\mathrm{d}X_{t,n} = (X_{t,1} - X_{t,n})\mathrm{d}t + \mathrm{d}\,W_{t,n}$$

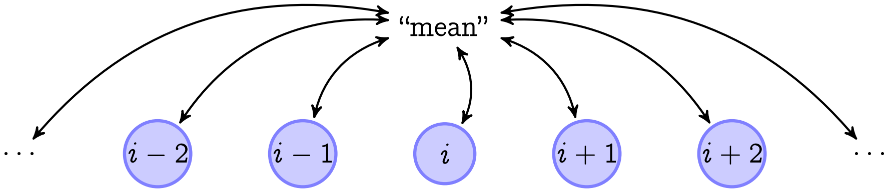
with initial I.I.D. random variables $X_{0,i}$, independent of standard Brownian motions $(W_{\cdot,i})$, $1 \leq i \leq n$.

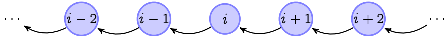
We will consider a discrete-time version later.

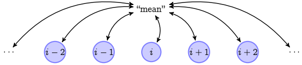
For comparison, on the same probability space, we consider a typical mean-field interacting system where each particle is attracted towards the mean, defined by

$$\mathrm{d}X_{t,i} = \Big(\frac{1}{n}\sum_{j=1}^{n} X_{t,j} - X_{t,i}\Big)\mathrm{d}t + \mathrm{d}\,W_{t,i}\,; \quad t \ge 0\,, \quad i = 1,\ldots,n\,.$$

The particle $X_{\cdot,i}$ at node i is *directly* attracted towards the mean $(X_{\cdot,1} + \cdots + X_{\cdot,n})/n$ of the system.







Directed chain/Cascade   Mean-field interaction

Two types of interactions

Questions.

Q1. What is the essential difference between these two types of interacting systems for large $n \to \infty$?

Q2. Can we detect the type of interaction from the particle behavior at one node?

Q3. Are the optimization problem (e.g., optimal stopping) and the answers different between two types?

Motivation.

- Effects of graph (network) structure: mean-field type and local directed chain dependence.
- Systemic risk problems in financial network: effects of network structure on default cascades.

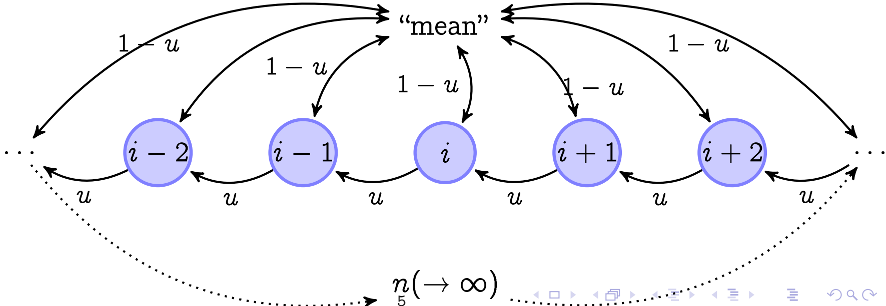
Let us introduce an interpolated system of linear equations:

$$\mathrm{d}X_{t,i} = \Big(u \cdot X_{t,i+1} + (1-u) \cdot \frac{1}{n}\sum_{j=1}^{n} X_{t,j} - X_{t,i}\Big)\mathrm{d}t + \mathrm{d}\,W_{t,i}\,,$$

$$\mathrm{d}X_{t,n} = \Big(u \cdot X_{t,1} + (1-u) \cdot \frac{1}{n}\sum_{j=1}^{n} X_{t,j} - X_{t,n}\Big)\mathrm{d}t + \mathrm{d}\,W_{t,n}$$

for $t \ge 0$, $i = 1, \ldots, n-1$ with the initial $X_{0,i}$, $1 \le i \le n$, and for fixed $u \in [0,1]$.

- What is the limit as $n \to \infty$?



An Infinite-dimensional McKean-Vlasov equation

Given $u \in [0,1]$, let us consider a pair $(X_t^{(u)}, \widetilde{X}_t^{(u)}, t \ge 0)$ on $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$ with a mixture distribution

$$F_t^{(u)}(\cdot) := u \cdot \delta_{\widetilde{X}_t^{(u)}}(\cdot) + (1-u) \cdot \mathcal{L}_{X_t^{(u)}}(\cdot)$$

of the Dirac measure $\delta_{\widetilde{X}_t^{(u)}}$ and the identical marginal law $\mathcal{L}_{X_\cdot^{(u)}} \equiv \mathcal{L}_{\widetilde{X}_\cdot^{(u)}} = \mathrm{Law}(X_\cdot^{(u)})$ and a McKean-Vlasov equation

$$\boxed{\mathrm{d}X_t^{(u)} = b(t, X_t^{(u)}, F_t^{(u)})\,\mathrm{d}t + \mathrm{d}B_t\,; \quad t \ge 0\,,}$$

where Brownian motion $(B_t, t \ge 0)$ is independent of $\widetilde{X}_\cdot^{(u)}$ and $X_0^{(u)}$, and $b : \mathbb{R}_+ \times \mathbb{R} \times \mathcal{M}(\mathbb{R}) \to \mathbb{R}$ is a measurable function. Here $\mathcal{M}(\mathbb{R})$ is a family of probability measures on $\mathbb{R}$.

- two *extreme* cases $u = 0, 1$.

Weak existence and uniqueness

Proposition. Suppose $b : [0, \infty) \times \mathbb{R} \times \mathcal{M}(\mathbb{R}) \to \mathbb{R}$ is Lipschitz. in the sense that there exists a measurable function $\tilde{b} : [0, \infty) \times \mathbb{R} \times \mathbb{R}$. such that b is represented as

$$b(t, x, \mu) = \int_{\mathbb{R}} \tilde{b}(t, x, y)\mu(\mathrm{d}y) ; \quad t \in [0, \infty) , \ x \in \mathbb{R} , \ \mu \in \mathcal{M}(\mathbb{R}) ,$$

and for every $T > 0$ there exists a constant $C_T > 0$ such that

$$|\tilde{b}(t, x_1, y_1) - \tilde{b}(t, x_1, y_2)| \le C_T(|x_1 - x_2| + |y_1 - y_2|) ; \quad t \ge 0 .$$

With the same constant C_T, let us also assume that $\tilde{b}$ is of linear growth, i.e.,

$$\sup_{0 \le s \le T} |\tilde{b}(s, x, y)| \le C_T(|x| + |y|) ; \quad x, y \in \mathbb{R} .$$

Then for each $u \in [0,1]$ there exists a weak solution $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$, $(X_\cdot^{(u)}, \widetilde{X}_\cdot^{(u)}, B_\cdot)$, unique in distribution for

$$\mathrm{d}X_t^{(u)} = b(t, X_t^{(u)}, F_t^{(u)})\,\mathrm{d}t + \mathrm{d}B_t\,; \quad t \geq 0\,, \quad \text{with}$$

$$F_t^{(u)}(\cdot) := u \cdot \delta_{\widetilde{X}_t^{(u)}}(\cdot) + (1-u) \cdot \mathcal{L}_{X_t^{(u)}}(\cdot)\,,$$

$$\mathcal{L}_{X_\cdot^{(u)}} \equiv \mathcal{L}_{\widetilde{X}_\cdot^{(u)}}\,, \quad \boldsymbol{\sigma}(B_\cdot) \perp\!\!\!\perp \boldsymbol{\sigma}(\widetilde{X}_\cdot^{(u)}, X_0^{(u)})\,.$$

In addition, assume $\mathbb{E}[|X_0|] < \infty$. Then, there exists a constant $c\,(> 0)$, such that

$$\mathbb{E}[\sup_{0 \leq t \leq T} |\widetilde{X}_t^{(u)}|] = \mathbb{E}[\sup_{0 \leq t \leq T} |X_t^{(u)}|] \leq (\mathbb{E}[|X_0|] + c)e^{c\,T}\,. \square$$

- Proof is based on a contraction argument for a Picard iteration under the Wasserstein metric $W_1(\cdot,\cdot)$ (cf. Sznitman ('91), Graham ('92)).

- Some extensions.

- When $u = 0$, set $(X_\cdot^\bullet, \widetilde{X}_\cdot^\bullet) := (X_\cdot^{(0)}, \widetilde{X}_\cdot^{(0)})$

$$\mathrm{d}X_t^\bullet = b(t, X_t^\bullet, \mathcal{L}_{X_t^\bullet})\,\mathrm{d}t + \mathrm{d}B_t\,; \quad t \ge 0\,,$$

and the corresponding copy $\widetilde{X}_\cdot^\bullet$ disappears. In particular, if $\widetilde{X}_0^\bullet$ is independent of $X_0^\bullet$, then $\widetilde{X}_\cdot^\bullet$ is independent of $X_\cdot^\bullet$.

- When $u = 1$, set $(X_\cdot^\dagger, \widetilde{X}_\cdot^\dagger) := (X_\cdot^{(1)}, \widetilde{X}_\cdot^{(1)})$

$$\mathrm{d}X_t^\dagger = b(t, X_t^\dagger, \delta_{\widetilde{X}_t^\dagger})\,\mathrm{d}t + \mathrm{d}B_t\,; \quad t \ge 0\,,$$

where $\widetilde{X}_\cdot^\dagger$ has the same law as $X_\cdot^\dagger$, independent of Brownian motion, i.e., $\mathrm{Law}(X_\cdot^\dagger) = \mathrm{Law}(\widetilde{X}_\cdot^\dagger)$ and
$\boldsymbol{\sigma}(\widetilde{X}_t^\dagger, t \ge 0) \perp\!\!\!\perp \boldsymbol{\sigma}(B_t, t \ge 0)$.
The corresponding non-linear contribution from the law $\mathrm{Law}(X_\cdot^\dagger)$ of $X_\cdot^\dagger$ disappears.

Let us consider finite systems $(X^{(u)}_{t,i}\,,\ t \ge 0\,,\ i = 1, \ldots, n)\,,$ $n \in \mathbb{N}$ defined by the system of stochastic differential equations

$$\mathrm{d}X^{(u)}_{t,i} = b(t, X^{(u)}_{t,i}, \widehat{F}^{(u)}_{t,i})\mathrm{d}t + \mathrm{d}W_{t,i}\,; \quad t \ge 0\,, \quad i = 1, \ldots, n-1\,,$$

where

$$\widehat{F}^{(u)}_{t,i}(\cdot) := u \cdot \delta_{X^{(u)}_{t,i+1}}(\cdot) + (1-u) \cdot \frac{1}{n} \sum_{j=1}^{n} \delta_{X^{(u)}_{t,j}}(\cdot)\,, \quad i = 1, \ldots, n-1$$

with the boundary particle

$$\mathrm{d}X^{(u)}_{t,n} = b\Big(t, X^{(u)}_{t,n}, u \cdot \delta_{X^{(u)}_{t,1}} + (1-u) \cdot \frac{1}{n} \sum_{j=1}^{n} \delta_{X^{(u)}_{t,j}}\Big)\mathrm{d}t + \mathrm{d}W_{t,n}\,.$$

- Here $W_{\cdot,i}$, $i \in \mathbb{N}$ are standard independent Brownian motions on a filtered probability space, independent of the initial values $X_{0,i}^{(u)}$, $i = 1, \dots, n$.
- We assume $X_{0,i}^{(u)}$ are I.I.D with $\mathbb{E}[|X_{0,1}^{(u)}|^2] < +\infty$.

- It is natural to write $X_{\cdot,n+1}^{(u)} \equiv X_{\cdot,1}^{(u)}$, because

$\mathrm{Law}(X_{\cdot,i}^{(u)}) = \mathrm{Law}(X_{\cdot,1}^{(u)})$, $i = 1, \dots, n$ and

$$\mathrm{Law}(X_{\cdot,i}^{(u)}, X_{\cdot,i+1}^{(u)}) = \mathrm{Law}(X_{\cdot,1}^{(u)}, X_{\cdot,2}^{(u)}); \quad i = 1 \dots, n.$$

Joint and Marginal empirical distributions

Let us assign the weight $1/n$ to the Dirac measure at $(X_{t,i}^{(u)}, X_{t,i+1}^{(u)})$ for $i = 1, \ldots, n$, and consider the law of the joint empirical measure process

$$\mathrm{M}_{t,n} := \frac{1}{n}\sum_{i=1}^{n} \delta_{(X_{t,i}^{(u)}, X_{t,i+1}^{(u)})}, \quad \text{with} \quad \mathrm{m}_{t,n} := \frac{1}{n}\sum_{i=1}^{n} \delta_{X_{t,i}^{(u)}},$$

$0 \le t \le T$ in the space $\mathcal{M}(\Omega_1)$ of probability measures on the topological space $\Omega_1 := D([0,T], (\mathcal{M}(\mathbb{R}^2), \|\cdot\|_1))$ of càdlàg functions on $[0,T]$ equipped with the Skorokhod topology,

where $(\mathcal{M}(\mathbb{R}^2), \|\cdot\|_1)$ is the space of p.m.'s on $\mathbb{R}^2$ equipped with the metric $\|\mu - \nu\|_1 := \sup_f \int_{\mathbb{R}^2} f(x)\mathrm{d}(\mu-\nu)(x)$. Here the supremum is taken over the bounded Lipschitz functions $f : \mathbb{R}^2 \to \mathbb{R}$ with $\sup_{x\in\mathbb{R}^2} |f(x)| \le 1$ and $\sup_{x,y\in\mathbb{R}^2} |f(x) - f(y)|/\|x-y\| \le 1$.

Law of Large Numbers

By the construction the sequence of the law of the initial empirical measure converges to the Dirac measure concentrated in M_0 (say), i.e.,

$$\mathrm{Law}(\mathrm{M}_{0,n}) \xrightarrow[n\to\infty]{} \delta_{\mathrm{M}_0} \quad \text{weakly in} \quad \mathcal{M}((\mathcal{M}(\mathbb{R}^2), \|\cdot\|_1)).$$

We denote by $\mathrm{m}_0(\mathrm{d}y) := \mathrm{M}_0(\mathbb{R} \times \mathrm{d}y) = \mathrm{M}_0(\mathrm{d}y \times \mathbb{R})$ the marginal of M_0.

Proposition. Fix $u \in [0,1]$. Under the same assumptions for the functional b the law of empirical measure process $\mathrm{M}_{\cdot,n}$ converges in $\mathcal{M}(\Omega_1)$ to the Dirac measure concentrated in the deterministic measure-valued process $\mathrm{M}_t, 0 \le t \le T$, as $n \to \infty$, i.e.,

$$\lim_{n\to\infty} \mathrm{Law}(\mathrm{M}_{t,n}, 0 \le t \le T) = \delta_{(\mathrm{M}_t, 0 \le t \le T)} \quad \text{in} \quad \mathcal{M}(\Omega_1).$$

The marginal laws of $\mathrm{M}_\cdot$ are the same, i.e., $\mathrm{M}_t(\mathbb{R}\times \mathrm{d}y) = \mathrm{M}_t(\mathrm{d}y\times\mathbb{R}) =: \mathrm{m}_t(\mathrm{d}y)$, $0\le t\le T$, and the joint $\mathrm{M}_\cdot$ and its marginal $\mathrm{m}_\cdot$ satisfy the integral equation

$$\int_{\mathbb{R}} g(x)\mathrm{m}_t(\mathrm{d}x) = \int_{\mathbb{R}} g(x)\mathrm{m}_0(\mathrm{d}x) + \int_0^t [\mathcal{A}_s(\mathrm{M})g]\,\mathrm{d}s\,; \quad 0\le t\le T$$

for every test function $g\in C_c^2(\mathbb{R})$, where

$$\begin{aligned}\mathcal{A}_s(\mathrm{M})g &:= u\int_{\mathbb{R}^2}\widetilde{b}(s,y_1,y_2)g'(y_1)\mathrm{M}_s(\mathrm{d}y_1\mathrm{d}y_2)\\ &\quad + (1-u)\int_{\mathbb{R}^2}\widetilde{b}(s,y_1,y_2)g'(y_1)\mathrm{m}_s(\mathrm{d}y_1)\mathrm{m}_s(\mathrm{d}y_2)\\ &\quad + \frac{1}{2}\int_{\mathbb{R}} g''(y_1)\mathrm{m}_s(\mathrm{d}y_1)\,; \quad 0\le s\le T\,.\end{aligned}$$

Moreover, $\mathrm{M}_\cdot$ is the joint distribution of the solution pair $(X_\cdot, \widetilde{X}_\cdot)$ unique in the sense of distribution with the common marginal $\mathrm{m}_\cdot = \mathrm{Law}(X_\cdot) = \mathrm{Law}(\widetilde{X}_\cdot)$. $\square$

- Proof uses the martingale problem (Oelschläger ('84)).

Fluctuation

- Let us also consider $\overline{X}_{t,i}$, $t \ge 0$, $i = 1, \ldots, n+1$, $n \ge 1$, defined recursively from the pair $(\overline{X}_{\cdot,n}, \overline{X}_{\cdot,n+1}) := (X^{(u)}_{\cdot}, \widetilde{X}^{(u)}_{\cdot})$ of the solution to the infinite-dimensional McKean-Vlasov stochastic equation

$$\mathrm{d}\overline{X}_{t,n} = b(t, \overline{X}_{t,n}, u \cdot \delta_{\overline{X}_{t,n+1}} + (1-u) \cdot \mathcal{L}_{\overline{X}_{t,n}}) \mathrm{d}t + \mathrm{d}W_{t,n} ; \quad t \ge 0 ,$$

and then for $j = n-1, n-2, \ldots, 1$, given $\overline{X}_{\cdot,j+1}$, we solve

$$\mathrm{d}\overline{X}_{t,j} = b(t, \overline{X}_{t,j}, u \cdot \delta_{\overline{X}_{t,j+1}} + (1-u) \cdot \mathcal{L}_{\overline{X}_{t,j}}) \mathrm{d}t + \mathrm{d}W_{t,j} ; \quad t \ge 0$$

with the distributional restrictions for each pair $(\overline{X}_{\cdot,j}, \overline{X}_{\cdot,j+1})$. We set the common law $\mathrm{m}^* = \mathrm{Law}(\overline{X}_{\cdot,i})$ for $i = 1, \ldots, n+1$, and we also assume the initial values are the same as $X^{(u)}_{0,i} = \overline{X}_{0,i}$, $i = 1, \ldots, n$ almost surely.

Proposition. In addition, let us assume that the marginal distribution $\mathrm{m}_t(dy) = \mathrm{m}_t^*(dy)$ of $(X_t^{(u)}, t \ge 0)$ has the density $m_t(\cdot)$ (i.e., $\mathrm{m}_t(dy) = m_t(y)dy$, $y \in \mathbb{R}$) with $\int_{\mathbb{R}} |y|^2 \mathrm{m}_0(dy) < \infty$ and assume there exists a constant C_T such that

$$\left| \widetilde{b}(t, x_1, y_1) \cdot \frac{m_t(x_1)}{m_t(y_1)} - \widetilde{b}(t, x_2, y_2) \cdot \frac{m_t(x_2)}{m_t(y_2)} \right| \le C_T(|x_1 - x_2| + |y_1 - y_2|)$$

for every $(x_i, y_i) \in \mathbb{R}^2$, $i = 1, 2$, $0 \le t \le T$ and

$$\left| \widetilde{b}(t, x, y) \cdot \frac{m_t(x)}{m_t(y)} \right| \le C_T(|x| + |y|)$$

for every $(x, y) \in \mathbb{R}^2$, $0 \le t \le T$.

Then for the difference between two systems

$$\sup_{n \ge 1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbb{E}[\sup_{0 \le s \le T} |X_{s,i}^{(u)} - \overline{X}_{s,i}|] < \infty .$$

A detection problem of mean-field

Suppose that we observe only one sample path $X_\cdot^{(u)}$ with

$$\mathrm{d}X_t^{(u)} = b(t, X_t^{(u)}, u \cdot \delta_{\widetilde{X}_t^{(u)}} + (1-u) \cdot \mathcal{L}_{X_t^{(u)}})\mathrm{d}t + \mathrm{d}B_t\,; \quad t \ge 0$$

but we do not observe the other particle $\widetilde{X}^{(u)}$.

Here $1-u$ is the size of effect from the distribution of $X_\cdot^{(u)}$

Q2. Can we detect $u \in [0,1]$ only from the one sample path?

A. In general, it seems difficult, however, we can resolve it when b is linear.

Let us take a linear functional $b(t, x, \mu) := -\int_{\mathbb{R}} (x - y)\mu(\mathrm{d}y)$ for $t \ge 0$, $x \in \mathbb{R}$, $\mu \in \mathcal{M}(\mathbb{R})$ of mean-reverting type.

$$\boxed{\mathrm{d}X_t^{(u)} = -(u\,(X_t^{(u)} - \widetilde{X}_t^{(u)}) + (1-u)(X_t^{(u)} - \mathbb{E}[X_t^{(u)}]))\,\mathrm{d}t + \mathrm{d}B_t}$$

for $t \ge 0$ and each $u \in [0, 1]$.

Setting a fixed initial value $X_0^{(u)} = 0$ for simplicity, we see

$$\mathbb{E}[X_t^{(u)}] = \mathbb{E}[\widetilde{X}_t^{(u)}] = \mathbb{E}[X_t^{\bullet}] = \mathbb{E}[X_t^{\dagger}] = 0\,, \quad t \ge 0\,,\ u \in [0, 1]\,,$$

with an explicitly solvable Gaussian pair $(X^{(u)}(t), \widetilde{X}^{(u)}(t))$ for $t \ge 0$, $u \in [0, 1]$.

The iteration scheme, when $u = 1$.

For simplicity let us set $X_i^\dagger(0) = 0$ and $u = 1$. Given $X_2^\dagger(\cdot)$, we define

$$X_1^\dagger(t) = \int_0^t e^{-(t-s)} X_2^\dagger(s)\mathrm{d}s + \int_0^t e^{-(t-s)}\mathrm{d}B_1(s),$$

and also, given $X_3^\dagger(\cdot)$, we define

$$X_2^\dagger(s) = \int_0^s e^{-(s-v)} X_3^\dagger(v)\mathrm{d}u + \int_0^s e^{-(s-v)}\mathrm{d}B_2(v)$$

for $t \ge 0$, and hence substituting $X_2^\dagger(\cdot)$ into the first one,

$$X_1^\dagger(t) = \int_0^t e^{-(t-s)}\mathrm{d}B_1(s) + \int_0^t \int_0^s e^{-(t-v)}\mathrm{d}B_2(v)\mathrm{d}s$$

$$+ \int_0^t e^{-(t-s)} \int_0^s e^{-(s-v)} X_3^\dagger(v)\mathrm{d}v$$

for $t \ge 0$.

By the product rule for semimartingales, we observe

$$\int_0^t \int_0^s e^u (s-u)^{k-1} \mathrm{d}B(u)\mathrm{d}s = \int_0^t e^u \frac{(t-u)^k}{k} \mathrm{d}B(u),$$

for $k \in \mathbb{N}$, $t \geq 0$, and hence

$$\int_0^t \int_0^s e^u \mathrm{d}B(u)\mathrm{d}s = \int_0^t e^u (t-u)\mathrm{d}B(u),$$

$$\int_0^t \int_0^{s_k} \cdots \int_0^{s_1} e^u \mathrm{d}B(u)\mathrm{d}s_1 \cdots \mathrm{d}s_k = \int_0^t e^u \frac{(t-u)^k}{k!} \mathrm{d}B(u)$$

for $k \in \mathbb{N}$, $t \geq 0$. Thus for the above example we have

$$X_1^\dagger(t) = \sum_{k=0}^{\infty} \int_0^t e^{-(t-v)} \cdot \frac{(t-v)^k}{k!} \mathrm{d}B_{k+1}(v)$$

for $t \geq 0$.

$$X_1^\dagger(t) = \sum_{k=0}^{\infty} \int_0^t e^{-(t-v)} \cdot \frac{(t-v)^k}{k!} \mathrm{d}B_{k+1}(v)$$

is a centered, Gaussian process with covariances

$$\mathbb{E}[X_1^\dagger(s) X_1^\dagger(t)] = e^{-(s+t)} \sum_{k=0}^{\infty} \int_0^s \frac{e^{2v}}{(k!)^2} (s-v)^k (t-v)^k \mathrm{d}v$$

$$= e^{-(t-s)} \int_0^s e^{-2v} I_0(2\sqrt{(t-s+v)v}) \mathrm{d}v$$

for $0 \le s \le t$, where $I_0(\cdot)$ is the modified Bessel function of the first kind with parameter 0, i.e.,

$$I_\nu(x) := \sum_{k=0}^{\infty} \left(\frac{x}{2}\right)^{2k+\nu} \frac{1}{\Gamma(k+1)\Gamma(\nu+k+1)}$$

for $x > 0$, $\nu \ge -1$.

In particular,

$$\mathrm{Var}(X_1^\dagger(t)) = \int_0^t e^{-2v} I_0(2v)\mathrm{d}v = te^{-2t}(I_0(2t) + I_1(2t)) < \infty$$

(it grows as $O(t^{1/2})$ for large t, also,

$$\mathbb{E}[X_1^\dagger(s)X_1^\dagger(s+t)] = O(e^{-(t-2\sqrt{(t+s)s})}t^{-1/4}).)$$

Thus $X_1^\dagger(\cdot)$ is *not stationary*.

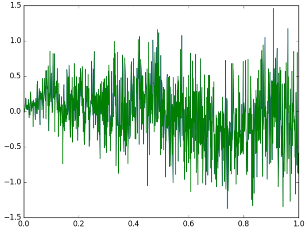
The (marginal) distribution of $X_k^\dagger(\cdot)$, $k \in \mathbb{N}$ is the same as $X_1^\dagger(\cdot)$, and hence, we may compute (at least numerically)

$$\mathbb{E}[X_1^\dagger(t)X_2^\dagger(u)] = \int_0^t e^{(t-s)}\mathbb{E}[X_2^\dagger(s)X_2^\dagger(u)]\mathrm{d}s$$

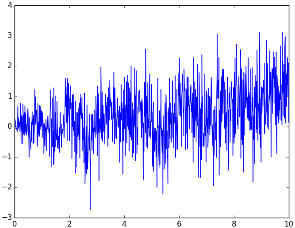
$$= \int_0^t e^{(t-s)}\mathbb{E}[X_1^\dagger(s)X_1^\dagger(u)]\mathrm{d}s$$

and recursively, $\mathbb{E}[X_1^\dagger(t)X_k^\dagger(u)]$, $k \in \mathbb{N}$ for $0 \le t, u < \infty$.

Sample path of $(X_1^\dagger(\cdot), X_2^\dagger(\cdot))$ generated from the covariance structure.



$X_1^\dagger(\cdot)$



$X_2^\dagger(\cdot)$

$$X_1^\dagger(t) = \int_0^t e^{-(t-s)} X_2^\dagger(s)\mathrm{d}s + \int_0^t e^{-(t-s)}\mathrm{d}B_1(s)\,,$$

for $t \geq 0$ and $\mathrm{Law}(X_1^\dagger(\cdot)) = \mathrm{Law}(X_2^\dagger(\cdot))$.

In general, we have

$$X_t^{(u)} = \int_0^t e^{-(t-s)} u \widetilde{X}_s^{(u)} \mathrm{d}s + \int_0^t e^{-(t-s)} \mathrm{d}B_s \, ,$$

$$\widetilde{X}_t^{(u)} = \int_0^t \sum_{k=0}^{\infty} \mathfrak{p}_{0,k}(t-s; u) \, \mathrm{d} W_{s,k} \, ,$$

$$\mathfrak{p}_{0,k}(t-s; u) := \frac{u^k (t-s)^k}{k!} e^{-(t-s)} \, ,$$

where $(W_\cdot^k, k \geq 0)$ is a sequence of independent, one-dimensional standard Brownian motions, independent of the Brownian motion $B(\cdot)$.

- Note that the integrand $\mathfrak{p}_{0,k}(t-s;u)$, $k \in \mathbb{N}_0$ is a (taboo) transition probability $\mathbb{P}(M(t-s) = k | M(0) = 0)$ of a Markov chain $M(\cdot)$ in the state space $\mathbb{N}_0$ with generator matrix $\boldsymbol{Q} = (q_{i,j})_{i,j \in \mathbb{N}_0}$ with $q_{i,i+1} = u \in [0,1]$, $q_{i,i} = -1$ and $q_{i,j} = 0$ for the other entries $j \neq i, i+1$.

Thus we interpret $\mathfrak{p}_{0,k}(t-s;u)$ as $(0,k)$-element of the $\mathbb{N}_0 \times \mathbb{N}_0$-dimensional matrix exponential $e^{(t-s)\boldsymbol{Q}}$, i.e.,

$$\big(\,\mathfrak{p}_{i,j}(t-s;u) := \mathbb{P}(M(t-s) = j \,|\, M(0) = i)\,, i,j \in \mathbb{N}_0\,\big)$$

$$\equiv \big((e^{(t-s)\boldsymbol{Q}})_{i,j}, i,j \in \mathbb{N}_0\big)\,; \quad t \ge s \ge 0\,.$$

Then we have a Feynman-Kac representation formula

$$\widetilde{X}^{(u)}_t = \mathbb{E}^M\Big[\int_0^t \sum_{k=0}^{\infty} \mathbf{1}_{\{M(t-s)=k\}} \mathrm{d}\, W_{s,k} \,\Big|\, M(0) = 0\Big]\,; \quad t \ge 0\,,$$

where the expectation is taken with respect to the probability induced by the Markov chain $M(\cdot)$, independent of the Brownian motions $(W_{\cdot,k}, k \in \mathbb{N}_0)$.

$$\mathrm{d}\boldsymbol{X}_t = \boldsymbol{Q}\boldsymbol{X}_t\,\mathrm{d}t + \mathrm{d}\boldsymbol{W}_t\,,$$

where $\boldsymbol{X}_\cdot := (X^{(u)}_{\cdot,k}, k \in \mathbb{N}_0)$ with $\boldsymbol{X}_0 = \boldsymbol{0}$, and $\boldsymbol{W}_\cdot := (W_{\cdot,k}, k \in \mathbb{N}_0)$ with the backward Kolmogorov equation

$$\frac{\mathrm{d}}{\mathrm{d}t} e^{t\boldsymbol{Q}} = \boldsymbol{Q}\, e^{t\boldsymbol{Q}}\,; \quad t \geq 0\,.$$

Thus, by Itô's formula we directly verify

$$\mathrm{d}\Big(\int_0^t e^{(t-s)\boldsymbol{Q}}\mathrm{d}\boldsymbol{W}_s\Big) = \Big(\boldsymbol{Q}\int_0^t e^{(t-s)\boldsymbol{Q}}\mathrm{d}\boldsymbol{W}_s\Big)\,\mathrm{d}t + \mathrm{d}\boldsymbol{W}_t\,; \quad t \geq 0\,,$$

and hence

$$\boldsymbol{X}_t = \int_0^t e^{(t-s)\boldsymbol{Q}}\,\mathrm{d}\boldsymbol{W}_s\,; \quad t \geq 0\,,$$

is a solution.

- We may compute the variance-covariances.

- Although Q has the specific form here, it is easy to see that in general, the Feynman-Kac formula still holds for the infinite-dimensional OU process with a class of generators Q which form a Banach algebra
(e.g., the generator of the discrete-state, compound Poisson processes, Friedman ('71)).

- More generally, it is connected to stochastic evolution equation (see e.g., Dawson ('72), Da Prato & Zabczyk ('92), Kallianpur & Xiong ('95), Batt, Kallianpur, Karandikar, & Xiong ('98), Athreya, Bass & Perkins ('05) for more general results in Hilbert spaces).
For more recently elaborated work in the similar direction see Chong & Klüppelberg ('12), Ramanan ('18).

Asymptotic dichotomy

The asymptotic behaviors of their variances as $t \to \infty$ are dichotomous

$$\mathrm{Var}(X_t^{(u)}) = \int_0^t e^{-2v} I_0(2uv)\mathrm{d}v = \begin{cases} O(1), & u \in [0,1), \\ O(\sqrt{t}), & u = 1, \end{cases}$$

with $$\mathrm{Var}(X_t^{(0)}) = \frac{1-e^{-2t}}{2},$$

$$\mathrm{Var}(X_t^{(1)}) = te^{-2t}(I_0(2t) + I_1(2t))$$

for $t \geq 0$.

Weighted by Poisson probabilities

- An interpretation of

$$X_1^{\dagger}(t) = \sum_{k=0}^{\infty} \int_0^t e^{-(t-v)} \cdot \frac{(t-v)^k}{k!} \mathrm{d}B_{k+1}(v)$$

$$=: \sum_{k=0}^{\infty} \int_0^t \mathfrak{p}_k(t-v) \mathrm{d}B_{k+1}(v)$$

for $t \geq 0$:

Suppose $N(s)$, $0 \leq s \leq t$ is a Poisson process with rate 1, independent of $(B_k(\cdot), k \in \mathbb{N})$. Then

$$X_1^{\dagger}(t) = \mathbb{E}\Big[\sum_{k=0}^{\infty} \int_0^t \mathbf{1}_{\{N(t-v)=k\}} \mathrm{d}B_{k+1}(v) \Big| \mathcal{F}(t) \Big],$$

where $\mathcal{F}(t) := \sigma(B_k(s), 0 \leq s \leq t, k \in \mathbb{N})$, $t \geq 0$.

If we replace the Poisson probability by compound Poisson probability, i.e.,

$$\widetilde{N}(t) := \sum_{k=1}^{N(t)} \xi_k \, ,$$

where $(\xi_k, k \in \mathbb{N})$ are I.I.D. integer-valued R.V.'s with $\mathbb{P}(\xi_1 = i) = p_i$, $1 \le i \le q$, $\sum_{i=1}^{q} p_i = 1$ for some $q \in \mathbb{N}$, independent of $N(\cdot)$ and $(B_k(\cdot), k \in \mathbb{N})$, then

$$\widetilde{X}_1^{\dagger}(t) := \mathbb{E}\Big[\sum_{k=0}^{\infty} \int_0^t \mathbf{1}_{\{\widetilde{N}(t-u)=k\}} \mathrm{d}B_{k+1}(u) \Big| \mathcal{F}(t) \Big]$$

$$= \sum_{k=0}^{\infty} \int_0^t \widetilde{\mathfrak{p}}_k(t-u) \mathrm{d}B_{k+1}(u) \, ,$$

where

$$\widetilde{\mathfrak{p}}_k(t) := \frac{\partial^k}{\partial z^k} \Big[\exp \Big(\sum_{i=1}^{q} p_i t (z^i - 1) \Big) \Big] \Big|_{z=0}$$

for $k \in \mathbb{N}$, $t \ge 0$

corresponds to the modified matrix

$$\widetilde{Q} := \begin{pmatrix} -1 & p_1 & p_2 & \cdots & p_q & 0 & \cdots \\ 0 & -1 & p_1 & p_2 & \cdots & p_q & \ddots \\ & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \end{pmatrix},$$

and

$$\mathrm{d}X_k^{\dagger}(t) = \Big(- X_k^{\dagger}(t) + \sum_{i=1}^{q} p_i X_{i+k}^{\dagger}(t) \Big) \, \mathrm{d}t + \mathrm{d}B_k(t)$$

with $X_k^{\dagger}(0) = 0$ for $k \in \mathbb{N}$, $t \geq 0$.

- In particular, if $q = 2$, $p_1 = p_2 = 1/2$, then

$$\widetilde{\mathfrak{p}}_k(t) = \sum_{j=0}^{\lfloor k/2 \rfloor} \frac{e^{-t} t^{k-j}}{2^{k-j}(k-2j)!\, j!} ; \quad t \geq 0, k \in \mathbb{N}.$$

If we replace Poisson process by the $\mathbb{Z}$-valued, continuous-time Markov chains $M(\cdot)$, we obtain

- Another example: if we take

$$\widetilde{Q} := \begin{pmatrix} \ddots & \ddots & \ddots & & & \\ & 1/2 & -1 & 1/2 & & \\ & & 1/2 & -1 & 1/2 & \\ & & & \ddots & \ddots & \ddots \end{pmatrix}$$

then the solution can be represented by

$$\widetilde{X}_i^{\dagger}(t) = \mathbb{E}\Big[\sum_{k=-\infty}^{\infty} \int_0^t \mathbf{1}_{\{M_i(t-s)=k\}} \mathrm{d}B_k(s) \Big| \mathcal{F}(t) \Big], \quad i \in \mathbb{Z}$$

with continuous-time, simple symmetric random walk $M_i(\cdot)$ on $\mathbb{Z}$ with $M_i(0) = i \in \mathbb{Z}$.

- Connection to SPDEs with a discrete Laplacian for this case, while for the OU case it is connected to a discrete gradient.

- Another modification: now with repulsions we consider

$$
\begin{aligned}
\mathrm{d}X_1(t) &= (X_1(t) - X_2(t))\mathrm{d}t + \mathrm{d}B_1(t)\,, \\
\mathrm{d}X_2(t) &= (X_2(t) - X_3(t))\mathrm{d}t + \mathrm{d}B_2(t)\,, \\
&\cdots
\end{aligned}
$$

We may use the same reasoning in this case to obtain

$$
X_1(t) = \int_0^t \sum_{k=0}^{\infty} e^{t-s} \cdot \frac{(-1)^k (t-s)^k}{k!} \mathrm{d}B_{k+1}(s)
$$

with exponentially growing variance

$$
\mathrm{Var}(X_1(t)) = te^{2t}(I_0(2t) - I_1(2t))\,; \quad t \geq 0\,.
$$

Now let us take $\mathcal{F}_{\cdot} = (\boldsymbol{\sigma}(X_s, \widetilde{X}_s), 0 \le s \le \cdot)$.
Thanks to Girsanov theorem, the log Radon-Nykodim derivative of the solution $\mathbb{P}^{(u)}$ w.r.t. the Wiener measure $\mathbb{P}_0$ is

$$\log \frac{\mathrm{d}\mathbb{P}^{(u)}}{\mathrm{d}\mathbb{P}_0}\Big|_{\mathcal{F}_T} = \int_0^T (X_t - u\widetilde{X}_t)\mathrm{d}X_t + \frac{1}{2}\int_0^T (X_t - u\widetilde{X}_t)^2 \mathrm{d}t\,.$$

Thus given $\mathcal{F}_T^X$, the observer may maximizes the conditional log likelihood function

$$\mathbb{E}\Big[-\log\Big(\frac{\mathrm{d}\mathbb{P}^{(u)}}{\mathrm{d}\mathbb{P}_0}\Big|_{\mathcal{F}_T}\Big) \Big| \mathcal{F}_T^X \Big]$$

with respect to u, and formally obtain a unique maximizer

$$\widehat{u} := \Big(\int_0^T \mathbb{E}[\widetilde{X}_t^2 | \mathcal{F}_T^X] \mathrm{d}t \Big)^{-1} \cdot \mathbb{E}\Big[\int_0^T X_t \widetilde{X}_t \mathrm{d}t + \int_0^T \widetilde{X}_t \mathrm{d}X_t \,\Big|\, \mathcal{F}_T^X \Big]$$

as an estimator of u.

If we replace $\widetilde{X}_.$ by $X_.$, then we obtain a modified estimator

$$\widehat{u}_m := \Big(\int_0^T X_t^2 \mathrm{d}t\Big)^{-1} \cdot \Big(\int_0^T X_t^2 \mathrm{d}t + \int_0^T X_t \mathrm{d}X_t\Big)$$

$$= 1 - \Big(2\int_0^T X_t^2 \mathrm{d}t\Big)^{-1} (T - X_T^2).$$

$$\lim_{T\to\infty} \widehat{u}_m = 1 - \sqrt{1-u^2} \le u \in [0,1].$$

Another typical method of estimation of u is known as the method of moments. We may obtain the method of moments estimator by matching the second moment in the limit, i.e.,

$$\widehat{u}_M = \Big[1 - \Big(\frac{2}{T}\int_0^T X_t^2 \mathrm{d}t\Big)^{-1/2}\Big]^{1/2}.$$

$$\lim_{T\to\infty} \widehat{u}_M = u \in [0,1].$$

Thus this method of moments estimator $\widehat{u}_m$ is asymptotically consistent to the value u as $T \to \infty$.

A Discrete-time version — modified AR(1)

Let us consider its discrete-time version, namely, for $a \in (0,1)$, $u \in [0,1]$ the solution pair $(X_n, \widetilde{X}_n)$, $n = 0,1,2,\ldots$ which satisfies the following recursive equation for $n = 1,2,\ldots$

$$\boxed{X_n = aX_{n-1} + (1-a)(u\widetilde{X}_{n-1} + (1-u)\mathbb{E}[X_{n-1}]) + \varepsilon_n}$$

where we assume $\mathrm{Law}(X_\cdot) \equiv \mathrm{Law}(\widetilde{X}_\cdot)$ and $\widetilde{X}_\cdot$ is independent of the noise ε_n, $n \in \mathbb{N}$. We shall find the joint distribution of $(X_n, \widetilde{X}_n)$, $n \in \mathbb{N}$.

For simplicity, let us assume $\underline{X_0 = 0 = \widetilde{X}_0}$.

Then it is reduce to $\mathbb{E}[X_n] = \mathbb{E}[\widetilde{X}_n] = 0$ and hence

$$X_n = aX_{n-1} + (1-a)u\widetilde{X}_{n-1} + \varepsilon_n\,; \quad n = 1,2,\ldots$$

with distributional constraints $\mathrm{Law}(X_\cdot) \equiv \mathrm{Law}(\widetilde{X}_\cdot)$.

Recursively substituting from $X_0 = \widetilde{X}_0 = 0$, we have

$$X_1 = \varepsilon_1 , \quad X_2 = aX_1 + (1-a)u\widetilde{X}_1 + \varepsilon_2 ,$$

$$X_3 = aX_2 + (1-a)u\widetilde{X}_2 + \varepsilon_3 = \cdots ,$$

$$X_n = u \sum_{k=1}^{n-1} a^{k-1}(1-a)\widetilde{X}_{n-k} + \sum_{k=0}^{n-1} a^k \varepsilon_{n-k} .$$

Thanks to the constraints $\mathrm{Law}(X_\cdot) \equiv \mathrm{Law}(\widetilde{X}_\cdot)$, we solve

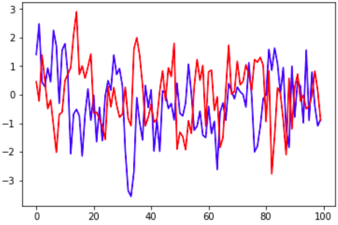
$$\boxed{\begin{aligned} X_n &= \sum_{0 \le \ell \le k \le n-1} \binom{k}{\ell} u^\ell (1-a)^\ell a^{k-\ell} \varepsilon_{n-k,\ell} , \\ \widetilde{X}_n &= \sum_{0 \le \ell \le k \le n-1} \binom{k}{\ell} u^\ell (1-a)^\ell a^{k-\ell} \varepsilon_{n-k,\ell+1} ; \quad n \in \mathbb{N} , \end{aligned}}$$

where $\varepsilon_{n,m}$, $n, m \in \mathbb{N}$ are independently, identically distributed noise with $\varepsilon_{n,0} = \varepsilon_n$, $n \in \mathbb{N}$.

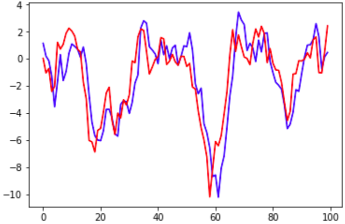
Simulated sample path $(X_n, \widetilde{X}_n)$

$$X_n = aX_{n-1} + (1-a)u\widetilde{X}_{n-1} + \varepsilon_n \,; \quad n = 1, 2, \ldots$$

with distributional constraints $\mathrm{Law}(X_\cdot) \equiv \mathrm{Law}(\widetilde{X}_\cdot)$.



$u = 0.2$



$u = 1$

The sample path $(X_n, \widetilde{X}_n)$, $n = 1, \ldots, 100$
with $a = 0.5$, $X_0 = 0 = \widetilde{X}_0$ and I.I.D. $N(0,1)$ noise $\varepsilon_\cdot$.

Asymptotic dichotomy

The variance-covariances can be calculated, e.g.,

$$\mathbb{E}[X_n^2] = \sum_{k=0}^{n-1}\sum_{\ell=0}^{k}\binom{k}{\ell}^2 u^{2\ell}(1-a)^{2\ell}a^{2(k-\ell)}$$

$$= \sum_{k=0}^{n-1} u^k(1-a)^k\,{}_2F_1\Big(-k,-k,1;\frac{a^2}{(1-a)^2}\Big),$$

$$\mathbb{E}[X_n\widetilde{X}_n] = \sum_{k=1}^{n-1}\sum_{\ell=1}^{k}\binom{k}{\ell}\binom{k}{\ell-1}u^{2\ell-1}(1-a)^{2\ell-1}a^{2(k-\ell)+1}$$

for $n\in\mathbb{N}$. Here ${}_2F_1(\cdot)$ is the Gauss hypergeometric function. Particularly, we have as $n\to\infty$

$$\mathrm{Var}(X_n) = \begin{cases} O(1) & \text{if } u\in[0,1)\,,\ a\in[0,1)\,,\\ O(\sqrt{n}) & \text{if } u=1\,,\ a\in(0,1)\,,\\ O(n) & \text{if } u=1\,,\ a=0\,,\\ & \text{or if } u\in(0,1]\,,\ a=1\,. \end{cases}$$

Another formulation — Cointegration

Another simple discrete-time version we may consider

$$X_n - a u \widetilde{X}_n = \alpha (X_{n-1} - a u \widetilde{X}_{n-1}) + \varepsilon_n \, ; \quad n = 1, 2, \ldots$$

with distributional constraints $\mathrm{Law}(X_\cdot) \equiv \mathrm{Law}(\widetilde{X}_\cdot)$. We assume $\widetilde{X}_\cdot$ is independent of the noise ε_n , $n \in \mathbb{N}$.

This is a type of cointegration model between $(X_\cdot, \widetilde{X}_\cdot)$, i.e.,

$$Y_n = \alpha \, Y_{n-1} + \varepsilon_n \, , \quad Y_n := X_n - a u \widetilde{X}_n$$

for $n \in \mathbb{N}$.

We assume $|a| \le 1$, $|\alpha| \le 1$, $u \in [0, 1]$.

Under the assumption $X_0 = 0 = \widetilde{X}_0$, $|a| < 1$, $|\alpha| \in (0, 1]$, $u \in [0, 1]$ the solution can be represented as

$$X_n = \sum_{\ell=0}^{\infty} \sum_{k=0}^{n} \alpha^k (au)^\ell \varepsilon_{n-k,\ell},$$

$$\widetilde{X}_n = \sum_{\ell=0}^{\infty} \sum_{k=0}^{n} \alpha^k (au)^\ell \varepsilon_{n-k,\ell+1}.$$

This case provides stationary series if $|au| < 1$, $|\alpha| < 1$. If $|a| < 1$ with $u = 1 = \alpha$, the variance grows linearly like a random walk.

- Ongoing project: Statistical inference problems, unit root test, partial information.

Optimal stopping problem and reflected BSDE

Given the solution $(X_t^{(u)}, 0 \le t \le T)$ to the infinite dimensional McKean-Vlasov equation, we shall consider the following reflected backward stochastic differential equation

$$Y_t = \xi + \int_t^T f(s, Y_s, Z_s)\mathrm{d}s + K_T - K_t - \int_t^T Z_s \cdot \mathrm{d}B_s + M_T - M_t \,,$$

$$Y_t \ge L_t \,; \quad 0 \le t \le T \,, \quad \int_0^T (Y_t - L_t)\mathrm{d}K_t = 0 \,,$$

where $K.$ is adapted, increasing, cont., $M.$ is a cont., local martingale orthogonal to the BM $B.$ with $M_0 = 0$, and

$$L_t := h(X_t^{(u)}) \,; \; 0 \le t \le T \,, \quad \xi := g(X_T^{(u)})$$

with a measurable function g with at most linear growth, a continuous function h with at most linear growth with $g(\cdot) \ge h(\cdot)$.

We define

$$S^2_{\mathbf{F}} := \{(Y_t)_{0 \le t \le T}, \text{ continuous, adapted } \mathbb{E}[\sup_{0 \le t \le T} |Y_t|^2] < \infty\},$$

$$L^2_{\mathbf{F}} := \{(Z_t)_{0 \le t \le T}, \text{progressively meas.}, \mathbb{E}\Big[\int_0^T |Z_t|^2 \mathrm{d}t\Big] < \infty\},$$

$$M^2_{\mathbf{F}} := \{(K_t)_{0 \le t \le T}, \text{adapted, cont., increasing}, \mathbb{E}[K_T^2] < +\infty\},$$

and assume that there is a constant $C > 0$ such that for every $t, y_1, y_2 \in \mathbb{R}$, $z_1, z_2 \in \ell^2(\mathbb{R})$

$$|g(t, y_1, z) - g(t, y_2, z_2)| \le C(|y_1 - y_2| + \|z_1 - z_2\|_2)$$

and $g(\cdot, 0, 0) \in L^2_{\mathbf{F}}$.

Proposition. The reflected BSDE admits a unique adapted solution (Y, Z, K, M) in $S^2_{\mathbf{F}} \times L^2_{\mathbf{F}} \times M^2_{\mathbf{F}} \times L^2_{\mathbf{F}}$.

The solution $Y_\cdot$ is the value function of optimal stopping time problem

$$Y_t = \operatorname*{ess\,sup}_{\tau \in \mathcal{T}_t} \mathbb{E}\Big[\int_t^\tau f(s, Y_s, Z_s) \mathrm{d}s + h(X_t^{(u)}) \cdot \mathbf{1}_{\{\tau < T\}} + g(X_T^{(u)}) \cdot \mathbf{1}_{\{\tau = T\}} \Big| \mathcal{F}_t \Big]$$

for $t \in [0, T]$, where $\mathcal{T}_t$ is the family of stopping times dominated by T and greater than or equal to t.

Proof is based on El Karoui, Peng & Quenez ('97a-b). $\square$

- Extension to the mean-field, (reflected) BSDE of Buckdhan, Djehiche, Li & Peng ('09), Li & Luo ('12).
- Interaction between $(X_\cdot^{(u)}, \widetilde{X}_\cdot^{(u)})$.

Summary:

- Two kinds of systems with different interactions: Directed chain/Cascade or Mean-field interaction
- Interpolation of these two systems leads us an infinite-dimensional McKean-Vlasov equation:
  - Well-posed : Weak solution exists and unique in law.
  - Particle approximation : LLN of joint empirical measure and $\sqrt{n}$ estimate.
- Case of linear functional :
  - Connection to the infinite-dimensional OU process
  - Detection problem.
  - Discrete-time case.
- Optimal stopping and reflected BSDE.

- Detering, Fouque & Ichiba (2018) "An infinite-dimensional McKean-Vlasov stochastic equation" arXiv: 1805.01962 Preprint.

- Part of research is supported by NSF DMS-1313373 and DMS DMS-1615229.

アルゴリズム取引の実際

足立高徳

首都大学東京 大学院経営学研究科

2018 年 8 月 6 日 (月)
松本・経済統計キャンプ 2018
信州大学

はじめに

アルゴリズム取引とは何か？

- コンピュータを使って，つまりプログラムにしたがって自動的に取引を行う手法をアルゴリズム取引という．
- その中でも特に高速に取引を行う手法を高頻度取引あるいは HFT と呼ぶ．
- どのくらい早いか？ 1,000 分の 1 秒 (1 ミリ秒) に数 10 個の注文を出す．

  しかし，今やニューヨーク，ロンドン，シンガポールではマイクロ秒 (100 万分の 1 秒) の世界に突入している．

どのくらい流行っているか？

- 2012 年のデータによると米では全取引の 5 割，欧州では 4 割.
- 2013 年には，日本 (東京証券取引所) では 26%を占めているという研究が発表された.
- しかし，現状では注文がアルゴリズムによるものか，人によるものかの区別は確実には判定できない.
- 最近の実感では，東証の 6，7 割の注文はアルゴによるものと思える.

今までの取引とどこが違うのか？

- 従来人が行ってきた取引を機械に代わってやらせる範囲では，取引のアイデアそのものはそれほど変わらない.
- しかし直観が利用できない.
- そのため，あらかじめ決めておいた取引規則に曖昧な箇所があった場合，曖昧な規則では想定外の状況に対してどんな振る舞いをするかがまったくわからなくなる.
- アルゴリズム取引では曖昧性がまったく許されない.
- どこまでも科学的に取引を行わないといけない.
- Symbolic AI を超えたらどうなるか？

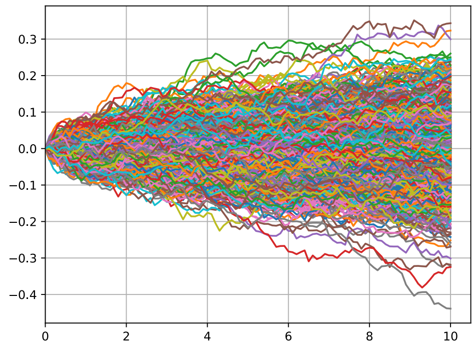
アルゴ・ビジネスの階層

- アルファの特定
- 戦略の策定
- アルゴリズムの実装
- オペレーション

不確実性と投資

投資の肝

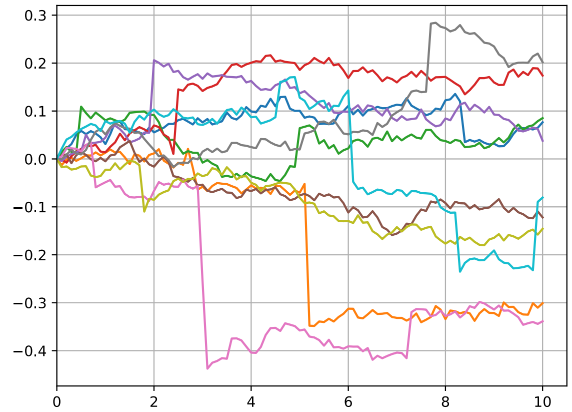
- 将来の不確実性を前に，どうやって身を処するか，というのが投資の肝．
- 不確実性の拡がり（全）情報の拡大 ..



$\ldots, \mathcal{F}_{10}, \ldots, \mathcal{F}_{20.5}, \ldots, \mathcal{F}_{45}, \ldots$

ジャンプがある不確実さ

- 将来の変化は，必ずしも連続とは限らない．



$\ldots, \mathcal{F}_{t-1}, \mathcal{F}_t, \mathcal{F}_{t+1}, \ldots$

状況と制御

- 時刻 t の状況
  $\mathcal{F}_s(s \leq t)$ で決まる変動．(適合過程)

  例：資産価格

- 時刻 t の制御
  $\mathcal{F}_s(s < t)$ で決まる変動．(可予測過程)

  例：取引量

確率微分方程式

状況の変動は確率微分方程式で記述する．

$$\underset{(m\times1)}{d\mathbf{X}^{\mathbf{u}}_t} = \mu(t, \underset{(m\times1)}{\mathbf{X}^{\mathbf{u}}_t}, \mathbf{u}_t)dt + \underset{(m\times n)}{\sigma(t,\mathbf{X}^{\mathbf{u}}_t,\mathbf{u}_t)}\underset{(n\times1)}{d\mathbf{W}_t} + \underset{(m\times\ell)}{\gamma(t,\mathbf{X}^{\mathbf{u}}_t,\mathbf{Y},\mathbf{N}^{\mathbf{u}}_{t-,\mathbf{X}^{\mathbf{u}_t}},\mathbf{u}_t)}\underset{(\ell\times1)}{d\mathbf{N}^{\mathbf{u}}_{t,\mathbf{X}^{\mathbf{u}_t}}}$$

この方程式の解 $\mathbf{X}^{\mathbf{u}}_t$ を引数として持つ関数 $f(t, \mathbf{X}^{\mathbf{u}}_t)$ の積分方法について述べたのが有名な伊藤の公式である．

伊藤清先生 (1915-2008)

HJB 方程式

- 制御方法を考えるのが戦略である．
- はっきりと数値で表された目標があるとする．この数値を表す式を値関数と呼ぶ．

$$H(\underset{(m\times1)}{\mathbf{x}}) = \sup_{\mathbf{u}\in\mathcal{A}} \mathbb{E}\Big[G(\underset{(m\times1)}{\mathbf{X}^{\mathbf{u}}_T}) + \int_0^T F(s, \underset{(m\times1)}{\mathbf{X}^{\mathbf{u}}_s}, \underset{(p\times1)}{\mathbf{u}_s})ds\Big]$$

- HJB 方程式は，値関数（目的関数）を最大化するように制御する確率過程を決定する．
- この値関数を目標にする，という点が後述するディープ・ラーニング（深層学習）に適しているかもしれない．
- しかし，みなが同じ情報を使っていたら，戦略も似たようなものになる．
- 結局，取引コストを上回るような収益を上げるのは難しい．

利益の源泉

- 実際のところ $\mathcal{F}_t$ の情報は使い切れていない．
- そこから，まだ見つけられていない利益の源泉となる情報 (アルファと呼ぶ) を見つけ，
- それを使って
  - 取引速度を変えたり
  - 売買タイミングをずらす

  などの方法で，より高い収益を狙うことが重要になる．
- 有意なアルファの発生イベントは，トリガーと呼ばれる．
- このトリガーによって取引アルゴリズムが参照するパラメタを変動させることになる．

説明変数と被説明変数

- 株価や収益率を表す確率変数を被説明変数という．
- この被説明変数をいくつかの別の確率変数の組み合わせで表現しようと試みたとする．
- 例：日経平均を，S & P500 とドル円為替レートで説明する．

$$r_t^{Nikkei} = \alpha + \beta_{SP500} r_t^{SP500} + \beta_{USD\_YEN} r_t^{USD\_YEN} + \varepsilon_t$$

- このように説明するために用いられる確率変数を説明変数と呼ぶ．
- 説明変数によって推定した被説明変数の推定値と実現値の差 ε_t を誤差と呼ぶ．
- この誤差も確率変数で，その分布は通常平均が 0 である．つまり $\mathbb{E}[\varepsilon_t] = 0$.

アルファとは何ものか？ (1)

- $\mathcal{F}_t$ 適合な確率過程 Y_t のとき，以下のように定義する．

$$\Delta Y_t := \frac{Y_t - Y_{t-h}}{Y_{t-h}}$$

- $h > 0$ は時間地平
- Y_t が証券の価格変動を表しているならば，ΔY_t は直近 h 単位時間の間の収益率
- ΔY_t も，$\mathcal{F}_t$ 適合
- K 個の説明変数で ΔY_{t+h} を線形回帰する．

$$\Delta Y_{t+h} = \alpha + \beta_1 X_{1,t-} + \beta_2 X_{2,t-} + \cdots + \beta_K X_{K,t-} + \varepsilon_t \quad (1)$$

- (1) の右辺は可予測過程
- 左辺は適合過程ですらない

アルファとは何ものか？ (2)

- (1) のような上手い近似を得られたとき，アルファを特定した，という．
- アルファを導く説明変数は，取引の時間地平によって探し方が大きく異なる．
- （まっとうな）アルファは可予測過程である．
- つまり制御に用いることができる．

超短期アルファ

HFT でのアルファのソースは？

- 板情報 (Limit Order Book)
- 高頻度テキストデータ (Twitter など)

板情報 (LOB)

HFT では，アルファは板情報から探すことが多い．

09:00:08.280

100(1)	1006	
300(2)	1005	
1500(2)	1004	
400(1)	1003	
200(1)	1002	
500(2)	1001	
1000(1)	1000	
	999	
	998	1500(4)
	997	400(1)
	996	700(2)
	995	900(3)
	994	1000(1)

Figure: ザラバ中の板情報

LOB を利用した説明変数の具体例 (1)

- **Mid price:** P_i
- **Ask side:** $\{P^a_{i,\ell}, s^a_{i,\ell}\}_\ell$ where $P_i < P^a_{i,1} < P^a_{i,2} < \ldots$
- **Bid side:** $\{P^b_{i,\ell}, s^b_{i,\ell}\}_\ell$ where $P_i > P^b_{i,1} > P^b_{i,2} > \ldots$

$$x_i := \frac{\sum_\ell s^b_{i,\ell} e^{-\lambda_i (P_i - P^b_{i,\ell})}}{\sum_\ell s^a_{i,\ell} e^{-\lambda_i (P^a_{i,\ell} - P_i)}} \tag{2}$$

- 多重線形性に注意しながら，こうした説明変数を 5 個程度 (それより多いのはいけない，つまり $K <= 5$) 探す．
- 実際はパラメタ λ_i の値を決めるところが大変難しい．シミュレーションの計算時間との妥協の産物になることが多い．
- しかし一旦 λ_i が決まったなら，少なくとも 1 ヶ月は動かしてはいけない．

Momentum を考慮した説明変数

クロスセクショナルな説明変数 $x_{i,t}$ が与えられたとする．
この時，減衰率 μ_i を使って

$$y_{i,t} := \int_0^t e^{-\mu_i (t-s)} x_{i,s} ds$$

と定義する．(再帰的計算に向いた手法)

- λ_i や μ_i を時変 (確率変数) にして，もっと原始的なパラメタで表現する．
- さらにその原始的なパラメタを遺伝的アルゴリズムで求める．
- . . .

LOB に関する最近の研究

- [Minami et al., 2008]
  日経 225 の銘柄の LOB をベクトルとみなし，無作為に抽出した 10000 ティックをクラスター分析により 768 通りのパターンに分類．パターンの時系列変化から価格変動確率分布を計算した．
- [Cont et al., 2010], [Vinkovskaya, 2014]
  LOB の各板を待ち行列とみなし，点過程モデルで表現した．
- [Kato et al., 2014]
  日経先物の 1 秒毎の LOB から複合 Poisson 分布と仮定した各板の注文到着強度を最尤法により推定し，それらが最良気配からの乖離に対して冪乗則があることを示した．その上で MI 関数が S 形状であることを示した．

つぎの価格変化が上昇である確率 ([Cont et al., 2010])

$$P^{\mathrm{up}}_{\mathbf{x}} = \mathcal{L}^{-1}\{\hat{F}_{U_{\mathbf{x}}}(s)\}(0).$$

ただし

$$\begin{aligned}\hat{F}_{U_{\mathbf{x}}}(s) :=& \frac{1}{s}\Big\{(1 - \frac{\Lambda_{\mathbf{x}}}{\Lambda_{\mathbf{x}} + s})\hat{f}_{\sigma^A_{\mathbf{x}}}(\Lambda_{\mathbf{x}} + s) + \frac{\Lambda_{\mathbf{x}}}{\Lambda_{\mathbf{x}} + s}\Big\} \\ & \Big\{(1 - \frac{\Lambda_{\mathbf{x}}}{\Lambda_{\mathbf{x}} - s})\hat{f}_{\sigma^B_{\mathbf{x}}}(\Lambda_{\mathbf{x}} - s) + \frac{\Lambda_{\mathbf{x}}}{\Lambda_{\mathbf{x}} - s}\Big\},\end{aligned}$$

$$\hat{f}_{\sigma^A_{\mathbf{x}}}(s) := \Big(\frac{-1}{\lambda(k^S(\mathbf{x}))}\Big)^{|x_{k^A(\mathbf{x})}|} \prod_{i=1}^{|x_{k^A(\mathbf{x})}|} \Phi^{\infty}_{j=i} \frac{-\lambda(k^S(\mathbf{x}))(\mu + j\theta(k^S(\mathbf{x})))}{\lambda(k^S(\mathbf{x})) + \mu + j\theta(k^S(\mathbf{x})) + s},$$

$$\hat{f}_{\sigma^B_{\mathbf{x}}}(s) := \Big(\frac{-1}{\lambda(k^S(\mathbf{x}))}\Big)^{|x_{k^B(\mathbf{x})}|} \prod_{i=1}^{|x_{k^B(\mathbf{x})}|} \Phi^{\infty}_{j=i} \frac{-\lambda(k^S(\mathbf{x}))(\mu + j\theta(k^S(\mathbf{x})))}{\lambda(k^S(\mathbf{x})) + \mu + j\theta(k^S(\mathbf{x})) + s},$$

$$\Lambda_{\mathbf{x}} := \sum_{1 \le j < k^S(\mathbf{x})} \lambda(j).$$

Cont チームの仕事

待ち行列理論 (queueing theory) を使う．([Cont et al., 2010])

- 指値注文，成行注文，キャンセルなどをイベント型と呼ぶ．
- Bid/Ask の各レベル (板) 毎にイベント型 i のイベントがやってくる到着時刻 T_n^i を定式化する．
- つまり，イベント型 i と時刻 t 毎に intensity $\lambda_i(t)$ を持つポアソン過程と考える．(Hawkes 過程)

$$\lambda_i(t) := \theta_i + \sum_{j=1}^{J} \delta_{ij} \sum_{T_n^j \le t} e^{-\kappa_i (t - T_n^j)}$$

- さらに励起感応度 δ_{ij} や減衰率 κ_i を bid-ask スプレッドに依存させるなどの工夫もある．([Vinkovskaya, 2014])

こうした理論 (LOB) モデルは，説明変数の特定に利用できる．
が，それだけではない．．

LOB 理論モデルがあると何が嬉しいか？

- 従来のシミュレーション方法
  - 過去データ (historical data) に対して，直接アルファ・モデルをキャリブレートしていく．
  - 過去データは大きく，アルファ・モデルのパラメタ数は多いので，大量の時間がかかる．
  - あるいは，十分な種類のパラメタ値を試せない．
- LOB 理論モデルを使ったシミュレーション方法
  - 過去データに対して，LOB 理論モデルをキャリブレートしていく．
  - 過去データは大きいが，パラメタ数は比較的少ないので，モデルさえ良ければよい近似が得られる．
  - LOB 理論モデルに対して，アルファ・モデルをキャリブレートしていく．
  - もし解析的手法が使えるならば，LOB 理論モデルに対するシュミレーションは高速に行えるので，パラメタ数が多いアルファ・モデルも十分な種類のパラメタ値で試せる．
  - 過去データで直接キャリブレートすることに対する不安定性を緩和できる．

戦略

アルファの有効利用

有意なアルファを持っていても，戦略がミスマッチならば，折角のアルファの力も十分に発揮できない．

- 注文送出のシグナルとしてのみ利用するわけではない．
- 取引コストと期待リターンのトレードオフ．(bid-ask スプレッド・コスト)
- (一般に）市場インパクトの最小化を狙う．(permanent and/or temporary impact)
- 情報隠蔽
- 特に株価指数裁定取引では，トラッキング・エラー Δ_t をコストを抑えながら小さくするのに欠かせない．

ペア・トレード

- co-move するふたつのアセット S^1 と S^2 がある．
- 以下の X_t が 0 の周りで変動するように β_1 と β_2 を決める．

$$X_t := \beta_1 S_t^1 + \beta_2 S_t^2$$

ただし $\beta_1\beta_2 < 0$.

- T を時間地平とする．$t < T$ に対して

$$\mathcal{T}[t, T] := \{\tau : \Omega \to [t, T] \mid \mathbb{F} - \text{停止時刻}\}$$

- $\tau \in \mathcal{T}[t, T]$ に対して，作動基準 (*performance criterion*)

$$H^\tau(t, x) := \mathbb{E}[e^{-r(\tau-t)}G(X_\tau) \mid X_t = x]$$

- 値関数 (*value function*)

$$H(t, x) := \sup_{\tau\in\mathcal{T}[t,T]} H^\tau(t, x)$$

最適停止問題：理論的解法

- 確率過程 X_t が以下の確率微分方程式を満たすとする．

$$dX_t = \mu(t, X_t)dt + \sigma(t, X_t)dW_t$$

- OU 過程としてもよい．

定理 (DPE [Touzi, 2013])

$H \in \mathcal{C}([0, T], \mathbb{R})$，$G : \mathbb{R} \to \mathbb{R}$ が連続関数とすると，任意の $(t, x) \in [0, T] \times \mathbb{R}$ に対して以下が成り立つ．

$$\left\{\left(\left(\frac{\partial}{\partial t} + \mathcal{L}_{t,x} - r\right)H(t, x)\right) \vee \left(G(x) - H(t, x)\right)\right\} = 0$$

ただし

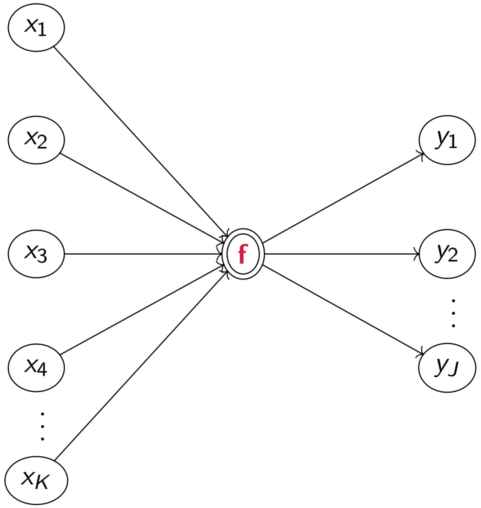
$$\mathcal{L}_{t,x} := \mu(t, t)\frac{\partial}{\partial x} + \frac{1}{2}\sigma^2(t, x)\frac{\partial^2}{\partial x^2}.$$

統計的裁定取引：現場での解法

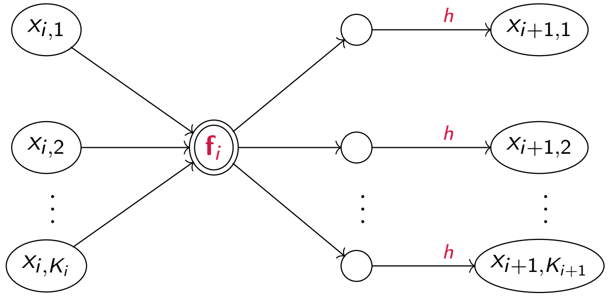
- S^1 と S^2 のアルファ，α^1 と α^2 を特定しているとする．
- $\tilde{X}_t := \beta_1 S_t^1(1 + \alpha_t^1) + \beta_2 S_t^2(1 + \alpha_t^2)$
- $\alpha_t := \tilde{X}_t - X_t$ をシグナルとして，最適停止時刻を決定する．

教師あり学習のおさらい

ニューロン $\mathbf{f} : \mathbb{R}^K \to \mathbb{R}^J$



層 $\mathbb{R}^{K_i} \to \mathbb{R}^{K_{i+1}}$

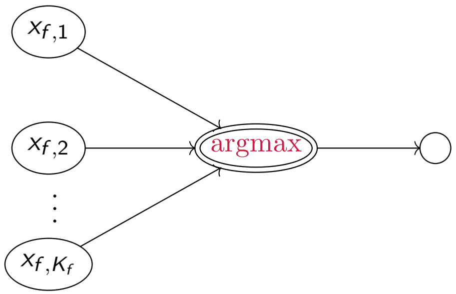


ここで h は，いわゆる活性化関数 (*activation function*) と呼ばれる関数で，このノートでは出力スコアを $[0,1]$ 区間に標準化するために，シグモイド関数を使う．

ニューラル・ネットワークの基本構造

ニューラル・ネットワーク (*neural network*) は，層を重ねたもので，その最終段の層は，予測か分類か，という目的によって異なったものになる．

分類の場合の最終段層



つまり，一番高確率のクラスを選ぶわけだが．．

アフィン層

関数 $\mathbf{f}_i$ が以下のような形をしている時，その層はアフィン層 (*affine layer*) と呼ばれる．

$$\underset{(K_{i+1}\times 1)}{\mathbf{x}_{i+1}} := \underset{(K_{i+1}\times K_i)}{\mathbf{W}_i^T}\underset{(K_i\times 1)}{\mathbf{x}_i} + \underset{(K_{i+1}\times 1)}{\mathbf{b}_i}$$

言い換えれば

$$\underset{(1\times K_{i+1})}{\mathbf{x}_{i+1}^T} := \underset{(1\times K_i)}{\mathbf{x}_i^T}\underset{(K_i\times K_{i+1})}{\mathbf{W}_i} + \underset{(1\times K_{i+1})}{\mathbf{b}_i^T}.$$

$\mathbf{b}_i$ を $\mathbf{W}_i$ の第 1 列に吸収させることによってバイアス項を消去できるが，実はこれはそれほど素晴らしいアイデアではない．というのも，いくつかのアフィン層を重ねていくと新たな複雑さが導入されてしまうのだ．

ソフトマックス層

関数 $\mathbf{f}_i : \mathbb{R}^{K_i} \to \mathbb{R}^{K_i}$ (つまり，$K_{i+1} = K_i$) は，以下のように定義されているとき，ソフトマックス層 (*softmax layer*) と呼ばれる．

$$x_{i+1,j} := \frac{e^{x_{i,j}}}{\sum_{k=1}^{K_i} e^{x_{i,k}}}.$$

明らかに，ソフトマックス関数 $\mathbf{f}_i$ は，K_i 個の点に散らばる確率分布を与える．

さらに偏導関数は以下で与えられる点に注意されたい．

$$\frac{\partial x_{i+1,j}}{\partial x_{i,k}} = \begin{cases} x_{i+1,j}(1 - x_{i+1,j}) \text{ if } j = k, \\ -x_{i+1,k}x_{i+1,j} \text{ if } j \neq k. \end{cases}$$

推定結果の評価

われわれの問題は，入力データに応じて K_f 個のクラスの中のひとつを選ぶ，というものである．このときの教師データは，以下のような $(K_f \times 1)$ 次元のベクトル $\mathbf{t}$ で表現される実現分布となる (one-hot representation).

$$t_k = \begin{cases} 1 \text{ もし } k \text{ が実現クラスならば}, \\ 0 \text{ そうでなければ}. \end{cases}$$

ここで，推定値が $(K_f \times 1)$ 次元のベクトル $\mathbf{p}$ で以下のような性質を持つ，すなわち確率分布とみなせるとする．

$$p_k > 0\,(k = 1, 2, \ldots, K_f), \ \sum_{i=1}^{K_f} p_i = 1.$$

このとき，交差エントロピー誤差 (*cross entropy error*) を以下のように定義する．

$$H(\mathbf{t}, \mathbf{p}) := -\mathbf{t}^T \log(\mathbf{p}).$$

われわれのゴールは，最終段層の出力で観測される $L := H(\mathbf{t}, \mathbf{p})$ を最小化することである．

学習方法

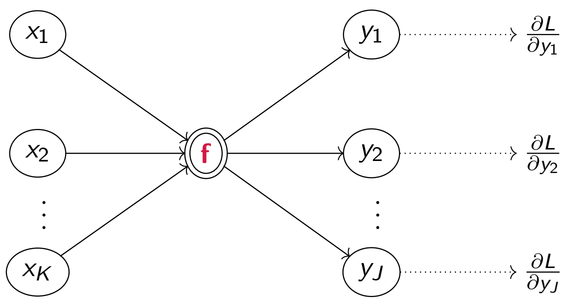
- ニューラル・ネットワークは，L の値に応じて重み行列 $\mathbf{W}_i$ とバイアス・ベクトル $\mathbf{b}_i$ を更新することによって学習する．
- 各ノード x ごとに $\frac{\partial L}{\partial x}$ を計算する．
- アフィン層のなかで $\frac{\partial L}{\partial w_{k,j}}$ と $\frac{\partial L}{\partial b_j}$ を計算する．
- 重みとバイアスを以下のようにして更新する．

$$w_{k,j} := w_{k,j} - \eta \frac{\partial L}{\partial w_{k,j}},$$

$$b_j := b_j - \eta \frac{\partial L}{\partial b_j}$$

ただし η は学習率 (*learning rate*) と呼ばれる定数である．

偏微分係数 $\frac{\partial L}{\partial x}$ の後退伝搬



$$\frac{\partial L}{\partial x_k} = \sum_{j=1}^{J} \frac{\partial L}{\partial y_j} \frac{\partial y_j}{\partial x_k} = \frac{\partial L}{\partial \mathbf{y}^T} \frac{\partial \mathbf{y}}{\partial \mathbf{x}}.$$

計算の必要があるのは，青字の偏微分係数だけ．

最終段層における $\frac{\partial L}{\partial x}$

最終段層 $\mathbb{R}^K \to \mathbb{R}^J$ では，ソフトマックス関数を使って確率分布に変換した後，交差エントロピー誤差を適用してロス関数 L を計算する．今，

$$\mathbf{p} := \mathrm{softmax}(\mathbf{x}), \quad L := H(\mathbf{t}, \mathbf{p})$$

とすると，以下を得る．

$$\frac{\partial p_j}{\partial x_k} = \begin{cases} p_k(1 - p_k) \text{ if } j = k, \\ -p_k p_j \text{ if } j \neq k \end{cases} \qquad \frac{\partial L}{\partial p_j} = -\frac{t_j}{p_j}.$$

すなわち

$$\frac{\partial L}{\partial x_k} = \sum_{j=1}^{J} \frac{\partial L}{\partial p_j} \frac{\partial p_j}{\partial x_k} = p_k \left(\frac{\partial L}{\partial p_k} - \sum_{j=1}^{J} p_j \frac{\partial L}{\partial p_j} \right) = p_k - t_k.$$

したがって

$$\frac{\partial L}{\partial \mathbf{x}} = \mathbf{p} - \mathbf{t}.$$

最終段層の実装

```
def forward(self, x, t):
  self.t = t
  self.y = softmax(x)
  self.loss = cross_entropy_error(self.y, self.t)

  return self.loss

def backward(self, dout=1):
  batch_size = self.t.shape[0]
  if self.t.size == self.y.size:
    dx = (self.y - self.t) / batch_size
  else:
    dx = self.y.copy()
    dx[np.arange(batch_size), self.t] -= 1
    dx = dx / batch_size

  return dx
```


アフィン層 (バッチ版)

$$\underset{(N\times J)}{\mathbf{Y}} := \underset{(N\times K)}{\mathbf{X}} \underset{(K\times J)}{\mathbf{W}} + \underset{(N\times J)}{\begin{bmatrix} \mathbf{b}^T \\ \vdots \\ \mathbf{b}^T \end{bmatrix}}$$

ここで

$$\underset{(N\times J)}{\mathbf{Y}} := \begin{bmatrix} \mathbf{y}_1^T \\ \vdots \\ \mathbf{y}_N^T \end{bmatrix}, \text{ and } \underset{(N\times K)}{\mathbf{X}} := \begin{bmatrix} \mathbf{x}_1^T \\ \vdots \\ \mathbf{x}_N^T \end{bmatrix}.$$

したがって

$$y_{n,j} = \sum_{k=1}^{K} x_{n,k} w_{k,j} + b_j.$$

以下に注意せよ．

$$\frac{\partial y_{n,j}}{\partial x_{n,k}} = w_{k,j}, \quad \frac{\partial y_{n,i}}{\partial w_{k,j}} = \begin{cases} x_{n,k} \text{ if } i = j, \\ 0 \text{ if } i \neq j \end{cases} \qquad \frac{\partial y_{n,i}}{\partial b_j} = \begin{cases} 1 \text{ if } i = j, \\ 0 \text{ if } i \neq j \end{cases}.$$

アフィン層の偏微分係数

$\frac{\partial L}{\partial \mathbf{Y}}$ の値がすでにわかっているとき，以下のようにして他の偏微分係数を計算できる．

$$\frac{\partial L}{\partial x_{n,k}} = \sum_{j=1}^{J} \frac{\partial L}{\partial y_{n,j}} \frac{\partial y_{n,j}}{\partial x_{n,k}} = \sum_{j=1}^{J} \frac{\partial L}{\partial y_{n,j}} w_{k,j}. \text{ Hence } \underset{(N\times K)}{\frac{\partial L}{\partial \mathbf{X}}} = \underset{(N\times J)}{\frac{\partial L}{\partial \mathbf{Y}}} \underset{(J\times K)}{\mathbf{W}^T}.$$

$$\frac{\partial L}{\partial w_{k,j}} = \sum_{n=1}^{N} \sum_{i=1}^{J} \frac{\partial L}{\partial y_{n,i}} \frac{\partial y_{n,i}}{\partial w_{k,j}} = \sum_{n=1}^{N} \frac{\partial L}{\partial y_{n,j}} x_{n,k}.$$

ゆえに

$$\underset{(K\times J)}{\frac{\partial L}{\partial \mathbf{W}}} = \underset{(K\times N)}{\mathbf{X}^T} \underset{(N\times J)}{\frac{\partial L}{\partial \mathbf{Y}}}.$$

$$\frac{\partial L}{\partial b_j} = \sum_{n=1}^{N} \sum_{i=1}^{J} \frac{\partial L}{\partial y_{n,i}} \frac{\partial y_{n,i}}{\partial b_j} = \sum_{n=1}^{N} \frac{\partial L}{\partial y_{n,j}}. \text{ Hence } \underset{(J\times 1)}{\frac{\partial L}{\partial \mathbf{b}}} = \underset{(J\times N)}{\frac{\partial L}{\partial \mathbf{Y}}^T} \underset{(N\times 1)}{\mathbf{1}}.$$

アフィン層の実装

```
def forward(self, x):
  self.x = x
  y = self.x.dot(self.W) + self.b
  return y

def backward(self, dy):
  dx = dy.dot(self.W.T)
  self.dW = self.x.T.dot(dy)
  self.db = np.sum(dy, axis=0)
  return dx
```


教師あり学習を使ったアルファ探索

板情報 (LOB)

HFT では，アルファは板情報から探すことが多い．

```
09:00:00.090500
     126000(23)  199.9
      34000(11)  199.8
      10000( 2)  199.7
      25000( 2)  199.6
      56000(10)  199.5
       7000( 5)  199.4
                 199.3*     30000( 8)
                 199.2      33000(20)
                 199.1     110000(32)
                 199.0     629000(197)
                 198.9     102000(11)
                 198.8      38000( 9)
                 198.7      53000( 8)
```

Figure: ザラバ中の板情報: 6502(東芝) 2017/2/17

板の高頻度データ

time	kehai	ita-4	ita-3	ita-2	ita-1	ita0	ita1	ita2	ita3	it
32400.0905	199.3	-10000	-25000	-56000	-7000	30000	33000	110000	629000	1020(
32400.1048	199.5	-126000	-34000	-10000	-25000	-56000	0	30000	33000	1100(
32400.1513	199.3	-10000	-25000	-56000	0	35000	33000	110000	629000	1020(
32400.1536	199.6	-713000	-126000	-34000	-10000	-25000	0	0	35000	330(
32400.2268	199.6	-713000	-126000	-34000	-10000	-25000	0	0	35000	330(
32400.2276	199.6	-713000	-126000	-34000	-10000	-25000	0	0	35000	330(
32400.2322	199.4	-34000	-10000	-25000	0	5000	35000	33000	110000	6290(
32400.2506	199.6	-713000	-126000	-34000	-10000	-15000	0	5000	35000	330(
32400.2972	199.6	-713000	-126000	-34000	-10000	-15000	0	5000	35000	330(
32400.3106	199.6	-713000	-126000	-34000	-10000	-15000	0	5000	35000	340(
32400.3512	199.6	-713000	-126000	-34000	-10000	-15000	0	5000	35000	340(
32400.3671	199.5	-126000	-34000	-10000	-15000	5000	5000	35000	34000	1100(

1秒ごとの板データ

time	kehai	ita-4	ita-3	ita-2	ita-1	ita0	ita1	ita2	ita3	ita4
32401	199.2	-12000	0	0	0	29000	115000	628000	102000	57000
32402	199.2	-1000	-134000	0	0	5000	115000	629000	102000	38000
32403	199.2	0	-135000	0	0	11000	9000	628000	105000	38000
32404	199.2	0	-134000	0	0	9000	0	626000	102000	38000
32405	199.2	0	-133000	0	0	5000	2000	658000	108000	40000
32406	199.2	-34000	-126000	0	0	5000	2000	659000	111000	40000
32407	199.1	-124000	-6000	0	0	6000	669000	114000	42000	63000
32408	199.1	-124000	-4000	0	0	6000	669000	114000	46000	64000
32409	199.7	-21000	-748000	-126000	-49000	-21000	0	0	0	6000
32410	199.2	0	-25000	-1000	0	19000	9000	666000	112000	56000
32411	199.5	-126000	-43000	0	0	-25000	16000	7000	11000	11000
32412	199.3	-2000	0	-7000	0	20000	16000	9000	669000	112000
32413	199.4	-48000	-3000	0	-9000	-4000	16000	16000	11000	670000

板データから教師データを作成する方法 (1)

- p_i を i 番目の気配値とする．
- 与えれた時間地平 h，たとえば 10 秒，に対して，以下のようにして収益ベクトル $\Delta\mathbf{v}$ と教師ベクトル $\mathbf{v}$ を作る．

$$\Delta r_i := \begin{cases} \frac{p_{i+h}}{p_i} - 1 \text{ if } i + h \le n, \\ 0 \text{ if } i + h > n, \end{cases}$$

$$v_i := \begin{cases} 1 \text{ if } \Delta r_i > 0, \\ 0 \text{ if } \Delta r_i = 0, \\ -1 \text{ if } \Delta r_i < 0. \end{cases}$$

- 残念ながら，これはうまくいかない．
  何故か？

板データから教師データを作成する方法 (2)

- 例えば，流動性の低い銘柄の場合，
- このような短い期間 ($h = 10$ 秒) では，教師ベクトル $\mathbf{v}$ のほとんどすべての要素が中央のケース，つまり 0 になってしまう．
- すると，全要素が 0 のベクトルが大抵の場合最良の予測になってしまう．
- 一旦，ロス関数の出力ベクトルの全要素が 0 になるように重みが調整されてしまうと，もはやそれ以上更新されなくなってしまう．

機械学習の pros と cons

- Pros:
  モデル・フリー．つまり特定の確率分布を前提としない
- Cons:
  テスト期間のレジメが，トレーニング期間のそれとは違うかもしれない．

機械学習は本当に学習しているのか？

- レジメ・チェンジは，トレーニング期間に学んだ経験をぶち壊してしまうかもしれない．
- AI の高速な推論は，レジメ・チェンジのペースを早めるかもしれない．
- 明日に有用な知識は，何も残らないようになってしまうかもしれない．

アルファとリスク管理 (1)

- 統計的分布を知っていれば，テストはできる．
- しかしながら一般的には，NN の状態空間はとてつもなく大きく複雑で，たとえばロス関数の適当な分布を見つけることができない．
- つまり NN を使って生成したアルファをもとにしたアルゴリズム取引に使える信頼できるリスク管理手法を発見することは難しい．

アルファとリスク管理 (2)

“投資とは，(不確実かもしれない) 将来の価値のために，確実な現在の価値を犠牲にすることである”

– ウィリアム・シャープ

1. 確率分布がわかったあとでのリスク計測
2. 分布が定かでないときのリスク計測（モデル・リスク）
3. モデルがあまりに複雑で，リスクを計測できない

リーマン・ショックのときの状況を調べた際には，２番めのリスクが中心課題となった．
しかしながら，AI で駆動されたアルゴリズム取引が市場を支配するようになる近未来では，３番目のリスクが最も大きな問題となるかもしれない．

HFT の現状と未来

ミリ秒の世界

競争相手より，より早く取引せよ．一番速いものがすべての利益の 80%を獲得し，2 番めが 15%を獲り，残りは端数しか得られない．

would sell their grandmothers for a microsecond

スピードを上げる (Latency を減らす) には，以下の 3 つの方法がある．

- Box
  - できるだけ速いマシン．
  - GPU (安価な高速グラフィック・チップ) の利用．
  - FPGA (ゲートレベルでのチップ化) の利用．
- Logic
  - ロボットの高速化
  - nanosec (10^{-9} 秒) への挑戦 (`timespec::tv_nsec` を使う)
- Line
  - ホップ数の削減
  - 配線の直線化 (同じ部屋の中，あるいは建物間など)
  - マイクロ波取引
  - コロケーション (JPX では今はとてもポピュラー)

どんな状況だったのか？

- SIP (証券情報処理機構) が NBBO (全国最良気配値) を提供していたが...
- SIP は大変遅かった．
- 結果，市場の広範囲に渡って「無知」を醸成してしまった．
- この問題は，システムを改善することによって解決できるかに見えた．
- しかしながら，新しいタイプの戦場が現出した．ダークプール (私設取引システム) である．

約定比率の低下が意味すること

- 約定比率が下がってきた (around 2004, 2005)
- TK - 約定速度が極端に遅く，約定が返ってくる前にアルファの賞味期限を超えることが多くなる．(取引所の問題)
- NY - 速度の問題もあるが，より複雑な状況．
  - 彼らは ECN を持っている
  - BATS の約定率がとてもよい．何故？
  - 回線裁定取引 (a.k.a. front running)
- ダークプールの出現により，東京でさえも事態はより複雑になってきた．

統計的 HFT vs. 決定論的 HFT

- **Good HFT**
  統計的 (statistical) HFT. (taking market risk)
  Ex.) electric market makers
- **Bad (or at least Unpopular) HFT**
  決定論的 (deterministic) HFT. (taking no market risk)
  Ex.) front runners (latency artbitragers)

HFT は勝ち続けるというのは本当か？

多くの ECN があることを利用して US で行われているある種の決定論的 HFT ならば可能かもしれない.
が，統計的 HFT ではもちろん不可能.

本当に規制は必要なのか？
(もっと進んだ理論に基づく) テクノロジーで解決できないのか？

AI とは別の HFT の未来

- HFT 性悪説?
  - ハッカーとクラッカー
  - HFT と フロントランニング
- HFT のコモディティ化
  - 時間解像度はどんどん細かくなっていき，今の HFT も「あんなものでも HFT と言っていた時代だったのだな」と振り返られるようになるだろう.
  - HFT を特別の取引として規制する意味がなくなる.

極限の HFT : 時間解像度

- 時間解像度 (取引速度)

HFT	$1 - 100$sec
VHFT	$10^{-3} - 1$sec
UHFT	$10^{-6} - 10^{-3}$sec
SHFT	$10^{-9} - 10^{-6}$sec
...	

- 時間解像度がプランク時間に近づいていったらどうなるか？
  - イベントの前後関係を特定できなくなる？
    (Heisenberg の不確定性原理)
  - Local self-financing property が崩れる？

極限の HFT : 刻み値

- 2003 年の NYSE での刻み値縮小では，取引ボリュームが一気に数割増えた.
- 2014 年 7 月の TSE での刻み値縮小は，大型株ではボリュームが増えたが，板が薄くなり取引がしづらくなった.
- 2015 年 9 月には TSE は，いくつかの銘柄の刻み値を「50 銭」から「1 円」に戻した.
- しかし，将来的には時間解像度の縮小と同様に，刻み値も縮小していくだろう.

極限の HFT：メモリ

- オンメモリで処理する場合
  - 2003 年．NYSE の刻み値縮小の影響で一日の tick データが 2GB を超えた．
  - 当時の 32 ビットマシンで直接アクセスできるアドレス空間を超えたので，姑息な手段で耐える必要があった．
  - 64 ビットマシンのアドレス空間の限界は 8EB (8×10^{18}byte)．
  - すぐに 128 ビットマシンが登場するだろうから，こちらは大丈夫だろう．
- クラウドに置いたデータを処理する場合
  - 通信速度がネックになって問題が発生する (?)

極限の HFT：リスク管理

- 微小時間のリスク管理
  - 2010 年 5 月のフラッシュ・クラッシュ
  - イベントの観測順序の乱れ
  - 量子リスク理論 (?)

アルゴ取引の現状と未来

アルゴ取引の問題点

- 相場操縦と紙一重の戦略をアルゴがとることがある．
- 見せ玉の要件は約定の意思がないことであり，目的達成と同時に見せ玉をキャンセルしていることが重要なポイントである．
- でも，AI に意思があるのか？
- 特に相手もアルゴのときはどうか？
- 逆に人がやった場合でも相手がアルゴのときは？
- アルゴを誘引って何だろう？
- そんなタコなアルゴリズムを書く方が悪いのではない？

アルゴ取引と規制

- 一方，もし人はダメでアルゴは OK ということならば，アルゴと人を区別できなくてはいけない．
- しかし，現状ではこれは難しい．
  究極的には Turing test しかない．でもこれも，人のふりをされたら判定できないだろう．
- それでも，東証ではすでに 6 − 7 割はアルゴと思われるから，この問題は避けて通れない．
- 結局，ゲームなのだから，もはや禁じ手にする合理的な理由は見つけられないのかもしれない．
- 規制は理論に基づいてなされなくてはいけない．
- そうでなければ，いたちごっこになってしまうだけのように思える．

公正な市場

- 人間とアルゴが同じ市場で競争するのは公平か？
- 2016 年暮．Goldman Sachs は 600 名いたトレーダーを 2 名に減らし，代わりに 200 名の計算機エンジニアが保守するアルゴ取引プログラムを雇った．
  (MIT Technology Review 2017 年 2 月 7 日)
- IEX - 350 マイクロ秒遅れの PTS
- ボロ儲けするなど過去の話？

強化学習は有効か？

- 価値 (報酬) 関数がはっきりしていること．
- 高精度よりも低精度 → パターンを探すのが得意
- 例：Twitter のテキストからパターンを探す．
- しかし高精度な数値計算を要求するのにどれだけ有効か？
- アルファ探索には有効だろう．(リサーチ)
- 時間地平が長い戦略にも有効だろう．(リアルタイム)

Alpha Go Zero : ML without Big Data

我々は今まで機械学習（ML）とビッグデータ解析をある意味不可分と考え，それ故 ML には大量の計算リソースが必要と思い込んでいた．
Alpha Go Zero の画期的なところは，（特定の分野における）人類の知識を含む大量のデータを使わずに，アルゴリズムのみで今まで人類が到達不能だった知識（技術）に至った，ということだろう．その副産物として，必要な計算量の減少があるけれど，これは見方を変えれば，我々人類の知識がいかにオーバーヘッドの多い体系であるか，ということだと思う．

例えば，抽象化に関する何らかの指針をルールとして定めることで，まずは群論などの特定の分野から始め，次第に universal mathematics とでも呼ぶべき人類の過去の知見に依存しない数学 (インフラ) を創り出すことができるかもしれない．

教師なし学習の利用と問題点

- エコノミストを AI に置き換える，つまりアクティブ運用のコストを下げる．
- AI の出してきた提言の背後に因果律を探そうという試みは有効か？ (NHK 特集 2017 年 7 月 22 日)
- 因果性がわからない世界を金融業界はどこまで許容できるか？ (超多段式グレンジャー因果律？)
- (経済) 理論のない，データだけの世界．
  誰がそれを信じる？
- 自己勘定でやる分には問題ないが，お客様の資金を種銭とするのであれば顧客説明責任の観点が依然障壁としてある．
- 黒魔術に支配された取引になってしまわないか？

ブレイクスルー

- 量子コンピューティングは，金融市場を効率化する．
- Deep learning のインパクト
- エコノミスト vs. ニューラル・ネットワーク

金融業界の AI 化

- PwC identifies three waves of automation between now and the mid-2030s: algorithm wave, augmentation wave and autonomy wave
- In the first wave up to the early 2020s, relatively few jobs will be automated but **financial services could be relatively highly impacted**
- Up to 30% of existing jobs could be impacted by the mid-2030s, with the transport, manufacturing and retail sectors particularly affected
- There should be broadly offsetting job gains provided investment is made in retraining

(PwC Press-Release 2018 年 2 月 6 日)

正しいはしご?

“このはしごはとても高く，何段あるかわからない．
ただ，AI の歴史は誤ったはしごに登っては下りることの繰り返しだった．
『正しいはしご』にたどり着いた意味は，大きい．”

– デミス・ハサビス (ディープ・マインド)

(日本経済新聞 2017 年 6 月 4 日朝刊)

“人間はミクロにみると，ニューラルネットワークを使っています．一方で，現象でみれば言語を使っているわけです．
言語は意識のもとで成り立っています．
だから論理構造がちゃんとしてないといけない．”

– 甘利俊一 (理化学研究所)

(BITA デジマラボ 2017 年 4 月 16 日)

報われる社会

- アルゴリズム，AI によってスマートになった金融機能は，多くが人の手を離れて AI によって運営される．その結果..
- 金融業に従事する残った人たちも，もはや適切な手数料以上の収入をそこから得ることはできなくなる，
- 製造業や農業，福祉や教育，芸術や文化的活動をはじめとする，いわゆる金融技術を駆使する業種とは異なる実業の人たちが報われる社会が実現する，
- アルゴリズム取引の先に続く将来が，そんな社会に向かうことになれば素晴らしい．

References I

Cont, R., Stoikov, S., and Talreja, R. (2010).
A stochastic model for order book dynamics.
Operations Research, 58(3):549–563.

Goodfellow, I., Bengio, Y., and Courville, A. (2016).
Deep Learning.
MIT Press.

Hagan, M. T., Demuth, H. B., Beale, M. H., and Jesús, O. D. (2016).
Neural network design.
Martin Hagan, 2nd edition.

Kato, T., Ogihara, T., and Takada, H. (2014).
リミット・オーダー・ブックに対する実証分析と金融市場に見られる冪乗則について.
Preprint at FMA2014, Kyoto University.

Minami, R., Kumegawa, M., and Yoon, H. (2008).
株式市場における売買注文状況と日中価格変動の解析.
Preprint at the 22nd Conf. of Japanese Soc. for Artificial Intelligence.

Rojas, R. (1996).
Neural networks.
Springer.

References II

Silver, D. (2017).
Mastering the game of Go without human knowledge.
Nature, 550:354–359.

Touzi, N. (2013).
Optimal Stochastic Control, Stochastic Target Problems, and Backward SDE.
Springer.

Vinkovskaya, E. (2014).
A point process model for the dynamics of limit order books.
Ph.D. thesis, Columbia University.

巣籠悠輔 (2017).
詳解ディープラーニング.
マイナビ出版.

斎藤康毅 (2016).
ゼロから作る *Deep Learning*.
O'Reilly.

足立高徳 (2018).
アルゴリズム取引.
朝倉書店.

宣伝

Local SIML Estimation of Some Brownian Functionals

Naoto Kunitomo [1]

Meiji University

2018 August

[1] A joint work with Seisho Sato.

Outline

Estimation of Brownian Functionals

Let

$$Y(t_i^n) = X(t_i^n) + \epsilon_n v(t_i^n) \ \ (i = 1, \cdots, n) \tag{1}$$

be the (one dimensional) observed (log-)price at t_i^n $(0 = t_0^n \le t_1^n \le \cdots \le t_n^n = 1)$ and $v(t_i^n)$ $(= v_i)$ be a sequence of i.i.d. random variables with $\mathbf{E}[v_i] = 0$ and $\mathbf{E}[v_i^2] = \sigma_v^2$ (> 0). We assume that

$$\epsilon_n = \frac{1}{n^{\delta}} \ , \tag{2}$$

where δ (≥ 0) is a constant. When $\delta = 0$, it is the micro-market noise model, while it is the high-frequency financial model without noise when $\delta = \infty$. When $0 < \delta < \infty$, it corresponds to the small-noise model. The underlying continuous-time Brownian martingale is

$$X(t) = X(0) + \int_0^t \sigma_s dB_s \ (0 \le s \le t \le 1) \ , \tag{3}$$

which is independent of $v(t_i^n)$, σ_s is the (instantaneous) volatility function and B_s is the Brownan motion.

We assume that when the volatility process is is a diffusion type process as

$$\sigma_t^2 = \sigma_0^2 + \int_0^t \mu_s^\sigma ds + \int_0^t \omega_s^\sigma dB_s^\sigma \ (0 \le s \le t \le 1) \ , \qquad (4)$$

where μ_s^σ and ω_s^σ are the drift and diffusion coefficients and B_s^σ is another Brownian motion, which may be correlated with B_s.

The problem of our original interest is how to estimate Brownian functionals of the form

$$V(g, 2r) = \int_0^1 g(s) \sigma_s^{2r} ds \qquad (5)$$

for any positive integer r and a known function $g(s)$ from a set of observations of $Y(t_i^n)$ $(i = 1, \cdots, n)$. We denote $V(2r) = V(g, 2r)$ when $g(s) = 1$ $(0 \le s \le 1)$.

Example 1 : When r=1, the integrated volatility is given by

$$V(2) = \int_0^1 \sigma_s^2 ds \ .$$

Example 2 : The asymptotic variance of the SIML estimator of integrated volatility $V(2)$ is given by

$$2V(4) = 2\int_0^1 \sigma_s^4 ds \ .$$

No-Micro-Market-noise Case

For simplicity, we take $t_j^n - t_{j-1}^n = 1/n$ ($j = 1, \cdots, n$) and $t_0^n = 0$. We divide $(0, 1]$ into $b(n)$ sub-intervals and in every interval we allocate $c(n)$ observations. First, we consider the sequence $c^*(n)$ such that $c^*(n) \to \infty$ and we can take $b(n) \to \infty$ and $b(n) \sim n/c^*(n)$. A typical choice of observations in each interval would be $c^*(n) = [n^\gamma]$ $(0 < \gamma < 1)$, whereupon $b(n) \sim n^{1-\gamma}$. Because there are some extra observations (n is not equal to $b(n)c^*(n)$) and $b(n)$ is a positive integer, we adjust the number of terms in each interval $c(n) = c^*(n) + (\text{several terms})$. We can ignore the effects of extra terms because they are asymptotically negligible and $b(n)c(n) = n$.

When there does not have micro-market noise, we simply use the log-return process $r_j = y(t_j^n) - y(t_{j-1}^n)$ from the log-price process $y(t_j^n)$. We order the data r_j and denote as $r_{k,(i)}$ $(k = 1, \cdots, c(n); i = 1, \cdots, b(n))$.
When $p = 1$, let the $2r-$th moment in the i-th interval be

$$M^*_{2r,(i)} = \sum_{k=1}^{c(n)} [r_{k,(i)}]^{2r} \ .$$

Then we define the local realized moment (LRM) estimator of $V^*(2r)$ by

$$\hat{V}^*(2r) = \frac{n^{r-1}}{a_r} \sum_{i=1}^{b(n)} M^*_{2r,(i)}$$

where

$$a_r = \frac{2r!}{r!\ 2^r} \ .$$

When $r = 1$, it is the realized volatility (RV).

Proposition 1 : Assume that there is no micro-market noise, i.e. $\epsilon_n = 0$ with $p = 1$ and $r \geq 1$. Also assume that $Y(t_i^n) = X(t_i^n)$, $v(t_i^n)$ is a sequence of i.i.d. random variables with $\mathbf{E}[v_i^{4r}] < \infty$ and σ_s $(0 \leq s \leq 1)$ is bounded.
(i) As $n \longrightarrow \infty$

$$\hat{V}^*(2r) - V(2r) \xrightarrow{p} 0 \,. \tag{6}$$

(ii) As $n \longrightarrow \infty$

$$\sqrt{n}\left[\hat{V}^*(2r) - V(2r)\right] \stackrel{\mathcal{L}-s}{\rightarrow} N\left[0, W\right] \,, \tag{7}$$

where $\mathcal{L} - s$ means the stable convergence and

$$W = c_r^* \int_0^1 \left[\sigma_x(s)\right]^{4r} ds \,, \tag{8}$$

where c_r^* $(= a_{2r}/a_r^2 - 1)$ is a positive constant.

Local SIML Estimation

By following Chapter 3 of Kunitomo et al. (2018), we transform $c(n) \times p$ matrix $\mathbf{Y}_{c(n,(i)}$ to $c(n) \times p$ matrix $\mathbf{Z}_{n,(i)}$ $(= (\mathbf{z}_k^{'}(i)))$ $(i = 1, \cdots, b(n))$ by

$$\mathbf{Z}_{c(n),(i)} = h_{c(n)}^{-1/2} \mathbf{P}_{c(n)} \mathbf{C}_{c(n)}^{-1} \left(\mathbf{Y}_{c(n),(i)} - \bar{\mathbf{Y}}_{0,(i)} \right)$$

where $h_{c(n)} = 1/c(n)$, $c(n) \times c(n)$ matrices

$$\mathbf{C}_{c(n)}^{-1} = \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 \\ -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & 0 & \cdots \\ 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & -1 & 1 \end{pmatrix} ,$$

$$\mathbf{P}_{c(n)} = (p_{jk}) \, , \; p_{jk} = \sqrt{\frac{2}{c(n) + \frac{1}{2}}} \cos \left[\frac{2\pi}{2c(n)+1} (k - \frac{1}{2})(j - \frac{1}{2}) \right]$$

The initial conditions are $\bar{\mathbf{Y}}_0(i) = \mathbf{1}_{c(n)} \cdot \mathbf{y}_0^{'}(i)$. and we have the spectral decomposition $\mathbf{C}_{c(n)}^{-1}\mathbf{C}_{c(n)}^{'-1} = \mathbf{P}_{c(n)}\mathbf{D}_{c(n)}\mathbf{P}_{c(n)}^{'}$, where $\mathbf{D}_{c(n)}$ is a diagonal matrix with the k-th element

$$d_k = 2\left[1 - \cos(\pi(\frac{2k-1}{2c(n)+1}))\right] \ (k = 1, \cdots, c(n)).$$

and

$$a_{k,c(n)} = c(n)d_k = 4c(n)\sin^2\left[\frac{\pi}{2}\left(\frac{2k-1}{2c(n)+1}\right)\right] \ (k = 1, \cdots, n) \ .$$

When $p = 1$, let the $2r-$th moment in the i-th interval be

$$M_{2r,(i)} = \frac{1}{m_c} \sum_{k=1}^{m_c} [z_{k,(i)}]^{2r} \ .$$

Then we define the LSIML estimator of $V(2r)$ by

$$\hat{V}(2r) = \frac{b(n)^{r-1}}{a_r} \sum_{i=1}^{b(n)} M_{2r,(i)}$$

where

$$a_r = \frac{2r!}{r!\ 2^r} \ .$$

In particular, $a_1 = 1, a_2 = 3$ and $a_3 = 15$.

Asymptotic Properties of Local SIML

(i) The Case of $r = 1$

First, we consider the asymptotic behavior of the quantity $(1/m_c)\sum_{k=1}^{m_c} z_{k,(i)}^2$. in the i-th interval $I_c(i) = ((i-1)\frac{c(n)}{n}, i\frac{c(n)}{n}]$ $(i = 1, \cdots, b(n))$.

By using the analogous arguments as Chapter 5 of Kunitomo et. al (2018) to the interval $((i-1)\frac{c(n)}{n}, i\frac{c(n)}{n}]$,

$$\sqrt{m_c}\sum_{k,l=1}^{c(n)}[c_{kl}r_k r_l - \delta(k,l)\int_{t_{k-1}^n}^{t_k^n}\sigma_s^2 ds](\frac{n}{c(n)}) = O_p(1) ,$$

where r_k are hidden returns in the i-th interval and $t_k^n - t_{k-1}^n = 1/n$.

The asymptotic bias term for $\sum_{i=1}^{b(n)}[1/m_c]\sum_{k=1}^{m_c} z_{k,(i)}^2$ becomes

$$\mathrm{AB}_n = b(n)\frac{\pi^2}{3}\frac{(m_c)^2}{c(n)}[\epsilon_n]^2\sigma_v^2 \ .$$

Because the normalizing factor of the above terms is $\sqrt{m_c}b(n)$, we find that

$$\mathbf{Var}\left[\frac{1}{m_c}\sum_{k=1}^{m_c} z_k^2(i) - \int_{s\in((i-1)\frac{c(n)}{n},i\frac{c(n)}{n}]}\sigma_s^2 ds\right] = O(\frac{1}{m_c b(n)^2}) \ .$$

Then we have the relation that

$$\left[\sum_{i=1}^{b(n)} \frac{1}{m_c} \sum_{k=1}^{m_c} z_k^2(i)\right] - \int_0^1 \sigma_s^2 ds \xrightarrow{p} 0 ,$$

provided that $\max\{\frac{1}{b(n)}, \frac{1}{m_c}\} \longrightarrow 0$ and $b(n)\frac{(m_c)^2}{c(n)}[\epsilon_n]^2 \longrightarrow 0$. For the asymptotic normality without any asymptotic bias term, a sufficent condition would be

$$\sqrt{m_c b(n)} b(n) \frac{(m_c)^2}{c(n)} [\epsilon_n]^2 \longrightarrow 0 .$$

Proposition 2 : When $r = 1$ and $p = 1$, assume that $v(t_i^n)$ is a sequence of i.i.d. random variables with $\mathbf{E}[v_i^4] < \infty$, σ_s $(0 \leq s \leq 1)$ is bounded and $\alpha^* > 0$ and $\alpha_2^* > 0$.
(i) For $m_c = [c(n)^\alpha]$ and $0 < \alpha < \min\{0.5, \alpha_1^*\}$, as $n \longrightarrow \infty$

$$\hat{V}(2) - V(2) \xrightarrow{p} 0 \,. \tag{9}$$

(ii) For $m_c = [c(n)^\alpha]$ and $0 < \alpha < \min\{0.4, \alpha_1^*\}$, as $n \longrightarrow \infty$

$$\sqrt{m_c b(n)} \left[\hat{V}(2) - V(2) \right] \stackrel{\mathcal{L}-s}{\rightarrow} N\left[0, W\right] \,, \tag{10}$$

where

$$W = 2 \int_0^1 \left[\sigma_x(s)\right]^4 ds \,. \tag{11}$$

If we take $\delta = 0.0$ and $\gamma = 2/3$, then the above condition for consistency implies $0 < \alpha < 1/4$.

(ii) The Case of $r = 2$

The main bias term $m_c^2/c(n)$ in each interval of the sample size $c(n)$

$$\frac{1}{m_c}\sum_{k=1}^{m_c} a_{k,c(n)} = O(\frac{1}{m_c}\times\frac{m_c^3}{c(n)}) = O(\frac{m_c^2}{c(n)}) \ .$$

The main signal part of $(1/m_c)\sum_{k=1}^{m_c} z_k^4$ is given by

$$\sum_{j_1,j_2,j_3,j_4=1}^{c(n)} [\frac{4}{\sqrt{m_c}}\sum_{k=1}^{m_c} s_{k,j_1}s_{k,j_2}s_{k,j_3}s_{k,j_4}]r_{j_1}r_{j_2}r_{j_3}r_{j_4} \quad ,$$

where $s_{jk} = \cos\theta_{jk}$ and $\theta_{jk} = \frac{2\pi}{2m_c+1}(j-\frac{1}{2})(k-\frac{1}{2})$ $(j = 1,\cdots,c(n); k = 1,\cdots,m_c)$.

There are many terms for its asymptotic variance such as

$$\sum_{i_1=i_2\neq i_3=i_4}\sum_{j_1=j_2\neq j_3=j_4}[(\frac{4}{m_c}\sum_{k=1}^{m_c}s_{k,i_1}^2 s_{k,i_3}^2][(\frac{4}{m_c}\sum_{k'=1}^{m_c}s_{k',j_1}^2 s_{k',j_3}^2]$$

$$\times \quad \mathbf{E}(r_{i_1}^2)\mathbf{E}(r_{i_3}^2)\mathbf{E}(r_{j_1}^2)\mathbf{E}(r_{j_3}^2) \ .$$

The main bias term comes from the main noise part and in the interval $I_c(i)=((i-1)\frac{c(n)}{n},i\frac{c(n)}{n}]$ becomes

$$\frac{1}{m_c}\sum_{k=1}^{m_c}\sum_{j_1,j_2,j_3,j_4=1}^{c(n)} b_{k,j_1}b_{k,j_2}b_{k,j_3}b_{k,j_4}\mathbf{E}[v_{j_1}v_{j_2}v_{j_3}v_{j_4}] \ .$$

Hence when $r=2$, the typical bias term $[m_c^4/c(n)^2]$ in each sub-interval because

$$\frac{1}{m_c}\sum_{i_1=j_1\neq i_2=j_2} b_{k,i_1}^2 b_{k,i_2}^2=\frac{1}{m_c}\sum_{k=1}^{m_c}a_{k,c(n)}^2=O(\frac{m_c^4}{c(n)^2}) \ ,$$

where

$$a_{k,c(n)}=4c(n)\sin^2\left[\frac{\pi}{2}\frac{2k-1}{2c(n)+1}\right] \ .$$

Then the condition for consisitency becomes

$$b(n)[b(n)\frac{(m_c)^4}{c(n)^2}][\epsilon_n]^4 \longrightarrow 0\ .$$

If we set $c(n) = n^{\gamma}$, $b(n) = n^{1-\gamma}$ and $m_c = [c(n)]^{\alpha}$, then

$$b(n)^2[\frac{(m_c)^4}{c(n)^2}][\epsilon_n]^4 = n^{2(1-\gamma)-2\gamma+4\gamma\alpha-4\delta} = n^{2[1-2\gamma+2\gamma\alpha-2\delta]}\ .$$

The condition for the asymptotic normality without bias becomes

$$\sqrt{m_c b(n)}b(n)^2[\frac{(m_c)^4}{c(n)^2}][\epsilon_n]^4 = n^{\frac{1-\gamma+\alpha\gamma}{2}+2(1-2\gamma)-2\gamma+4\gamma\alpha-4\delta}\ .$$

Theorem 3 : When $p = 1$ and $r \geq 1$, assume that $v(t_i^n)$ is a sequence of i.i.d. random variables with $\mathbf{E}[v_i^{4r}] < \infty$ and σ_s $(0 \leq s \leq 1)$ is bounded. We set $\alpha_{1r}^* = [2\gamma + 2\delta - 1]/[2\gamma]$ and $\alpha_{2r}^* = [(4r+1)\gamma - (1+2r) + 4r\delta]/[(4r+1)\gamma]$.
(i) For $m_c = [c(n)^\alpha]$ and $0 < \alpha < \min\{0.5, \alpha_{1r}^*\}$ $((\alpha_{1r}^* > 0)$, as $n \longrightarrow \infty$

$$\hat{V}(2r) - V(2r) \xrightarrow{p} 0\ . \tag{12}$$

(ii) For $m_c = [c(n)^\alpha]$ and $0 < \alpha < \min\{0.4, \alpha_{2r}^*\}$ $(\alpha_{2r}^* > 0)$, as $n \longrightarrow \infty$

$$\sqrt{m_c b(n)} \left[\hat{V}(2r) - V(2r) \right] \stackrel{\mathcal{L}-s}{\rightarrow} N[0, W]\ , \tag{13}$$

where

$$W = c_r^* \int_0^1 [\sigma_x(s)]^{4r}\, ds\ , \tag{14}$$

where c_r^* is a positive constant.

In particular, when $r = 1$, $c_1^* = 2$. In this case, we have Proposition 2. In the general case, $c_r^* = a_{2r}/a_r^2 - 1$ and $c_2^* = 105/3^2 - 1$ when $r = 2$.
It is because

$$a_{2r} = \frac{4r!}{2r!2^r} = \frac{4r!}{4r \cdot 2(2r-1)\cdots 2} = (4r-1)(4r-3)\cdots 1 \ ,$$

for instance. It is interesting to find that the form of the asymptotic variance for the LSIML estimation is the same as RV when there is no micro-market noise.

An Optimal Choice of α and γ

Theorem 4 : When $p = 1$ and $r \geq 1$, assume that $v(t_i^n)$ is a sequence of i.i.d. random variables with $\mathbf{E}[v_i^{4r}] < \infty$ and σ_s $(0 \leq s \leq 1)$ is bounded. An optimal choice of $m_c = [c(n)^{\alpha}]$ and $c(n) = n^{\gamma}$ (with $\epsilon_n = n^{-\delta}$) to minimize MSE when n is large, is given by

$$-1 + \gamma(1-\alpha) = 2[(1-2\gamma)r + 2r\alpha\gamma] - 4r\delta \ , \qquad (15)$$

which means the choice as

$$\alpha^* = \frac{(4r+1)\gamma + 4r\delta - 2r - 1}{(4r+1)\gamma} = 1 + \frac{4r\delta - 2r - 1}{(4r+1)\gamma} \ . \qquad (16)$$

For instance, when $\delta = 0$, $\alpha^* = 1 - 3/[5\gamma]$ for $r = 1$ and $\alpha^* = 1 - 5/[9\gamma]$ for $r = 2$. When $\delta = 0$ and we take α^*, then the MSE is proportionalto $n^{-1+\gamma^*(1-\alpha^*}$), which is

$$\mathrm{MSE} \sim n^{\frac{-2r}{4r+1}} \quad .$$

When $r = 1$, we find that $2r/[4r+1] = 2/5$, which is the same as the asymptotic order of the SIML estimation. Moreover, when $r = 2$, we have $2r/[4r+1] = 4/9$.

Simulations

In our simulations we set $b(n) = [n^{1-\gamma}]$, $c(n) = [n^{\gamma}]$ and the number of replications is 3,000. Also we have investigated several cases in which the instantaneous volatility function σ_s^2 is given by

$$\sigma_s^2 = \sigma_0^2 \left[a_0 + a_1 s + a_2 s^2 \right] ,$$

where a_i $(i = 0, 1, 2)$ are constants and we have some restrictions such that $\sigma_s > 0$ for $s \in [0, 1]$.

Table 1 : Estimation of integrated fourth-order functional $(a_0 = 1.0, a_1 = 0.0, a_2 = 0.0; \sigma_u^2 = 0.0005,\ b(n) = 5, c(n) = 521)$

n=2,605	$V(2) = 2.00$	$V(4) = 4.0$
mean	2.009	4.053
Var	0.134	2.837
AV	0.133	2.843

Table 2 : Estimation of integrated fourth-order functional $(a_0 = 1.0, a_1 = 0.0, a_2 = 0.0; \sigma_u^2 = 0.0005,\ b(n) = 10, c(n) = 1,000)$

n=10,000	$V(2) = 2.00$	$V(4) = 4.0$
mean	2.023	4.056
Var	0.092	1.973
AV	0.089	1.895

Table 3 : Estimation of integrated fourth-order functional $(a_0 = 6.0, a_1 = -24.0, a_2 = 24.0; \sigma_u^2 = 0.0005,\ b(n) = 7, c(n) = 355)$

n=2,485	$V(2) = 2.00$	$V(4) = 7.2$
mean	2.014	6.973
Var	0.342	30.824
AV	0.343	36.549

Table 4 : Estimation of integrated fourth-order functional $(a_0 = 6.0, a_1 = -24.0, a_2 = 24.0; \sigma_u^2 = 0.0005,\ b(n) = 10, c(n) = 1,000)$

n=10,000	$V(2) = 2.00$	$V(4) = 7.2$
mean	2.023	7.167
Var	0.160	15.093
AV	0.160	17.056

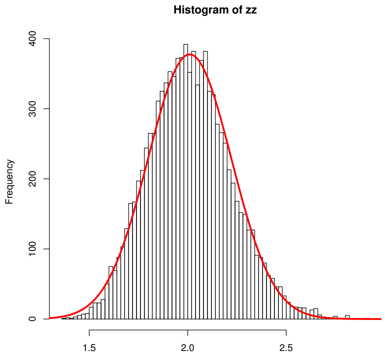


Figure 1: Normalized Histogram and Normalized Distribution

Generalizations

Let

$$\mathbf{Y}(t_i^n) = \mathbf{X}(t_i^n) + \epsilon_n \mathbf{v}(t_i^n) \ \ (i = 1, \cdots, n)$$

be the (p-dimensional) observed (log-)prices $\mathbf{Y}(t_i^n) = (Y_j(t_i^n))$ at t_i^n $(0 = t_0^n \leq t_1^n \leq \cdots \leq t_n^n = 1)$ and $\mathbf{v}(t_i^n)$ $(= (v_j(t_i^n)))$ be a sequence of i.i.d. random vectors with $\mathbf{E}[\mathbf{v}(t_i^n)] = 0$. The general form of the SDE for the $p-$dimensional continuous-time stochastic processes is given by

$$d\mathbf{X} = \boldsymbol{\mu}_t dt + \boldsymbol{\sigma}_t d\mathbf{B}_t \ \ ,$$

and

$$\mathbf{X}(t) = \mathbf{X}(0) + \int_0^t \boldsymbol{\mu}(s) ds + \int_0^t \boldsymbol{\sigma}(s) d\mathbf{B}_s \ \ .$$

Kunitomo (2018) has proposed one way to determine the underlying number of hidden volatilities under the micro-market noise.

When the volatility parameters follow a diffusion type process as

$$\boldsymbol{\sigma}_{ij}(t) = \boldsymbol{\sigma}_{ij}(0) + \int_0^t \boldsymbol{\mu}_{ij}^{\sigma}(s)ds + \int_0^t \boldsymbol{\omega}_{ij}^{\sigma}(s)d\mathbf{B}_s^{\sigma} \; (0 \leq s \leq t \leq 1) \; ,$$

where $\boldsymbol{\mu}_{ij}(s)$ and $\boldsymbol{\omega}_{ij}^{\omega}(s)$ are the drift and $1 \times q_2$ diffusion coefficients and $\mathbf{B}_s^{\sigma}$ is another $q_2 \times 1$ Brownian motion vector, which may be correlated with $\mathbf{B}_s$.

The class of diffusion processes can be extended to the class of Itô semi-martingales, which is given by

$$\begin{aligned}\mathbf{X}(t) &= \mathbf{X}(0)+\int_0^t \boldsymbol{\mu}_s ds+\int_0^t \mathbf{C}_x(s)\, d\mathbf{B}(s) \\ &\quad +\int_s\int_{\|\delta(s,x)\|<1} \delta(s,\mathbf{x})(\boldsymbol{\mu}-\boldsymbol{\nu})(ds,d\mathbf{x}) \\ &\quad +\int_s\int_{\|\delta(s,x)\|\geq 1} \delta(s,\mathbf{x})\mu(ds,d\mathbf{x})\ .\end{aligned}$$

(Ikeda and Watanabe (1989), and Jacod and Protter (2012).)

There is an interesting problem whether we need the jump terms in addition to the diffusion processes when we have the micro-market noise. Kunitomo and Kurisu (2017) have shown that the effects of micro-market noise is significant and the validity of existing procedure is questionable. When $p = 1$, a test statistics for the presence of jump terms may be

$$\sqrt{m_c b(n)}[V_k(4) - V_1(4)]$$

where $V_k(4)$ is constructed from $|y(t_j^n) - y(t_{j-k}^n)|^4$ $(k \geq 2)$ and $V_1(4)$ is constructed from $|y(t_j^n) - y(t_{j-1}^n)|^4$.

References

- Ait-Sahalia, Y. and J. Jacod (2014), *High-Frequency Financial Econometrics*, Princeton University Press.
- Jacod, J., Y. L., Per A. Mykland, M. Podolskijc, and M. Vetter (2009), "Microstructure noise in the continuous case: The pre-averaging approach," *Stochastic Processes and their Applications*, 119 (2009) 2249?2276.
- Kunitomo, N. (2018), "Detecting Factors of Quadratic Covariation under Micro-market Noises," Unpublished Manuscript.
- Kunitomo, N. and Kurisu, D. (2017), "Effects of Jump and Noise in High-Frequency Financial Econometrics," *Asia-Pasific Financial Markets*, Springer.
- Kunitomo, N., S. Sato and D. Kurisu (2018), *Separating Information Maximum Likelihood Method for High-Frequency Financial Data*, Springer.

Estimation for affine term structure with smooth transition

Shingo Mukunoki [1] Kosuke Oya [2]

[1]Graduate School of Economics, Osaka University

[2]Graduate School of Economics & MMDS, Osaka University

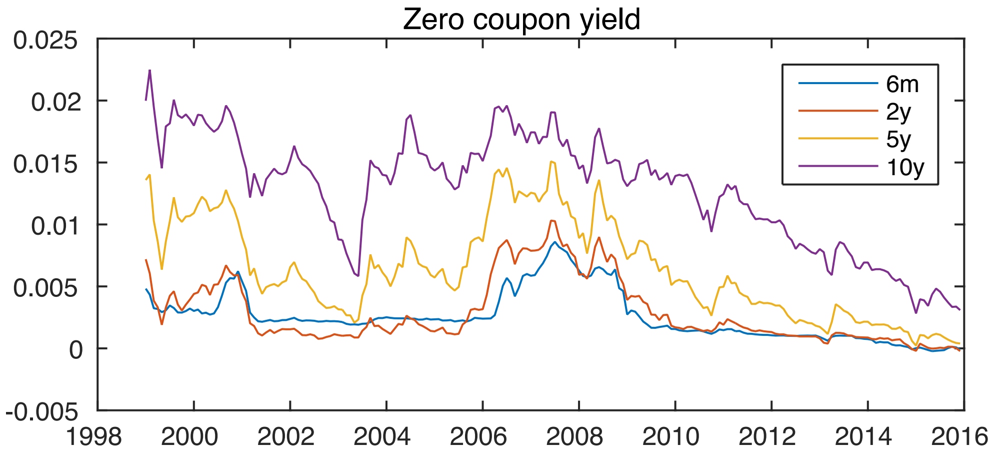
松本・経済統計キャンプ 2018@信州大学 2018 年 8 月 6 日

科研プロジェクト「新しい時系列計量分析の理論と応用」

Outline

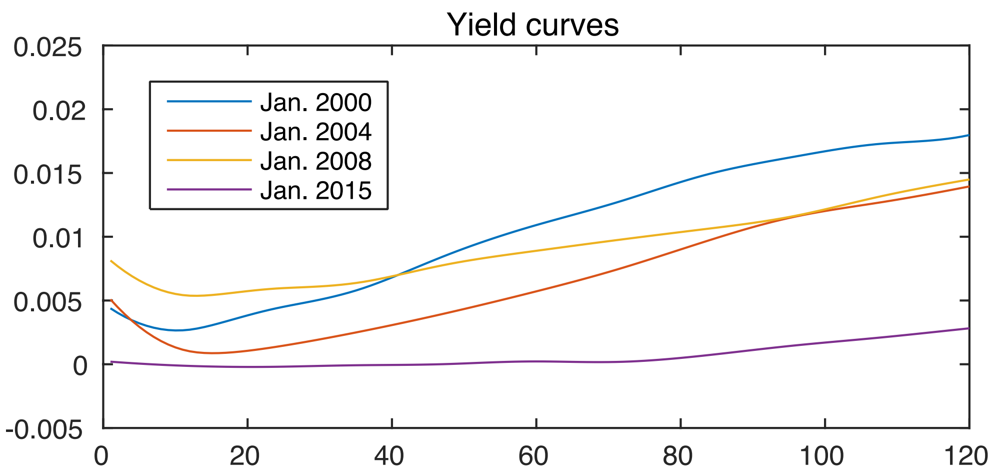
Introduction

- Term structure models provide the relationship among yields with different time to maturity.
- Japanese government bond (JGB) from Mar.1999 to Dec.2015.



Yield curves

- Yield curves take various structures.



Affine term structure

- Affine term structre model tells us how investors price risk under no arbitrage assumption.
- Vasicek (1977), Duffie and Kan(1996), etc.

$$
\begin{aligned}
y_t^{(n)} &= a_n + b_n' f_t, \\
\beta_n &= \left\{ I_m + \Phi^{\mathbb{Q}\prime} + \cdots + (\Phi^{\mathbb{Q}\prime})^{n-1} \right\} \beta_1 \\
\alpha_n &= n\alpha_1 + \left\{ \beta_1' + \beta_2' + \cdots + \beta_{n-1}' \right\} c^{\mathbb{Q}} - \frac{1}{2} \left\{ \beta_1' \Sigma \beta_1 + \cdots + \beta_{n-1}' \Sigma \beta_{n-1} \right\} \\
a_n &= -\frac{1}{n}\alpha_n, \quad b_n = -\frac{1}{n}\beta_n
\end{aligned}
$$

Difficulties in estimation of canonical affine term structure models

- Unobserved factors
- Highly non-linear and badly behaved objective function

Previous studies and Ours

- Hamilton and Wu (2012)
  - point out that the parameter estimates correspond to a local maxima of the likelihood surface.
  - the minimum chi-square estimation is adopted which is one of the two stage estimations.
- Diez de los Rios (2015)
  - uses principal components as factors and applies asymptotic least square (ALS) estimation.
  - assumes factors (principal components) follow VAR process.
- This study
  - introduces more general framework to the market price of risk, that is, factor follows smooth-transition VAR (STVAR) process (the term structure model with smooth transition market price of risk).

Canonical affine term structure model

Assume followings:

- m-dimensional factor vector f_t is assumed to follow VAR process under risk-neutral measure $\mathbb{Q}$.

$$f_t = c^{\mathbb{Q}} + \Phi^{\mathbb{Q}} f_{t-1} + v_t^{\mathbb{Q}}, \quad v_t^{\mathbb{Q}} \sim i.i.d.N(0, \Sigma) \tag{1}$$

  - It is possible to take more than one for the lag order of VAR.
  - f_t captures variations in all the varibales relate to bond price.
- Short rate r_t is assumed to be the linear function of factors

$$r_t = \alpha_1 + \beta_1' f_t. \tag{2}$$

- No-arbitrage: Bond price $p_t^{(n)}$ at t with n periods to maturity,

$$p_t^{(n)} = \exp(-r_t) E_t^{\mathbb{Q}} \left[p_{t+1}^{(n-1)} \right]$$

Intro. Canonical ATSM Estimation and data Our research Summary and future research References References

Bond price and yield with n periods to maturity

- Stochastic discount factor is represented as follow

$$m_{t+1} = \exp\left(-r_t - \frac{1}{2}\lambda_t' \Sigma^{-1} \lambda_t - \lambda_t' \Sigma^{-1} v_{t+1} \right), \quad \lambda_t = \Lambda + \Lambda_f f_t$$

where λ_t is the market price of risk.

Bond price at time t with maturity n is given by

$$p_t^{(n)} = \exp(-r_t) E_t^{\mathbb{Q}} \left[p_{t+1}^{(n-1)} \right] = E_t[\, m_{t+1} \, p_{t+1}^{(n-1)} \,] = \exp\left(\alpha_n + \beta_n f_t \right)$$

where $\alpha_n = \alpha_{n-1} + \beta_{n-1}' c^{\mathbb{Q}} - \frac{1}{2}\beta_{n-1}' \Sigma \beta_{n-1} + \alpha_1$ and $\beta_n = \beta_1 + (\Phi^{\mathbb{Q}})' \beta_{n-1}$

Yield: $y_t^{(n)} = -\dfrac{1}{n} \log p_t^{(n)} = -\dfrac{1}{n} \alpha_n - \dfrac{1}{n} \beta_n' f_t = a_n + b_n' f_t, \qquad r_t = y_t^{(1)}$

Factor under physical measure $\mathbb{P}$

$$f_t = \underbrace{(c^{\mathbb{Q}} + \Lambda)}_{c} + \underbrace{(\Phi^{\mathbb{Q}} + \Lambda_f)}_{\Phi} f_{t-1} + v_t, \quad v_t \sim i.i.d.N(0, \Sigma) \text{ under } \mathbb{P}. \tag{3}$$

Canonical Model for estimation

For m-dimensional factor f_t and $y_t = (y_t^{(1)}, \ldots, y_t^{(N)})'$,

$$f_t = c + \Phi f_{t-1} + v_t, \qquad v_t \sim i.i.d.N(0, \Sigma)$$
$$y_t = a(\theta) + b(\theta) f_t + \eta_t, \quad \eta_t \sim i.i.d.N(0, \Omega), \quad E[v_t \eta_s] = 0 \text{ for } \forall\ t, s$$

where $a(\theta) = (a_1(\theta), \ldots, a_N(\theta))'$, $b(\theta) = (b_1(\theta), \ldots, b_N(\theta))'$

$$b_n(\theta) = -\frac{1}{n}\left\{I_m + \Phi^{\mathbb{Q}\prime} + \cdots + (\Phi^{\mathbb{Q}\prime})^{n-1}\right\}\beta_1$$
$$a_n(\theta) = -\alpha_1 + \{b_1(\theta)' + 2b_2(\theta)' + \cdots + (n-1)b_{n-1}(\theta)'\} c^{\mathbb{Q}}/n$$
$$- \left\{b_1(\theta)'\Sigma b_1(\theta) + \cdots + (n-1)^2 b_{n-1}(\theta)'\Sigma b_{n-1}(\theta)\right\}/2n.$$

Parameter to be interest.

$$\theta = \left(c', vec(\Phi)', vech(\Sigma)', c^{\mathbb{Q}\prime}, vec(\Phi^{\mathbb{Q}})', \alpha_1, \beta_1', vech(\Omega)'\right)'$$

Previous research: HW (2012) and Diez de los Rios (2015)

- 1st step: Estimate auxiliary parameters $\pi = (\tau', \pi_{\mathbb{P}}')'$

$$\tau = (c', vec(\Phi)', vech(\Sigma)', vech(\Omega)')', \quad \pi_{\mathbb{P}} = (a', vec(b)')'.$$

State space model

$$f_t = c + \Phi f_{t-1} + v_t \quad v_t \sim i.i.d.N(0, \Sigma)$$
$$y_t = a + b\, f_t + \eta_t \quad \eta_t \sim i.i.d.N(0, \Omega)$$

- 2nd step: Estimate parameters of interest $\theta = (\tau', \theta_{\mathbb{Q}}')'$

$$\theta_{\mathbb{Q}} = \left(c^{\mathbb{Q}\prime}, vec(\Phi^{\mathbb{Q}})', \alpha_1, \beta_1'\right)'$$

Objective function (W_{HW}, W_{ALS}: weights)

$[\hat{\pi} - g(\theta)]' W_{HW} [\hat{\pi} - g(\theta)]$,	HW estimation
$[\gamma(\hat{\pi}) - \Gamma(\hat{\pi})\theta]' W_{ALS} [\gamma(\hat{\pi}) - \Gamma(\hat{\pi})\theta]$,	ALS estimation in Diez... (2015)

- $\pi = g(\theta)$ used in HW (2012) is highly non-linear in θ.
- $\gamma(\pi) = \Gamma(\pi)\,\theta$ in Diez... (2015) is based on the following restrictions

$$\alpha_n = \alpha_{n-1} + \beta'_{n-1}c^{\mathbb{Q}} - (1/2)\beta'_{n-1}\Sigma\beta_{n-1} + \alpha_1,$$

$$\beta_n = \beta_1 + (\Phi^{\mathbb{Q}})'\beta_{n-1}.$$

These recursion can be rewritten as

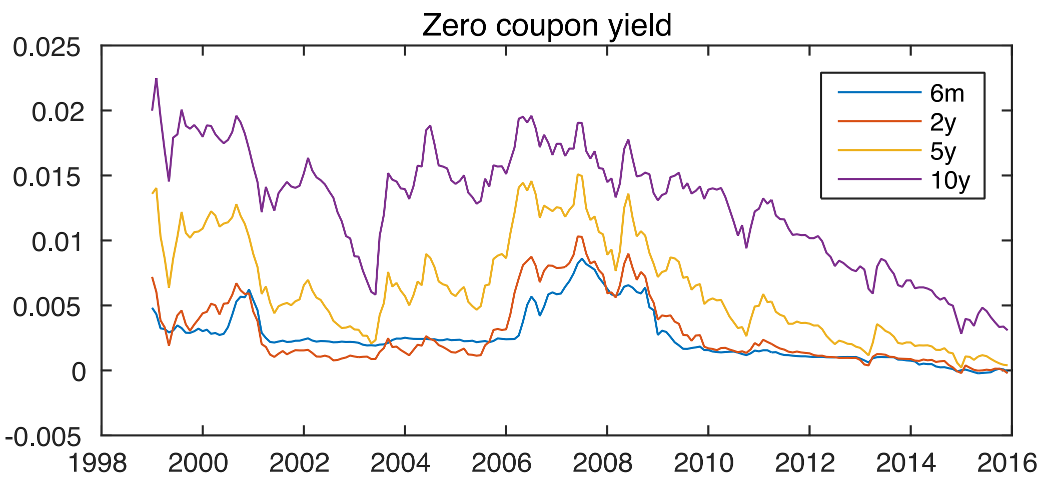
$$\begin{cases} n\, a_n - (n-1)a_{n-1} - a_1 + (1/2)(n-1)^2 b'_{n-1}\Sigma b_{n-1} = (n-1)b'_{n-1}c^{\mathbb{Q}}, \\ n\, b_n - b_1 = (n-1)^2 b'_{n-1}\Phi^{\mathbb{Q}} \end{cases}$$

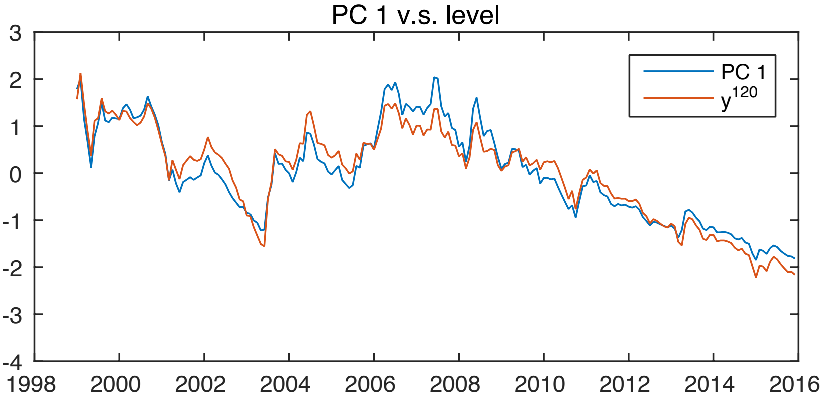
$$a_1 = -\alpha_1, \quad b_1 = -\beta_1.$$

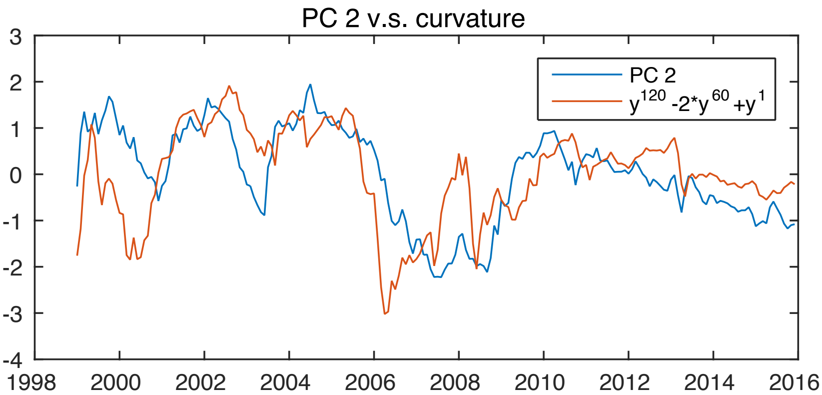
- The estimation method in Diez ...(2015) doesn't require numerical optimization.
- Diez ...(2015) adopts principal components for factors and assumes that the factors follow VAR process.

JGB yields

- Monthly data from March 1999 to December 2015.
- Time to maturity: 1month to 120 months.
- Data source
  Mar. 1999 to Dec. 2011: Kikuchi and Shintani (2012)
  Jan. 2012 to Dec. 2015: We compute by the same method.







	prop.	cum.
PC 1	93.44	93.44
PC 2	5.56	99.00
⋮	⋮	⋮

the first two principal components explain 99 % of the variation in bond yields.

Our research

To allow the variability in the risk sensitivity of investors, we introduce a following extension for the market price of risk.

$$\lambda_t = \Lambda + \Lambda_f f_t \ \rightarrow \lambda_t = \Lambda(t) + \Lambda_f(t) f_t$$

The coefficients of $\Lambda(t)$ and $\Lambda_f(t)$ transit smoothly through

$$\Lambda(t) = \Lambda^{(1)} + \Lambda^{(2)} H(t), \quad \Lambda_f(t) = \Lambda_f^{(1)} + \Lambda_f^{(2)} H(t),$$

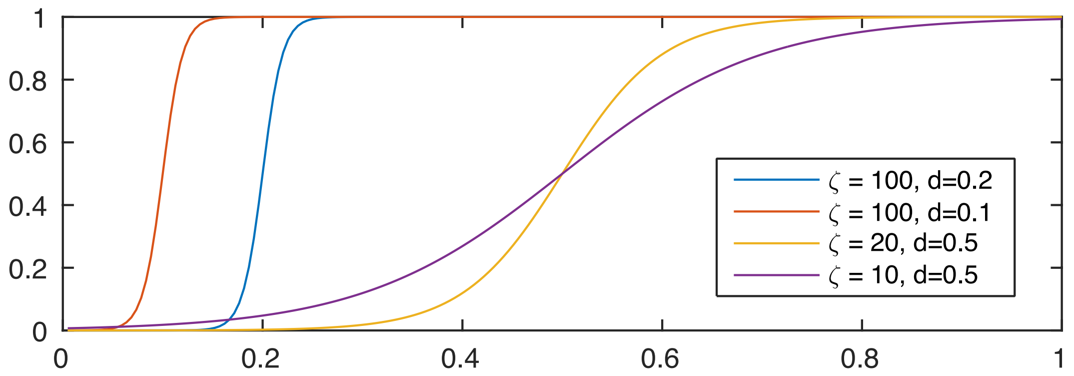
$H(t) = diag(h_{(1)}(t), \ldots, h_{(m)}(t))$ and $h_{(i)}(t) = \left\{1 + \exp\left(-\zeta_{(i)}(t/T - d_{(i)})\right)\right\}^{-1}$.

We will show later that factors follow STVAR under the assumption.

$$f_{t+1} = c(t) + \Phi(t) f_t + v_{t+1} \quad v_{t+1} \sim i.i.d.N(0, \Sigma)$$

Illustration of transition function

Plot; $h(t) = \{1 + \exp(-\zeta(t-d))\}^{-1}$



Factor process under $\mathbb{P}$

From the definition of stochastic discount factor m_{t+1},

$$
\begin{aligned}
E_t\left[\exp(k' f_{t+1})\right] &= E_t^{\mathbb{Q}}\left[\exp(k' f_{t+1}) \exp\left(\frac{1}{2}\lambda_t'\Sigma^{-1}\lambda_t + \lambda_t\Sigma^{-1}v_{t+1}\right)\right] \\
&= \exp\left(k'(c^{\mathbb{Q}} + \Phi^{\mathbb{Q}} f_t) + \frac{1}{2}\lambda_t'\Sigma^{-1}\lambda_t\right) E_t^{\mathbb{Q}}\left[\exp\left\{k' v_{t+1}^{\mathbb{Q}} + \lambda_t\Sigma^{-1}v_{t+1}\right\}\right] \\
&= \exp\left(k'(c^{\mathbb{Q}} + \Phi^{\mathbb{Q}} f_t) - \frac{1}{2}\lambda_t'\Sigma^{-1}\lambda_t\right) E_t^{\mathbb{Q}}\left[\exp\left\{(k' + \lambda_t'\Sigma^{-1})v_{t+1}^{\mathbb{Q}})\right\}\right] \\
&= \exp\left(k'(c^{\mathbb{Q}} + \Phi^{\mathbb{Q}} f_t + \lambda_t) + \frac{1}{2}k'\Sigma k\right) \\
&= \exp\left(k'\left\{c^{\mathbb{Q}} + \Lambda(t) + (\Phi^{\mathbb{Q}} + \Lambda_f(t)) f_t\right\} + \frac{1}{2}k'\Sigma k\right)
\end{aligned}
$$

Thus, the factor follows the following process under measure $\mathbb{P}$;

$$
f_{t+1} = \left\{c^{\mathbb{Q}} + \Lambda(t)\right\} + \left\{\Phi^{\mathbb{Q}} + \Lambda_f(t)\right\} f_t + v_{t+1} \quad v_{t+1} \sim i.i.d. N(0, \Sigma).
$$

Smooth transition VAR process, cf.) Teräsvirta (1994).

Reparameterize as follows,

$$f_{t+1} = c(t) + \Phi(t) f_t + v_{t+1}$$

where $c(t) = c^{(1)} + c^{(2)} H(t)$ and $\Phi(t) = \Phi^{(1)} + \Phi^{(2)} H(t)$

$H(t) = diag(h_{(1)}(t), \dots, h_{(m)}(t))$ and

$$h_{(i)}(t) = \left\{1 + \exp\left(-\zeta_{(i)}(t/T - d_{(i)})\right)\right\}^{-1}.$$

$$c^{(1)} = c^{\mathbb{Q}} + \Lambda^{(1)}, \quad c^{(2)} = \Lambda^{(2)}$$
$$\Phi^{(1)} = \Phi^{\mathbb{Q}} + \Lambda_f^{(1)}, \quad \Phi^{(2)} = \Lambda_f^{(2)}.$$

Intro. Canonical ATSM Estimation and data Our research Summary and future research References References

Linearity test cf.) Teräsvirta (1994) ex.) AR(1) v.s. STAR(1)

STAR(1) models

$$x_{t+1} = c^{(1)} + \phi^{(1)} x_t + (c^{(2)} + \phi^{(2)} x_t) h(t) + \varepsilon_{t+1}$$
$$h(t) = \{1 + \exp(-\zeta\,(t/T - d))\}^{-1}$$

- $H_0 : \zeta = 0$ implies AR model.
- However, if H_0 is true, parameters remain unidentified.

$$x_{t+1} = (c^{(1)} + c^{(2)}/2) + (\phi^{(1)} + \phi^{(2)}/2)\, x_t + \varepsilon_{t+1}$$

- To deal with the identification problem, consider the Taylor expansion of $h(t)$

$$h(t) = \frac{1}{2} + \frac{1}{4}\left(\frac{t}{T} - d\right)\zeta + R_t = \frac{1}{2} - \frac{1}{4}\, d\, \zeta + \frac{1}{4}\,\zeta\,\frac{t}{T} + R_t.$$

R_t: remainder term.

STAR(1) models can be rewritten as

$$\begin{aligned} x_{t+1} &= c^{(1)} + \phi^{(1)} x_t + (c^{(2)} + \phi^{(2)} x_t) h(t) + \varepsilon_{t+1}, \\ &= c^{(1)} + \phi^{(1)} x_t + (c^{(2)} + \phi^{(2)} x_t) \left(\frac{1}{2} - \frac{1}{4}\, d\, \zeta + \frac{1}{4} \zeta \frac{t}{T} + R_t \right) + \varepsilon_{t+1}, \\ &= c^{(1)*} + \phi^{(1)*} x_t + c^{(2)*} \frac{t}{T} + \phi^{(2)*} x_t \frac{t}{T} + \varepsilon^*_{t+1}. \end{aligned}$$

where

$$\begin{aligned} c^{(1)*} &= c^{(1)} + \left(\frac{1}{2} - \frac{1}{4} d\zeta \right) c^{(2)}, \quad \phi^{(1)*} = \phi^{(1)} + \left(\frac{1}{2} - \frac{1}{4} d\zeta \right) \phi^{(2)} \\ c^{(2)*} &= \frac{1}{4} \zeta\, c^{(2)}, \quad \phi^{(2)*} = \frac{1}{4} \zeta\, \phi^{(2)} \\ \varepsilon^*_{t+1} &= \varepsilon_{t+1} + (c^{(2)} + \phi^{(2)} x_t) R_t \end{aligned}$$

We use equivalent hypothesis; $H_0 : c^{(2)*} = \phi^{(2)*} = 0$.

Intro. Canonical ATSM Estimation and data Our research Summary and future research References References

Our model and linearity test

STVAR models

$$f_{t+1} = c^{(1)} + \Phi^{(1)} f_t + (c^{(2)} + \Phi^{(2)} f_t)\, H(t) + v_{t+1}$$

where $H(t) = diag(h_{(1)}(t), \ldots, h_{(m)}(t))$ and

$$h_{(i)}(t) = \left\{ 1 + \exp\left(-\zeta_{(i)} (t/T - d_{(i)}) \right) \right\}^{-1}.$$

....

We test linearity based on the third Taylor expansion.

$H_0 : \zeta_{(i)} = 0,\ i = 1, \ldots, m$: VAR

$H_1 : {}^{\exists} i,\ \zeta_{(i)} \neq 0$: STVAR

LM test statistics	upper 5% point
67.14	28.87

MMDS Center for Mathematical Modeling and Data Science Osaka University

Our model and ALS estimation

State space model

$$y_t = a + b f_t + \eta_t, \quad \eta_t \sim i.i.d.N(0, \sigma_\eta^2 P_\perp P_\perp')$$
$$f_{t+1} = c(t) + \Phi(t) f_t + v_{t+1} \quad i.i.d.N(0, \Sigma)$$

Note that the OLS estimators satisfy the self-consistecy condition; $P'\hat{a} = 0$ and $P'\hat{b} = I_m$ since $f_t = P'y_t$.

- 1st step: Estimate auxiliary parameters $\pi_{ST} = (\tau_{ST}', \pi_{\mathbb{P}}')'$

$$\tau_{ST} = \left(c^{(1)\prime}, c^{(2)\prime}, vec(\Phi^{(1)})', vec(\Phi^{(2)})', vech(\Sigma)', \zeta', d', \sigma_\eta^2\right)'$$
$$\pi_{\mathbb{P}} = (a', vec(b)')'.$$

- 2nd step: Estimate parameters of interest $\theta_{ST} = (\tau_{ST}', \theta_{\mathbb{Q}}')'$

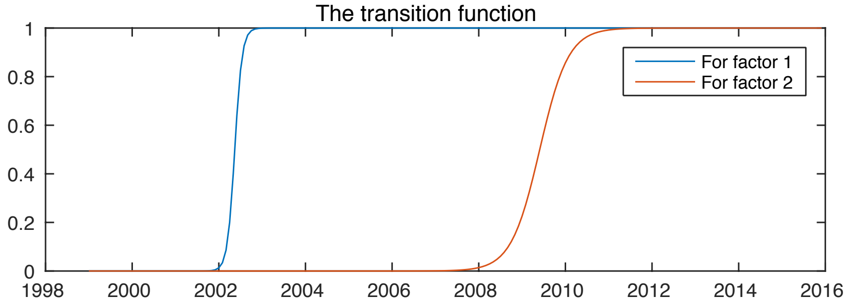
$$\theta_{\mathbb{Q}} = \left(c^{\mathbb{Q}\prime}, vec(\Phi^{\mathbb{Q}})', \alpha_1, \beta_1'\right)'$$

Obj. func.: $[\gamma(\hat{\pi}_{ST}) - \Gamma(\hat{\pi}_{ST})\theta_{ST}]' W [\gamma(\hat{\pi}_{ST}) - \Gamma(\hat{\pi}_{ST})\theta_{ST}]$

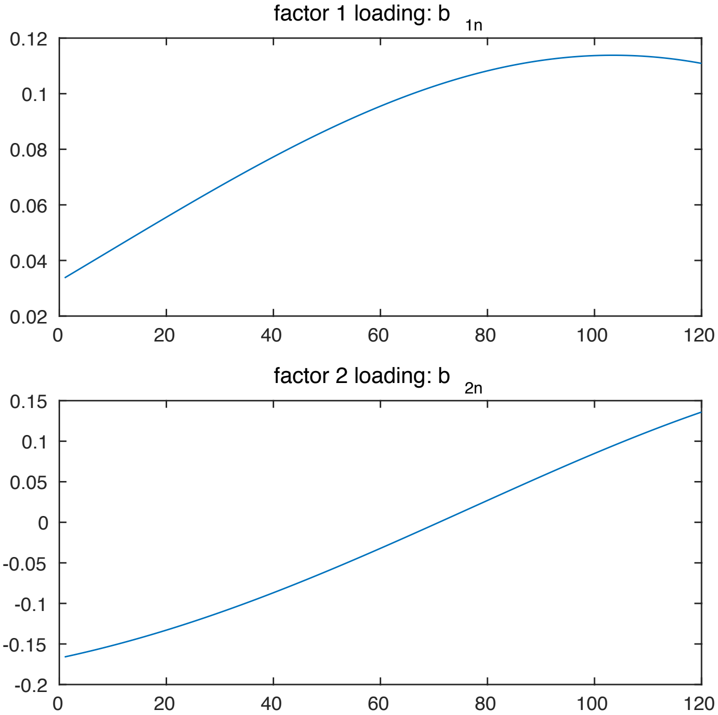
Empirical result

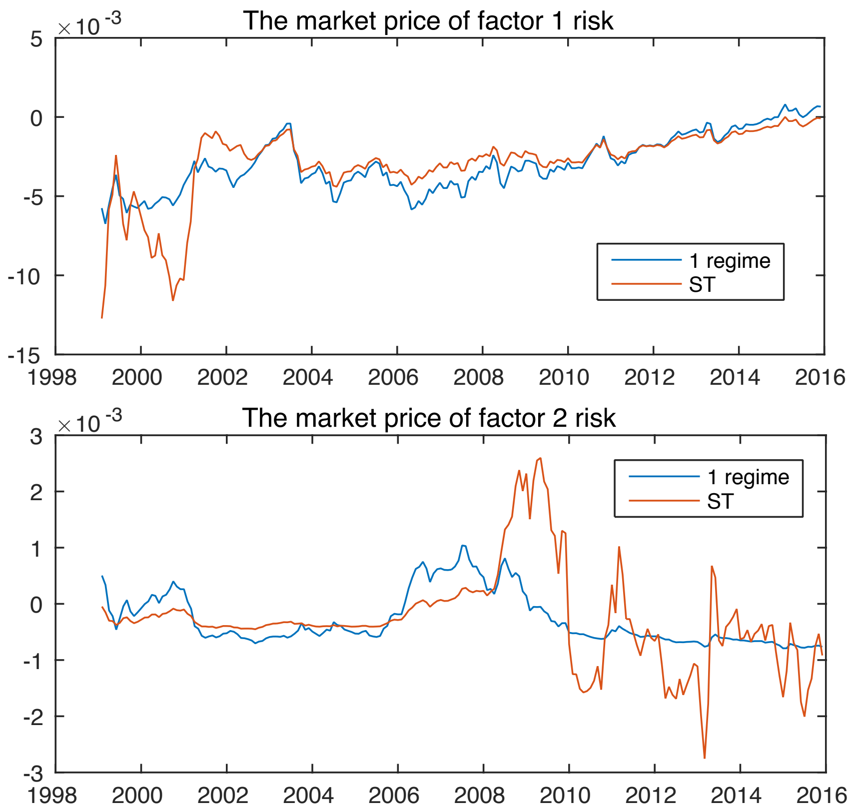
- Data
  - Monthly data of Japanese government bond yield from March 1999 to December 2015.
  - Time to maturity: 1month to 120 months.
- Model
  - 2 factors (The first two principle components)
  - 2 regimes for each factors

Estimated transition function



Factor loadings ($y_t^n = a_n + b_{1n}f_{1t} + b_{2n}f_{2t}$)



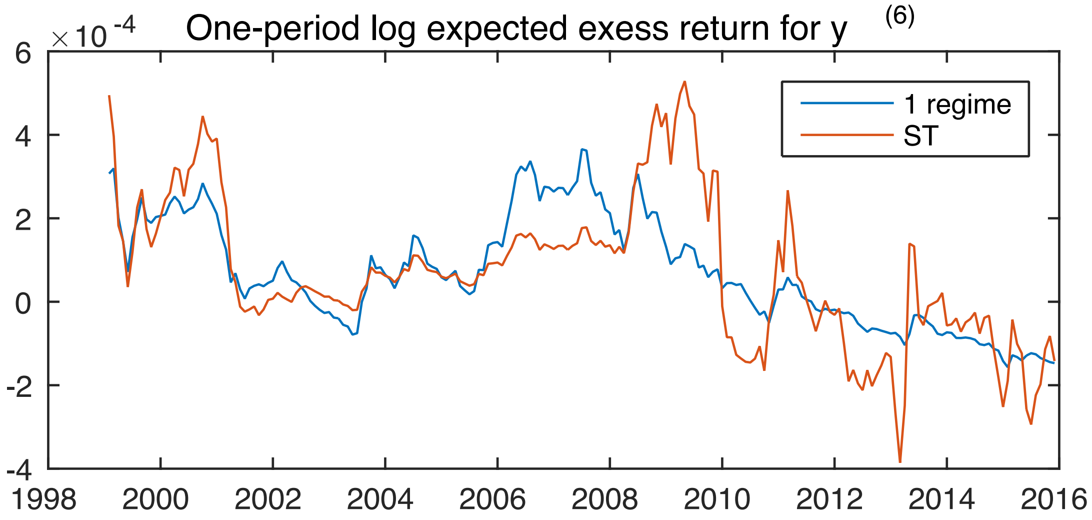


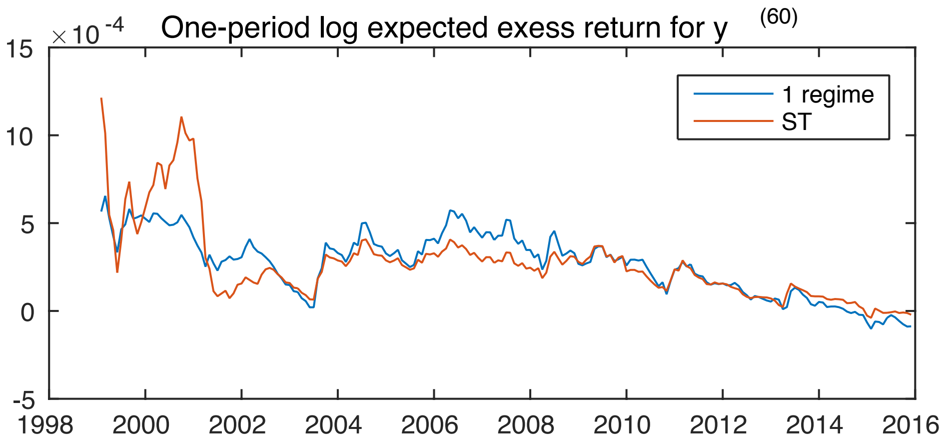
$$\lambda_t = \Lambda(t) + \Lambda_f(t) f_t$$

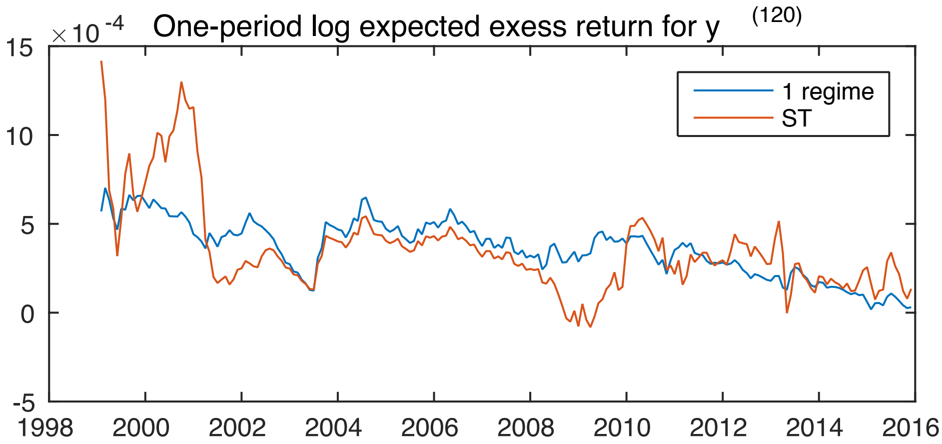
$$\lambda_t = \Lambda + \Lambda_f f_t$$

One period log expected excess return for n-month yield

$$\log \frac{E_t[p_{t+1}^{(n-1)}]}{p_t^{(n)}} - y_t^{(1)} = -b_n' \lambda_t$$







Summary

- We introduce more general framework to the market price of risk and provide the affine term structure model with smooth-transition.
- Our smooth transition model captures the variability in the market price of risk more than 1 regime model.

Future research

- Incorporation of more regimes to capture the various shapes of market price of risk.

References

Diez de los Rios, A. (2015), "A New Linear Estimator for Gaussian Dynamic Term Structure Models," *Journal of Business and Economic Statistics*, 33 (2), 282-295.

Hamilton, J. D. and Wu, J. C. (2012), "Identification and estimation of Gaussian affine term structure models," *Journal of Econometrics*, 168 (2), 315-331.

Kikuchi, K. and Shintani, K. (2012), "Comparison of estimation methods of zero coupon yield curve using JGB price data," IMES Discussion Paper, No. 2012-J-3 (in Japanese).

Teräsvirta, T. (1994), "Specification, Estimation and Evaluation of Smooth Transition Autoregressive Models," *Journal of the American Statistical Association*, 89 (425), 208-218.

Detecting Trend Factors in Non-stationary Errors-in-Variables Models

Naoto Kunitomo

Meiji University

August 2018

Outline

Motivations

- In many macro-times series we observe non-stationary trend, non-stationary seasonality and stationary cycles with measurement errors
- The official seasonal X-12-ARIMA uses the univariate Box-Jenkins method with moving averages
- In macro-consumption example there are several consumption time series which do not move in the same ways, but in similar ways
- We need to develop a new way to determine the number of factors of non-stationary trends

Non-Stationary Errors-in-Variables Model

Consider the non-stationary errors-in-variables representation

$$\mathbf{y}_i = \mathbf{x}_i + \mathbf{v}_i \ \ (i = 1, \cdots, n),$$

where $\mathbf{x}_i$ $(i = 1, \cdots, n)$ are a sequence of non-stationary $\mathbf{I}(1)$ process which satisfy

$$\Delta \mathbf{x}_i = (1 - \mathcal{L})\mathbf{x}_i = \mathbf{w}_i^{(x)}$$

with the lag-operator $\mathcal{L}\mathbf{x}_i = \mathbf{x}_{i-1}$, $\Delta = 1 - \mathcal{L}$,

$$\mathbf{w}_i^{(x)} = \sum_{j=0}^{\infty} \mathbf{C}_j^{(x)} \mathbf{e}_{i-j}^{(x)} \ ,$$

and $\mathbf{e}_i^{(x)}$ is a a sequence of i.i.d. random vectors with $\mathcal{E}(\mathbf{e}_i^{(x)}) = \mathbf{0}$ and $\mathcal{E}(\mathbf{e}_i^{(x)} \mathbf{e}_i^{(x)\prime}) = \mathbf{\Sigma}_e^{(x)}$ (positive-semi-definite).

The coefficients matrices $\mathbf{C}_j^{(x)}$ $(= c_{kl}^{(x)}(j))$ are absolutely summable $\sum_{j=0}^{\infty} \|\mathbf{C}_j^{(x)}\| < \infty$, where $\|\mathbf{C}_j^{(x)}\| = \max_{k,l=1,\cdots,p} |c_{kl}^{(x)}(j)|$ and $\mathbf{C}_j^{(x)} = (c_{kl}^{(x)}(j))$. The random vectors $\mathbf{v}_i$ $(i = 1, \cdots, n)$ are a sequence of stationary I(0) process with

$$\mathbf{v}_i = \sum_{j=0}^{\infty} \mathbf{C}_j^{(v)} \mathbf{e}_{i-j}^{(v)} ,$$

where the coefficient matrices $\mathbf{C}_j^{(v)}$ are absolutely summable $(\sum_{j=0}^{\infty} \|\mathbf{C}_j^{(v)}\| < \infty$, where $\|\mathbf{C}_j^{(v)}\| = \max_{k,l=1,\cdots,p} |c_{kl}^{(v)}(j)|$ and $\mathbf{C}_j^{(v)} = (c_{kl}^{(v)}(j)))$ and $\mathbf{v}_i$ are a sequence of i.i.d. random vectors with $\mathcal{E}(\mathbf{e}_i^{(v)}) = \mathbf{0}$, $\mathcal{E}(\mathbf{e}_i^{(v)} \mathbf{e}_i^{(v)\prime}) = \boldsymbol{\Sigma}_e^{(v)}$ (positive definite).

Let $\mathbf{f}_{\Delta x}(\mu)$ and $\mathbf{f}_v(\mu)$ be the spectral density $(p \times p)$ matrices of $\Delta \mathbf{x}_i$ and $\mathbf{v}_i$ $(i = 1, \cdots, n)$ as

$$\mathbf{f}_{\Delta x}(\mu) = \frac{1}{\pi}(\sum_{j=0}^{\infty} \mathbf{C}_j^{(x)} e^{2\pi i \mu j})\mathbf{\Sigma}_e^{(x)}(\sum_{j=0}^{\infty} \mathbf{C}_j^{(x)'} e^{-2\pi i \mu j}),$$

and

$$\mathbf{f}_v(\mu) = \frac{1}{\pi}(\sum_{j=0}^{\infty} \mathbf{C}_j^{(v)} e^{2\pi i \mu j})\mathbf{\Sigma}_e^{(v)}(\sum_{j=0}^{\infty} \mathbf{C}_j^{(v)'} e^{-2\pi i \mu j}),$$

where we set $\mathbf{C}_0^{(x)} = \mathbf{C}_0^{(v)} = \mathbf{I}_p$ and $i^2 = -1$ (see Chapter 7 of Anderson (1971) for instance.)
The spectral density matrix of the transformed vector process $\Delta \mathbf{y}_i$ $(= \mathbf{y}_i - \mathbf{y}_{i-1})$ is

$$\mathbf{f}_{\Delta y}(\mu) = \mathbf{f}_{\Delta x}(\mu) + (1 - e^{2i\mu}) f_v(\mu)(1 - e^{-2i\mu})$$

and we have the relation

$$\mathbf{f}_{\Delta y}(0) = \mathbf{f}_{\Delta x}(0) \quad .$$

We denote the long-run variance-covariance matrices of the trend components and the stationary components for $g, h = 1, \cdots, p$ as

$$\mathbf{\Omega}_x = \mathbf{f}_{\Delta x}(0) \; (= (\omega_{gh}^{(x)})) \; ,$$

and

$$\mathbf{\Omega}_v = f_v(0) \; = (\omega_{gh}^{(v)}) \; .$$

Macro-SIML Estimation of Common Trends

Consider the situation when each vectors $\Delta \mathbf{x}_i$ and $\mathbf{v}_i$ are independently, identically, and normally distributed (i.i.d.) as $N_p(\mathbf{0}, \mathbf{\Sigma}_x)$ and $N_p(\mathbf{0}, \mathbf{\Sigma}_v)$, respectively. We have the observations of an $n \times p$ matrix $\mathbf{Y}_n = (\mathbf{y}_i^{'})$ and set the $np \times 1$ random vector $(\mathbf{y}_1^{'}, \cdots, \mathbf{y}_n^{'})^{'}$. Given the initial condition $\mathbf{y}_0$, we have

$$\mathrm{vec}(\mathbf{Y}_n) \sim N_{n\times p}\left(\mathbf{1}_n \cdot \mathbf{y}_0^{'}, \mathbf{I}_n \otimes \mathbf{\Sigma}_v + \mathbf{C}_n \mathbf{C}_n^{'} \otimes \mathbf{\Sigma}_x\right) ,$$

where $\mathbf{1}_n^{'} = (1, \cdots, 1)$ and

$$\mathbf{C}_n = \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 \\ 1 & 1 & 0 & \cdots & 0 \\ 1 & 1 & 1 & \cdots & 0 \\ 1 & \cdots & 1 & 1 & 0 \\ 1 & \cdots & 1 & 1 & 1 \end{pmatrix}_{n \times n} .$$

Then, given the initial condition $\mathbf{y}_0$, the conditional log-likelihood function except a constant is

$$L_n^* = \log|\mathbf{I}_n \otimes \boldsymbol{\Sigma}_v + \mathbf{C}_n\mathbf{C}_n^{'} \otimes \boldsymbol{\Sigma}_x|^{-1/2}$$
$$-\frac{1}{2}[vec(\mathbf{Y}_n - \bar{\mathbf{Y}}_0)^{'}]^{'}[\mathbf{I}_n \otimes \boldsymbol{\Sigma}_v + \mathbf{C}_n\mathbf{C}_n^{'} \otimes \boldsymbol{\Sigma}_x]^{-1}[vec(\mathbf{Y}_n - \bar{\mathbf{Y}}_0)^{'}] \quad ,$$

where $\bar{\mathbf{Y}}_0 = \mathbf{1}_n \cdot \mathbf{y}_0^{'}$.
We use the K_n^*-transformation that from $\mathbf{Y}_n$ to $\mathbf{Z}_n$ $(=(\mathbf{z}_k^{'}))$ by $\mathbf{Z}_n = \mathbf{K}_n^*\left(\mathbf{Y}_n - \bar{\mathbf{Y}}_0\right)$, $\mathbf{K}_n^* = \mathbf{P}_n\mathbf{C}_n^{-1}$, where

$$\mathbf{C}_n^{-1} = \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 \\ -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & 0 & \cdots \\ 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & -1 & 1 \end{pmatrix}_{n\times n} \quad ,$$

and $\mathbf{P}_n = (p_{jk}^{(n)})$, $p_{jk}^{(n)} = \sqrt{\frac{2}{n+\frac{1}{2}}}\cos\left[\frac{2\pi}{2n+1}(k-\frac{1}{2})(j-\frac{1}{2})\right]$.

By using the spectral decomposition $\mathbf{C}_n^{-1}\mathbf{C}_n^{\prime -1} = \mathbf{P}_n\mathbf{D}_n\mathbf{P}_n^{\prime}$ and $\mathbf{D}_n$ is a diagonal matrix with the k-th element

$$d_k = 2[1 - \cos(\pi(\frac{2k-1}{2n+1}))] \ (k = 1, \cdots, n) \ .$$

Then the conditional likelihood function given the initial condition is proportional to

$$L_n = \sum_{k=1}^{n} \log |a_{kn}^{*}\mathbf{\Sigma}_v + \mathbf{\Sigma}_x|^{-1/2} - \frac{1}{2}\sum_{k=1}^{n} \mathbf{z}_k^{\prime}[a_{kn}^{*}\mathbf{\Sigma}_v + \mathbf{\Sigma}_x]^{-1}\mathbf{z}_k \ ,$$

where

$$a_{kn}^{*} \ (= d_k) \ = 4\sin^2\left[\frac{\pi}{2}\left(\frac{2k-1}{2n+1}\right)\right] \ (k = 1, \cdots, n) \ .$$

In the above representation, we have used two transformations on the nonstationary time series into the sequence of independent random variables $\mathbf{z}_k$ $(k = 1, \cdots, n)$ which follows $N_p(\mathbf{0}, \mathbf{\Sigma}_x + a_{kn}^{*}\mathbf{\Sigma}_v)$, and the coefficients a_{kn}^{*} is a dense sample of $4\sin^2(x)$ in $(0, \pi/2)$

It may be natural to use $\mathbf{z}_k\mathbf{z}_k^{'}$ to estimate $\mathbf{\Sigma}_x + a_{kn}^{*}\mathbf{\Sigma}_v$ since it is the variance-covariance matrix of $\mathbf{z}_k$. We notice that $a_{kn} \to 0$ as $n \to \infty$ for a fixed k. When k is small, a_{kn^*} is small and we can expect that $k = k_n$ depending n is still small when n is large. However, $a_m = (1/m_n)\sum_{k=1}^{m_n} a_{kn}^{*}$ is not small if m_n is near to n, which suggests the condition $m_n/n \to 0$ as $n \to \infty$. The separating information maximum likelihood (SIML) estimator of $\hat{\mathbf{\Sigma}}_x$ can be defined by

$$\mathbf{G}_m = \hat{\mathbf{\Sigma}}_{x,SIML} = \frac{1}{m_n}\sum_{k=1}^{m_n} \mathbf{z}_k\mathbf{z}_k^{'} \ .$$

Non-stationary Common Trends

Let $\mathbf{y}_i$ be the $i-$th observation of p-dimensional time series for $i = 1, \cdots, n$ and $\mathbf{y}_i = \mathbf{x}_i + \mathbf{v}_i$ and $\mathbf{Y}_n = (\mathbf{y}_i^{'})$ be an $n \times p$ $(p \geq 1)$ vector of observations. We assume that the vectors $\mathbf{x}_i$ satisfy

$$\mathbf{x}_i = \mathbf{x}_{i-1} + \mathbf{\Pi} \boldsymbol{\mu}_i \ ,$$

where $\mathbf{\Pi}$ is a $p \times q$ $(1 \leq q < p)$ matrix, $\boldsymbol{\mu}_i$ is a sequence of $q \times 1$ (i.i.d.) random vectors following $N_q(\mathbf{0}, \mathbf{\Sigma}_\mu)$ and $\mathbf{v}_i$ are i.i.d. (p-dimensional) random vectors following $N_p(\mathbf{0}, \mathbf{\Sigma}_v)$ with the (non-singular) variance-covariance matrix $\mathbf{\Sigma}_v$. We denote $\mathbf{\Pi}^* = \mathbf{\Pi} \mathbf{\Sigma}_\mu^{1/2}$. We note that a normalization for the matrix $\mathbf{\Pi}$ is needed. Since the rank of $\mathbf{\Pi}$ is q, then when $1 \leq q < p$ we can set a $p \times r$ matrix $\mathbf{B}$ such that

$$\mathbf{B}^{'} \mathbf{\Pi}^* = \mathbf{O}$$

and $r = p - q$.

One way to characterize the present formulation is to use

$$\mathcal{E}[\mathbf{z}_k \mathbf{z}_k^{'}] = \mathbf{\Sigma}_x + o(1) \ \text{for} k = 1, \cdots, m_n$$

$$\mathcal{E}[a_{kn}^{*-1} \mathbf{z}_k \mathbf{z}_k^{'}] = \mathbf{\Sigma}_v + \frac{1}{4}\mathbf{\Sigma}_x + o(1) \ \text{for} k = n+1-m_n, \cdots, n \ ,$$

where the rank of matrix $\mathbf{\Sigma}_x$ is $p - r \ (= q)$. Hence the present problem has a form of the reduced-rank problem asymptotically in the sense that n is large. Then it may be natural to consider the characteristic equation

$$\mathbf{G}_m \hat{\mathbf{B}}_{SIML} - \mathbf{H}\hat{\mathbf{B}}_{SIML}\mathbf{\Lambda} = \mathbf{0} \ ,$$

and $\mathbf{G}_m \ (= \hat{\Sigma}_{x.SIML}) = \frac{1}{m_n}\sum_{k=1}^{m_n} \mathbf{z}_k \mathbf{z}_k^{'}$, where

$$\mathbf{\Lambda} = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & 0 & \cdots \\ 0 & \cdots & & \\ 0 & \cdots & & \lambda_r \end{bmatrix} ,$$

with $0 < \lambda_1 < \cdots < \lambda_r < \cdots < \lambda_p$ and

$$|\mathbf{G}_m - \lambda \mathbf{H}| = 0 \ ,$$

where $\mathbf{H}$ is a non-singular matrix.

To estimate the variance-covariance matrix $\mathbf{\Sigma}_v$ for the noise vectors, there are two candidates. The first one we may use is given by

$$\hat{\Sigma}_{v.SIML}(1) = \frac{1}{2}[\frac{1}{n}\sum_{k=1}^{n}\mathbf{z}_k\mathbf{z}_k^{'} - \hat{\Sigma}_{x.SIML}] \ ,$$

and the second one is

$$\hat{\Sigma}_{v.SIML}(2) = \frac{1}{l_n}\sum_{k=n+1-l_n}^{n} a_{kn}^{-1}\mathbf{z}_k\mathbf{z}_k^{'} - \frac{1}{4}\hat{\Sigma}_{x.SIML} \ .$$

We call $\hat{\mathbf{B}}_{SIML}$ as the SIML estimator of $\mathbf{B}$. A simplified (consistent) estimation can be obtained by using the equations

$$\hat{\mathbf{\Sigma}}_{x.SIML} \times \hat{\mathbf{B}}_{SIL} = \mathbf{O} \ ,$$

that is,

$$\hat{\mathbf{\Sigma}}_{x.SIML} \times [\begin{array}{c} \mathbf{I}_r \\ -\hat{\mathbf{B}}_{2.SIL} \end{array}] = \mathbf{O} \ .$$

We can solve them as

$$\hat{\mathbf{B}}_{2.SIL} = \hat{\mathbf{\Sigma}}_{22x.SIML}^{-1}\hat{\mathbf{\Sigma}}_{21x.SIML} \ ,$$

Theorem 4.1 : In the non-stationary errors-in-variables model with $1 \leq q < p$, assume that (i) the sequences of coefficients $\mathbf{C}_i^{(x)}$ and $\mathbf{C}_v^{(v)}$ are squared summable, (ii) $\mathbf{e}_i^{(x)} = (e_{ji}^{(x)})$ and $\mathbf{e}_i^{(v)} = (e_{ji}^{(v)})$ are a sequence of independent random variables with $\mathcal{E}[e_{ji}^{(x)4}] < \infty$ and $\mathcal{E}[e_{ji}^{(v)4}] < \infty$ $(i, j = 1, \cdots, n; g, h = 1, \cdots, p)$, where $\mathbf{e}_i^{(x)} = (e_{ji}^{(x)})$ and $\mathbf{e}_i^{(v)} = (e_{ji}^{(v)})$ for $j = 1, \cdots, p$. We use $\mathbf{G}_m$ as the estimator $\mathbf{\Omega}_x$.

Then (i) For $m_n = [n^\alpha]$ and $0 < \alpha < 1$, as $n \longrightarrow \infty$

$$\hat{\mathbf{\Omega}}_x - \mathbf{\Omega}_x \xrightarrow{p} \mathbf{O} .$$

(ii) For $m_n = [n^\alpha]$ and $0 < \alpha < 0.8$, as $n \longrightarrow \infty$

$$\sqrt{m_n}\left[\hat{\omega}_{gh}^{(x)} - \omega_{gh}^{(x)}\right] \xrightarrow{\mathcal{L}} N\left(0, \omega_{gg}^{(x)}\omega_{hh}^{(x)} + \left[\omega_{gh}^{(x)}\right]^2\right) .$$

The covariance of the limiting distributions of $\sqrt{m_n}[\hat{\omega}_{gh}^{(x)} - \omega_{gh}^{(x)}]$ and $\sqrt{m_n}[\hat{\omega}_{kl}^{(x)} - \omega_{kl}^{(x)}]$ is given by $\omega_{gk}^{(x)}\omega_{hl}^{(x)} + \omega_{gl}^{(x)}\omega_{hk}^{(x)}$ $(g, h, k, l = 1, \cdots, p)$.

Theorem 4.2 : Assume the conditions for the asymptotic normality of the SIML estimator of $\hat{\mathbf{\Omega}}_x$ hold and $\mathbf{H}$ is a positive definite matrix. We take $m = m_n = n^{\alpha}$ $(0 < \alpha < 0.8)$. As $n \longrightarrow \infty$, $m \longrightarrow \infty$ and we have

$$c_m \left[\hat{\mathbf{B}}_2 - \mathbf{B}_2 \right] \xrightarrow{w} N_{rq} \left(\mathbf{0}, \mathbf{\Omega}_v^{*} \otimes [(\mathbf{O}, \mathbf{I}_q)\mathbf{\Omega}_x(\mathbf{O}, \mathbf{I}_q)^{'}]^{-1} \right) , \tag{1}$$

where $\mathbf{\Omega}_v^{*} = \mathbf{B}^{'}\mathbf{\Omega}_v\mathbf{B}$.

Theorem 4.3 : We take $m = m_n = n^{\alpha}$ $(0 < \alpha < 0.8)$. As $n \longrightarrow \infty$, $a_m \longrightarrow 0$, $c_m \longrightarrow \infty$, and

$$\sqrt{m}\left[(\frac{1}{a_m})\lambda_i - \lambda_{0i}^*\right] \xrightarrow{w} N(0, \frac{9}{5}d_{ii})$$

for $i = 1, \cdots, r$ $(= p - q)$ and $\sqrt{m}\left[(\frac{1}{a_m})\lambda_i - \lambda_{0i}^*\right]$ are asymptotically normal jointly, where $\boldsymbol{\Lambda}_0^* = (\mathrm{diag}\ \lambda_{0i})$. The covariances of $\sqrt{m}[(1/a_m)\lambda_i - \lambda_{0i}^*]$ and $\sqrt{m}[(1/a_m)\lambda_j - \lambda_{0i}^*]$ $(i,j = 1, \cdots, r)$ are given by $d_{ij} = \frac{9}{5}[\delta(i,j) + \omega^{ij}\omega_{ij}]$. Let $\lambda_i^* = \sqrt{m}[(\frac{1}{a_m})\lambda_i - \lambda_{0i}^*]$ $(i = 1, \cdots, r)$. Then we have

$$\begin{aligned}
\mathrm{Var}[\sum_{i=1}^{r} \lambda_i^*] &= \mathcal{E}\left[\sum_{i,i'=1}^{r} \lambda_i^* \lambda_{i'}^*\right] \\
&= \frac{a_m(2)}{a_m^2} \sum_{i,i'=1}^{r} \left[\delta(i,i') + \omega_{ii'}\omega^{ii'}\right] \\
&= \frac{9}{5} 2r\ .
\end{aligned}$$

Corollary 4.4 : We take $m = m_n = n^{\alpha}$ $(0 < \alpha < 0.8)$ and $\mathbf{H} = \hat{\mathbf{\Omega}}_v$. As $n \longrightarrow \infty$, $a_m \longrightarrow 0$, $c_m \longrightarrow \infty$,

$$\sqrt{\frac{m}{a_m(2)}}\left[\sum_{i=1}^{r}(\lambda_i - a_m)\right] \xrightarrow{w} N(0, 2r)\ .$$

and

$$\left[\frac{m}{a_m(2)}\frac{1}{2r}[\sum_{i=1}^{r}(\lambda_i - a_m)]^2\right] \xrightarrow{w} \chi^2(1)\ .$$

It may be natural to use the statistic for investigating the rank condition in the underlying population.
When we use $\mathbf{H} = \mathbf{I}_p$, let

$$\mathbf{\Lambda}^{**} = (\mathbf{B}^{'}\mathbf{H}\mathbf{B})\hat{\mathbf{\Omega}}_v^{*-1}\mathbf{\Lambda}^{*} \; (= (\lambda_{ij}^{**})) \; ,$$

and $\mathbf{\Lambda}^{*} = (\mathrm{diag}\lambda_i^{*})$. Then we replace $\mathbf{B}$ and $\mathbf{\Omega}_v^{*}$ by their consistent estimators, we have the following result.

Corollary 4.5 : We take $m = m_n = n^{\alpha}$ $(0 < \alpha < 0.8)$ and $\mathbf{H} = \mathbf{I}_p$. As $n \longrightarrow \infty$, $a_m \longrightarrow 0$, $c_m \longrightarrow \infty$,

$$\left[\frac{1}{2ra_m(2)}[\sum_{i=1}^{r}\lambda_{ii}^{**}]^2\right] \xrightarrow{w} \chi^2(1) \; .$$

For testing the hypothesis $H_0 \; : \; r = r_0$ ($r_0 \geq 1$ is a specified nbumber) against $H_A \; : \; r = r_0 - 1$, it may be reasonable to use the $r_0 - th$ smaller characteristic root and the rejection region can be constructed by the limiting normal distribution under H_0. Under H_A, the characteristic root $\lambda_{r_0} \xrightarrow{p} \infty$ and the test should be consistent. It is also possible to consider a sequence of local alternative hypothesis for the situation when the rank condition is true asymptotically.

Alternatively, we can use

$$R_0 = \sum_{i=1}^{r_0} \lambda_i \; .$$

It is straightforward to derive the asymptotic distribution of R_0.

The log-likelihood function is given by (-2) times

$$L_{1n}^* = \sum_{k=1}^{n} \log |a_{kn}^* \mathbf{\Sigma}_v + \mathbf{\Sigma}_x| + \sum_{k=1}^{n} \mathrm{tr}[a_{kn}^* \mathbf{\Sigma}_v + \mathbf{\Sigma}_x]^{-1} \mathbf{z}_k \mathbf{z}_k^{'} .$$

For the (-2) log-likelihood for the non-stationary trend factors may be approximated as

$$L_{2n}^* = \log |(\frac{1}{m}\sum_{k=1}^{m} a_{kn}^*)\mathbf{\Sigma}_v + \mathbf{\Sigma}^{(x)}| + \mathrm{tr}[(\frac{1}{m}\sum_{k=1}^{m} a_{kn}^*)\mathbf{\Sigma}_v + \mathbf{\Sigma}^{(x)}]^{-1}[\frac{1}{m}\sum_{k=1}^{m} \mathbf{z}_k \mathbf{z}_k^{'}] .$$

The number of trend parameters without constraint is $p(p+1)/2$, which is the number of parameters of the variance-covariance matrix. When $0 < q < p$, the number of free parameters is $q(q+1)/2 + q(p-q) = q(2p+1-q)/2$ (the number of coefficients $\mathbf{\Pi}$ is $p \times q$ and we can take an arbitrary coefficients $\mathbf{\Pi}_2$). For instance, when $p = 3$ it is 6 $(p = q)$, 5 $(q = p-1)$ and 3 $(q = p-2)$. Then we can define AIC by

$$\mathrm{AIC} = L_{1n}^*(\hat{\theta}_{ML}) + \frac{q(2p+1-q)}{2} \; (q = 1, \cdots, p),$$

Alternatively, we may use a non-stationary criterion

$$\mathrm{NAIC} = \sum_{i=1}^{p-q} \log[a_m + \lambda_i] + \frac{q(2p+1-q)}{2} \; (q = 1, \cdots, p).$$

We choose q^* such that AIC or NAIC is minimized with respect to q (or r) and we can determine the number of non-stationary factors.

Some remarks

- In many macro-times series we observe non-stationary trend, non-stationary seasonality and stationary cycles with measurement errors
- We develop a new way to determine the number of trend factors
- It is also straight-forward to develop a new way to determine the number of seasonal factors (Kunitomo (2018, unpublished))

References

- Akaike, H. (1973), "Information Theory and an extension of the maximum likelihood principle," 2nd International Symposium on Information Theory, B.N. Petrov and F. Csaki edited, Academiai Kiado, Budapest, 267-281.
- Anderson, T.W. (1971), *The Statistical Analysis of Time Series*, John-Wiley.
- Anderson, T.W. (1984), "Estimating Linear Statistical Relationships," *Annals of Statistics*, 12, 1-45.
- Engle, R. and C.W.J. Granger (1987), "Co-integration and Error Correction," *Econometrica*, Vol.55, 251-276.
- Kitagawa, G. (2010), *Introduction to Time Series Analysis*, CRC Press.
- Johansen, S. (1995), *Likelihood Based Inference in Cointegrated Vector Autoregressive Models*, Oxford UP.

- Kunitomo, N. (2018), "Common Seasonal Factors," In preparation.
- Kunitomo, N. , Sato and D. Kurisu (2018), *Separating Information Maximum Likelihood Estimation for High Frequency Financial Data*, Springer.
- Kunitomo, N. and S. Sato (2017), "Trend, Seasonality and Economic Time Series : the Non-stationary Errors-in-variables Models," SDS-4, MIMS, Meiji University, http://www.mims.meiji.ac.jp/publications/2017-ds.
- Kunitomo, N., N. Awaya and D. Kurisu (2017), "Some Properties of Estimation Methods for Structural Relationships in Non-stationary Errors-in-Variables Models," SDS-3, MIMS, Meiji University.
- Müller, U. and M. Watson (2018), "Long-run Covariability," *Econometrica*, 86-3, 775-804.

不等間隔観測の下でのノンパラメトリック空間回帰モデルに対する統計的推測

栗栖大輔

東工大 工学院 経営工学系

2018 年 8 月 6 日
データサイエンス・松本キャンプ 2018

Outline

- Introduction
- Assumptions
- Main results
- Simulations
- Conclusion

Introduction

本研究では以下で定義される空間回帰モデルを考える:

$$Y(s_j) = \mu(X(s_j)) + \sigma(X(s_j))V(s_j),\ j = 1, \dots, n.$$

ここで, $Y(s), X(s) \in \mathbb{R}$, $s \in \mathbb{R}^2$,

$\{X(s) : s \in \mathbb{R}^2\}$: strictly stationary spatial process (random field).

$V(s)$: i.i.d. random variables.

$\{(Y(s_j), X(s_j))\}_{j=1}^n$: 観測データ.

$\mu : \mathbb{R} \to \mathbb{R}$, $\sigma : \mathbb{R} \to [0, \infty)$.

Introduction

- 空間回帰モデルは geostatistical data の分析 (空間計量経済学, 環境データ, 地震データ etc.) で利用される.
- 例 1: インドの過去の土地収益の管理体制が現在の経済状況に与える影響の分析 (Robinson and Thawornkaiwong ('12, JoE)).
- 例 2: 海岸線付近のミネラルや有機物の濃度が海岸生態系に与える影響の分析 (Hallin et al. ('09, Bernoulli)).
- 例 3: 地震観測地点における深度とマグニチュード・震度の関係 (Fernández et al.('12, Environmetrics)).
- 例 4: 地球統計学 (分散不均一モデル, 鉱物の分布)

$$Y(s_j) = \beta X(s_j) + \sigma(X(s_j))V(s_j)\, j = 1, \ldots, n.$$

Nonparametric spatial regression models

データの観測の枠組みとして, domain expanding and infill (DEI) asymptotics を考える. $\{s_j\}_{j=1}^n$: データの観測地点.

- $\|\cdot\|$: Euclidean norm on $\mathbb{R}^2$.

$$\delta_{j,n} = \min\{\|s_k - s_j\| : 1 \le k \le n, k \ne j\},$$
$$\Delta_{j,n} = \max\{\|s_k - s_j\| : 1 \le k \le n, k \ne j\}.$$

- $\max_{1\le j\le n} \delta_{j,n} \to 0$ as $n \to \infty$.
- $\min_{1\le j\le n} \Delta_{j,n} \to \infty$ as $n \to \infty$.

特にデータが格子点上 ($\mathbb{Z}^d$) ではなく連続な空間上 ($\mathbb{R}^d$) で定義された空間過程モデルを想定する場合に相性がいい枠組み.

例 1 Brownian (or Gaussian) random fields.

例 2 Spatial Lévy-driven OU processes (Nguyen and Veraart ('17, SJS)).

例 3 Lévy-driven CARMA random fields (Brockwell and Matsuda ('17, JRSSB)) etc...

Nonparametric spatial regression models

本研究の目標は以下の 2 つ：

- DEI asymptotics のもとで平均関数 μ, 分散関数 σ^2 に対する多次元 CLT を導出.
- データ駆動型の推定量のバンド幅の選択方法を提案.

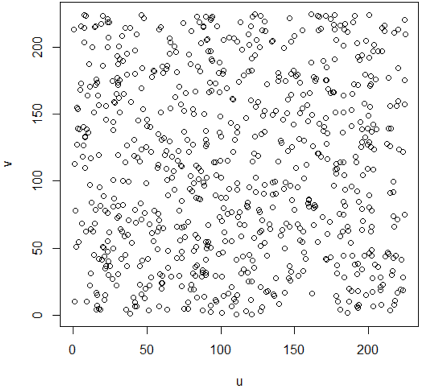


Figure: Simulated observation points which are randomly sampled from the lattice (u_j, v_k) with $u_j = 0.3 + (j-1) \times 0.3$ and $v_k = 0.6 + (k-1) \times 0.3$, $j, k = 1, \ldots, 750$.

Nonparametric spatial regression models

Previous works:

- Estimation (random field/spatial regression)
  - DEI asymptotics (random) : Matsuda and Yajima ('09, JRSSB).
  - (irregularly observed) lattice data : Jenish and Prucha ('09, JoE), Li ('16, SISP).
- Inference (random field/spatial regression/time series regression)
  - DEI asymptotics (deterministic) : Lu and Tjøstheim ('14, JASA).
  - (irregularly observed) lattice data : Hallin et al.('04, Ann. Statist.), Hallin et al.('09, Bernoulli), Robinson ('11, JoE), Robinson and Thawornkaiwong ('12, JoE).
  - regular observation: Zhang and Wu ('08, Ann. Statist.).

Assumptions

(A1): **Assumption on spatial process**

(i) $\{X(s), s \in \mathbb{R}^2\}$ is a strictly stationary spatial process, satisfying the α-mixing property that there exist a function φ such that $\varphi(t) \downarrow 0$ as $t \to \infty$, and a function $\psi : \mathbb{N}^2 \to [0, \infty)$ symmetric and increasing in each of its two arguments, such that

$$\alpha(\mathcal{B}(\mathcal{S}'), \mathcal{B}(\mathcal{S}'')) = \sup_{A \in \mathcal{B}(\mathcal{S}'), B \in \mathcal{B}(\mathcal{S}'')} |P(A \cap B) - P(A)P(B)| \leq \psi(\mathrm{Card}(\mathcal{S}'), \mathrm{Card}(\mathcal{S}''))\varphi(d(\mathcal{S}', \mathcal{S}'')), \quad (1)$$

where $\mathcal{S}', \mathcal{S}'' \subset \mathbb{R}^2$, $\mathcal{B}(\mathcal{S})$ be the Borel σ-field generated by $\{X_s, s \in \mathcal{S}\}$, and $d(\mathcal{S}', \mathcal{S}'') = \min\{\|s_k - s_j\| | s_k \in \mathcal{S}', s_j \in \mathcal{S}''\}$ for each $\mathcal{S}'$ and $\mathcal{S}''$.

(ii) For some constant $\gamma > \max\{1, 2\kappa/(2+\kappa)\}$ and some $\kappa > 0$,

$$\limsup_{m \to \infty} m^{4+3\gamma} \sum_{j=m}^{\infty} j \varphi^{\kappa/(2+\kappa)}(j) < \infty.$$

(A1): **Assumption on spatial process (conti.)**

(iii) $n\psi(1,n)\varphi(c_n) \to 0$ as $n \to \infty$, where $c_n = \{\delta_n^2 h^{\kappa/(2+\kappa)}\}^{-1/\gamma}$.

- 条件 (1) で $\psi \equiv 1$ のとき, random field は strong mixing であるという. Strong mixing condition は多くの確率過程 (diffusion process, Lévy-driven jump diffusion SDE etc.) や時系列モデル (ARMA, (nonlinear) ARCH, Markov chain etc.) に対して満たされる.
- 条件 (1) は連続時間確率過程や (非線形時系列) 時系列モデルに対する strong mixing condition の一般化.
- 実際この条件は多くの空間過程 (or 確率場) に対して満たされる (例えば Gaussian (or CARMA) random field など. 詳しくは Rosenbratt(1985), Guyon(1987) を参照).

Assumptions

(A2): **Assumption on bandwidths and sampling scheme**
As $n \to \infty$, the deterministic sequences $\{b_n\}_{n \geq 1}$ and $\{h_n\}_{\geq 1}$ satisfy the following conditions:

(i) $b_n, h_n \to 0$.

(ii) $n(b_n + h_n) \to \infty$, $\liminf_{n \to \infty} n(b_n^2 + h_n^2) > 0$, and $n(b_n^5 + h_n^5) = O(1)$.

(iii) $\delta_n^{-(2+2/\gamma)}(b_n + h_n)^{1-2\kappa/\{(2+\kappa)\gamma\}} \to 0$.

(A3): **Assumption on kernel function**

(i) The kernel function K is bounded, symmetric, and has a bounded support. Let $c_K = \int_{\mathbb{R}} z^2 K(z) dz$.

Assumptions

(A4): **Assumption on regression models**
Let f and $f_{j,k}$ be density functions of $X(s)$ and $(X(s_j), X(s_k))$ with $j \neq k$ respectively. For some $\epsilon > 0$, $U \subset \mathbb{R}^d$, U^ϵ denotes an ϵ-enlargement of U, that is, $U^\epsilon := \{x : \|x - y\| < \epsilon, y \in U\}$.

(i) $\inf_{x \in I^\epsilon} f(x) > 0$ and $\inf_{x \in I^\epsilon} \sigma(x) > 0$ for some compact set $I \subset \mathbb{R}$ and some $\epsilon > 0$ and $I \subset \mathbb{R}$.

(ii) The marginal density f and the joint density $f_{j,k}$ satisfy $|f_{j,k}(x, y) - f(x)f(y)| \leq C$ uniformly for $j \neq k$ and $(x, y) \in \mathbb{R}^2$ where C is a universal positive constant, and $\mu \in C^4(I^\epsilon)$, $\sigma \in C^2(I^\epsilon)$ where $I^{\otimes d} = \underbrace{I \times \ldots \times I}_{d}$. Here, for $S \subset \mathbb{R}$,

$$C^p(S) = \{f : \sup_{x \in S} |f^{(k)}(x)| < \infty, k = 0, 1, \ldots, p\}$$

is the set of functions having bounded derivatives on S up to order p, and $C^0(S)$ is the set of continuous functions on S.

(A4): **Assumption on regression models (conti.)**

(iii) $E[V(s_1)] = 0$, $E[V^2(s_1)] = 1$, and $E[|V(s_1)|^{8+4\kappa}] < \infty$. Here, $\kappa > 0$ is the constant which appear in Assumption (A1) (ii).

(iv) $\{X(s) : s \in \mathbb{R}^2\}$ and $\{V(s) : s \in \mathbb{R}^2\}$ are independent.

Main results

以下の推定量に対する極限定理を考える:

$$\widehat{\mu}_{b_n}(x) = \frac{1}{nb_n\widehat{f}_X(x)} \sum_{j=1}^{n} Y(s_j) K\left(\frac{x - X(s_j)}{b_n}\right),$$

$$\widehat{\sigma}^2_{h_n}(x) = \frac{1}{nh_n\tilde{f}_X(x)} \sum_{j=1}^{n} \{Y(s_j) - \widehat{\mu}^*_{b_n}(X(s_j))\}^2 K\left(\frac{x - X(s_j)}{h_n}\right).$$

ここで,

$$\widehat{f}_X(x) = \frac{1}{nb_n} \sum_{j=1}^{n} K\left(\frac{x - X(s_j)}{b_n}\right), \quad \tilde{f}_X(x) = \frac{1}{nh_n} \sum_{j=1}^{n} K\left(\frac{x - X(s_j)}{h_n}\right),$$

$$\widehat{\mu}^*_{b_n}(x) = 2\widehat{\mu}_{b_n}(x) - \widehat{\mu}_{\sqrt{2}b_n}(x).$$

Main results

- b_n, h_n: バンド幅, $b_n, h_n \to 0$, $n(b_n + h_n) \to \infty$ as $n \to \infty$.
- K: a kernel function with $\int K(x)dx = 1$.
- Bias reduction のため, $\widehat{\sigma}^2_{h_n}$ の定義で $\widehat{\mu}_{b_n}$ の代わりに $\widehat{\mu}^*_{b_n}$ ($\widehat{\mu}_{b_n}$ の jackknife version) を使っている. 各 $x \in \mathbb{R}$ に対して,

$$\widehat{\mu}_{b_n}(x) - \mu(x) = O_P(b_n^2),$$
$$\widehat{\mu}^*_{b_n}(x) - \mu(x) = O_P(b_n^4).$$

Main results

Proposition 1

Under Assumptions (A1), (A2), (A3) and (A4), for $-\infty < x_1 < x_2 < \cdots < x_N < \infty$, we have that

$$\sqrt{\frac{nb_n}{\|K\|_{L^2}^2}} \begin{pmatrix} \frac{\widehat{f}(x_1) - f(x_1) - b_n^2 c_K f''(x_1)/2}{\sqrt{f(x_1)}} \\ \vdots \\ \frac{\widehat{f}(x_N) - f(x_N) - b_n^2 c_K f''(x_N)/2}{\sqrt{f(x_N)}} \end{pmatrix} \xrightarrow{d} N(0, I_N).$$

where I_N is the $N \times N$ identity matrix.

Main results

- DEI asymptotics のもとではノンパラメトリック時系列データの解析でよく使われる small block-large block technique (Bernstein (1926, Ann. Math.)) が使えない.

  → 観測データが (regular) lattice 上のにある場合と異なり, データ間の距離が近づいていくため, 漸近的に独立になるような観測データのブロックが構成できない.

- 代わりに Bolthausen ('82, Ann. Probab.) による CLT の証明法 (特性関数の収束) を使う.

Main results

Proposition 2

Assume that G and H are functions such that

(i) $G \in C^0(\{x\}^\epsilon)$ for some $\epsilon > 0$.

(ii) $H : \mathbb{R} \to \mathbb{R}$ with $E[H(V(s_1))] = 0$, $E[H^2(V(s_1))] > 0$ and $E[|H(V(s_1))|^{4+2\kappa}] < \infty$.

Define

$$U_n = \frac{1}{\sqrt{nh_n}} \sum_{j=1}^{n} G(X(s_j))H(V(s_j))K\left(\frac{x - X(s_j)}{h_n}\right).$$

Then, under Assumptions (A1), (A2), (A3) and (A4), for $x \in I$ we have that

$$U_n \xrightarrow{d} N(0, f(x)G^2(x)V(H)\|K\|_{L^2}^2)$$

as $n \to \infty$, where $V(H) = E[H^2(V(s_1))]$.

Main results

Theorem 1

Under Assumptions (A1), (A2), (A3) and (A4), for $-\infty < x_1 < x_2 < \cdots < x_N < \infty$, we have that

$$\sqrt{\frac{nb_n}{\|K\|_{L^2}^2}} \begin{pmatrix} \frac{\sqrt{f(x_1)}(\widehat{\mu}_{b_n}(x_1) - \mu(x_1) - b_n^2 c_K \rho_\mu(x_\ell))}{\sqrt{\sigma^2(x_1)}} \\ \vdots \\ \frac{\sqrt{f(x_N)}(\widehat{\mu}_{b_n}(x_N) - \mu(x_N) - b_n^2 c_K \rho_\mu(x_\ell))}{\sqrt{\sigma^2(x_N)}} \end{pmatrix} \xrightarrow{d} N(0, I_N).$$

where $\rho_\mu(x) = \mu''(x)/2 + \{\mu(x)f''(x)/2 + \mu'(x)f'(x)\}/f(x)$ and I_N is the $N \times N$ identity matrix.

Main results

Theorem 2

Under Assumptions (A1), (A2), (A3) and (A4), and $E[V^4(s_1)] > 1$, for $-\infty < x_1 < x_2 < \cdots < x_N < \infty$, we have that

$$\sqrt{\frac{nh_n}{V_4\|K\|_{L^2}^2}} \begin{pmatrix} \dfrac{\sqrt{f(x_1)}(\widehat{\sigma}^2_{h_n}(x_1) - \sigma^2(x_1) - h_n^2 c_K \rho_\sigma(x_\ell))}{\sqrt{\sigma^4(x_1)}} \\ \vdots \\ \dfrac{\sqrt{f(x_N)}(\widehat{\sigma}^2_{h_n}(x_N) - \sigma^2(x_N) - h_n^2 c_K \rho_\sigma(x_\ell))}{\sqrt{\sigma^4(x_N)}} \end{pmatrix} \xrightarrow{d} N(0, I_N).$$

where $V_4 = E[V^4(s_1)] - 1$,
$\rho_\sigma(x) = \sigma'(x) + \sigma(x)\sigma''(x) + \{2\sigma(x)\sigma'(x)f'(x) + \sigma^2(x)f''(x)/2\}/f(x)$,
and I_N is the $N \times N$ identity matrix.

Main results

Proposition 1, Theorems 1, 2 の結果を利用して f, μ, σ^2 に対する $I = [a, b] \subset \mathbb{R}$, $-\infty < a < b < \infty$ 上での信頼バンドの構成方法を考える.

- V_4 を以下で推定する:

$$\widehat{V}_4 = \frac{\sum_{j=1}^{n} \widehat{V}^4(s_j) 1\{X(s_j) \in I\}}{\sum_{j=1}^{n} 1\{X(s_j) \in I\}} - 1, \ \ \widehat{V}(s_j) = \frac{Y(s_j) - \widehat{\mu}^*_{b_n}(X(s_j))}{\widehat{\sigma}_{h_n}(X(s_j))}.$$

- naive estimator $n^{-1} \sum_{j=1}^{n} \widehat{V}^4(s_j) - 1$ だと X の値が大きいところで推定精度が落ちるので finite sample performance が悪くなる.
- $\xi_1, \ldots, \xi_N$ を i.i.d. standard normal random variables とし, q_τ, $\tau \in (0, 1)$ を以下を満たす値とする:

$$P\left(\max_{1 \leq j \leq N} |\xi_j| > q_\tau \right) = \tau.$$

Main results

このとき, 各 design points $x_1, \dots, x_N \in I$ における f, μ, σ^2 の joint asymptotic $100(1-\tau)\%$ confidence intervals は以下で与えられる:

$$\widehat{C}_f(x_j) = \left[\widehat{f}_X(x_j) \pm \sqrt{\frac{\widehat{f}_X(x_j)\|K\|_{L^2}^2}{nb_n} q_\tau} \right], \ j = 1, \dots, N,$$

$$\widehat{C}_\mu(x_j) = \left[\widehat{\mu}_{b_n}(x_j) \pm \sqrt{\frac{\widehat{\sigma}_{h_n}(x_j)\|K\|_{L^2}^2}{nb_n \widehat{f}_X(x_j)} q_\tau} \right], \ j = 1, \dots, N,$$

$$\widehat{C}_{\sigma^2}(x_j) = \left[\widehat{\sigma}^2(x_j) \pm \sqrt{\frac{\widehat{\sigma}_{h_n}^4(x_j)\widehat{V}_4\|K\|_{L^2}^2}{nh_n \widehat{f}_X(x_j)} q_\tau} \right], \ j = 1, \dots, N.$$

これらを線形補完すればそれぞれの関数の confidence bands が得られる.

Simulations

- 観測地点の生成: lattice (u_j, v_k) with $u_j = 0.3 + (j-1) \times 0.3$ and $v_k = 0.6 + (k-1) \times 0.3$ for $j, k = 1, \ldots, n$ から n 個の locations (u_{j_ℓ}, v_{k_ℓ}) with $1 \leq j_\ell, k_\ell \leq n$, $\ell = 1, \ldots n$ を選び, $s_\ell = (u_{j_\ell}, v_{k_\ell})$ とする.
- 空間過程 $X(s)$ の生成: 以下の spatial moving average process からデータを発生させる:

$$X(s_j) = \sum_{\ell=-1}^{1} \sum_{m=-1}^{1} a_{\ell,m} Z_{\ell,m}, \ s_j = (u_\ell, v_m), \ j = 1, \ldots, n.$$

ここで, $Z_{\ell,m}$ は i.i.d. standard normal random variables, $a_{\ell,m}$ は 以下の行列 A の $(\ell+2, m+2)$ 成分:

$$A = \begin{pmatrix} 1/5 & 2/5 & -4/5 \\ -3/5 & -2/5 & -1/5 \\ -1/5 & 2/5 & -3/5 \end{pmatrix}.$$

Simulations

- $Y(s)$ の生成:

$$\mu(x) = 0.1 + 0.3x,\ \ \sigma^2(x) = 0.2 + 0.05x + 0.3x^2.$$

- $V(s_j) \sim i.i.d.N(0,1)$.
- Kernel function: Epanechnikov kernel $K(x) = \frac{3}{4}(1 - x^2)1\{|x| \leq 1\}$.
- Sample size $n = 750$.

Simulations

- 実際に推定量を計算する際にはバンド幅の値を決める必要がある．特に $n(b_n^5 + h_n^5) = o(1)$ なら，信頼バンドの構成にあたって漸近バイアスの推定が必要なくなる．
- 具体的には推定の意味で optimal なバンド幅よりも order の意味で小さいバンド幅を選ばなくてはならない．
- 一般には optimal bandwidth は f, μ, σ の滑らかさや mixing condition に出てくる κ 等の未知のパラメータに依存する．
- Bissantz et al. ('09, JRSSB) : deconvolution problem, Kurisu (2018a): estimation of Lévy meas. of Lévy-driven OU proc. のアイデアに基づくデータ駆動型のバンド幅の選択法を提案．

Simulations

1. Set a pilot bandwidth $b^P > 0$ and make a list of candidate bandwidths $b_\ell = \ell b^P / L$ for $\ell = 1, \ldots, L$.
2. Choose the smallest bandwidth b_ℓ ($\ell \geq 2$) such that the adjacent value $\max_{1 \leq j \leq N} |\hat{\mu}_{b_\ell}(x_j) - \hat{\mu}_{b_{\ell-1}}(x_j)|$ is smaller than $\tau \times \min\{\max_{1 \leq j \leq N} |\hat{\mu}_{b_\ell}(x_j) - \hat{\mu}_{b_{\ell-1}}(x_j)| : \ell = 2, \ldots, L\}$ for some $\tau > 1$.

In our simulation study, we choose $b^P = h^P = 1, L = 20$, and $\tau = 2$.

Simulations

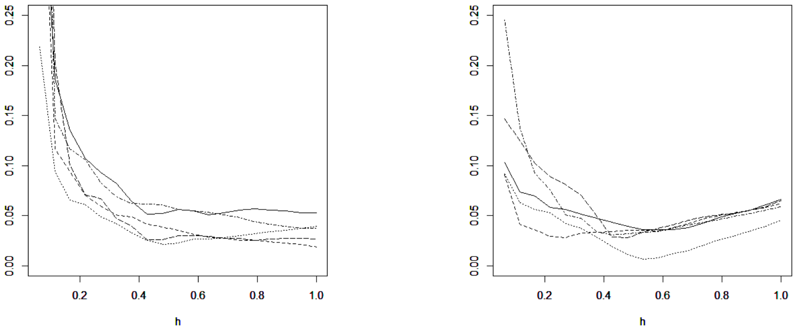
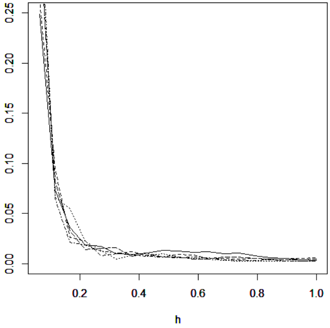


Figure: Discrete L^∞-distance between the true function μ and estimates $\hat{\mu}_{b_\ell}$ (left), and between the true function σ^2 and estimates $\hat{\sigma}^2_{h_\ell}$ (right) for different bandwidth values. We set $n = 750$, $I = [-0.5, 0, 5]$, and $x_j = -0.5 + (j-1) \times 0.1$, $j = 1, \ldots, 11$.

Simulations



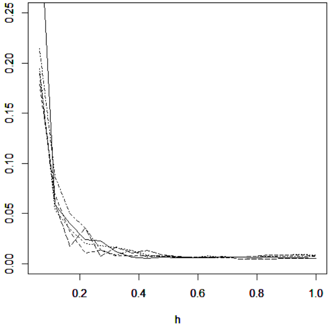


Figure: Discrete L^∞-distance between the estimates $\hat{\mu}_{b_\ell}$ (left), and between the estimates $\hat{\sigma}^2_{h_\ell}$ (right) for different bandwidth values. We set $n = 750$, $I = [-0.5, 0, 5]$, and $x_j = -0.5 + (j-1) \times 0.1$, $j = 1, \ldots, 11$.

Simulations

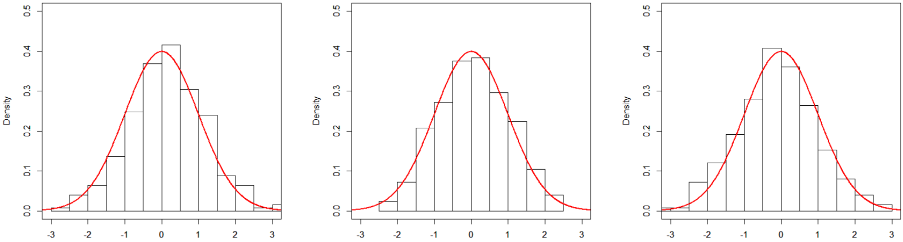


Figure: Normalized empirical distributions of estimates $\hat{\mu}_{b_n}(x)$ at $x = -0.25$(left), $x = 0$(center) and $x = 0.25$(right). The red line is the density of a standard normal distribution. We set $n = 750$ and the number of Monte Carlo iteration is 250.

Simulations

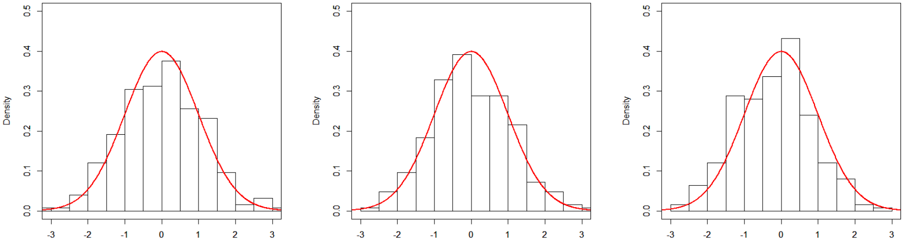


Figure: Normalized empirical distributions of estimates $\hat{\sigma}^2_{h_n}(x)$ at $x = -0.25$(left), $x = 0$(center) and $x = 0.25$(right). The red line is the density of a standard normal distribution. We set $n = 750$ and the number of Monte Carlo iteration is 250.

Simulations

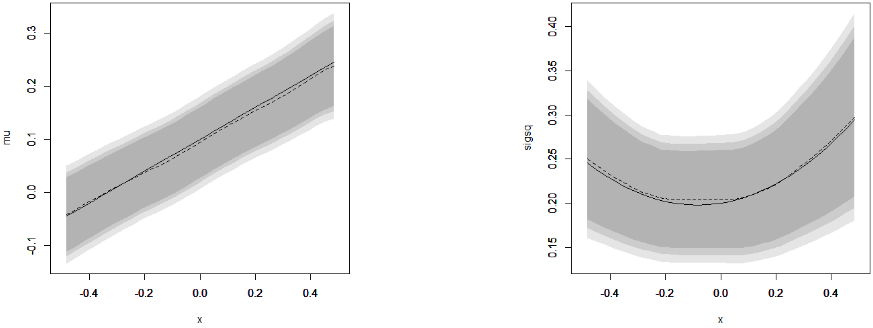


Figure: Estimates of μ (left) and σ^2 (right) together with 85%(dark gray), 95%(gray), and 99%(light gray) confidence bands. The solid lines correspond to the true functions. We set $n = 750$, $I = [-0.5, 0.5]$, and $x_j = -0.5 + (j-1) \times 0.1$, $j = 1, \ldots, 11$.

Conclusions

- DEI asymptotics の下でノンパラメトリック空間回帰モデルの平均・分散関数に対する多次元中心極限定理を導出した.
- 多次元中心極限定理を利用して f, μ, σ^2 に対する信頼バンドの構成方法を提案した.
- データ駆動型の推定量のバンド幅の選択法を提案した.
- 数値実験で finite sample performance を確認した.

Kurisu, D. (2018a). Nonparametric inference on Lévy measures of Lévy-driven Ornstein-Uhlenbeck processes under discrete observations. arXiv:1803.08671.

Kurisu, D. (2018b). On nonparametric inference for spatial regression models under domain expanding and infill asymptotics. arXiv:1804.09402.

参考文献 1

- Bernstein, S.N. (1926). Sur léstension du théorème limite du calcul des probabilités aux sommes de quantités dépendantes. *Ann. Math.* **21**, 1-59.
- Bissantz, N., Dümbgen, L., Holzmann, H. and Munk, A. (2007). Non-parametric confidence bands in deconvolution density estimation. *J. Roy. Stat. Soc. Ser. B Stat. Methodol.* **69** 483-506.
- Bolthausen, E. (1982). On the central limit theorem for stationary mixing random fields. *Ann. Probab.* **4**, 1047-1050.
- Brockwell, P.J. and Matsuda, Y. (2017), Continuous auto-regressive moving average random fields on $\mathbb{R}^n$. *J. Royal Statist. Soc. Ser. B* **79**, 833-857.
- Francisco-Fernández, M., Quintela-del-Río, A. and Fernández-Casal, R. (2012). Nonparametric methods for spatial regression. An application to seismic events. *Environmetrics* DOI:10.1002/env.1146.

参考文献 2

- Guyon, X. (1987). Estimation d'un champ par pseudo-vraisemblance conditionnelle: Étude asymptotique et application au cas Markovien. In Spatial Processes and Spatial Time Series Analysis. Proc. 6th FrancoBelgian Meeting of Statisticians (J.-J. Droesbeke et al., eds.) 1562. FUSL, Brussels.
- Jenish ,N. and Prucha, I. (2009). Central limit theorems and uniform laws of numbers for arrays of random fields. *J. Econometrics* **150**, 86-98.
- Nguyen, M. and Veraart, A.D. (2017). Spatio-temporal Ornstein-Uhlenbeck processes: theory, simulation and statistical inference. *Scand. J. Statist.* **44**, 46-80.
- Hallin, M., Lu, Z. and Yu, K. (2009). Local linear spatial quantile regression. *Bernoulli* **15**, 659-686.
- Li, L. (2016). Nonparametric regression on random fields with random design using wavelet method. *Stat. Inference Stoch. Process.* **19**, 51-69.

- Lu, Z. and Tjøstheim, D. (2014). Nonparametric estimation of probability density functions for irregularly observed spatial data. *J. Amer. Statist. Soc.* **109**, 1546-1564.
- Matsuda, Y. and Yajima, Y. (2009). Fourier analysis of irregularly spaced data on $\mathbb{R}^d$. *J. Royal Statist. Soc. Ser. B* **71**, 191-217.
- Robinson, P. (2011). Asymptotic theory for nonparametric regression with spatial data. *J. Econometrics* **165**, 5-19.
- Robinson, P. and Thawornkaiwong, S. (2012). Statistical inference on regression with spatial dependence. *J. Econometrics* **167**, 521-542.
- Rosenbratt, M. (1985). Stationary Sequences and Random Fields. Birkhäuser, Boston.
- Zhang, B.Z. and Wu, W.B. (2008). Confidence bands in nonparametric time series regression. *Ann. Statist.* **36**, 1854-1878.