SDS-2

釧路・経済統計キャンプ2017
(新しい時系列計量分析の理論と応用)

国友直人・北川 源四郎　編

November 2017

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

釧路・経済統計キャンプ2017

(新しい時系列計量分析の理論と応用) [1]

国友直人[2]

&

北川源四郎[3]

(共編)

2017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 2017 年 8 月 5 日に釧路公立大学において開催した研究集会 (明治大学先端数理科学インスティテュート (MIMS) 協賛）における講演をまとめたものである。

[2]明治大学政治経済学部

[3]明治大学先端数理科学インスティテュート (MIMS)

概要

経済統計・金融データ・政府統計データなどを主な分析対象とする統計的時系列分析ではなお幾つかの検討するべき基本的な問題が存在する。例えば「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たが、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実例である。国際的に連動している現代の経済では従来の計量分析ではほとんど考慮されていない経済変動を経験している。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である。また以前からよく知られているように、またマクロ経済データや高頻度金融時系列の分析などでは変数間の関係の分析が重要であるが、非定常多次元時系列分析にはなお課題が少なくない。

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」ではこうした社会的に重要な背景をも踏まえて、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、応用を検討する予定である。特にマクロ経済・金融現象の統計分析では、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、よりミクロ金融の分析ではジャンプを含む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、ジャンプ拡散確率過程の計量分析の方法を確立が望まれる。こうしたマクロ分野、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初年度であり、特にマクロ経済データの分析に応用可能性が高い時系列フィルタリングの理論と応用、時系列の因果性分析、高頻度金融データにおける因果性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。

2017 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：2017 年 8 月 5 日 (土)
会場：釧路公立大学第一会議室
オーガナイザー：国友直人協賛：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「時系列フィルタリングの発展と応用」北川源四郎
13:55〜14:45「Multi-scale analysis of lead-lag relationships in high-frequency financial markets」林高樹・小池祐太

< 休憩 >

< セッション II：経済統計・政府統計の統計学 >
Chair: 大屋幸輔
15:00〜15:40「日次・月次消費の状態推定問題」国友直人
15:40〜16:20「GDP 統計の見方について」佐藤整尚

< 休憩 >

< セッション III：時系列計量分析の統計学 >
Chair:生方雅人
16:30〜17:10「High-frequency Financial Data and G-Causality Analysis」大屋幸輔・木下亮
17:10〜17:50「Simultaneous multivariate point process models and G-Causality analysis of international financial markets」国友直人・栗栖大輔・粟屋直

時系列フィルタリングの発展と応用
ー 季節調整モデリングを中心に ー

明治大学 MIMS
北川源四郎

釧路・経済統計キャンプ 8/05/2017

MEIJI UNIVERSITY

目 次

センサス局季節調整研究プログラム

- X-11 variant of the Census Method II (Shiskin 1967)

- センサス局-NBER-ASA Conference (1978)

  A. Zellner, S. Kallek, J. Tukey, C. Granger, E.B. Dagum,
  J. Shiskin, P. Bloomfield, W. Cleaveland, D. Pierce, W. Wei,
  R. Engle, G. Tiao, E. Parzen, G. Chow, J. Geweke, G.E.P. Box,
  S. Hillmer

- センサス局-ASA Research Project (1980年前後)

  J. Durbin, A. Dempster, S. Hillmer
  D. Findley, B. Bell, W. Gersch, G. Kitagawa, A. Rave

Smoothness Prior Approach

平滑化の問題

$$y_n = f_n + \varepsilon_n, \quad n = 1, \ldots, N$$

y_n	観測値
f_n	未知パラメータ
ε_n	ノイズ（残差）

新NP問題：

観測値	$y_1, y_2, \ldots, y_n$	n 個
未知数	$t_1, t_2, \ldots, t_n$	n 個

正則化項付最小二乗法

Infidelity to the data

Infidelity to smoothness

$$\min_f \left[\sum_{n=1}^{N} (y_n - f_n)^2 + \lambda^2 \sum_{n=1}^{N} (\nabla^k f_n)^2 \right]$$

H. Akaike(1980) “Likelihood and the Bayes Procedure”, Bayesian Statistics, 1-13.

Whittaker (1923), Shiller (1973), Akaike(1980), Kitagawa-Gersch(1996)

MEIJI UNIVERSITY

ベイズの観点からの超パラメータ選択

$$\sum_{n=1}^{N}(y_n - f_n)^2 + \lambda^2 \sum_{n=1}^{N}(\nabla^k f_n)^2$$

$-1/(2\sigma^2)$ をかけて指数をとる

$$\exp\left\{-\frac{1}{2\sigma^2}\sum_{n=1}^{N}(y_n - f_n)^2\right\}\exp\left\{-\frac{\lambda^2}{2\sigma^2}\sum_{n=1}^{N}(\nabla^k f_n)^2\right\}$$

ベイズモデルによる解釈 $\theta = (\lambda^2, \sigma^2)$

$$\pi(f \mid y, \theta) \propto p(y \mid f, \theta)\pi(f \mid \theta)$$

事後分布

ABICによるθの決定

MEIJI UNIVERSITY

Akaike（1980）

時系列的解釈と状態空間モデル

$$\sum_{n=1}^{N}(y_n - t_n)^2 + \lambda^2 \sum_{n=2}^{N}(t_n - t_{n-1})^2$$

等価なモデル

$$t_n = t_{n-1} + v_n \qquad v_n \sim N(0, \tau^2)$$

$$y_n = t_n + w_n \qquad w_n \sim N(0, \sigma^2)$$

状態空間モデル

$$x_n = F x_{n-1} + G v_n$$

$$y_n = H x_n + w_n$$

$$\lambda^2 = \frac{\sigma^2}{\tau^2}$$

最尤法で自動的に決まる

Kitagawa（1981）, Kitagawa-Gersch (1984)

欠測値処理から状態空間モデルへ

時系列解析における状態空間モデルの利用

- １９７０年代までは時系列モデルは個別に研究されていた
- 状態空間モデル導入のきっかけ
  - ・統計的制御 (赤池, 1970)
  - ・ARMAモデルの最尤推定 (赤池, 1973)
  - ・欠測値の処理（同時分布の積分の自動化, 1984)
  - ・ベイズ型季節調整法(赤池)の状態空間表現 (1981)

$$\lambda^2 = \frac{\sigma^2}{\tau^2}$$

$$y_n = f(n) + \varepsilon_n, \quad n = 1, \ldots, N$$

$$\sum_{n=1}^{N} (y_n - t_n)^2 + \lambda^2 \sum_{n=2}^{N} (t_n - t_{n-1})^2$$

$$t_n = t_{n-1} + v_n \qquad v_n \sim N(0, \tau^2)$$

$$y_n = t_n + w_n \qquad w_n \sim N(0, \sigma^2)$$

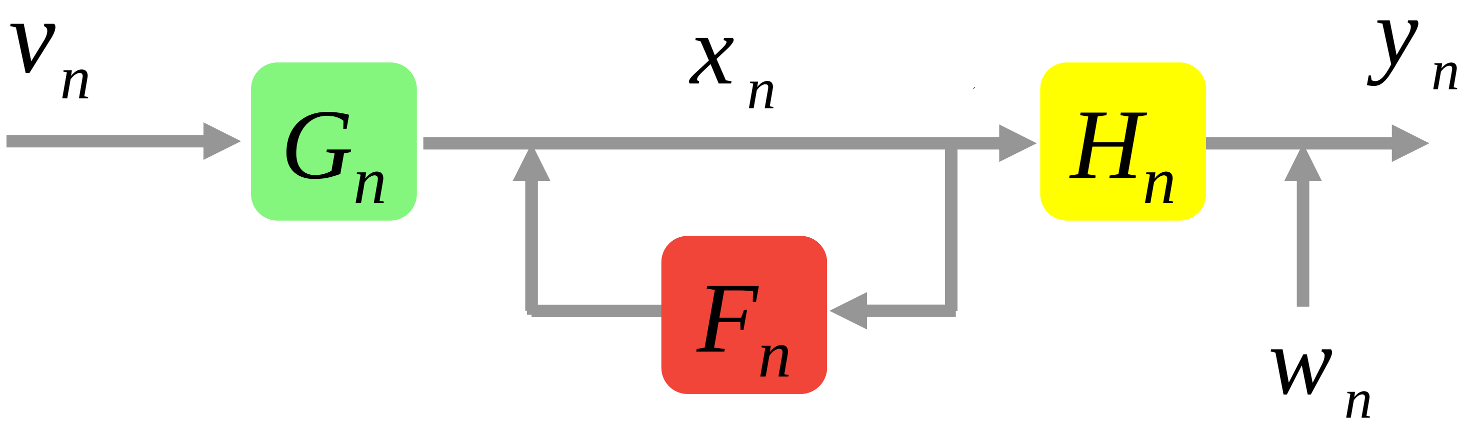
状態空間モデル

$$x_n = F_n x_{n-1} + G_n v_n \quad \text{状態モデル}$$

$$y_n = H_n x_n + w_n \quad \text{観測モデル}$$

y_n　時系列　　v_n　システムノイズ

x_n　状態ベクトル　　w_n　観測ノイズ



状態推定の問題

- 予測，長期予測
- 尤度計算，パラメータ推定
- 欠測値の補間・処理
- 異常値検出・処理
- 信号抽出

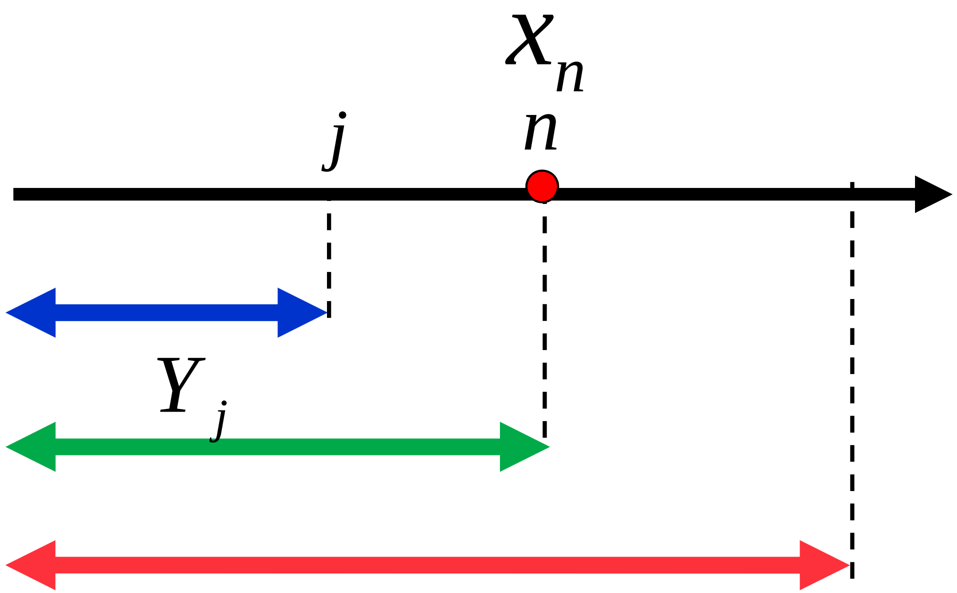
観測値（情報）$Y_j \equiv \{y_1, \dots, y_j\}$

情報 Y_j が与えられた下での状態 x_n の条件つき分布

for $j < n$ 予測

for $j = n$ フィルタ

for $j > n$ 平滑化



カルマンフィルタ

予測

$$x_{n|n-1} = F_n x_{n-1|n-1}$$

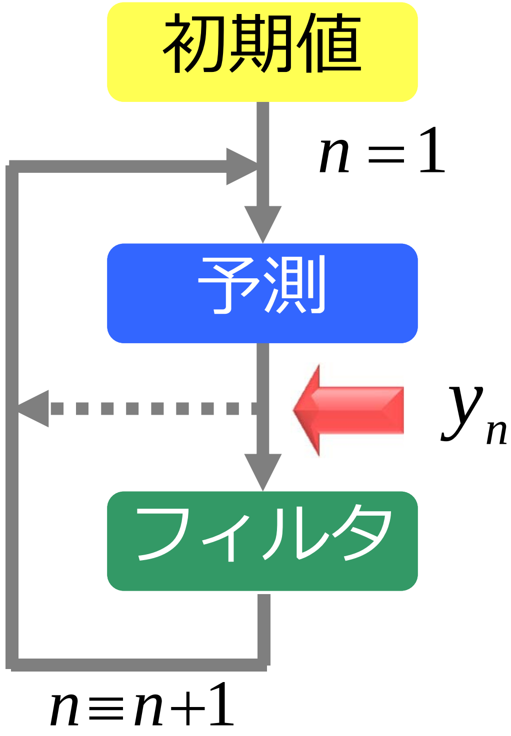
$$V_{n|n-1} = F_n V_{n-1|n-1} F_n^T + G_n Q_n G_n^T$$

フィルタ

$$K_n = V_{n|n-1} H_n^T (H_n V_{n|n-1} H_n^T + R_n)^{-1}$$

$$x_{n|n} = x_{n|n-1} + K_n (y_n - H_n x_{n|n-1})$$

$$V_{n|n} = (I - K_n H_n) V_{n|n-1}$$

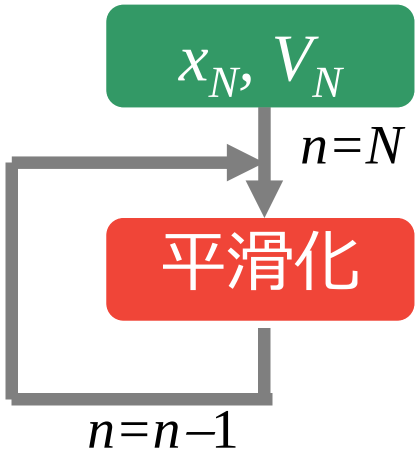


平滑化

$$A_n = V_{n|n} F_n^T V_{n+1|n}^{-1}$$

$$x_{n|N} = x_{n|n} + A_n (x_{n+1|N} - x_{n+1|n})$$

$$V_{n|N} = V_{n|n} + A_n (V_{n+1|N} - V_{n+1|n}) A_n^T$$



状態空間モデルの特長

- 透明性：仮定(主観)のモデルによる表現
- 客観性：尤度、AICによるモデル評価
  (パラメータ推定、モデル選択)
- 厳密性：端点処理、欠測処理
- 拡張性：サイクル、曜日効果、異常値、
  レベルシフト、外生変数
  他の問題への拡張可能性
- 一般化：非正規、非線形モデルへの一般化

季節調整のためのモデル

$$y_n = t_n + s_n + w_n$$

y_n: 観測値
t_n: トレンド成分
s_n: 季節成分
w_n: 不規則成分

トレンド成分のモデル

$$\Delta^k t_n = v_n$$

$$t_n = t_{n-1} + v_n$$ ほぼ一定値

$$t_n = 2t_{n-1} - t_{n-2} + v_n$$ 傾きがほぼ一定

季節成分のモデル化

季節成分の特性 $s_n \cong s_{n-p}$ （p:季節の長さ）

$$s_n = s_{n-p} + v_{2n}$$

$$\Delta_p^\ell s_n = v_{2n} \qquad \Delta_p = 1 - B^p$$

実際には、このモデルではトレンドと季節成分の分離はできない。

$$\Delta^\ell = (1 - B^p)^\ell$$
$$= (1 - B)^\ell (1 + B + \cdots + B^{p-1})^\ell$$

$$(1 + B + \cdots + B^{p-1})^\ell s_n = v_{2n}$$

季節調整のための状態空間モデル

$$y_n = t_n + s_n + w_n$$
$$\Delta^k t_n = v_n$$
$$s_n = -(s_{n-1} + \cdots + s_{n-p+1}) + u_n$$

y_n: 観測値
t_n: トレンド成分
s_n: 季節成分
w_n: 不規則成分

状態空間表現

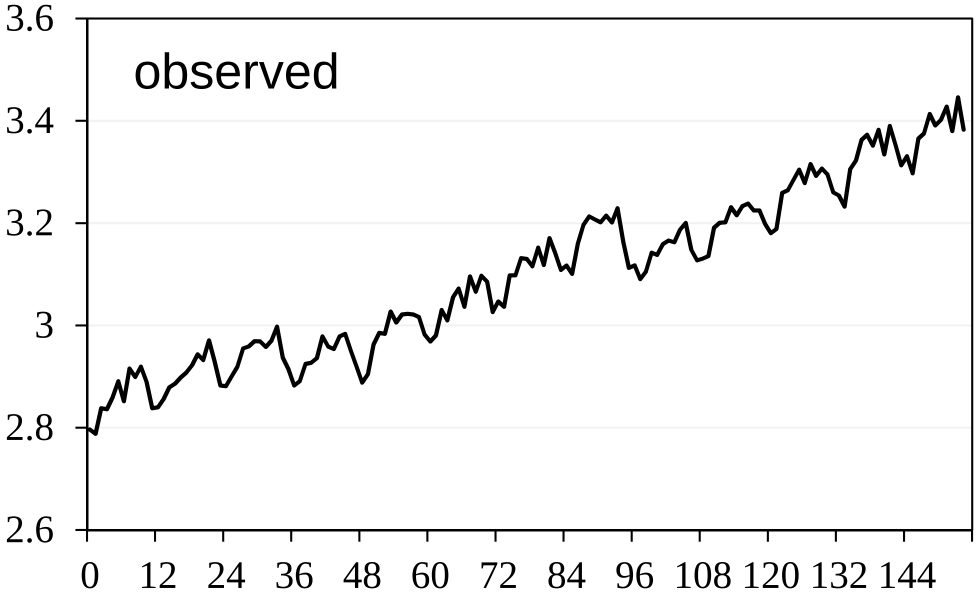
$$x_n = F x_{n-1} + G v_n$$
$$y_n = H x_n + w_n$$

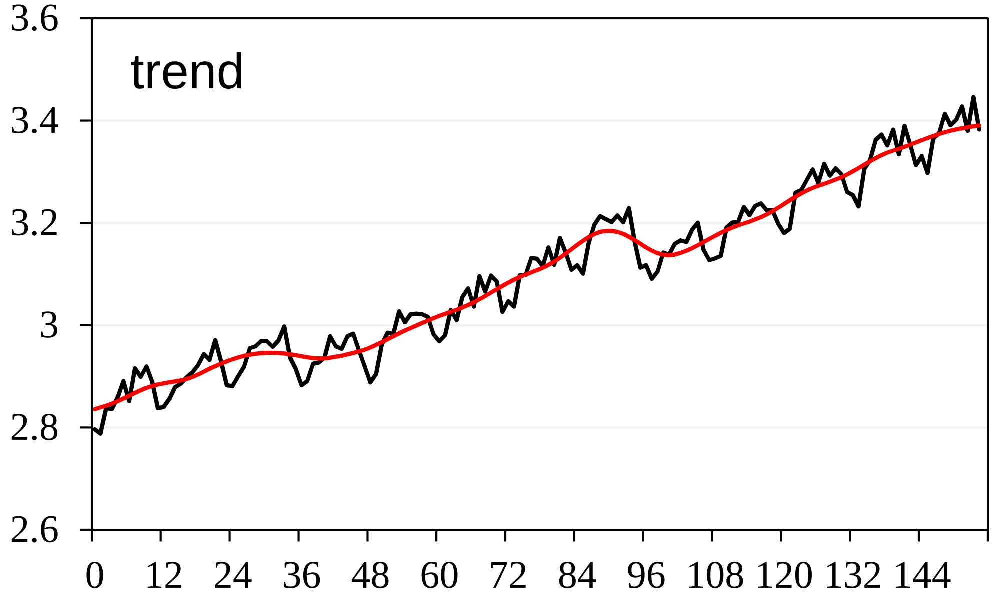
k=2, p=4 の場合　$x_n = [t_n, t_{n-1} | s_n, s_{n-1}, s_{n-2}]^T$

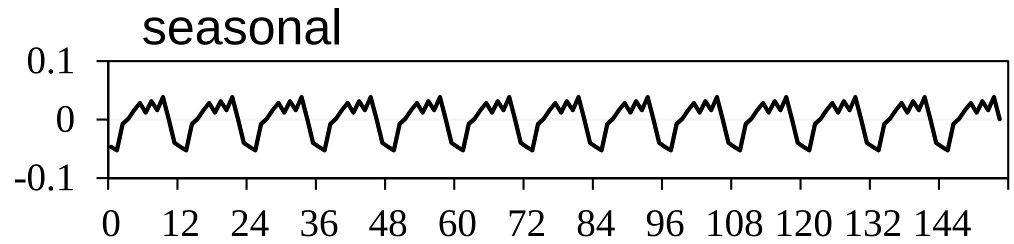
$$F = \begin{bmatrix} 2 & -1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & -1 & -1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}, \quad G = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

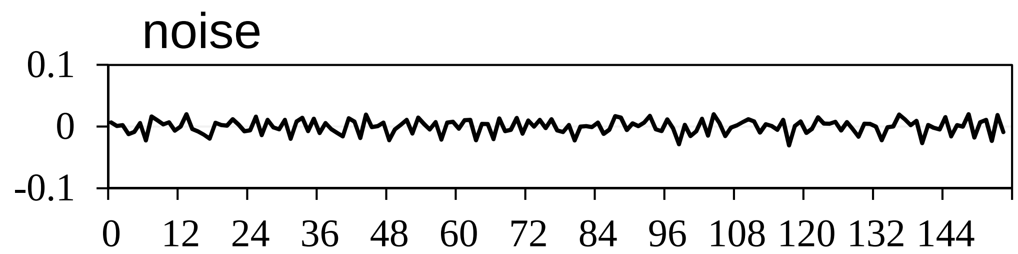
$$H = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 \end{bmatrix}$$

モデルによる季節調整







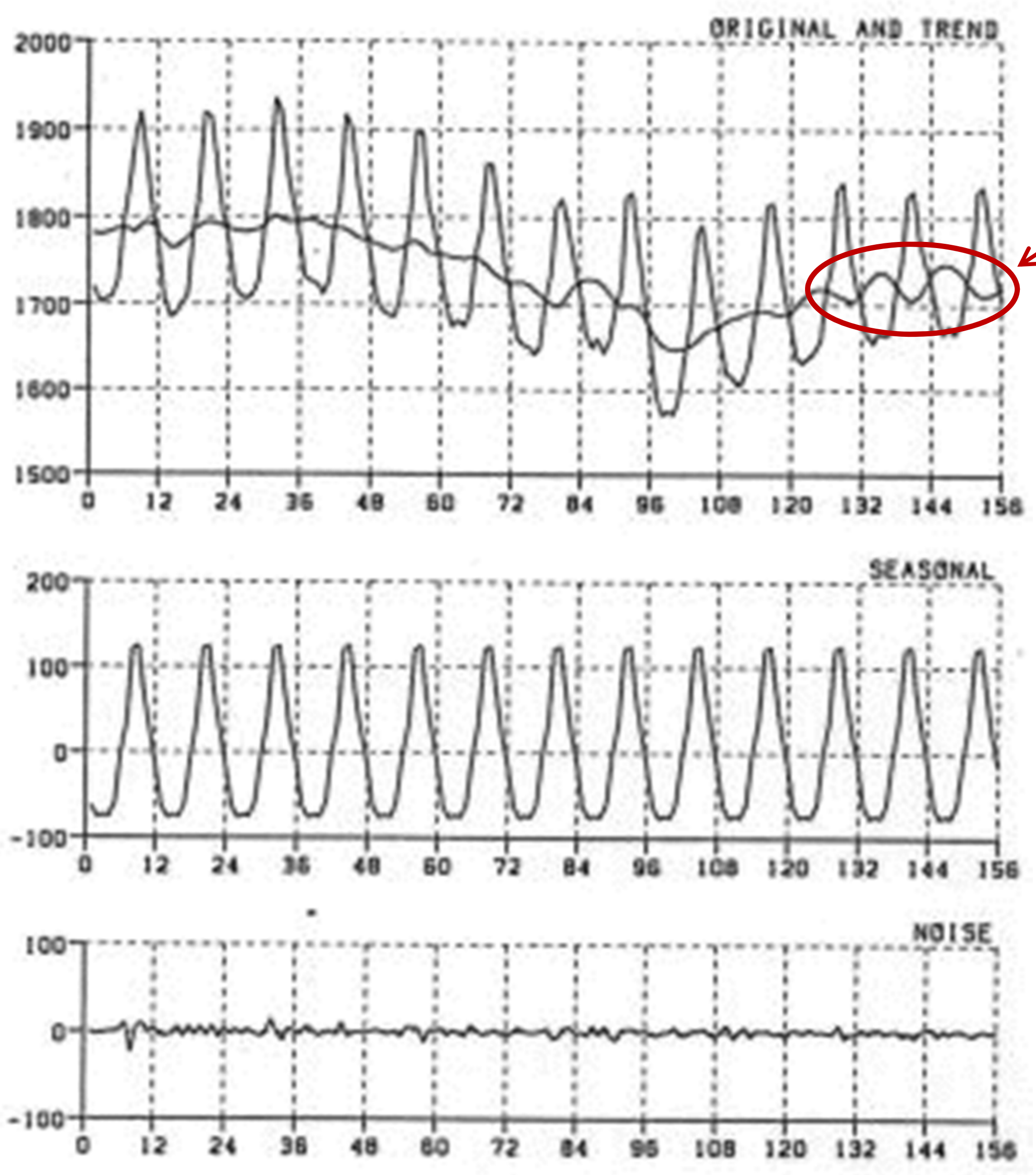


$$\sigma^2 = 0.150 \times 10^{-3}$$

$$\tau_1^2 = 0.0259, \tau_2^2 = 0.110 \times 10^{-7}$$

$$\ell = 342.532, \quad \mathrm{AIC} = -655.065$$

標準的季節調整の問題点



MEIJI UNIVERSITY

サイクル(定常成分)の抽出

$$y_n = t_n + s_n + \boxed{p_n} + w_n$$

定常成分モデル

$$p_n = a_1 p_{n-1} + \cdots + a_m p_{n-m} + r_n$$

MEIJI UNIVERSITY

定常成分つき季節調整モデルの状態空間表現

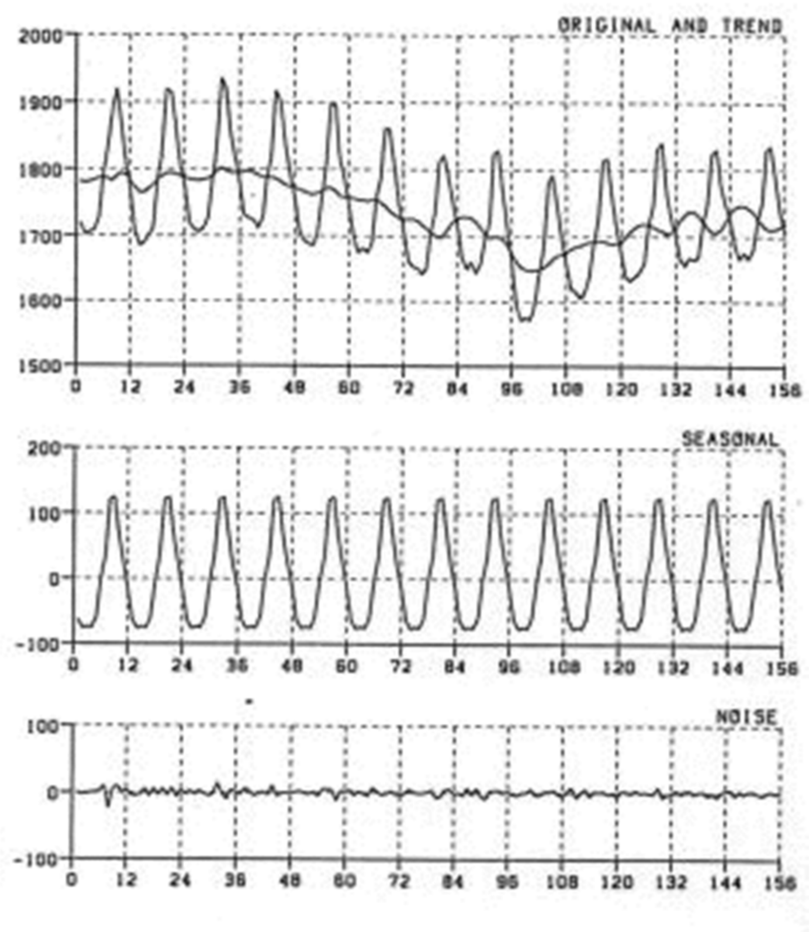
$$
\begin{aligned}
x_n &= F x_{n-1} + G v_n \\
y_n &= H x_n + w_n
\end{aligned}
$$

$$x_n = [t_n, t_{n-1} \mid s_n, s_{n-1}, \ldots, s_{n-10} \mid p_n, p_{n-1}]$$

$$
F = \begin{bmatrix}
2 & -1 & 0 & 0 & \cdots & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\
0 & 0 & -1 & -1 & \cdots & -1 & 0 & 0 \\
0 & 0 & 1 & 0 & \cdots & 0 & 0 & 0 \\
0 & 0 & 0 & \ddots & \ddots & \vdots & \vdots & \vdots \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & \cdots & 0 & a_1 & a_2 \\
0 & 0 & 0 & 0 & \cdots & 0 & 1 & 0
\end{bmatrix}, \quad
G = \begin{bmatrix}
1 & 0 & 0 \\
0 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 0 \\
\vdots & \vdots & \vdots \\
0 & 0 & 0 \\
0 & 0 & 1 \\
0 & 0 & 0
\end{bmatrix}
$$

$$H = \begin{bmatrix} 1 & 0 & 1 & 0 & \cdots & 0 & 1 & 0 \end{bmatrix}$$

標準的方法　　定常成分を加えた方法

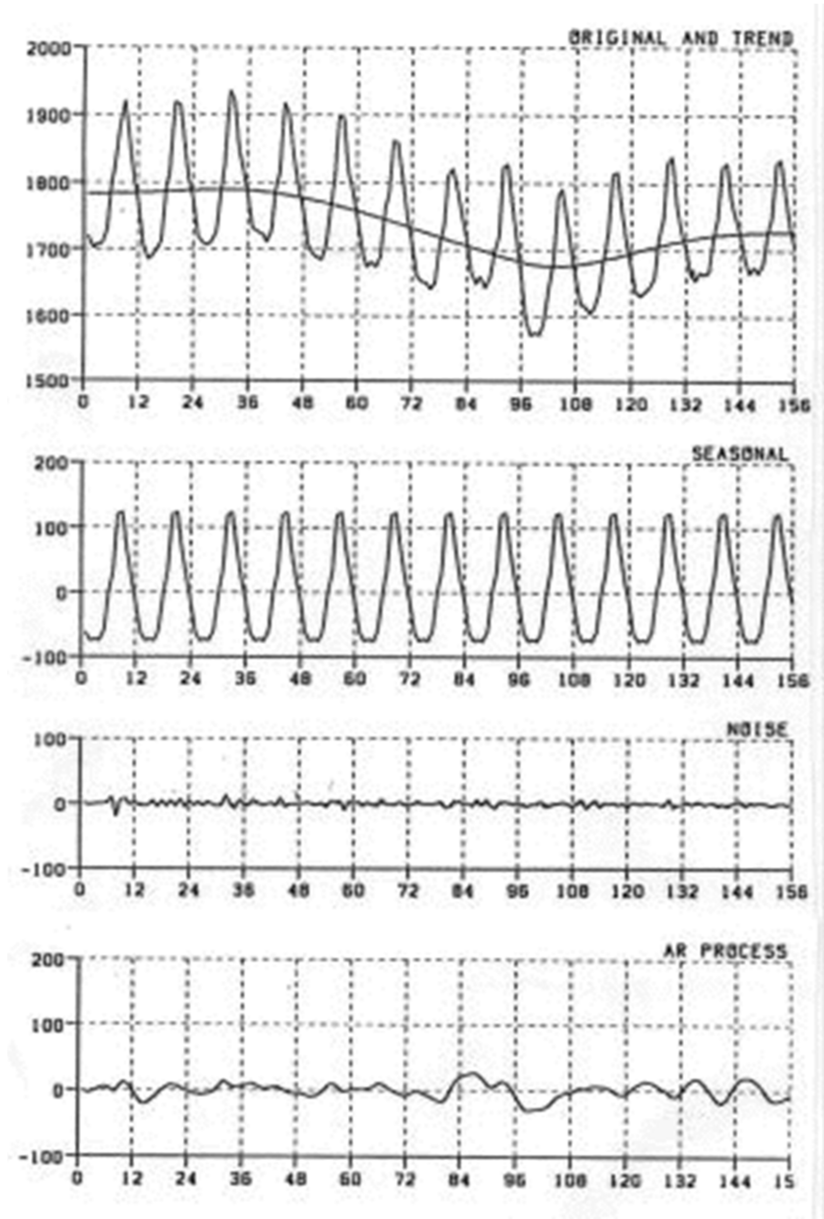


Trend

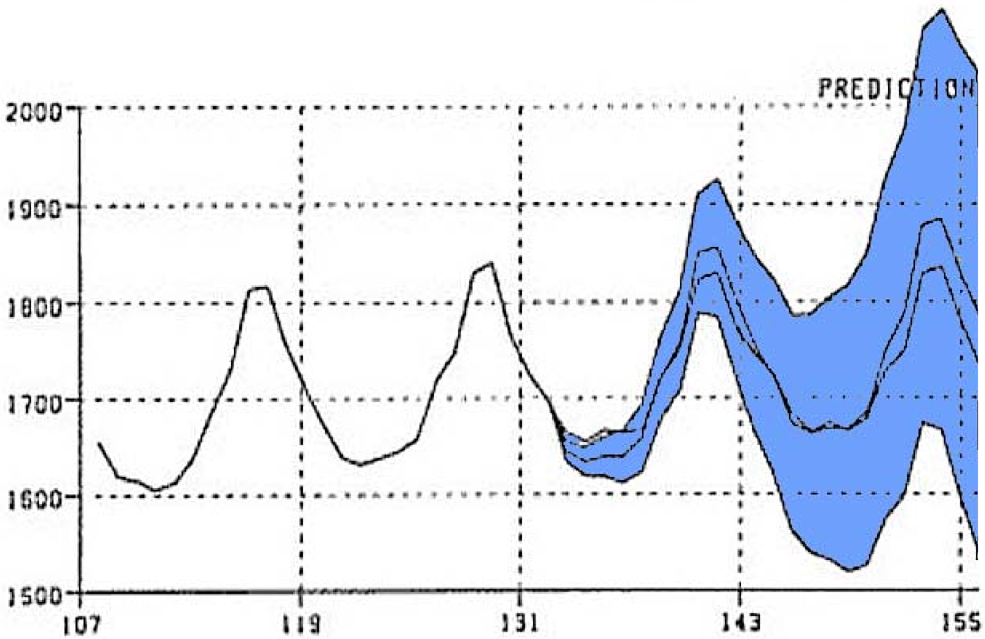
Seasonal

Noise

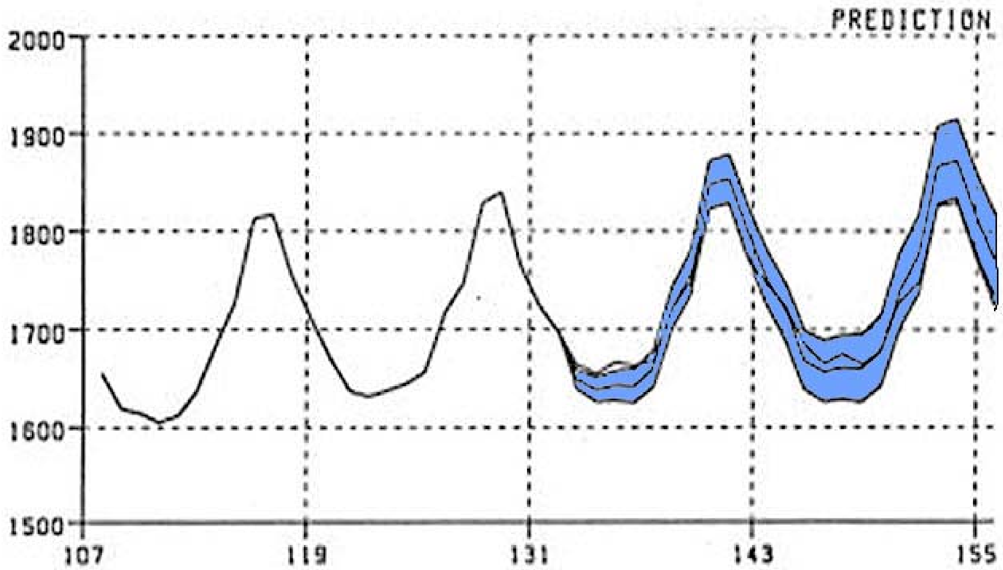
Cycle



長期予測の可能性



標準的モデル



Cycle成分を入れたモデル

短期予測精度と長期予測精度の両立

拡張可能性： 曜日効果

$$td_n = \beta_1 d_{n,1} + \cdots + \beta_7 d_{n,7}$$

$d_{n,j}$ 第 n 月の j 番目の曜日の数

β_j 曜日効果係数

$$\beta_1 + \cdots + \beta_7 = 0$$ 一意性の条件

$$\begin{aligned} td_n &= \beta_1 d_{n,1} + \cdots + \beta_6 d_{n,6} - (\beta_1 + \cdots + \beta_6) d_{n,7} \\ &= \beta_1 (d_{n,1} - d_{n,7}) + \cdots + \beta_6 (d_{n,6} - d_{n,7}) \end{aligned}$$

曜日調整のための季節調整モデル

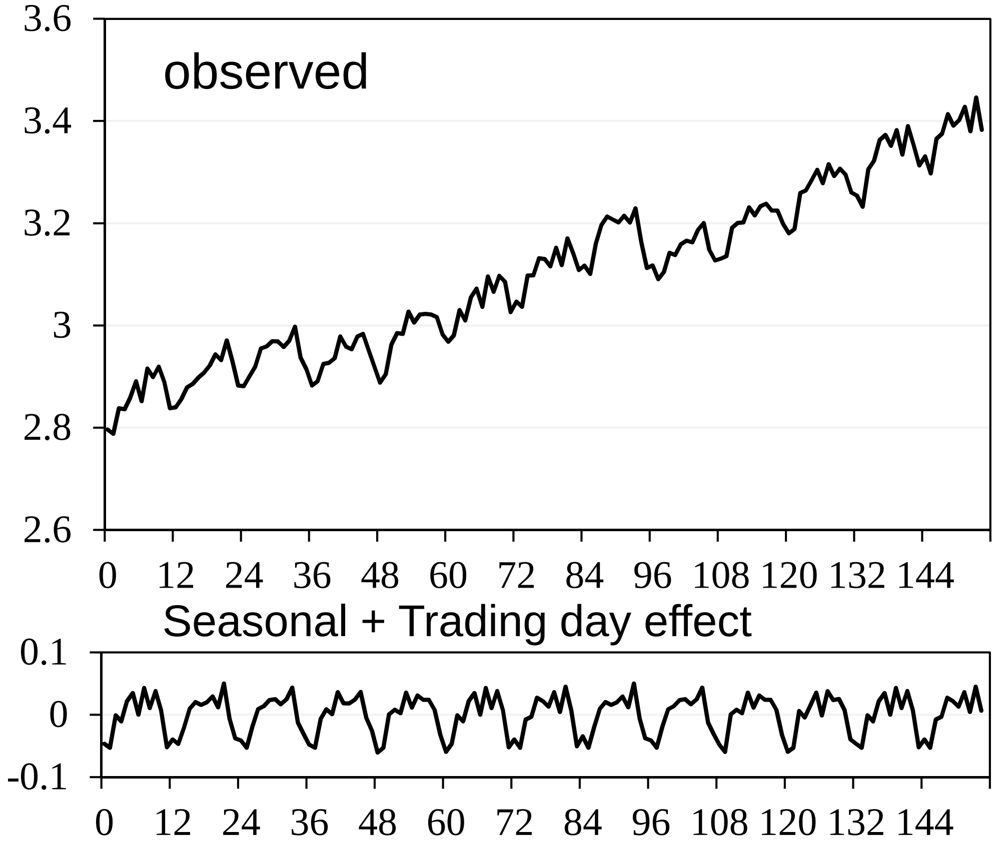
$$y_n = t_n + s_n + td_n + w_n$$

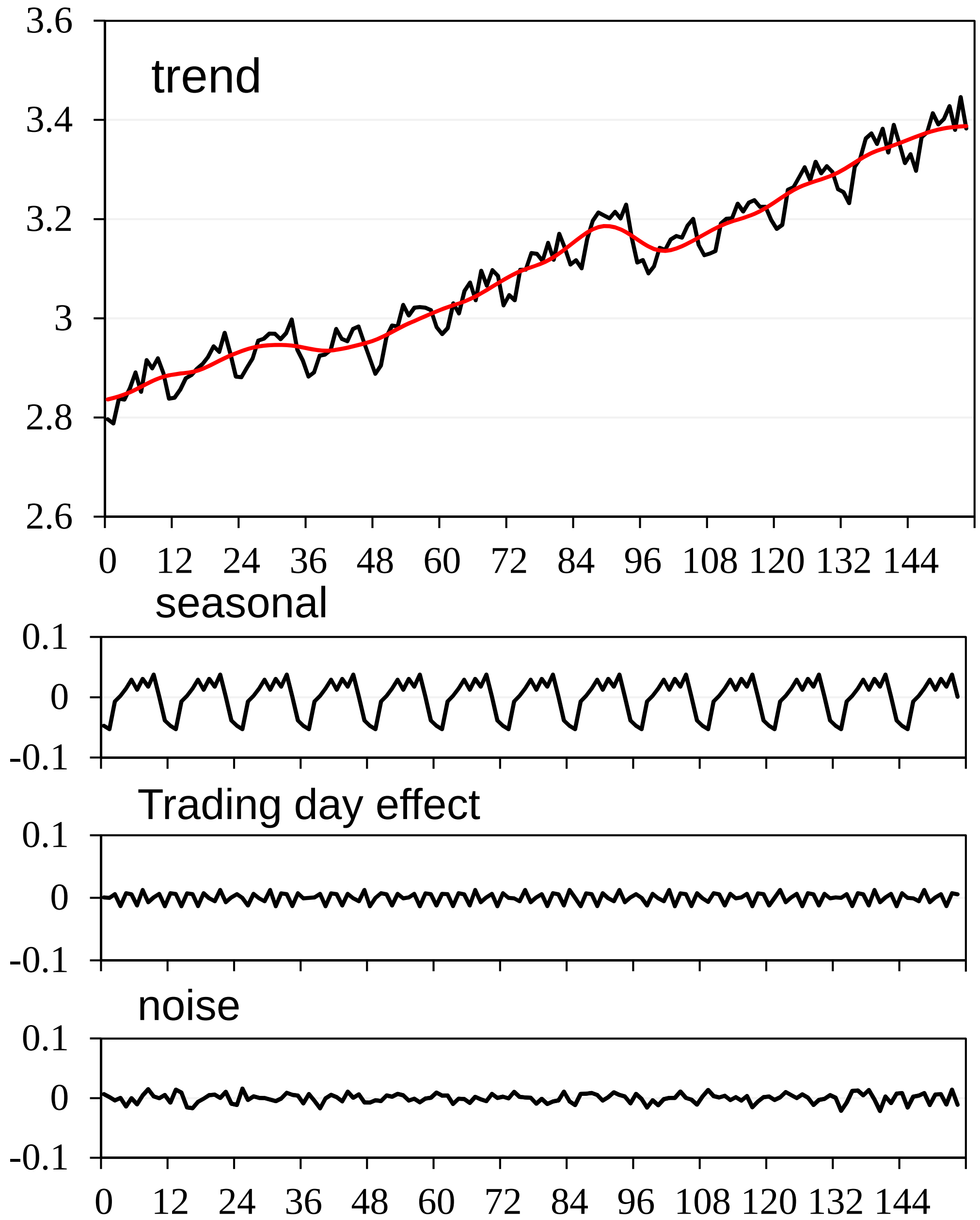
$$\Delta^k t_n = v_{1n}$$

$$s_n = -(s_{n-1} + \cdots + s_{n-p+1}) + v_{2n}$$

$$td_n = \beta_1 d_1^* + \cdots + \beta_6 d_6^*$$

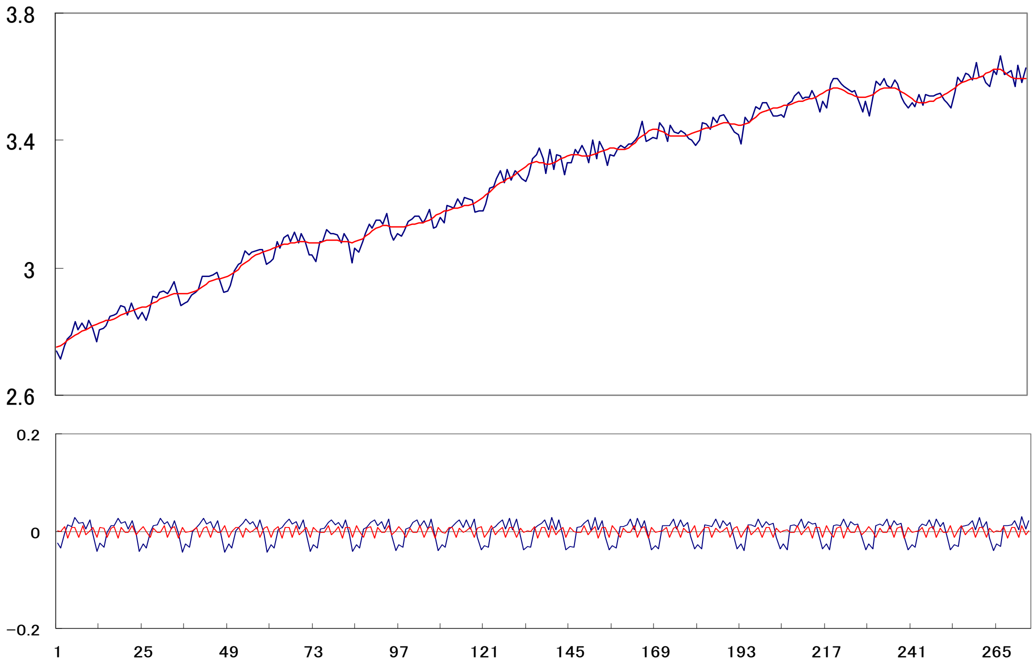
曜日調整

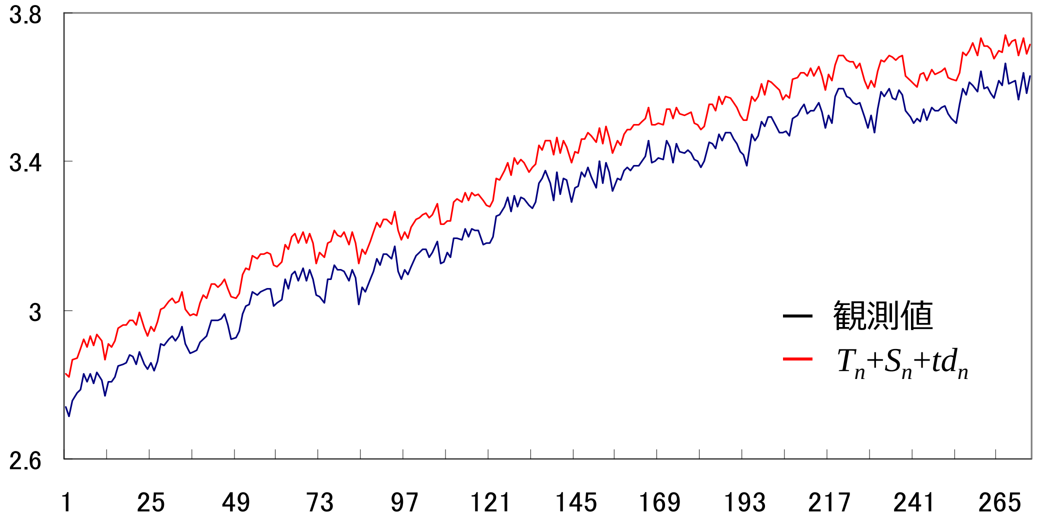


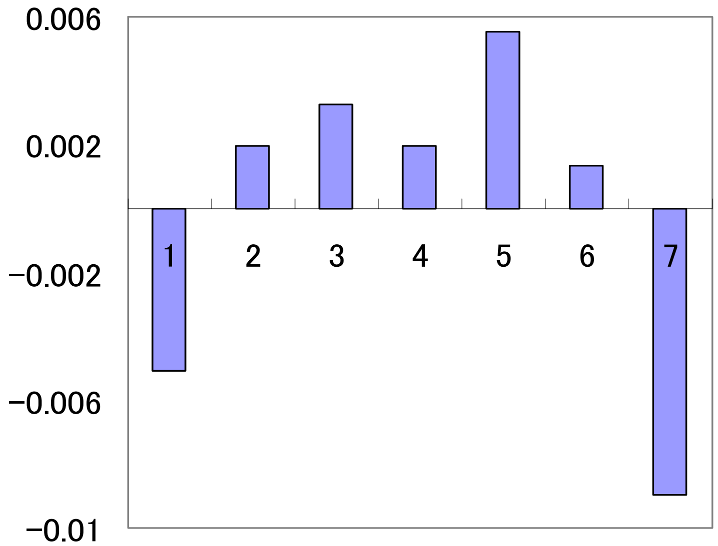


	Standard	Trading Day
σ^2	0.150×10^{-3}	0.761×10^{-4}
AIC	-655.065	-665.503

曜日効果







曜日効果の制約モデル

・$td_{\text{月}}=\cdot\cdot\cdot=td_{\text{金}}$

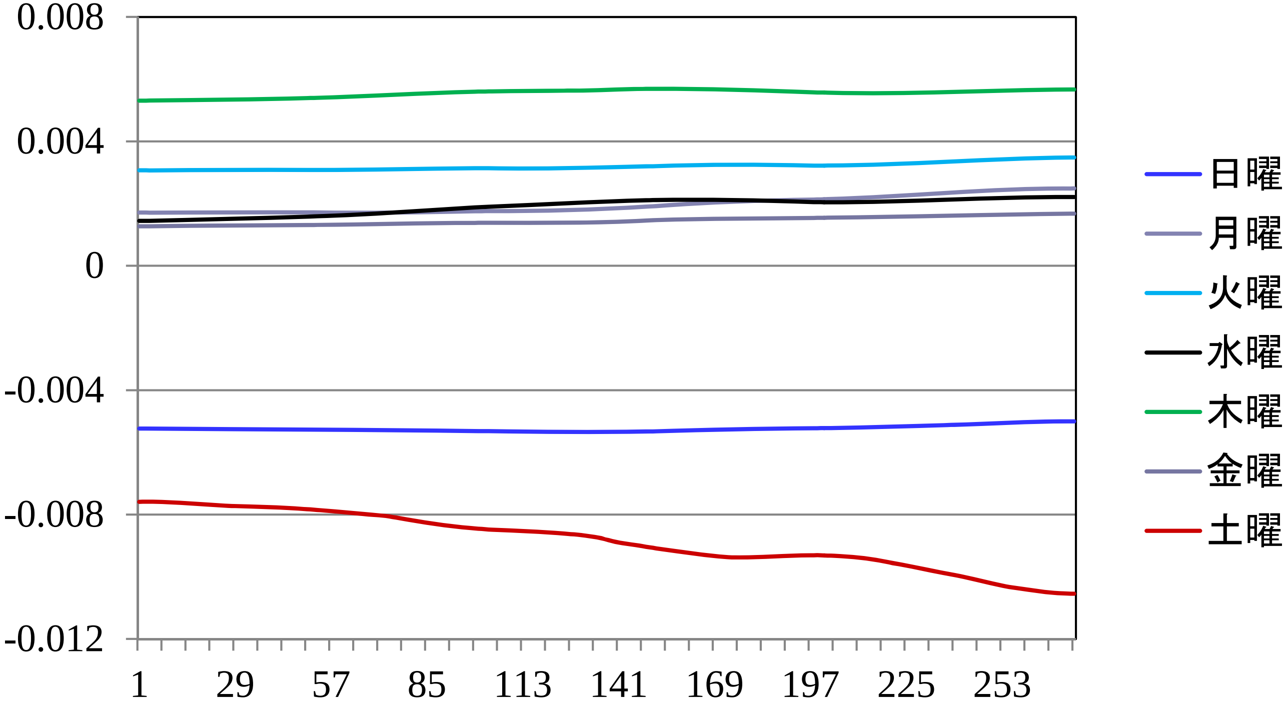
・$td_{\text{月}}=\cdot\cdot\cdot=td_{\text{金}}$, $td_{\text{土}}=td_{\text{日}}$

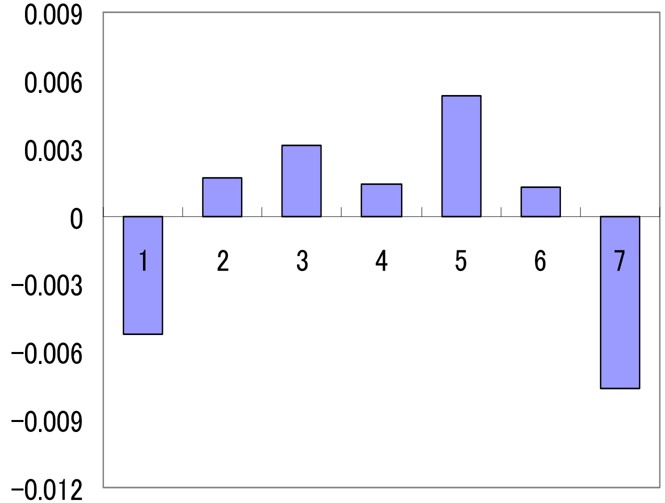
など

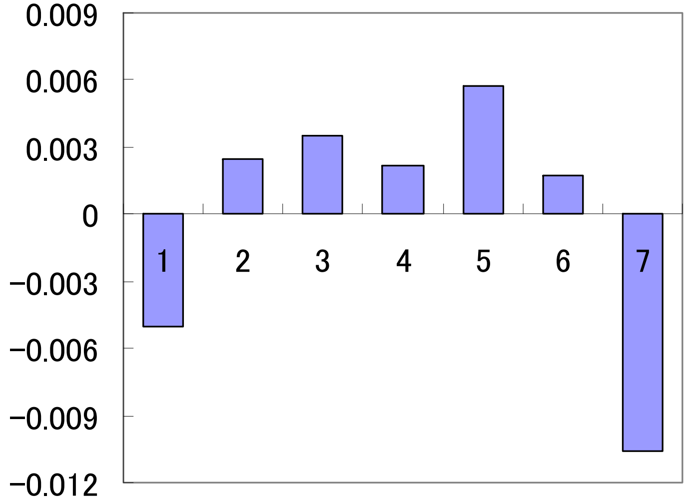
曜日効果の時間変化

曜日効果係数のモデル

$$\beta_{jn} = \beta_{j,n-1} + \varepsilon_{jn}$$







季節調整モデル（DECOMPの一般型）

$$y_n = t_n + s_n + p_n + td_n + r_n + w_n$$

t_n トレンド　p_n 定常変動　r_n 外生変数

s_n 季節成分　td_n 曜日効果　w_n ノイズ

成分モデル

$$\Delta^k t_n = v_{1n}$$

$$s_n = -(s_{n-1} + \cdots + s_{n-p+1}) + v_{2n}$$

$$p_n = a_1 p_{n-1} + \cdots + a_m p_{n-m} + v_{3n}$$

$$td_n = \beta_1 d_{n1} + \cdots + \beta_7 n$$

$$r_n = \gamma_1 r_{n1} + \cdots + \gamma_k r_{nk}$$

状態空間表現

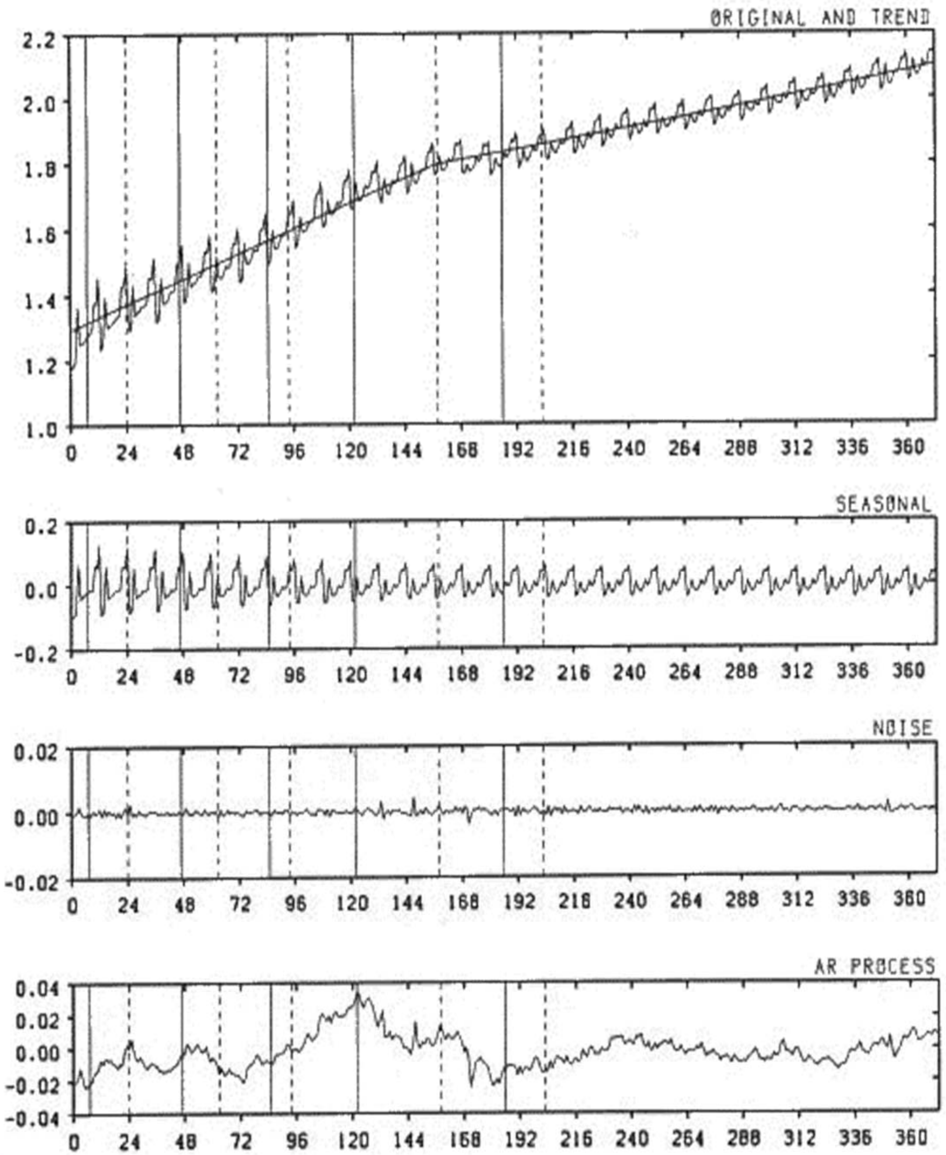
$$x_n = F x_{n-1} + G v_n$$

$$y_n = H x_n + w_n$$

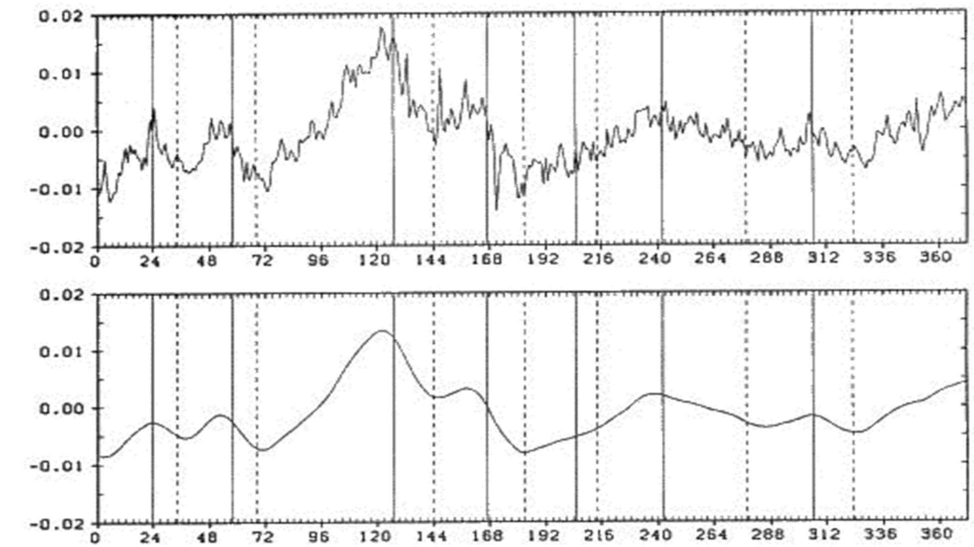
景気動向の推定

経企庁（福田）

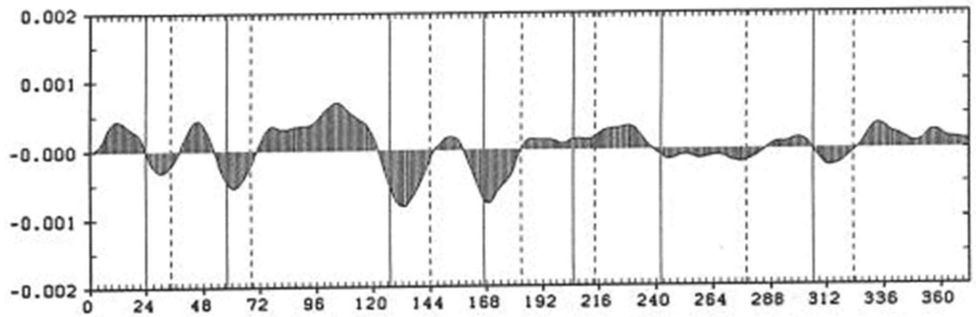
GDP系列の季節調整



GDP系列のCycle



Cycleの階差



Markov Switchingモデルへ

状態空間モデルによる非定常時系列解析

1. 平均非定常時系列
   - トレンド推定
   - 季節調整
2. 分散非定常時系列
   - 時変分散・ボラティリティ
3. 共分散非定常時系列
   - 時変スペクトル、時変係数ARモデル
4. 信号混合時系列
   - 地震波抽出、反射波抽出

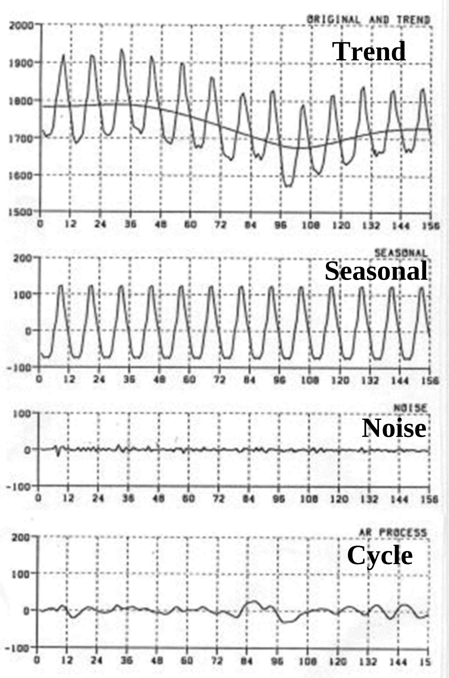
状態空間モデリングの水平展開)

観測モデル ＋ 成分モデル → 状態空間モデル → 非定常時系列解析

ベイズモデリングによりデータ数より多いパラメータの推定が可能に

季節調整モデル

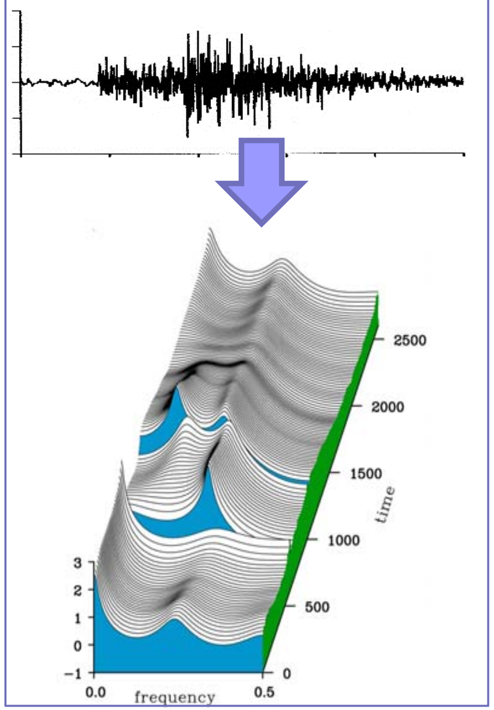
$$y_n = t_n + s_n + r_n + \varepsilon_n$$
$$\Delta^2 t_n = u_n$$
$$s_n = s_{n-p} + v_n$$
$$r_n = \sum_{j=1}^{\ell} a_j r_{n-j} + w_n$$



US Bureau of Labor Statistics

時変スペクトルモデル

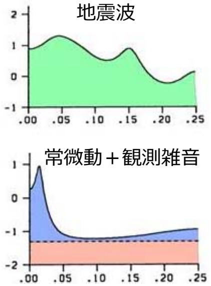
$$y_n = \sum_{i=1}^{m} a_{n,j} y_{n-j} + w_n$$
$$a_{n,j} = a_{n-1,j} + v_{n,j}$$
$$(j = 1, \ldots, m)$$

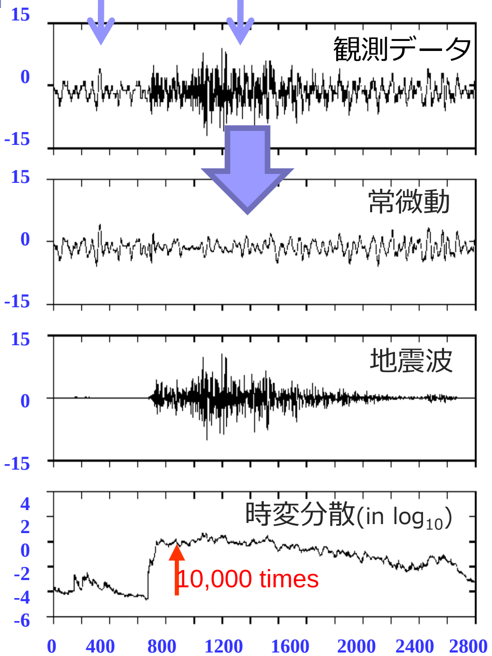


北大　高波鉄夫

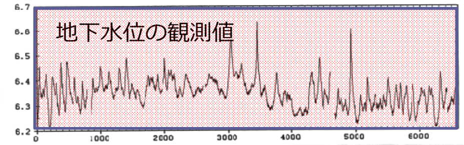
信号抽出モデル

$$y_n = r_n + s_n + w_n$$
$$r_n = \sum_{j=1}^{m} a_j r_{n-j} + u_n$$
$$s_n = \sum_{j=1}^{\ell} b_j s_{n-j} + v_n$$

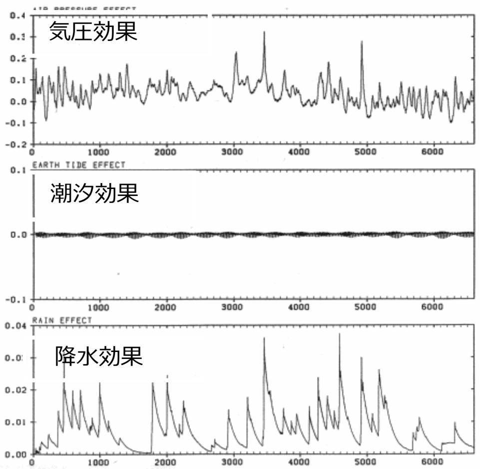


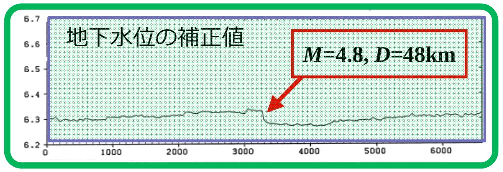


地下水位モデル



min AIC model
m=25, l=2, k=5





産総研　松本則夫

Trend + Stochastic Volatility model

日銀金融研

Model

Variance changes

$$y_n = t_n + w_n$$
$$y_n \sim N(t_n, \sigma_n^2)$$

Smoothness Priors

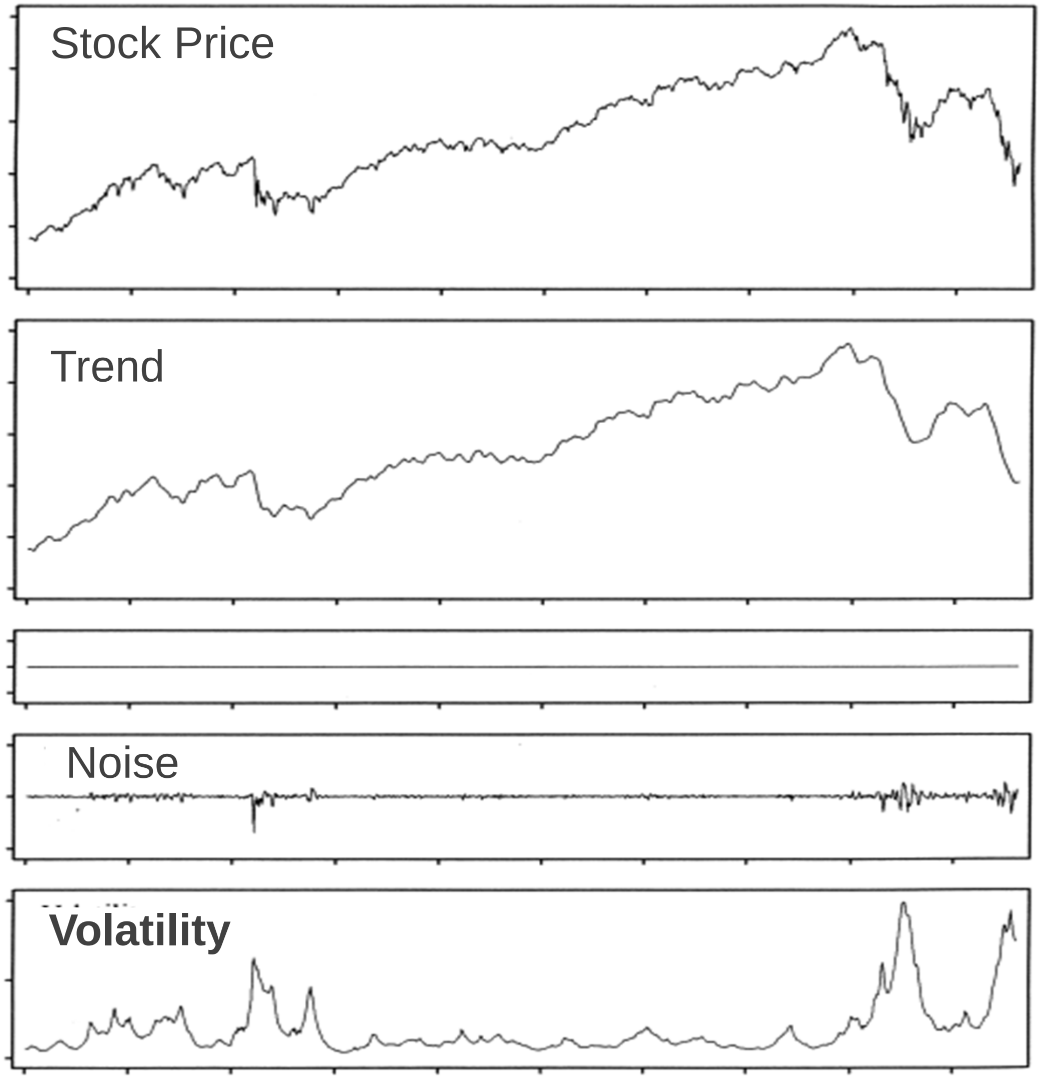
$$\Delta^k t_n = v_n$$
$$\Delta^\ell s_n = u_n$$

State space model(*k*=2,*l*=1)

$$x_n = \begin{bmatrix} t_n \\ t_{n-1} \\ s_n \end{bmatrix} = \begin{bmatrix} 2 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} t_{n-1} \\ t_{n-2} \\ s_{n-1} \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} v_n \\ u_n \end{bmatrix}$$

$$y_n = t_n + e^{s_n} w_n$$

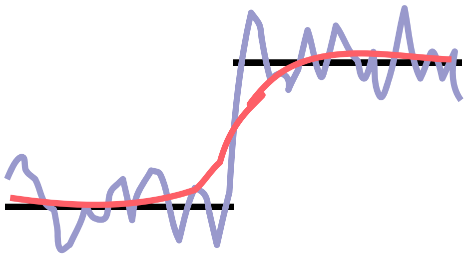
Nikkei 225 1/1/87-8/31/90



非線形・非ガウス型時系列モデリング

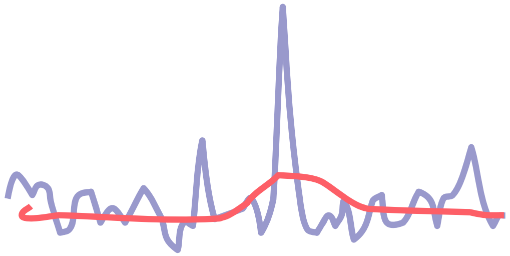
非ガウス型状態空間モデルへ（1984）

・構造変化

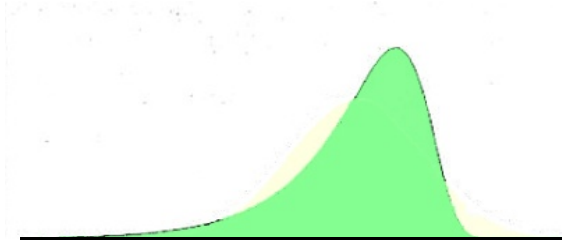


ゆっくりした変化＋急激な変化

異常値

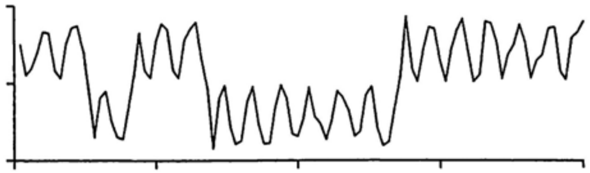


非ガウス型分布



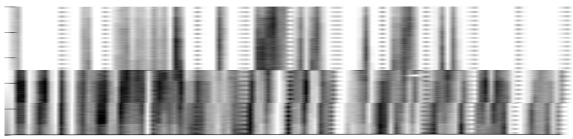
・ 非線形過程

$$x_n = f(x_{n-1}) + v_n$$

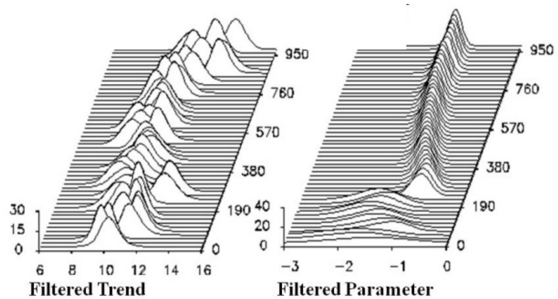


離散過程

ポアソン過程
二項過程



オンライン・
パラメータ推定

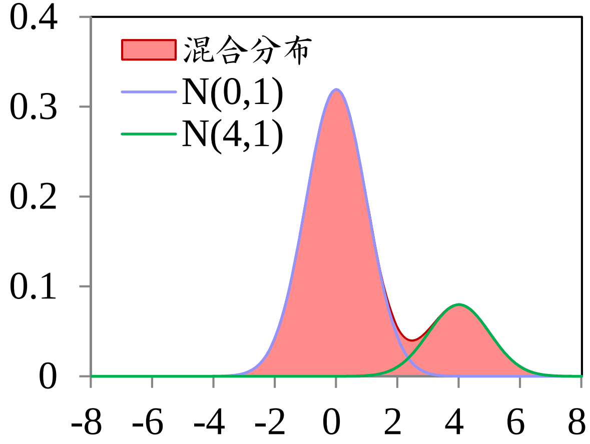


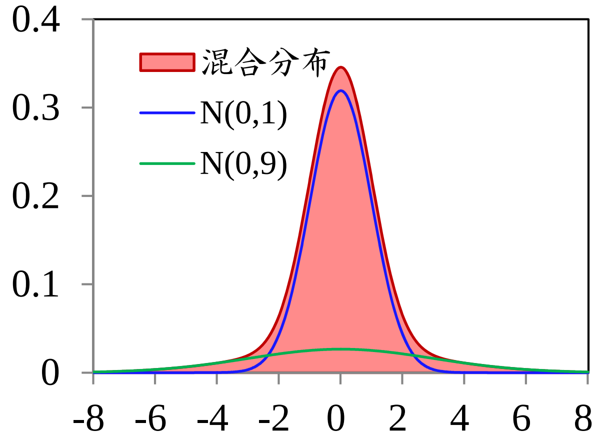
異常値検出から非ガウス・ベイズモデリングへ

- **異常値検出**

混合モデル → 非ガウス分布

組み合わせ数 → ベイズモデル





- **ロバスト推定と非ガウスモデル**

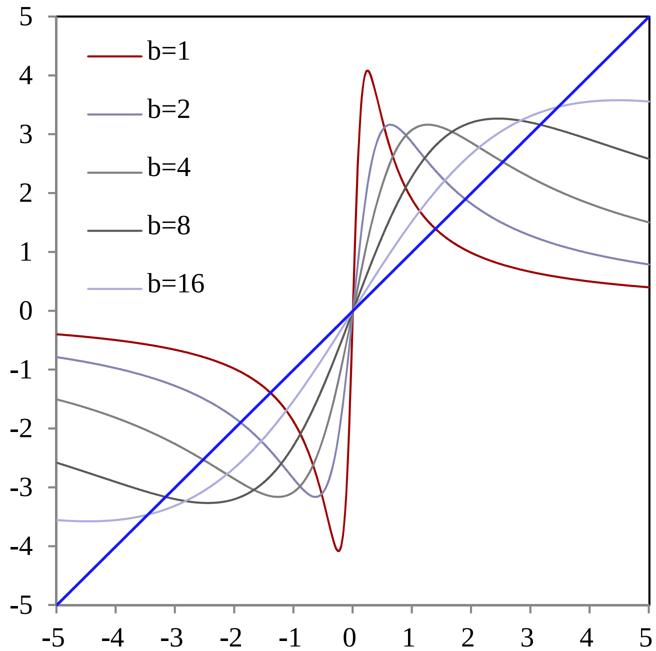
$$\sum_{i=1}^{n} \psi(x_i - \mu) = 0$$

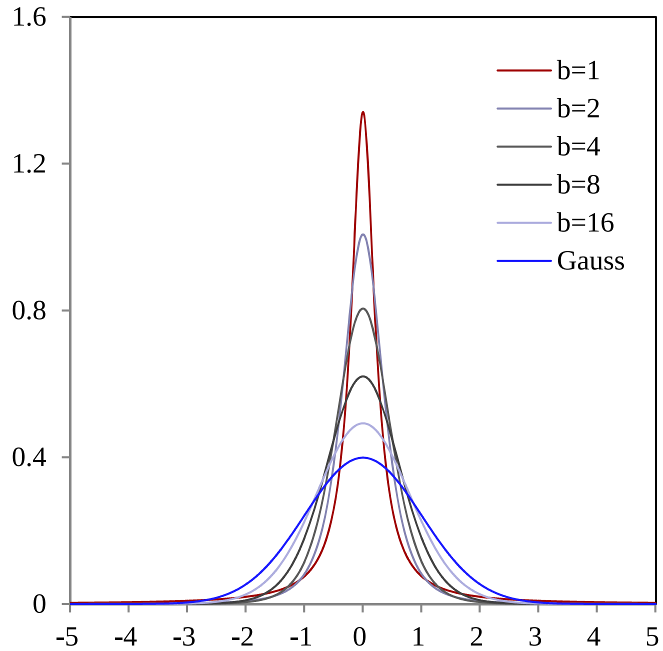
$$\psi(x) = -f'(x)/f(x)$$

Pearson分布族（Type IV）

$$f(x) = C\left\{(x-\mu)^2 + \tau^2\right\}^{-b}$$

$$C = \tau^{2b-1}\Gamma(b)/\Gamma(\tfrac{1}{2})\Gamma(b-\tfrac{1}{2})$$



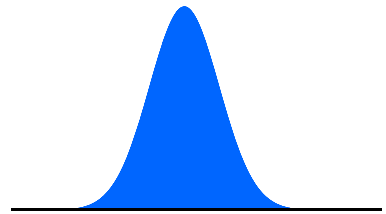


状態空間モデルの拡張

線形・ガウス型

$$x_n = Fx_{n-1} + Gv_n$$

$$y_n = Hx_n + w_n$$



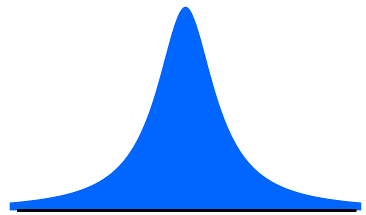
関数：非線形

分布：非ガウス型

非線形・非ガウス型

$$x_n = f(x_{n-1}, v_n)$$

$$y_n = h(x_n, w_n)$$



非ガウス型フィルタ・平滑化

一期先予測

$$p(x_n|Y_{n-1}) = \int_{-\infty}^{\infty} p(x_n|x_{n-1})p(x_{n-1}|Y_{n-1})dx_{n-1}$$

フィルタ

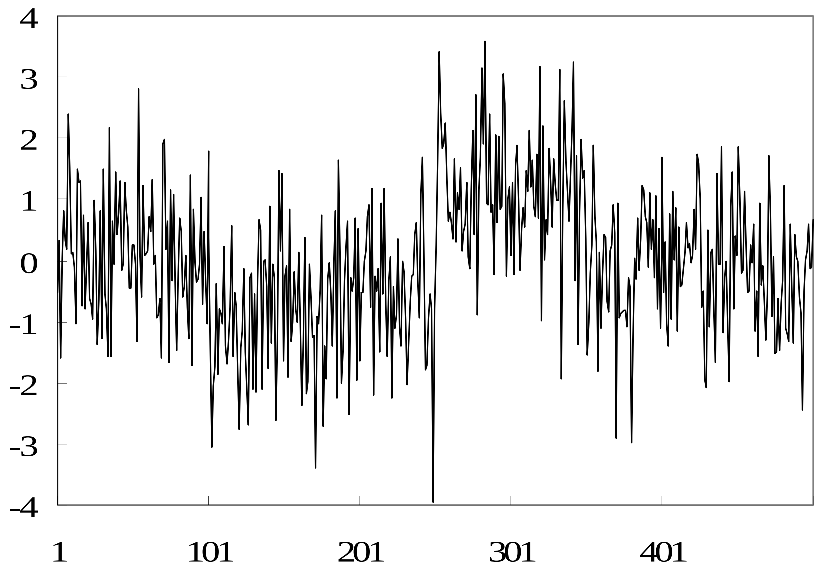
$$p(x_n|Y_n) = \frac{p(y_n|x_n)p(x_n|Y_{n-1})}{p(y_n|Y_{n-1})}$$

平滑化

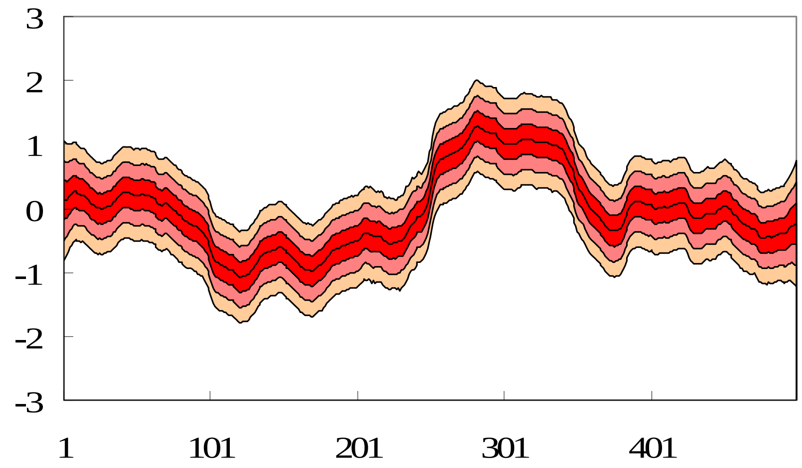
$$p(x_n|Y_N) = p(x_n|Y_n)\int_{-\infty}^{\infty} \frac{p(x_{n+1}|x_n)p(x_{n+1}|Y_N)}{p(x_{n+1}|Y_n)}dx_{n+1}$$

G. Kitagawa (1987): Non-Gaussian state-space modeling of nonstationary time series. *Journal of the American Statistical Association*, Vol.82, No.400,1032-1063 (with discussions)

非ガウス型平滑化



Kalman Smoother



Trend Model

$$t_n = t_{n-1} + v_n$$

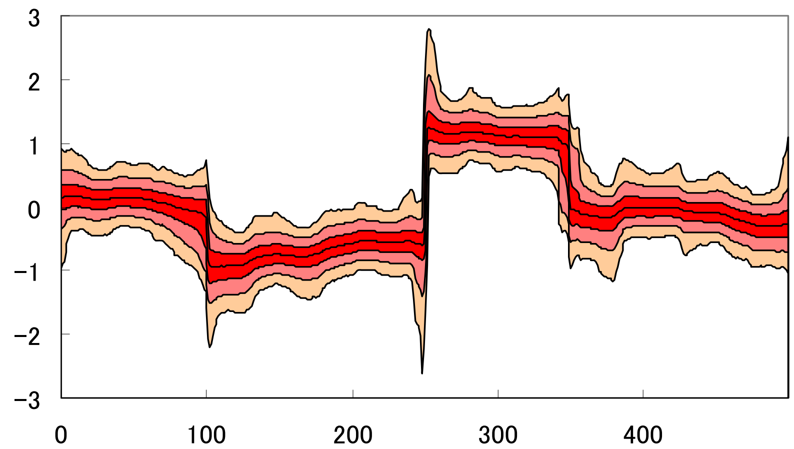
$$y_n = t_n + w_n$$

Noise Distribution

$$v_n \sim N(0, \tau^2) \text{ or } C(0, \tau^2)$$

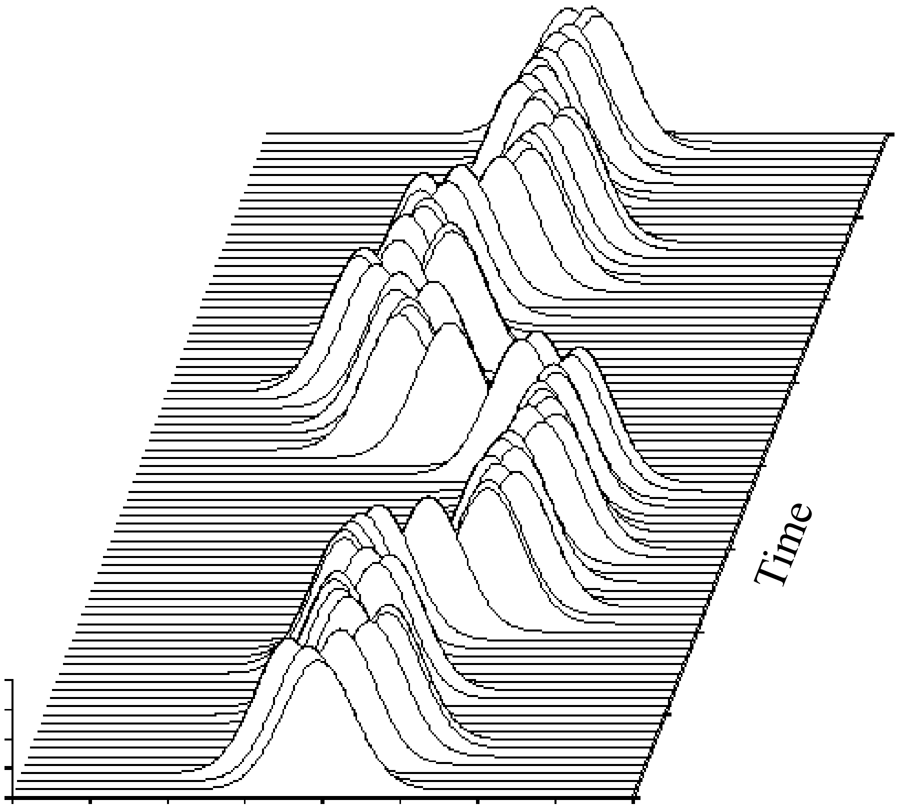
$$w_n \sim N(0, \sigma^2)$$

Exact Non-Gaussian Smoother

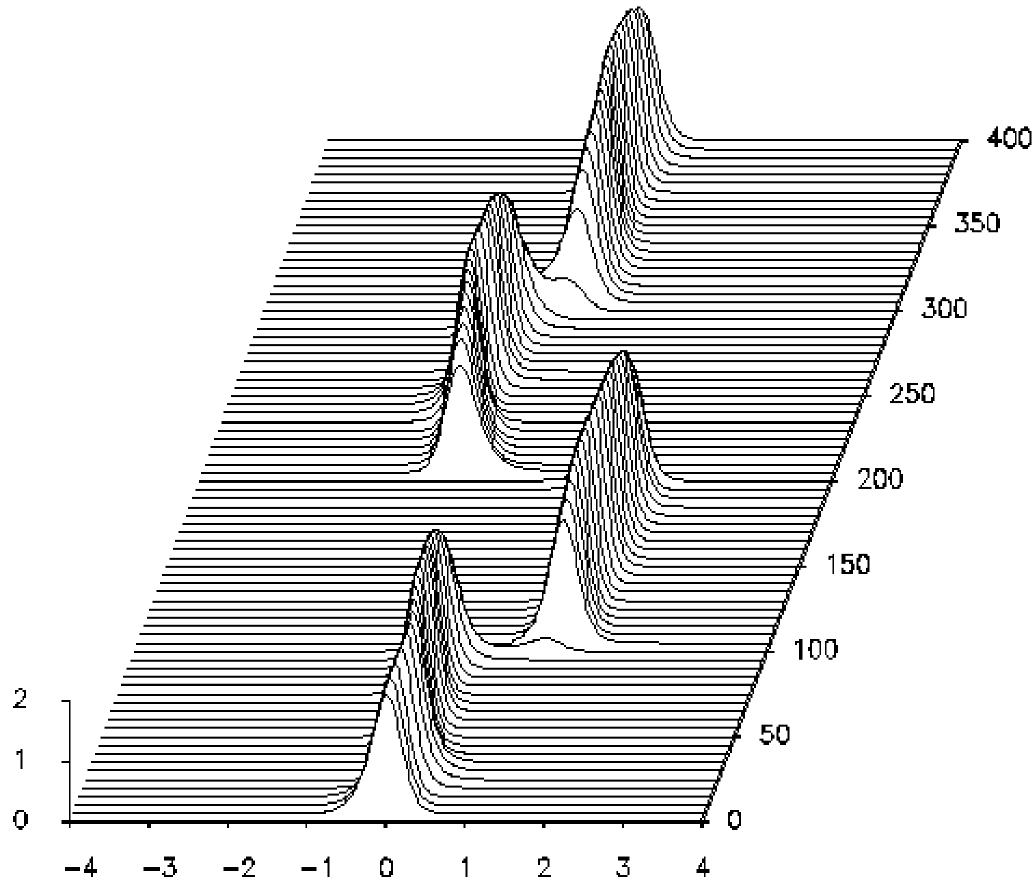


非ガウス型フィルタ・平滑化

Marginal Posterior Density



Gaussian Distribution



Cauchy Distribution

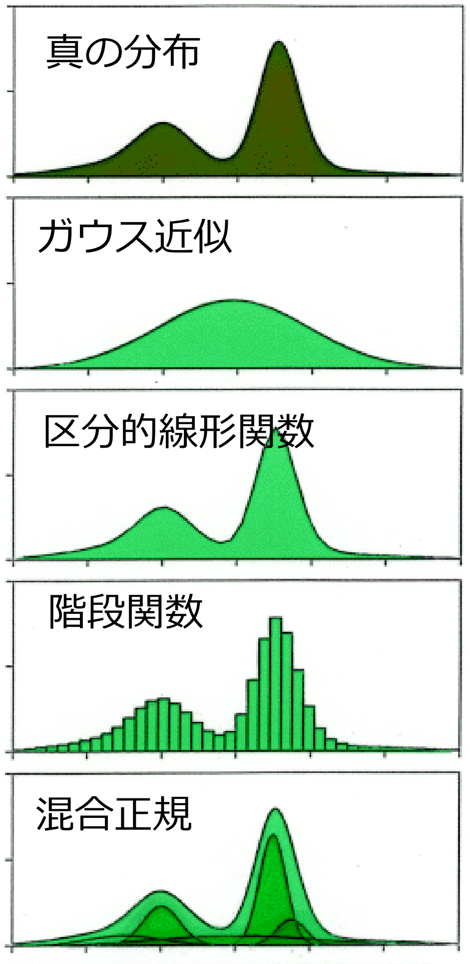
Gaussian-Sum Filter

季節調整モデルに非ガウス型平滑化の適用は無理:

ガウス和(混合正規)近似

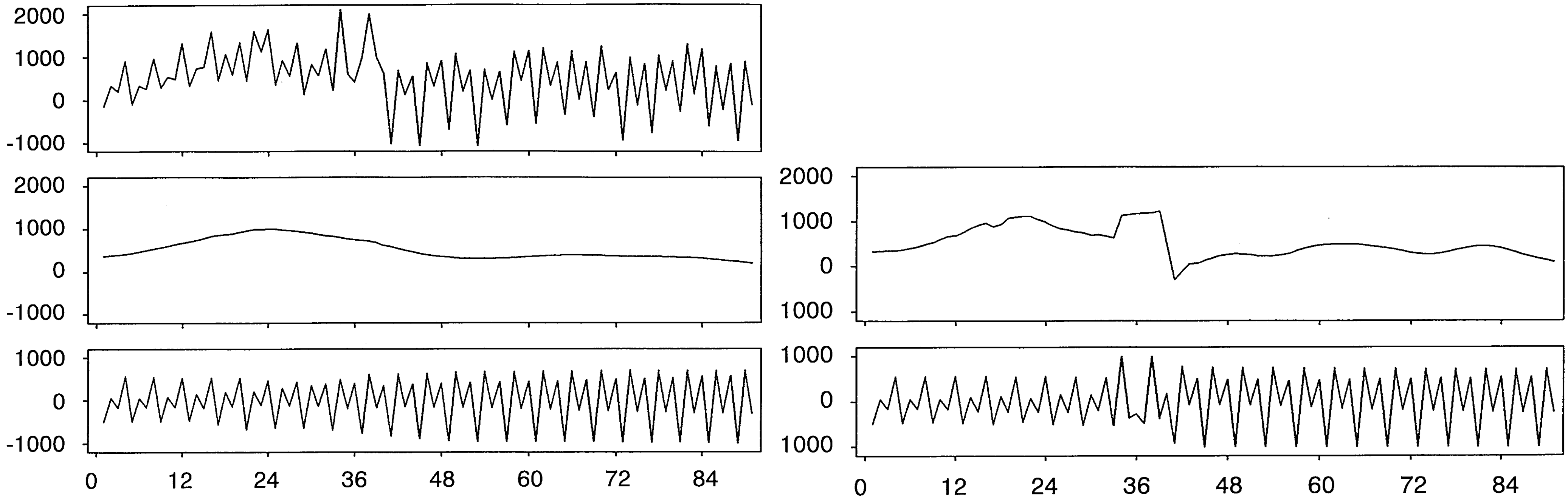
$$p(v_n) = \sum_{i=1}^{K_v} \alpha \varphi_{i\,i}(v_n) \qquad p(x_n \mid Y_{n-1}) = \sum_{k=1}^{L_n} \gamma_{kn} \varphi_{i\,k}(x_n \mid Y_{n-1})$$

$$p(w_n) = \sum_{j=1}^{K_w} \beta \varphi_{i\,j}(w_n) \qquad p(x_n \mid Y_n) = \sum_{\ell=1}^{M_n} \delta_{\ell kn} \varphi_{i\,\ell}(x_n \mid Y_n)$$



Gaussian-Sum Filter: Sorenson & Alspach (1971)
Gaussian-Sum Smoother: Kitagawa (1991,1994)

構造変化の検出

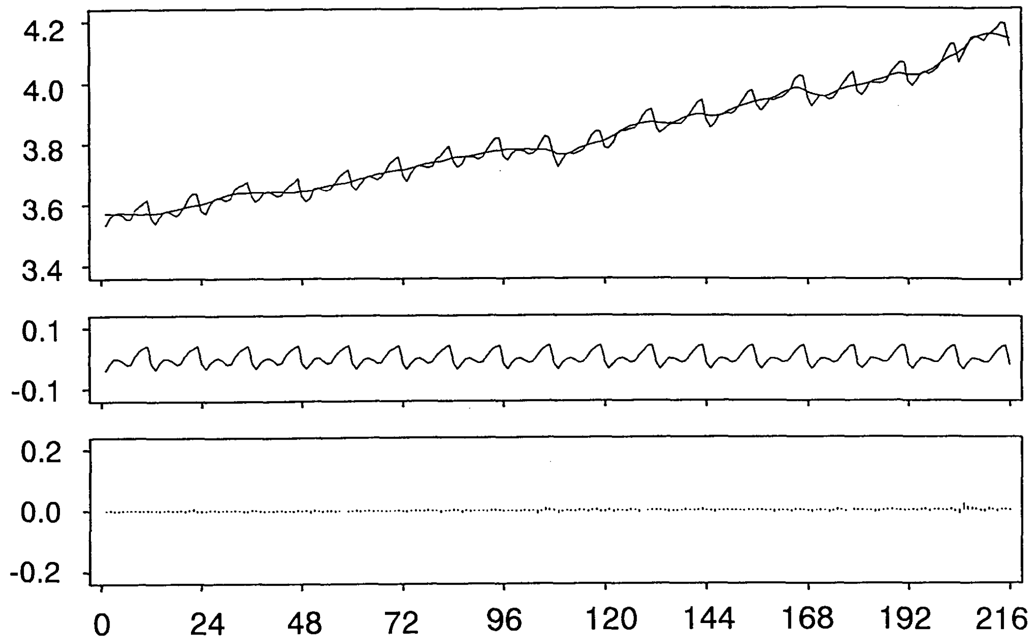


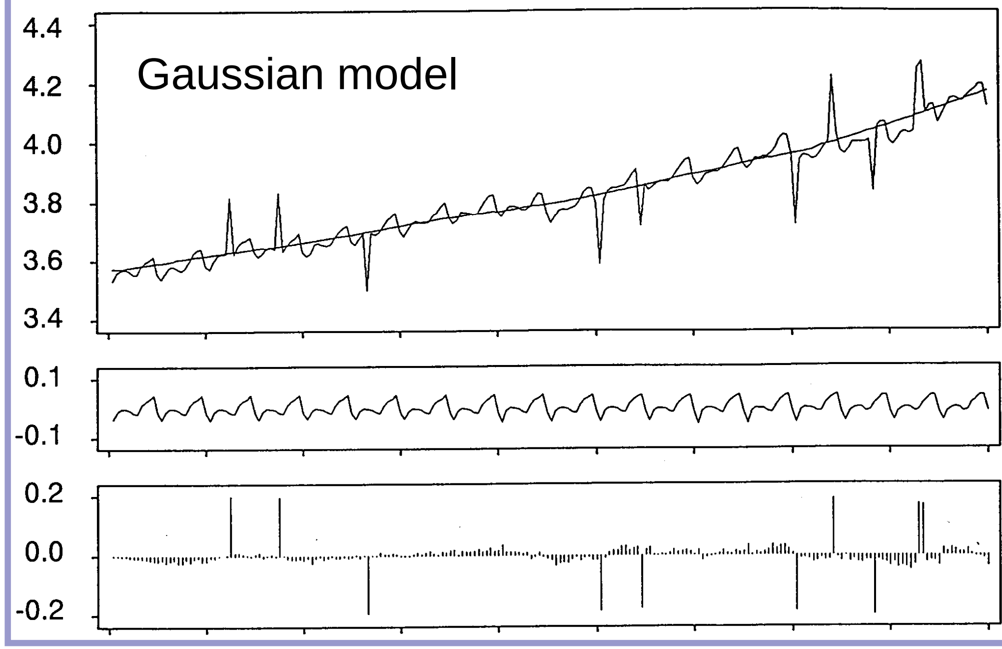
v_n : Gaussian

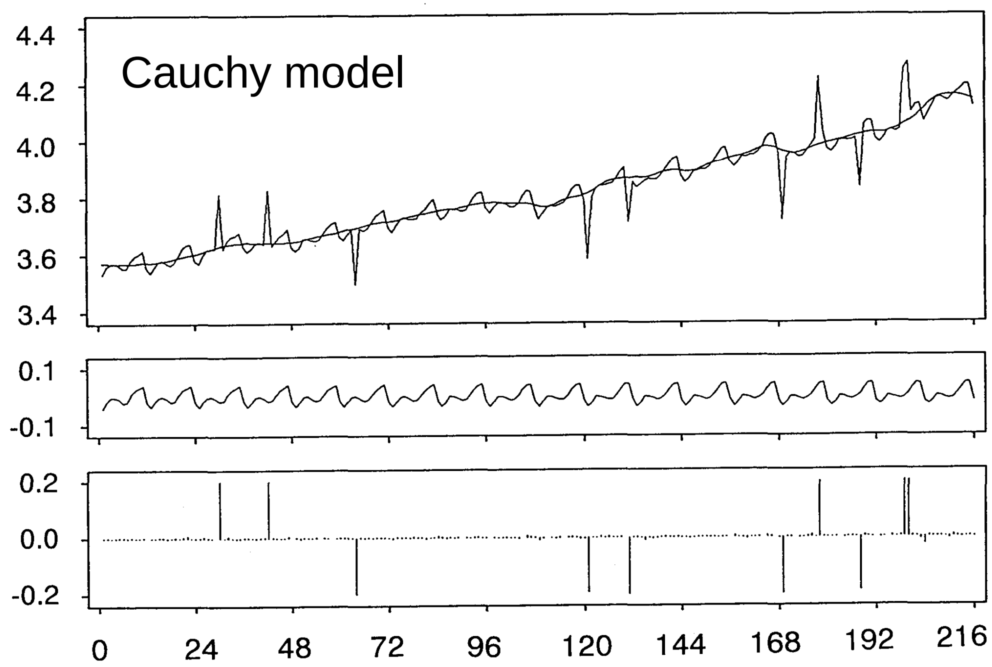
v_n : Gaussian Mixture

$$x_n = Fx_{n-1} + G\boxed{v_n}$$
$$y_n = Hx_n + w_n$$

異常値の処理（ロバストなトレンド推定）







W_n : Observation Noise

Distribution	AIC
Gaussian:	-701.41
Cauchy:	-1293.52

一般状態空間モデル

状態空間モデルの本質的な性質 (条件付き分布)

$$p(x_n \mid x_{n-1}, \ldots, x_0, y_{n-1}, \ldots, y_1) = F(x_n \mid x_{n-1})$$

$$p(y_n \mid x_n, \ldots, x_0, y_{n-1}, \ldots, y_1) = H(y_n \mid x_n)$$

一般状態空間モデル

$$x_n \sim F(\cdot \mid x_{n-1})$$

$$y_n \sim H(\cdot \mid x_n)$$

状態空間モデル

線形・ガウス型

$$x_n = Fx_{n-1} + Gv_n$$
$$y_n = Hx_n + w_n$$

関数：非線形
分布：非ガウス型

非線形・非ガウス型

$$x_n = f(x_{n-1}, v_n)$$
$$y_n = h(x_n, w_n)$$

条件付分布
離散状態・離散観測値

一般型

$$x_n \sim F(\cdot \mid x_{n-1})$$
$$y_n \sim H(\cdot \mid x_n)$$

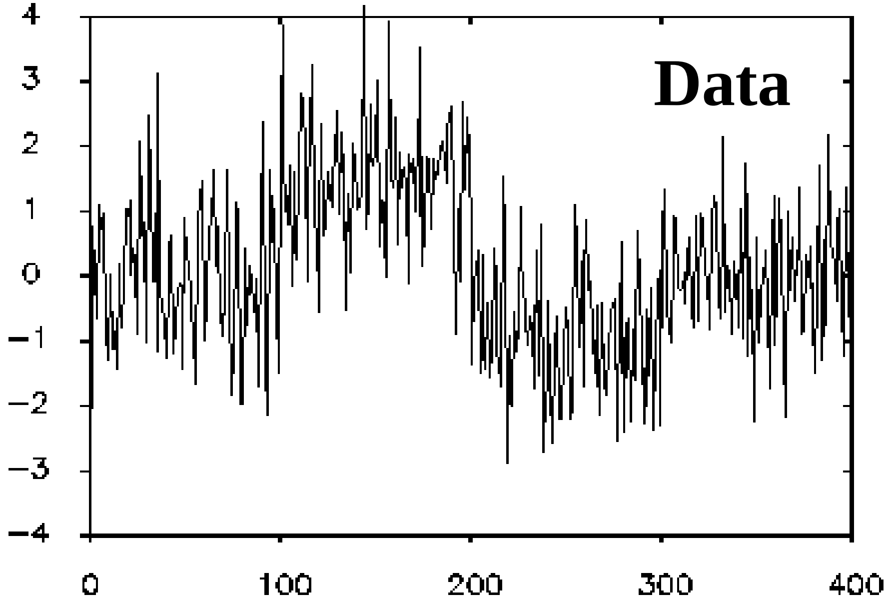
一般化状態空間モデリングの応用

1. 非ガウス型平滑化
   - レベルシフト（構造変化）
   - 非ガウス型季節調整
   - 確率的ボラティリティモデル
2. 非線形平滑化
   - トラッキング
   - Phase-unwrapping
3. 信号抽出問題
4. 計数データのモデリング
5. 自己組織型状態空間モデル

Gordon et al. (1993), Kitagawa (1996)

Doucet, de Freitas and Gordon (2001)　“Sequential Monte Carlo Methods in Practice”

レベルシフトの自動検出



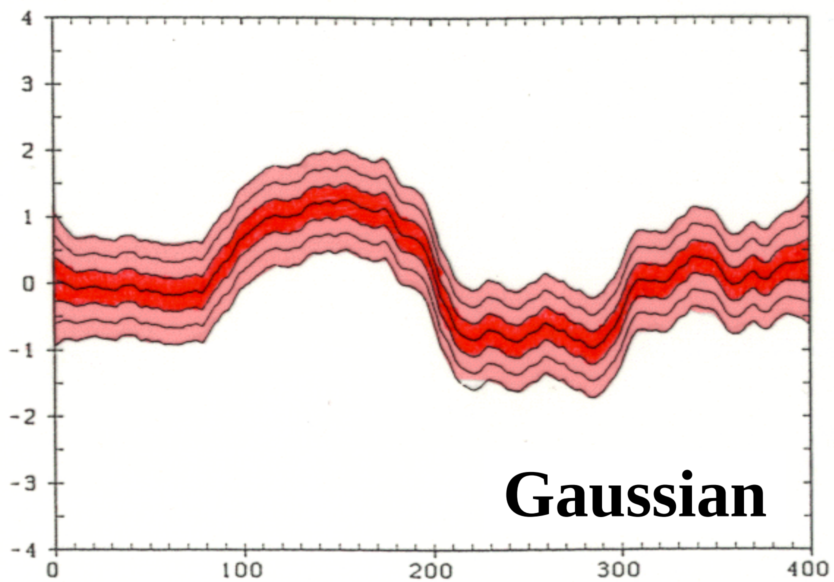
トレンドモデル

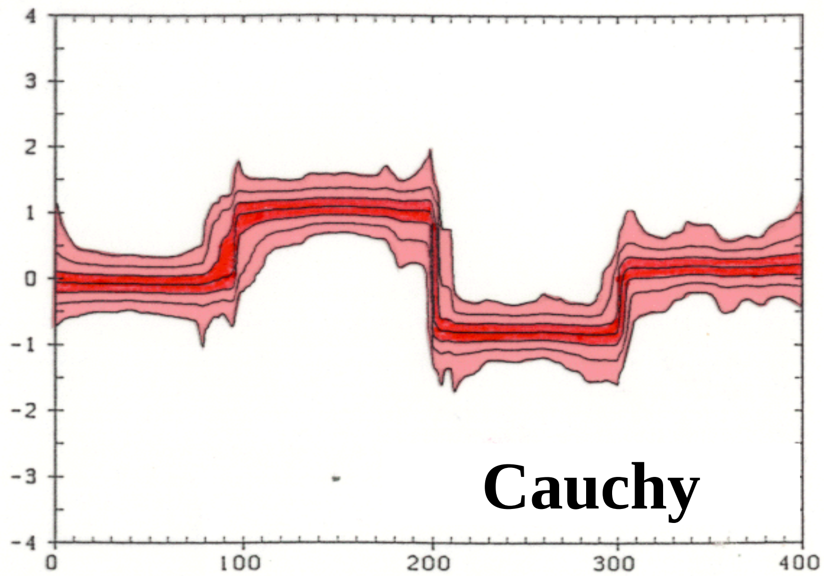
$$t_n = t_{n-1} + v_n$$
$$y_n = t_n + w_n$$

ノイズ分布

$v_n \sim C(0, \tau^2)$　コーシー分布

$w_n \sim N(0, \sigma^2)$　正規分布





MEIJI UNIVERSITY

自己組織型状態空間モデル

$$x_n \sim f(x \mid x_{n-1}, \theta_n)$$
$$y_n \sim h(y \mid x_n, \theta_n)$$

θ_n　　時変パラメータ

$$\theta_n = \theta_{n-1} + v_n$$

状態とパラメータの同時推定

"Although this extended Kalman filter approach appears perfectly straightforward, experiences has shown that with the usual state space model, it does not work well in practice."

Anderson and Moore (1979) p284

状態の拡大

$$z_n = \begin{pmatrix} x_n \\ \theta_n \end{pmatrix} \begin{matrix} \text{状態} \\ \text{パラメータ} \end{matrix}$$

$$z_n \sim F(z \mid z_{n-1})$$
$$y_n \sim H(y \mid z_n)$$

自己組織型フィルタ・平滑化

状態空間モデル

$$x_n = x_{n-1} + v_n$$
$$y_n = x_n + w_n$$

コーシー分布

$$p(v_n) = \frac{\tau}{\pi} \frac{1}{(v_n^2 + \tau^2)}$$

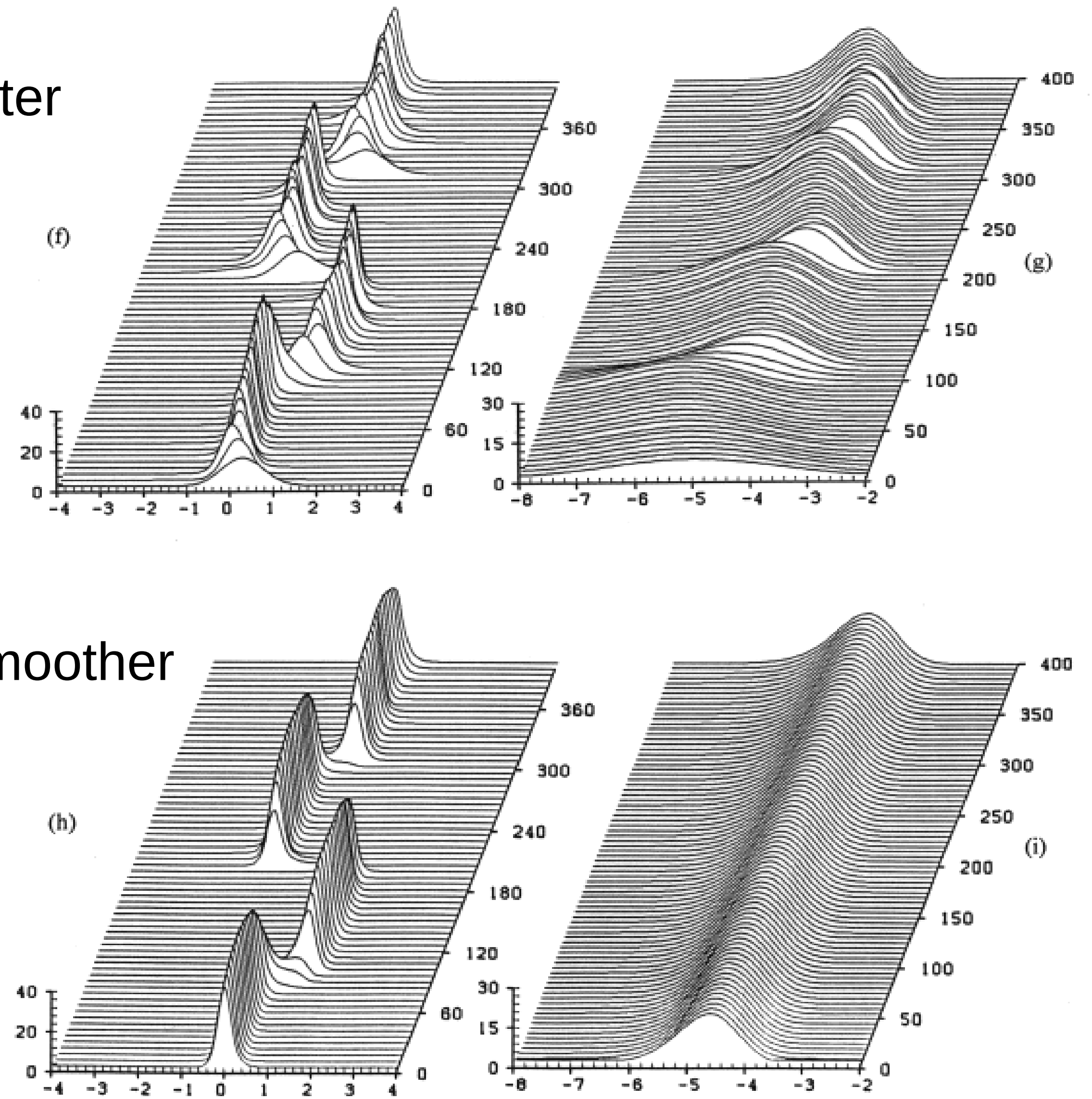
Augmented State Space Model

$$\begin{bmatrix} x_n \\ \log \tau_n^2 \end{bmatrix} = \begin{bmatrix} x_{n-1} \\ \log \tau_{n-1}^2 \end{bmatrix} + \begin{bmatrix} \tau_{n-1} \\ 0 \end{bmatrix} v_n$$

$$y_n = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} x_n \\ \log \tau_n^2 \end{bmatrix} + w_n$$

Filter

Smoother



Trend

$\log \tau^2$

自己組織型フィルタ・平滑化

状態空間モデル

$$x_n = x_{n-1} + v_n$$

$$y_n = x_n + w_n$$

$$p(v;\tau^2,b) = \frac{\Gamma(b)\tau^{2b-1}}{\Gamma(1/2)\Gamma(b-1/2)}\frac{1}{(v^2+\tau^2)^b}$$

自己組織型状態空間モデル

$$\begin{bmatrix} x_n \\ \theta_{n,1} \\ \theta_{n,2} \end{bmatrix} = \begin{bmatrix} x_{n-1} \\ \theta_{n-1,1} \\ \theta_{n-1,2} \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} v_n$$

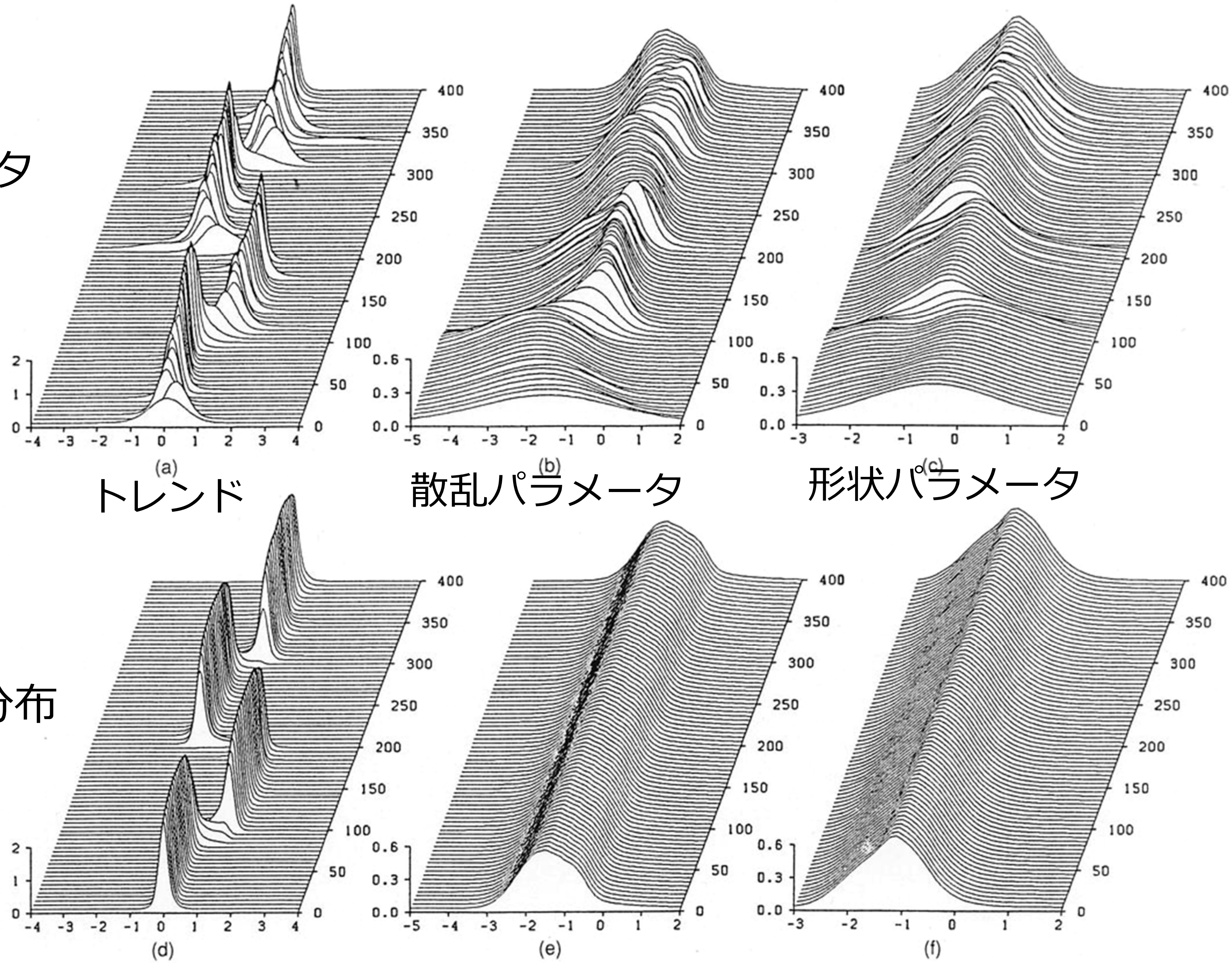
$$y_n = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_n \\ \theta_{n,1} \\ \theta_{n,2} \end{bmatrix} + w_n$$

$$\theta_{n,1} = \log \tau_n^2 - 3\theta_{n,2}$$

$$\theta_{n,2} = \log(b_n - 1/2)$$

自己組織型フィルタ・平滑化

フィルタ
分布



平滑化分布

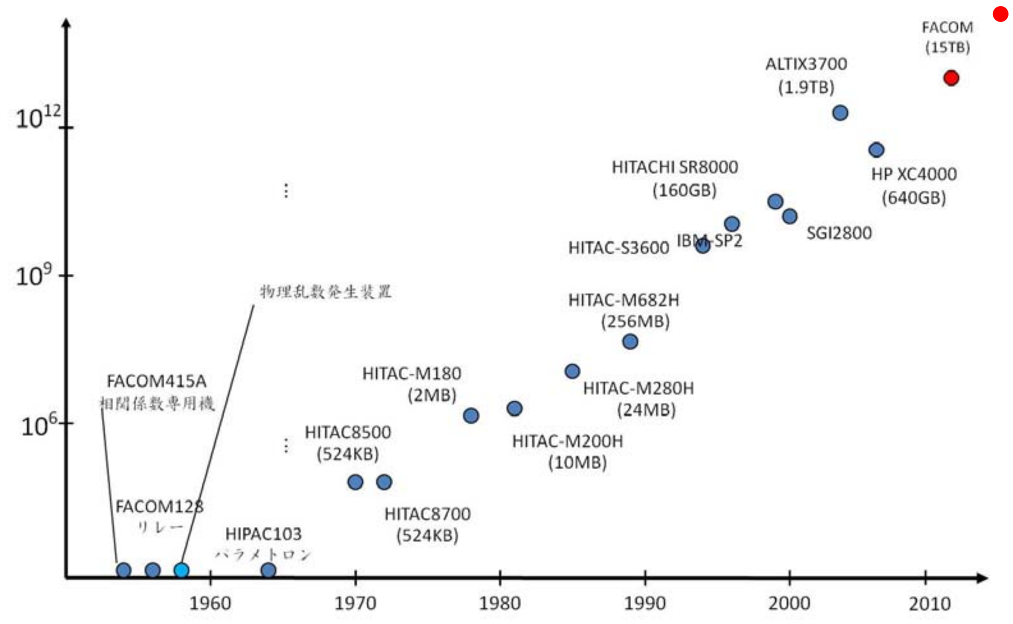
各時代の計算機

		速度	主記憶	
1977	TIMSAC78	1.5MIPS	1MB	
1984	Non-Gaussian Filter	8MIPS	10MB	
1992	Monte Carlo Filter	250MFLOPS	256MB	2GB拡張記憶
2010	統数研スパコン	34TFLOPS	12TB	
2011	京-computer	10PFLOPS	10PB	
2017	統数研	435TFLOPS	228TB	

計算機能力（速度,メモリ）

1984年 ⇨ 現在　5000万倍

1992年 ⇨ 現在　200万倍



分布の近似

0. ガウス近似

　　（拡張）カルマンフィルタ・平滑化

1. 区分的線形（階段）関数近似

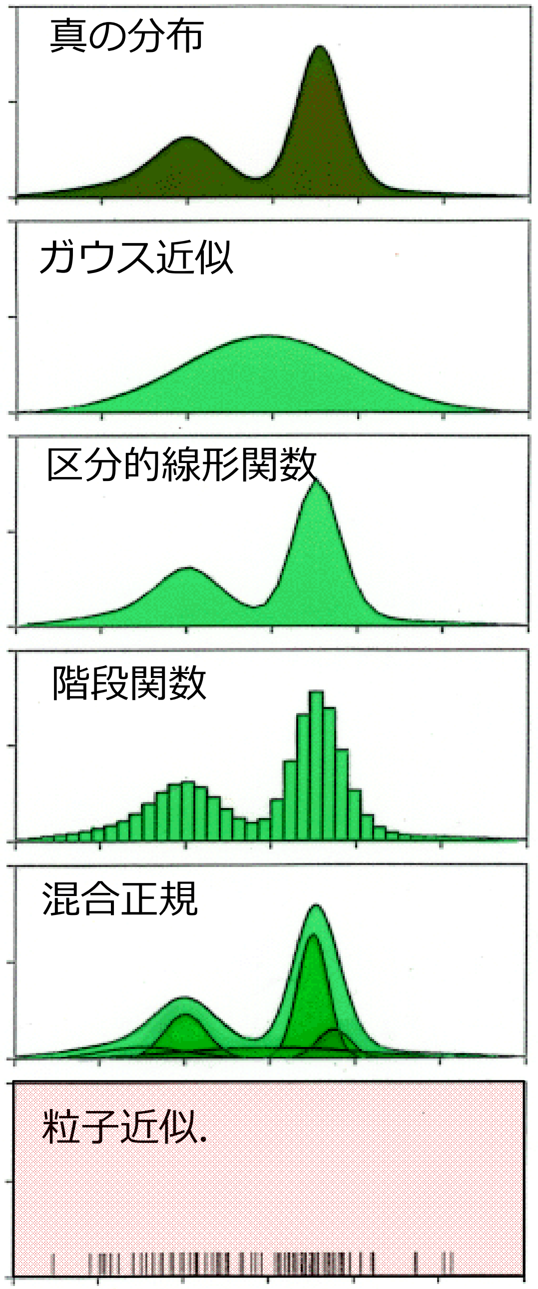
　　非ガウス型フィルタ・平滑化

2. 混合ガウス近似

　　ガウス和フィルタ・平滑化

3. 粒子近似

　　モンテカルロフィルタ・平滑化



粒子による分布の近似

$$\left\{p_n^{(1)},\cdots,p_n^{(m)}\right\}\sim p(x_n \mid Y_{n-1})$$ 予測分布

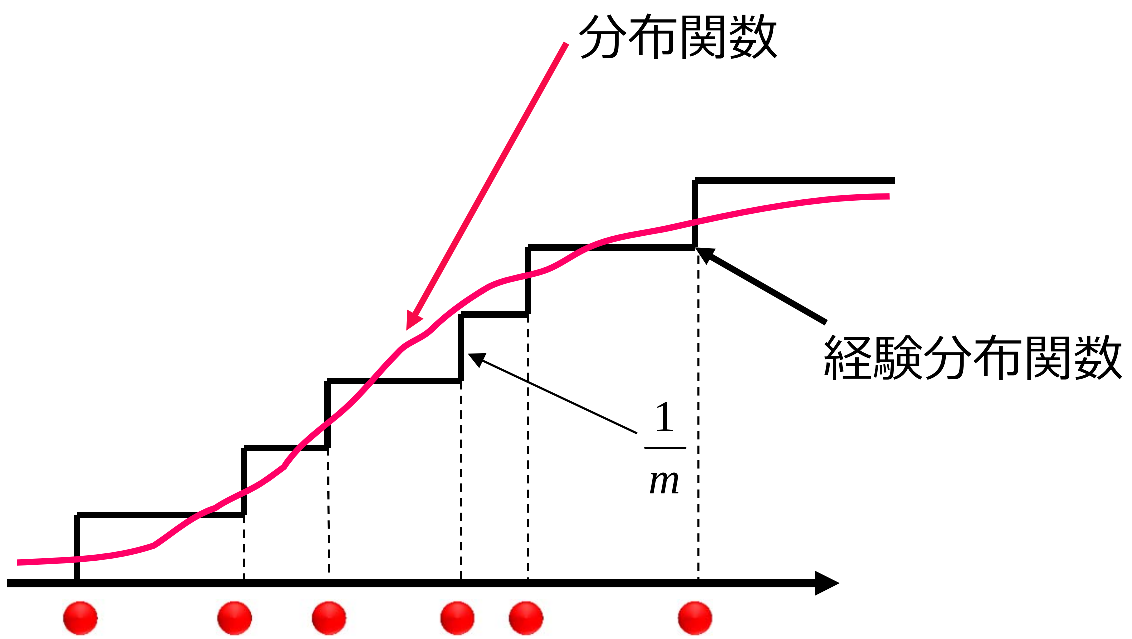
$$\left\{f_n^{(1)},\cdots,f_n^{(m)}\right\}\sim p(x_n \mid Y_n)$$ フィルタ分布

$$\left\{v_n^{(1)},\cdots,v_n^{(m)}\right\}\sim p(v_n)$$ ノイズ分布

$$\left\{p_n^{(1)},\cdots,p_n^{(m)}\right\}\sim p(x_n \mid Y_{n-1})$$

$$\Leftrightarrow \quad F_n(x)=\frac{1}{m}\sum_{j=1}^{m} I(x;p_n^{(j)})$$

$$\Leftrightarrow \quad \Pr(X_n = p_n^{(j)} \mid Y_{n-1}) = \frac{1}{m}$$



Prediction Step

$$x_n = F(x_{n-1}, v_n)$$

System Model

$$v_n^{(j)} \sim p(v)$$

$$f_{n-1}^{(j)} \sim p(x_{n-1}|Y_{n-1})$$

$$p_n^{(j)} = F(f_{n-1}^{(j)}, v_n^{(j)})$$

$$\begin{aligned} p(x_n | Y_n) &= \iint p(x_n, x_{n-1}, v_n | Y_{n-1}) dx_{n-1} dv_{n-1} \\ &= \iint p(x_n | x_{n-1}, v_n, Y_{n-1}) p(v_n | x_{n-1}, Y_{n-1}) p(x_{n-1} | Y_{n-1}) dx_{n-1} dv_{n-1} \\ &= \iint \delta(x_n - F(x_{n-1}, v_n)) p(v_n) p(x_{n-1} | Y_{n-1}) dx_{n-1} dv_{n-1} \end{aligned}$$

$$f_{n-1}^{(j)} \sim p(x_{n-1} | Y_{n-1})$$

$$v_n^{(j)} \sim p(v_n)$$

$$\Downarrow$$

$$p_n^{(j)} = F(f_{n-1}^{(j)}, v_n^{(j)}) \sim p(x_n|Y_{n-1})$$

Filtering Step (Resampling)

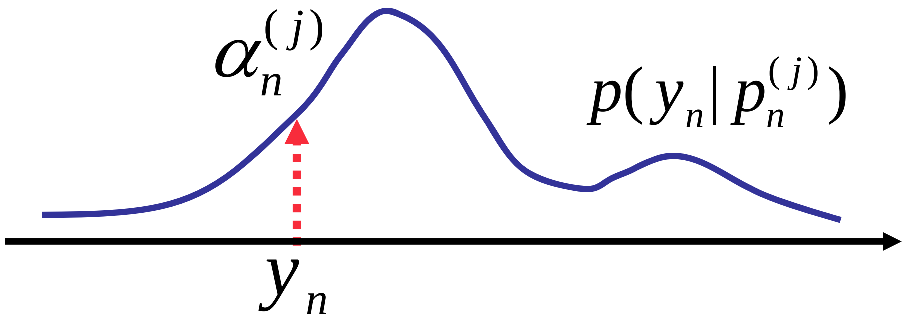
$\alpha_n^{(j)}$: Importance weight of particle $p_n^{(j)}$

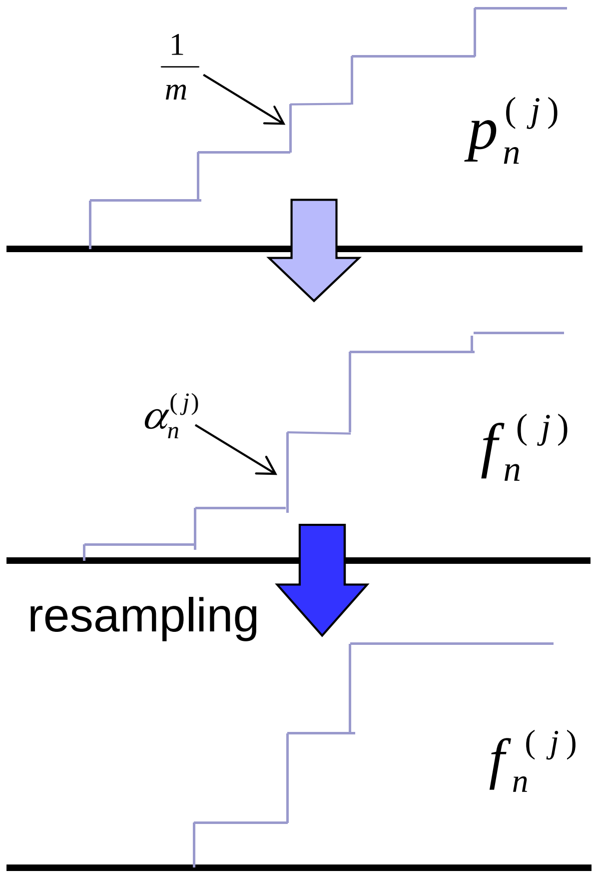
$$\alpha_n^{(j)} = p(y_n | X_n = p_n^{(j)})$$

$$\Pr(X_n = p_n^{(j)} | Y_n) = \Pr(X_n = p_n^{(j)} | Y_{n-1}, y_n)$$

$$= \frac{\Pr(y_n | X_n = p_n^{(j)}) \Pr(X_n = p_n^{(j)} | Y_{n-1})}{\sum_{i=1}^{m} \Pr(y_n | X_n = p_n^{(i)}) \Pr(X_n = p_n^{(i)} | Y_{n-1})}$$

$$= \frac{\alpha_n^{(j)} \frac{1}{m}}{\sum_{i=1}^{m} \alpha_n^{(j)} \frac{1}{m}} = \frac{\alpha_n^{(j)}}{\sum_{i=1}^{m} \alpha_n^{(j)}}$$





Sequential Monte Carlo Filter

システムノイズ

$$v_n^{(j)} \sim p(v) \qquad j = 1, \dots, m$$

予測分布

$$p_n^{(j)} = F(f_{n-1}^{(j)}, v_n^{(j)})$$

重要度 (ベイズ係数)

$$\alpha_n^{(j)} = p(y_n | p_n^{(j)})$$

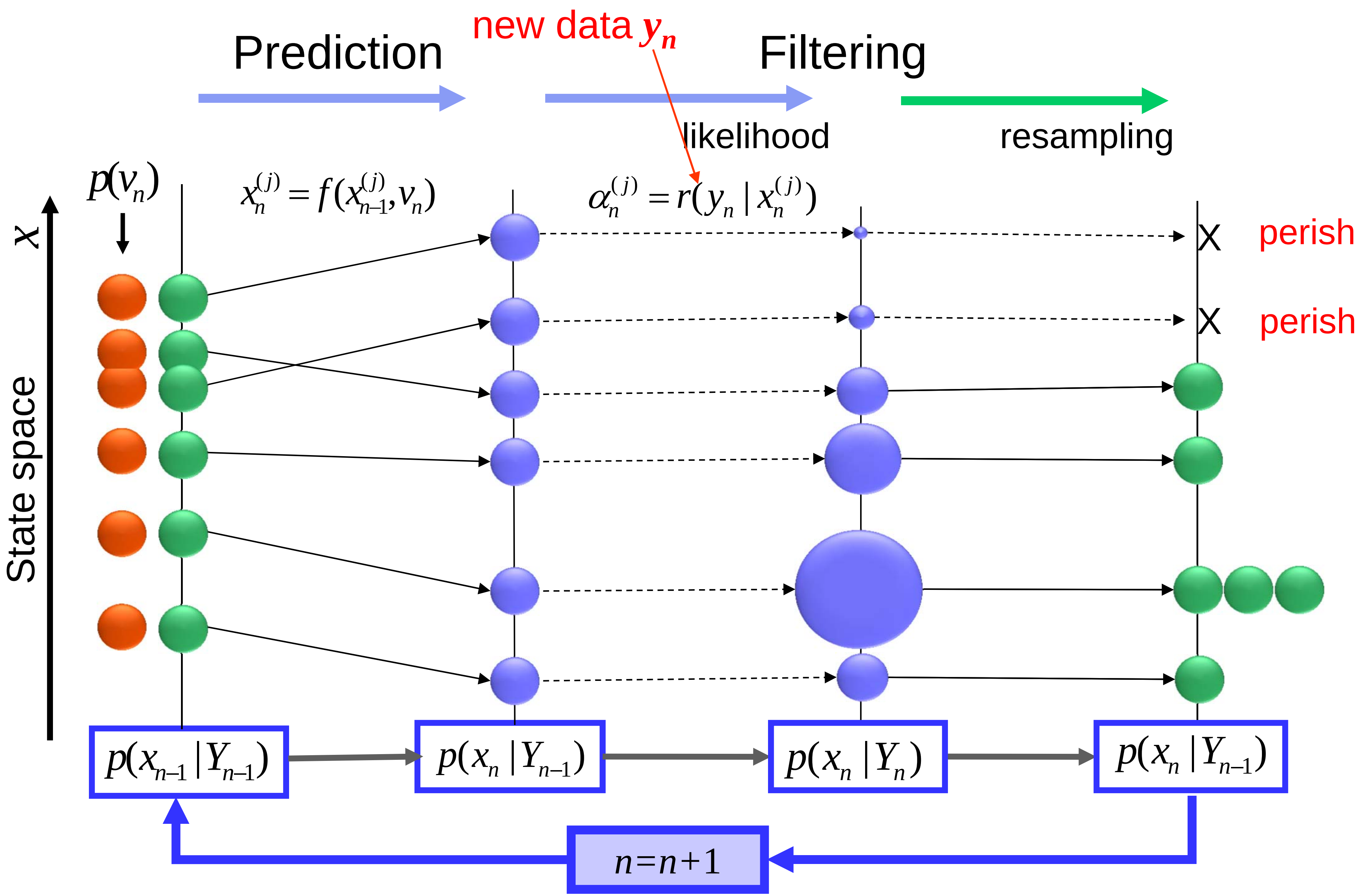
フィルタ分布のリサンプリング

$$\{p_n^{(j)}\} \Rightarrow \{f_n^{(j)}\}$$

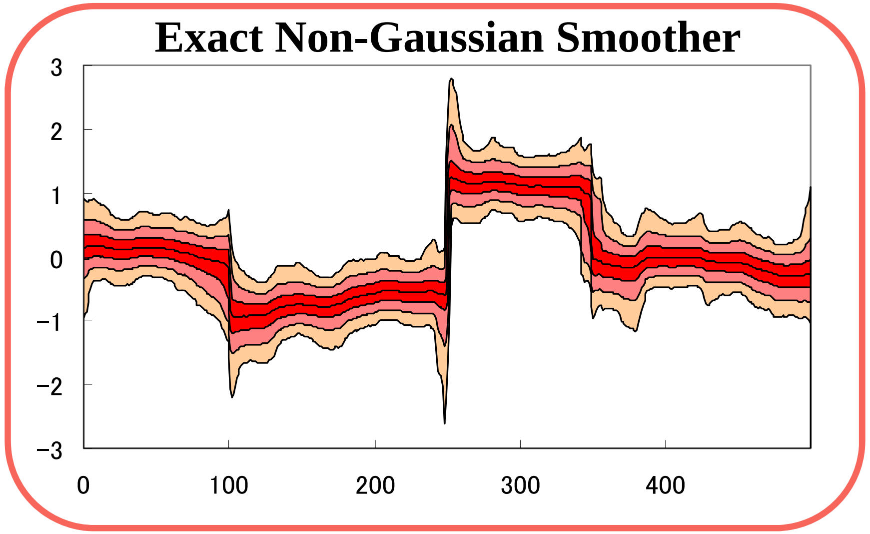
Gordon et al. (1993), Kitagawa (1996)

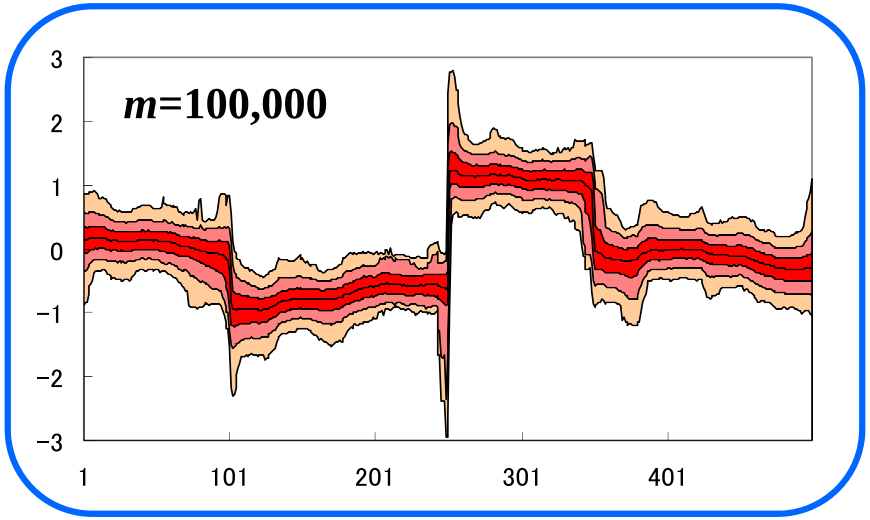
Doucet, de Freitas and Gordon (2001) "Sequential Monte Carlo Methods in Practice"

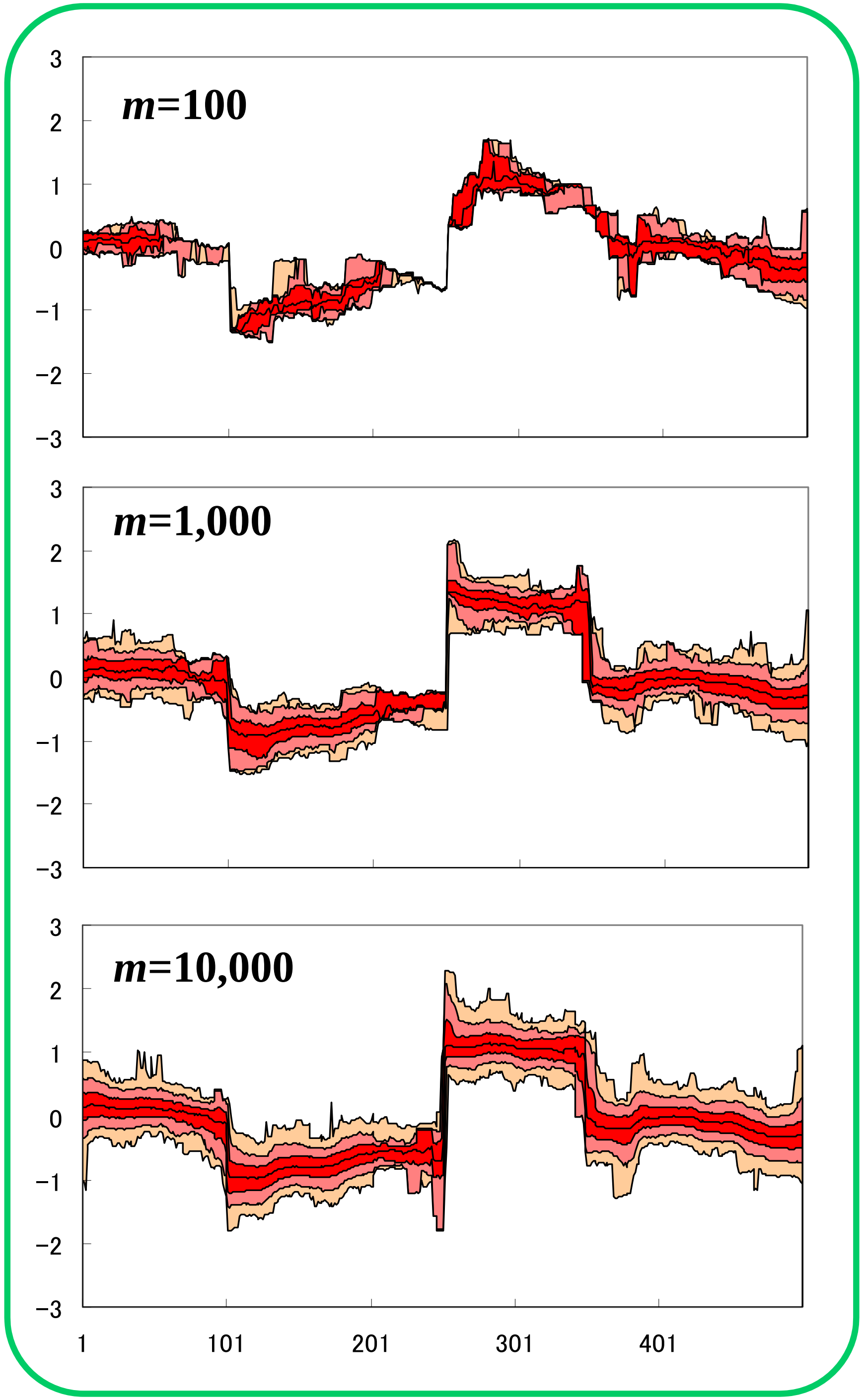
One Cycle of Monte Carlo Filtering



Single MCF







MEIJI UNIVERSITY

Self-Organizing State Space Model

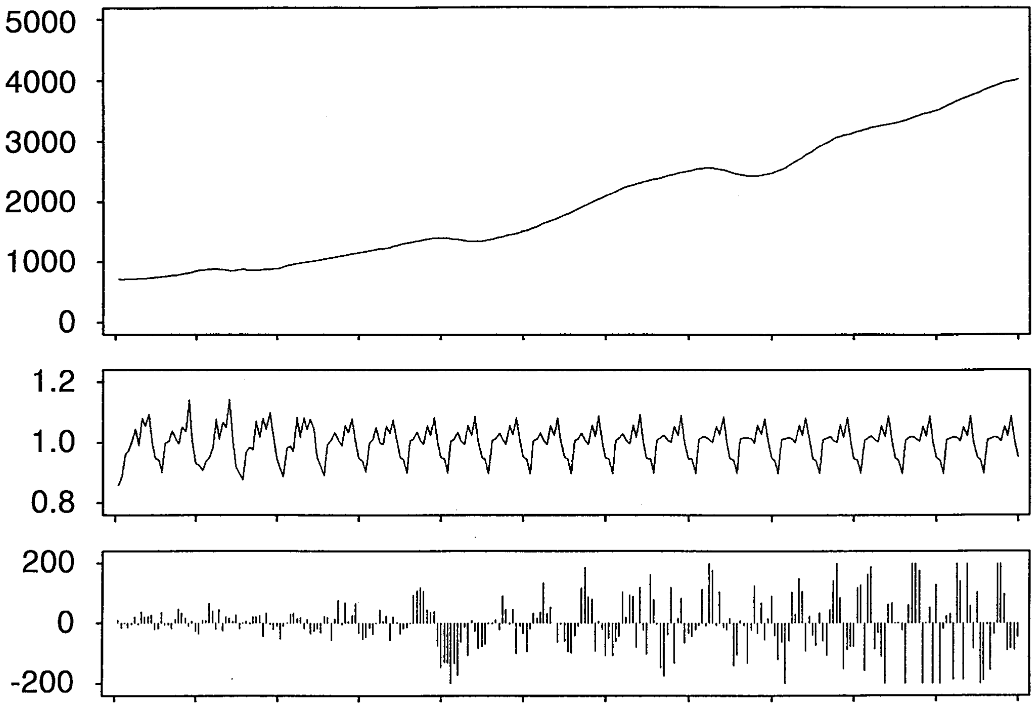
$$z_n = (x_n^T, \tau_1^2, \tau_2^2, \sigma^2)^T$$

$$z_n = \begin{bmatrix} F(x_{n-1}, v_n) \\ \tau_{n-1,1}^2 + u_{n1} \\ \tau_{n-1,2}^2 + u_{n2} \\ \sigma_n^2 + u_{n3} \end{bmatrix},$$

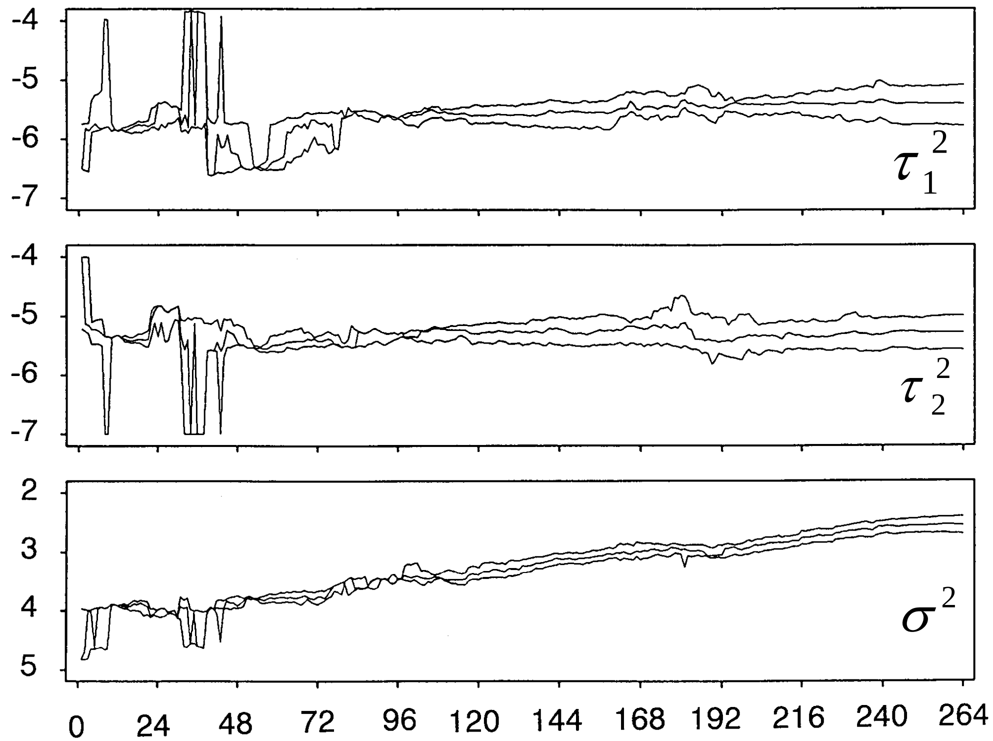
$$y_n = \exp\{T_n\}\exp\{S_n\}\exp\left\{\frac{\sigma_n^2}{2}\right\}w_n$$

Example： 状態とパラメータの同時推定

トレンド、季節成分、ノイズ



パラメータ　$\tau_1^2, \tau_2^2, \sigma^2$



加法モデルと乗法モデルの比較

$$y_n = t_n + s_n + w_n$$

加法モデル

$$y_n = t_n \times s_n \times w_n$$

乗法モデル

情報量規準　（$\log w_n$ が正規分布の場合）

$$\mathrm{AIC}' = \mathrm{AIC} + 2\sum_{i=1}^{n} \log y_i$$

乗法モデル

乗法型モデル

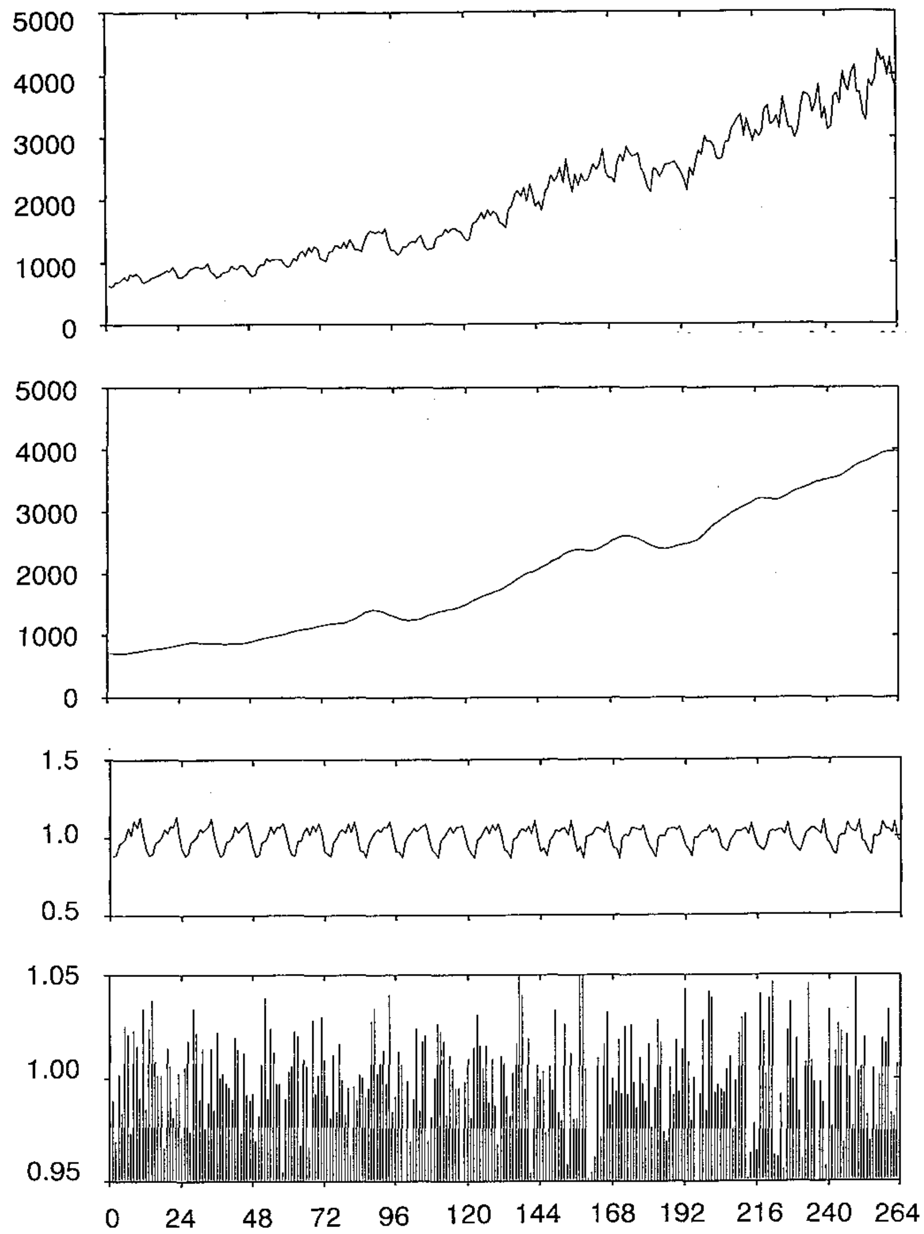
$$
h(x_n, w_n) = T_n \times S_n \times w_n
$$

$$
= \exp\{T_n\} \times \exp\{S_n\} \times w_n
$$

混合モデル

$$
= \exp\{T_n\} \times \exp\{S_n\} + w_n
$$

非線形モデルへ



計数データの季節調整

$$\mathrm{Prob}\ \{y_n = \ell \mid x_n\} = \frac{e^{-\lambda_n}\lambda_n^{\ell}}{\ell!}$$

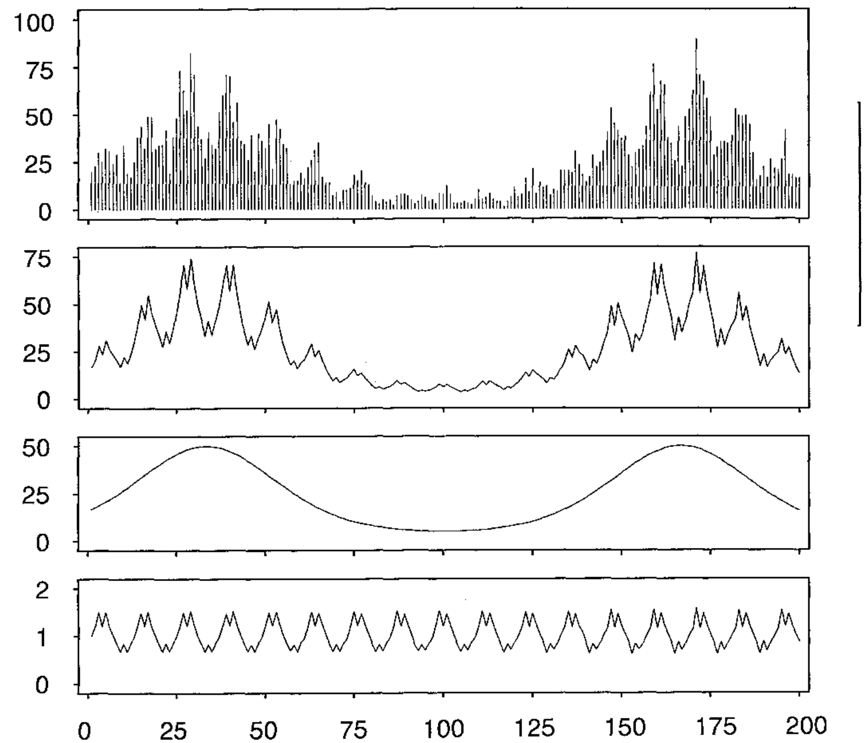
$$x_n = (T_n, T_{n-1}, S_n, \cdots, S_{n-p+2})^T$$

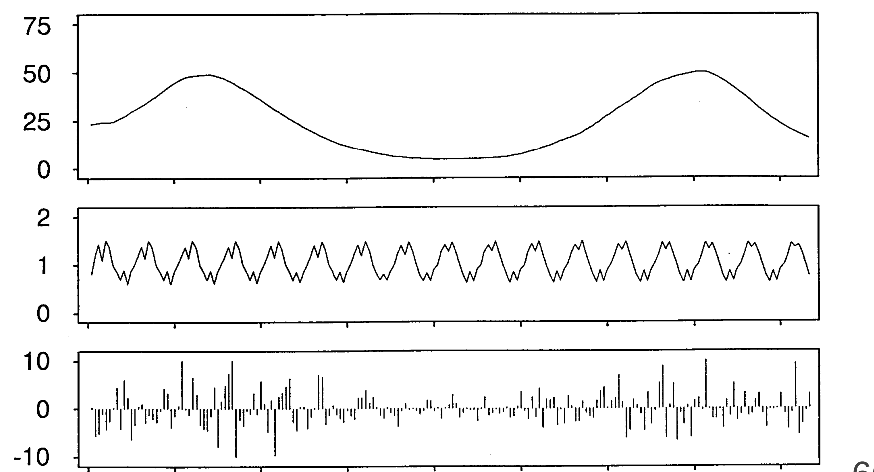
$$\lambda_n = \exp\{T_n\} \times \exp\{S_n\}$$

一般化状態空間モデル

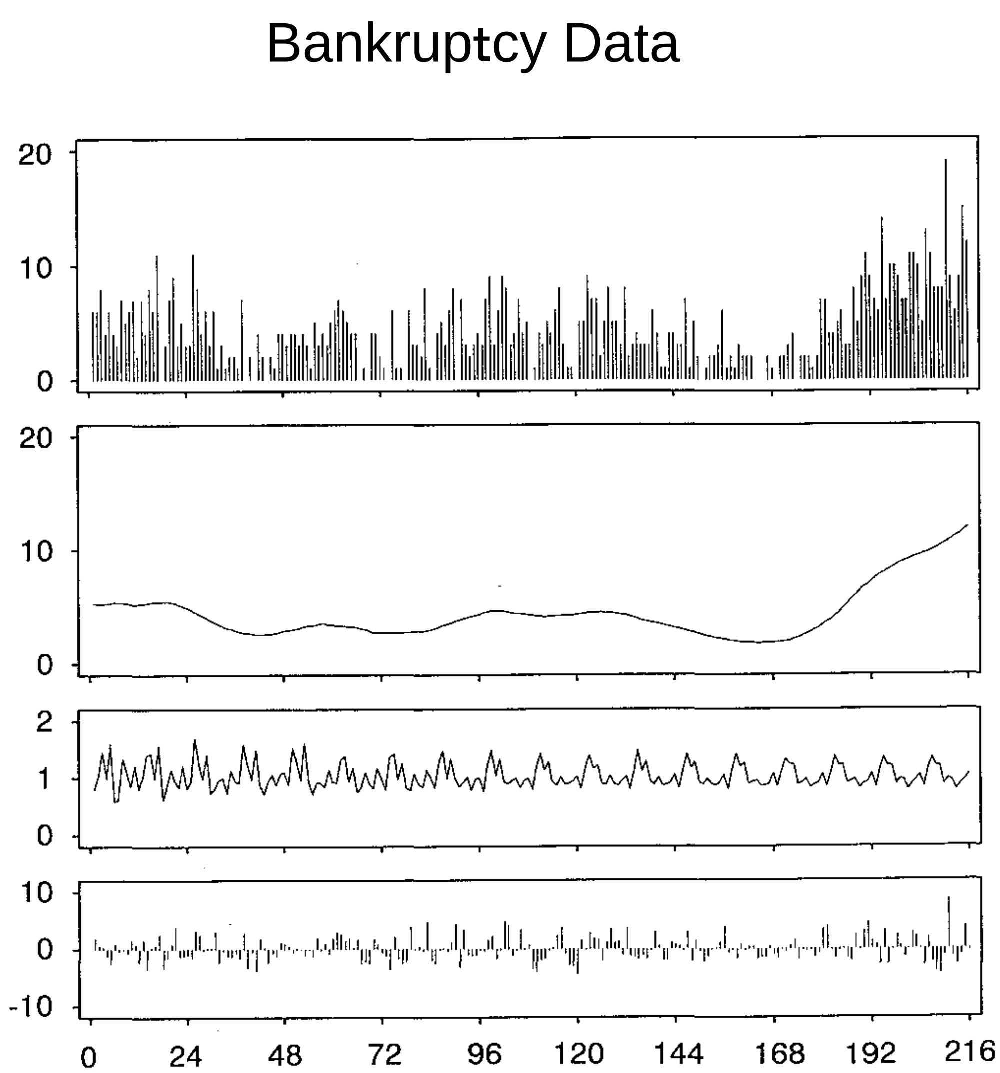
$$x_n = F(x_{n-1}, v_n)$$
$$y_n \sim H(\cdot \mid x_n)$$

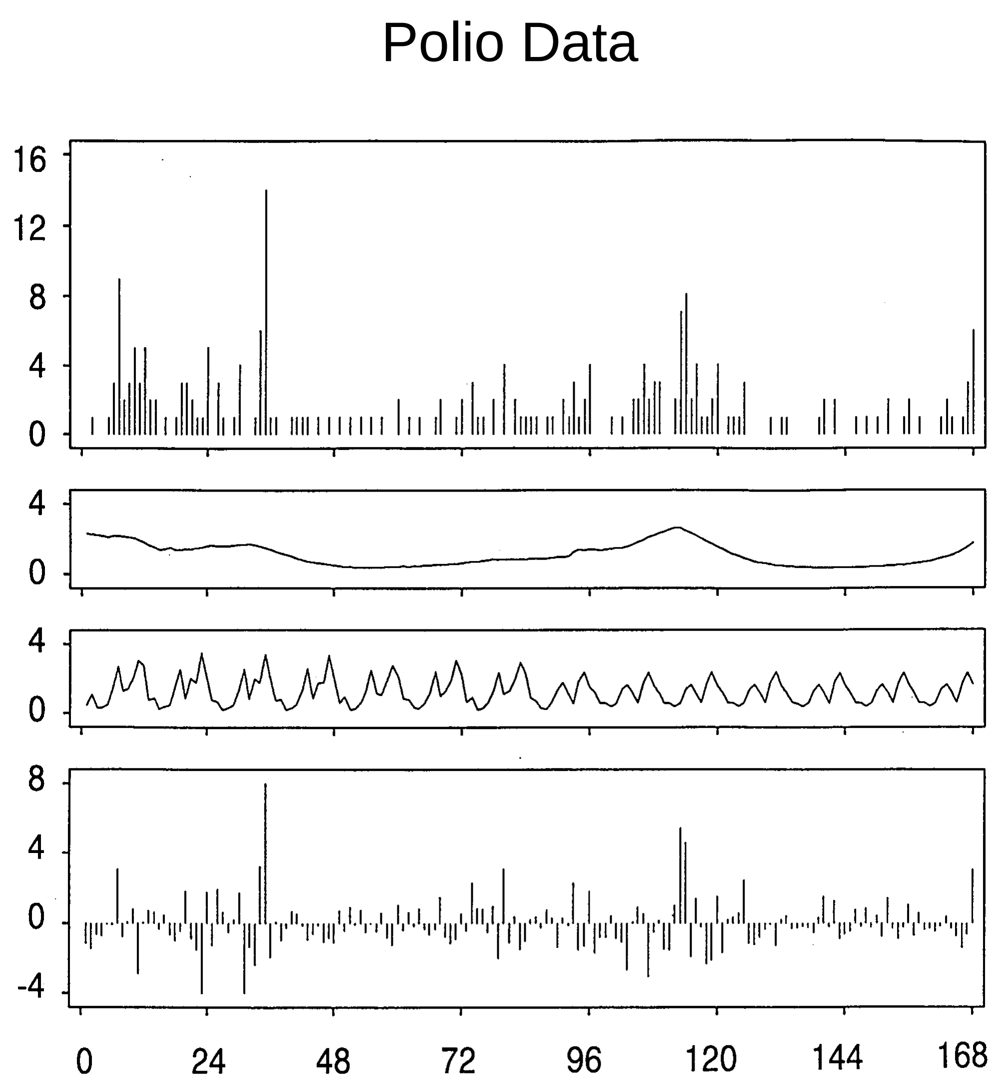
Test Data





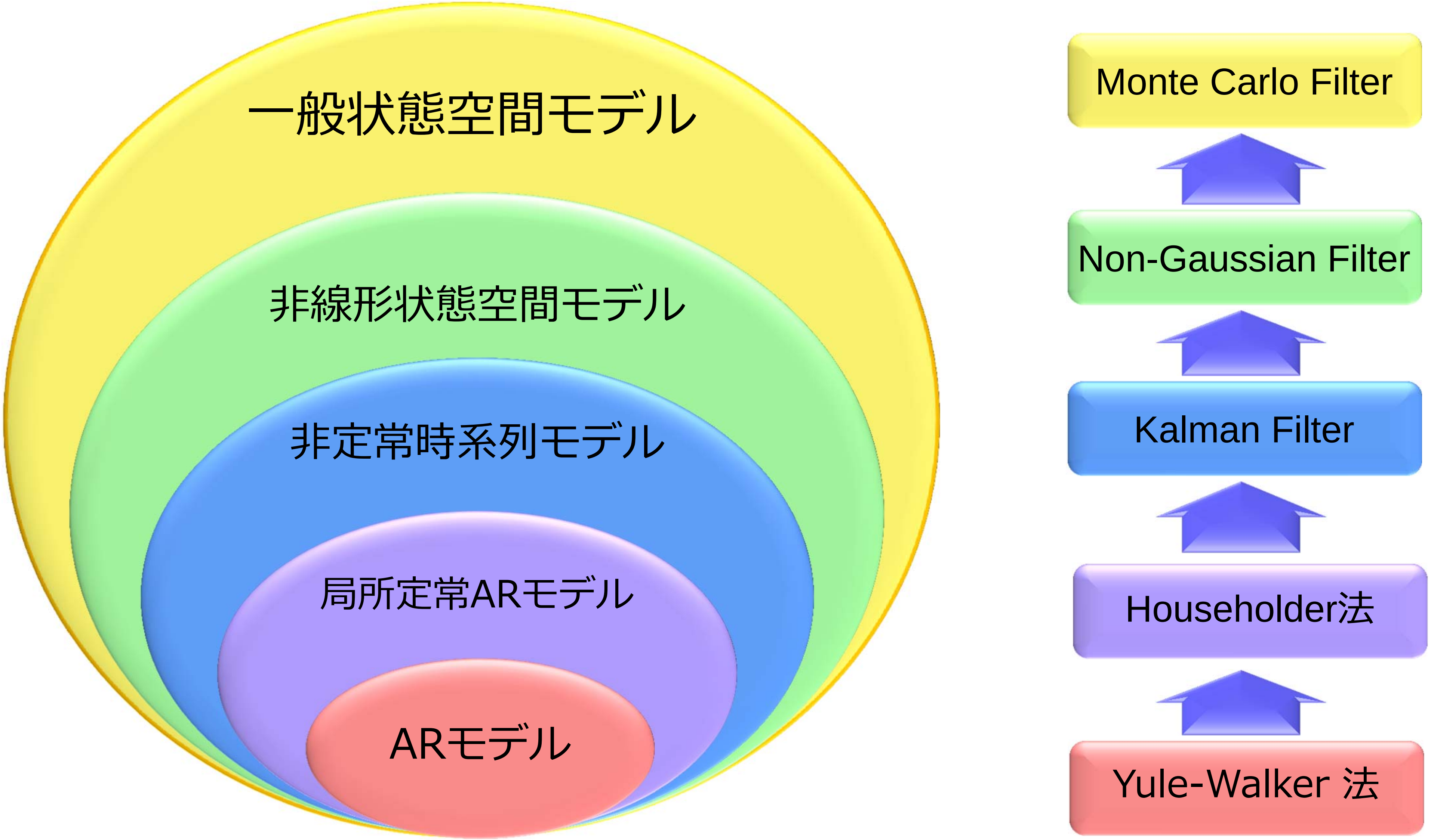
計数データの季節調整





No. of bankrupted Companies over 100M Yen
(Tokyo Shoko Research)

モデルの拡大・計算法の発展



DECOMP: 今後の課題

1．１変量版の改良

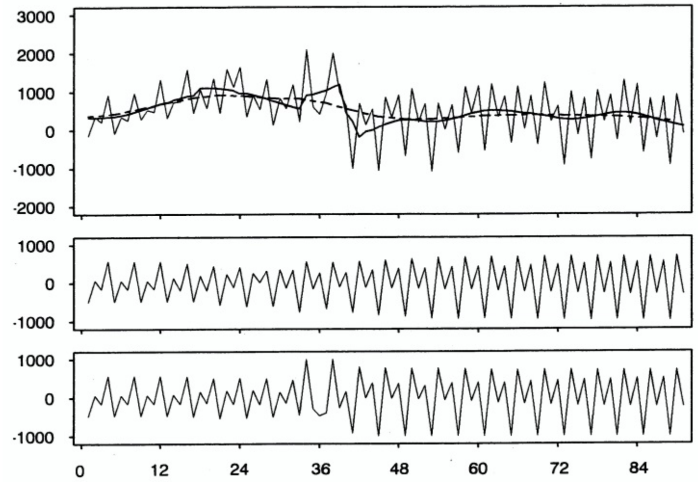
- DECOMPの改良
  - ✓ 季節成分モデル
    - ➢ 硬すぎる季節成分
  - ✓ 定常ARモデル
    - ➢ トレンドとの差別化による安定化
  - ✓ 曜日・祝日効果項
    - ➢ 月ー金の同一化などの制約モデル（開発ずみ）
  - ✓ 改定変動量の制御
    - ➢ Filterの利用
    - ➢ 変動制約項の導入
  - ✓ スペクトルDIP
    - ➢ 季節成分モデルの改良で解消か？

- 非ガウス型モデル
  - ✓ データの異常値の自動処理
  - ✓ 構造変化（トレンド、季節成分等の突然の変化）の自動処理

北川源四郎　季節調整プログラムDECOMPとその後の発展、
統計数理(1997)　第45巻，第2号 217-232

季節調整法（状態空間モデル）

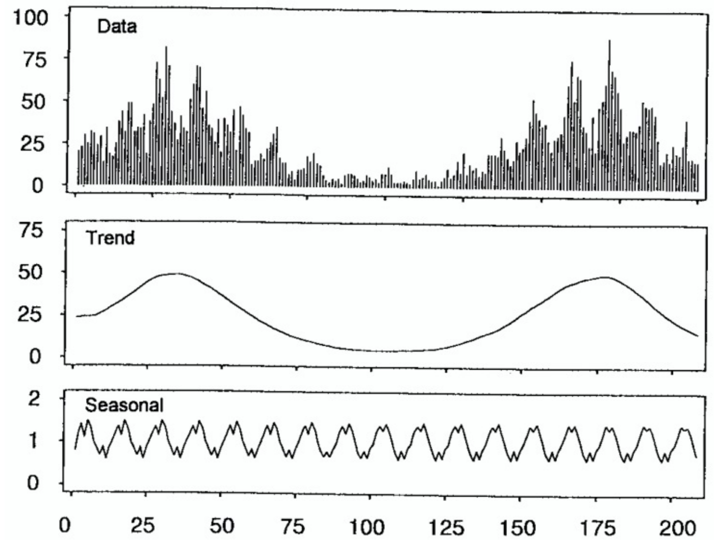
1変量：線形ガウス型
- DECOMP（TIMSAC-84） Information square root filter
- SEASON（岩波版） Kalman filter



1変量：非ガウス版

- **ガウス和フィルタ**
  - ➢ ノイズの混合ガウス和表現
  - 観測値中の異常値
  - トレンドや季節成分のジャンプ（突然の変化）
  - Kitagawa, G. (1989). Non-Gaussian seasonal adjustment. *Computers & Mathematics with Applications*, *18*(6-7), 503-514.
  - Kitagawa, G. (1994). The two-filter formula for smoothing and an implementation of the Gaussian-sum smoother. *Annals of the Institute of Statistical Mathematics*, *46*(4), 605-623.

- **粒子フィルタ**
  - ➢ 非ガウス型ノイズモデル：コーシー分布等
  - Kitagawa, G. (1996). Monte Carlo filter and smoother for non-Gaussian nonlinear state space models. *Journal of computational and graphical statistics*, *5*(1), 1-25.
  - ➢ 非線形モデル：積型モデル，混合モデル
  - ➢ 離散変数の季節調整
  - ➢ 自己組織化：状態とパラメータの同時推定



DECOMP: 今後の課題

2. 多変量版（２～１０変数程度）

- ソフトの多変量化自体はそれほど難しくないが、モデリングには多くの課題
- トレンド成分：共通トレンド・共和分など
- 多変量AR成分、季節成分の成分間のモデル化
- 解析ツール開発：

3. 超多変数版（１００変数）

- 定常カルマンフィルタの利用（分散共分散の計算不要）
- スパースモデリング（自動変数選択）
- 深層学習

4. ソフトウェア開発

多変量季節調整の問題点：トレンド

- 独立トレンドモデル：　特に問題はない

- トレンドの構造モデル（GDP推定用などの特定課題用）

$$T_{n1} = c_2 T_{n2} + \cdots + c_k T_{nk}$$

T_{nj}：　独立トレンド（j=2, …, k）

- トレンドの構造（因子）モデル

$$T_n = \sum_{j=1}^{k} C_j f_{nj} + v_n^{(T)}$$

$$\begin{bmatrix} T_{n1} \\ T_{n2} \\ \vdots \\ T_{n\ell} \end{bmatrix} = \begin{bmatrix} c_{11} & \cdots & c_{1k} \\ c_{21} & \cdots & c_{2k} \\ \vdots & \ddots & \vdots \\ c_{\ell 1} & \cdots & c_{\ell k} \end{bmatrix} \begin{bmatrix} f_{n1} \\ \vdots \\ f_{nk} \end{bmatrix} + \begin{bmatrix} v_{n1}^T \\ v_{n2}^T \\ \vdots \\ v_{n\ell}^T \end{bmatrix}$$

✓ 因子数 k の推定
✓ C_{ij}の推定、f_{nj} の推定

多変量季節調整の問題点： 多変量ARモデル

多変量m次AR成分モデル $$p_n = \sum_{j=1}^{m} A_j p_{n-j} + v_n^{(p)}$$

- 非線形最適化でパラメータ推定をする場合の制約充足の方法

・1変量の場合：

$$\left|a_m^m\right| < 1, \quad m = 1, \ldots, M$$
$$a_i^m = a_i^{m-1} - a_m^m a_{m-i}^{m-1}$$
$$\sigma_m^2 = \sigma_{m-1}^2 (1 - (a_m^m)^2)$$

$$\boxed{\begin{array}{l} -\infty < \theta < \infty \\ -1 < \dfrac{e^\theta - 1}{e^\theta + 1} < 1 \end{array}}$$

・多変量の場合：

$$p_n = \sum_{j=1}^{m} A_i^m p_{n-j} + v_n \qquad W_m = C_m - \sum_{i=1}^{m-1} A_i^{m-1} C_{m-i}$$
$$p_n = \sum_{j=1}^{m} B_i^m p_{n+j} + u_n \qquad V_m = C_0 - \sum_{i=1}^{m} A_i^m C_i^T, \qquad U_m = C_0 - \sum_{i=1}^{m} B_i^m C_i$$
$$v_n \sim N(0, V_m) \qquad A_m^m = W_m U_{m-1}^{-1}, \qquad B_m^m = W_m^T V_{m-1}^{-1}$$
$$u_n \sim N(0, W_m) \qquad A_i^m = A_i^{m-1} - A_m^m B_{m-i}^{m-1}, \qquad B_i^m = B_i^{m-1} - B_m^m A_{m-i}^{m-1}$$

✓ ｜A_m^mと B_m^mの固有値｜＜1のとき、多変量ARモデルは自動的に定常になるか

✓ parametrizationの方法？

Partially Linear State Space Model

State Vectors

$$z_n = \begin{bmatrix} x_n \\ \theta_n \end{bmatrix}$$

Linear state
Non-linear state

Partially Linear State Space Model

$$x_n = F_n(\theta_{n-1})x_{n-1} + G_n(\theta_{n-1})v_n \qquad v_n \sim N(0, Q_n(\theta_{n-1}))$$

$$\theta_n = J_n(\theta_{n-1}) + u_n \qquad u_n \sim N(0, S_n(\theta_{n-1}))$$

$$y_n = H_n(\theta_n)x_n + w_n \qquad w_n \sim N(0, R_n(\theta_{n-1}))$$

Basic Idea

$$p(x_n, \theta_n \mid Y_{1:n}) = p(x_n \mid \theta_n, Y_{1:n})\, p(\theta_n \mid Y_{1:n-1})$$

$$= \sum_{i=1}^{m} \alpha_n^{(i)} \delta(\theta_n - \theta_{n|n}^{(i)}) N(x_n \mid x_{n|n}^{(i)}, V_{n|n}^{(i)})$$

$N(x \mid \mu, V)$ Gaussian density with mean μ, covariance V.

Preparation for Rao-Blackwellization

$$z_n = \begin{bmatrix} x_n \\ \theta_n \end{bmatrix}$$

$$\begin{aligned}
p(z_n \mid Y_{1:n-1}) &= p(x_n, \theta_n \mid Y_{1:n-1}) \\
&= \iint p(x_n, x_{n-1}, \theta_n, \theta_{n-1} \mid Y_{1:n-1}) dx_{n-1} d\theta_{n-1} \\
&= \iint p(\theta_n \mid x_n, x_{n-1}, \theta_{n-1}) p(x_n \mid x_{n-1}, \theta_{n-1}) p(x_{n-1}, \theta_{n-1} \mid Y_{1:n-1}) dx_{n-1} d\theta_{n-1} \\
&= \int p(\theta_n \mid \theta_{n-1}) \boxed{\int p(x_n \mid x_{n-1}, \theta_{n-1}) p(x_{n-1} \mid \theta_{n-1}, Y_{1:n-1}) dx_{n-1}} p(\theta_{n-1} \mid Y_{1:n-1}) d\theta_{n-1} \\
&= \int p(\theta_n \mid \theta_{n-1}) q(x_n \mid \theta_{n-1}, Y_{1:n-1}) p(\theta_{n-1} \mid Y_{1:n-1}) d\theta_{n-1}
\end{aligned}$$

Kalman Predictor given $\theta_{n\text{-}1}$

$$\begin{aligned}
p(z_n \mid Y_{1:n}) &= p(x_n, \theta_n \mid y_n, Y_{1:n-1}) \\
&\propto p(y_n, x_n, \theta_n \mid Y_{1:n-1}) \\
&= p(y_n \mid x_n, \theta_n, Y_{1:n-1}) p(x_n, \theta_n \mid Y_{1:n-1}) \\
&= \boxed{p(y_n \mid x_n, \theta_n)} p(x_n, \theta_n \mid Y_{1:n-1})
\end{aligned}$$

Kalman Filter given θ_n

Rao-Blackwellized Particle Filter

0. Initialization

$$\theta_{0|0}^{(i)} \sim p_0(\theta), \qquad i = 1, \ldots, m$$

$$x_{0|0}^{(i)} \qquad \text{mean of } x_0^{(i)}$$

$$V_{0|0}^{(i)} \qquad \text{covariance of } x_0^{(i)}$$

1. Prediction

(1) Kalman filter prediction for x_n given $\theta_{n\text{-}1}$

$$x_{n-1|n-1}, V_{n-1|n-1} \quad \Rightarrow \quad x_{n|n-1}, V_{n|n-1}$$

(2) Prediction for θ_n

$$\theta_{n|n-1}^{(i)} = J_n(\theta_{n-1|n-1}^{(i)}) + u_n^{(i)}, \quad u_n^{(i)} \sim N(0, S_n(\theta_{n-1|n-1}^{(i)})) \quad i = 1, \ldots, m$$

2. Filter

(1) Kalman filter

$$x_{n|n-1}, V_{n|n-1} \xrightarrow{y_n} x_{n|n}, V_{n|n}$$

(2) Resampling

$$\theta_{n|n-1}^{(1)}, \ldots, \theta_{n|n-1}^{(m)} \xrightarrow{\alpha_n^{(1)}, \ldots, \alpha_n^{(m)}} \theta_{n|n}^{(1)}, \ldots, \theta_{n|n}^{(m)}$$

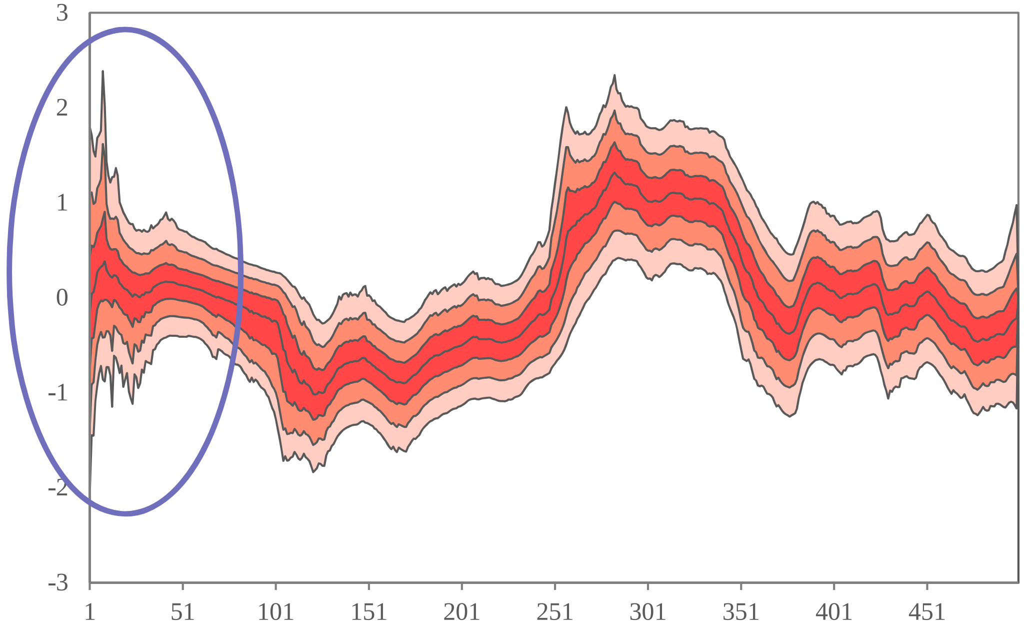
3. Fixed Interval Smoothing

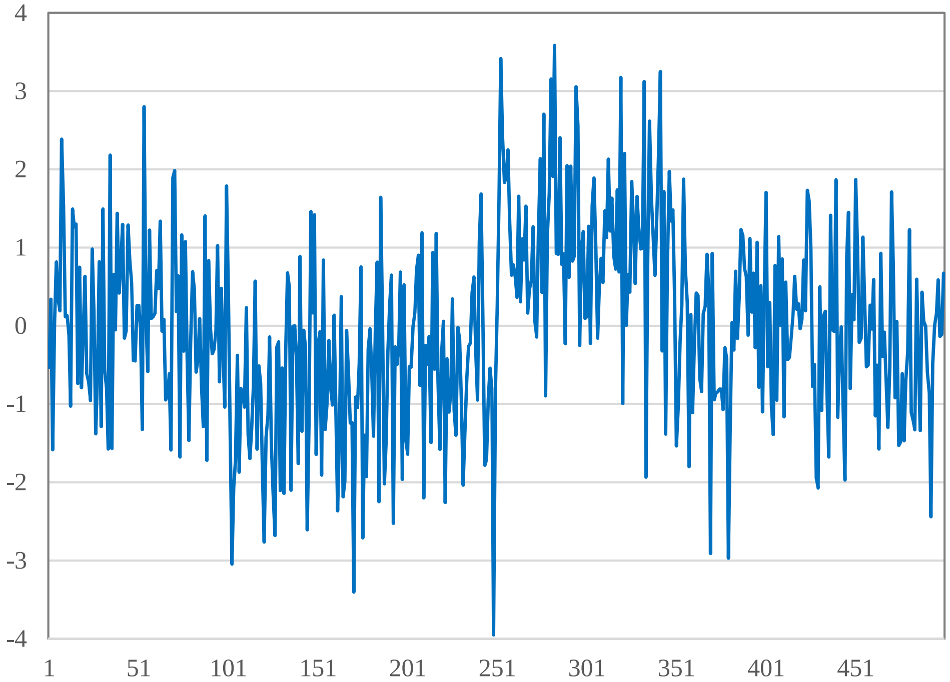
$$x_{n|N} = x_{n|n} + A_n(x_{n+1|N} - x_{n+1|n}) \qquad A_n = V_{n|n} F_{n+1}^T V_{n+1|n}^{-1}$$

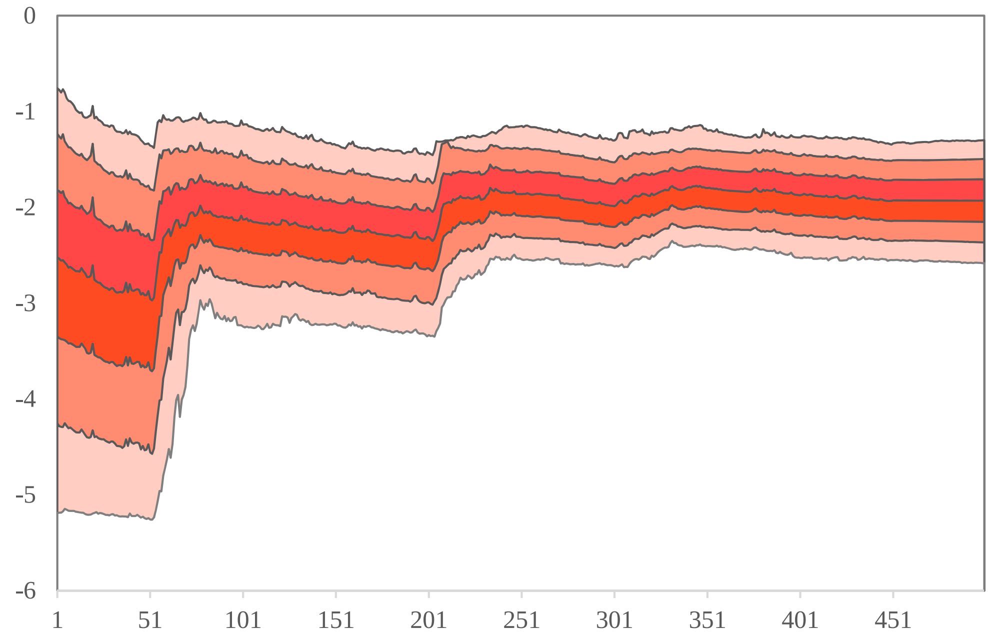
$$V_{n|N} = V_{n|n} + A_n(V_{n+1|N} - V_{n+1|n}) A_n^T$$

Partially Linear State Space Model

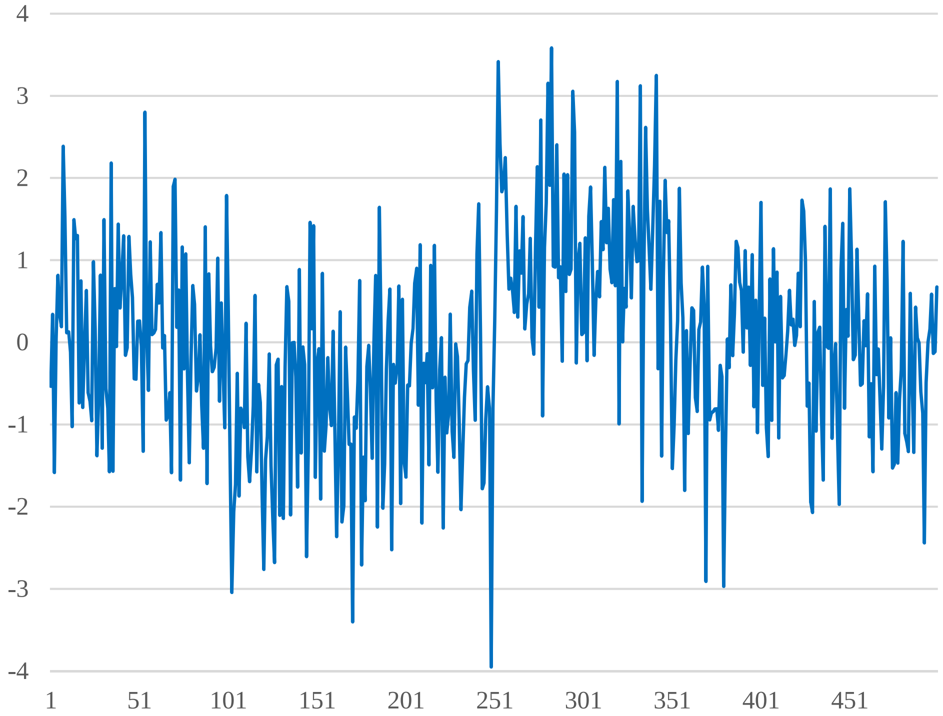
$$
\begin{aligned}
x_n &= x_{n-1} + v_n \qquad v_n \sim N(0, 10^{\theta_{n-1}}) \\
\theta_n &= \theta_{n-1} + u_n \qquad u_n \sim N(0, 10^{-4}) \\
y_n &= x_n + w_n \qquad\quad v_n \sim N(0, 1)
\end{aligned}
$$

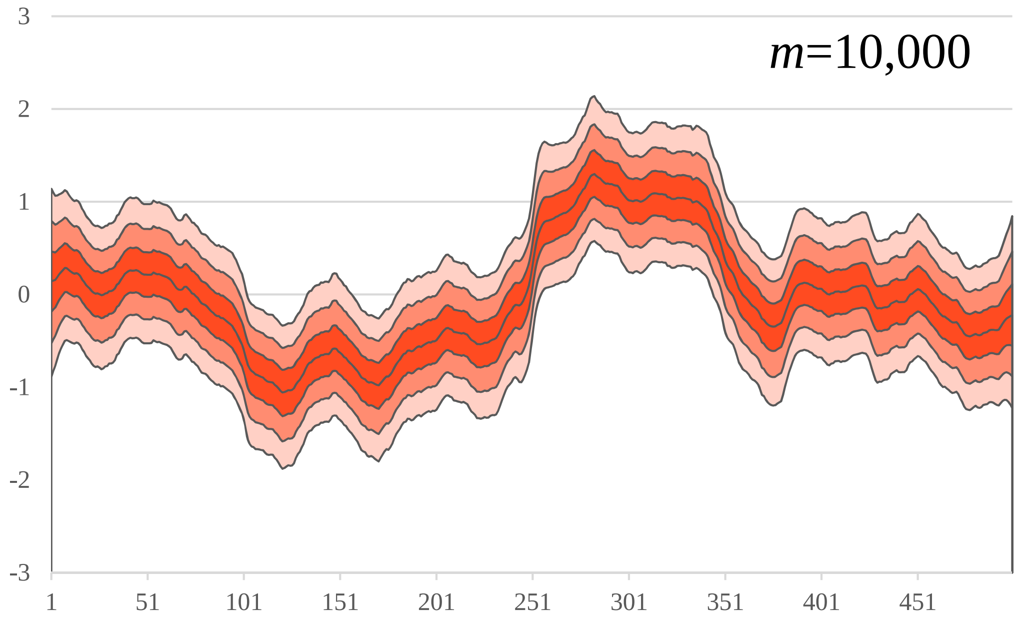


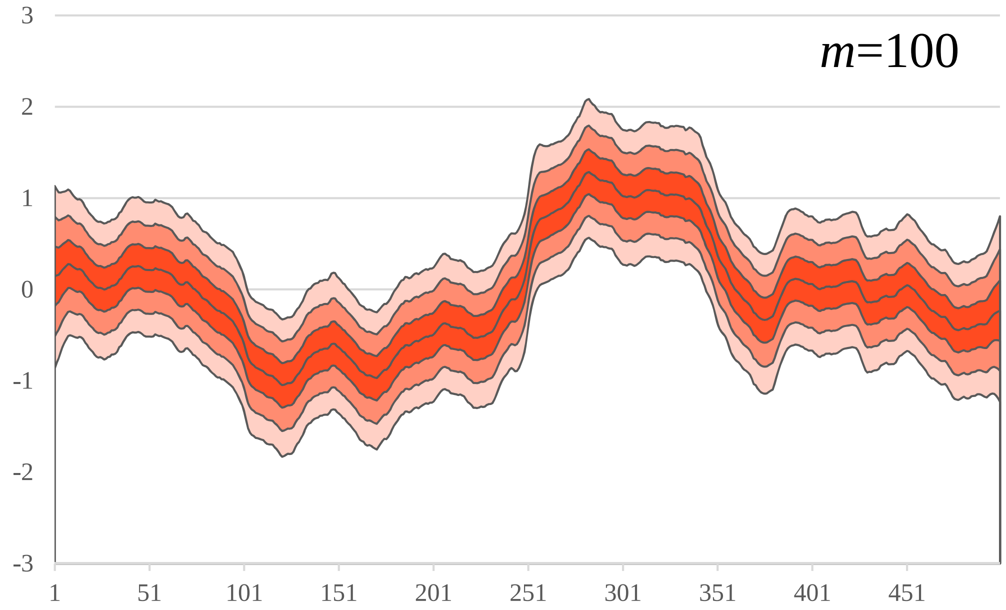




Fixed Parameter Case　$\theta_n = \theta_{n-1}$







- Randomでも等間隔でもよさそう
- 低次元の場合、m は少なくてもよい（例は2次元の場合）
- 高次元パラメータのときに直交配置のような方法が使えるか？

Multi-scale analysis of lead-lag relationships in high-frequency financial markets [1]

Takaki Hayashi

Keio University, Tokyo Metropolitan University, CREST JST

August 5, 2017

釧路 2017 経済統計キャンプ

[1] Joint work with Yuta Koike (Tokyo Metropolitan University)

Background

- **Lead-lag relationship**
  - Two time series are cross-correlated with each other at certain lags; "leader" and "lagger"
- Lead-lag relationships may occur perhaps because new information is absorbed into each security at different speeds
  - Across different assets
  - Across different trading venues
- **Ex.**: Stock index vs index futures (e.g. Kawaller et al., 1987)
  - A stock index consists of many individual stocks; it may be lagging behind the index futures
- Hoffmann, Rosenbaum, and Yoshida (2013) have proposed a model for lead-lag relationships in high-frequency financial data ("HRY model")

Background: HRY model

Hoffmann, Rosenbaum, and Yoshida (2013)

- $X^\nu = (X^\nu_t)_{t\in[-\delta,\infty)}$: the log-price process of the ν-th asset ($\nu = 1, 2$; $\delta > 0$)
- We suppose that $X = (X^1, X^2)$ is a 2-dimensional continuous Itô semimartingale defined on a stochastic basis $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[-\delta,\infty)}, P)$:

$$X_t = X_{-\delta} + \int_{-\delta}^{t} b_s ds + \int_{-\delta}^{t} \Sigma_s^{1/2} dW_s$$

  - W_s : a bivariate standard Wiener process
  - b_s: a bivariate càdlàg adapted process
  - Σ_s: 2×2-p.s.d. symmetric matrix-valued càdlàg adapted process

Background: HRY model

- X evolves on the interval $[0, T + \delta]$
- $0 \le t^\nu_1 < t^\nu_2 < \cdots < t^\nu_{n_\nu} \le T + \delta$: observation times for X^ν ($\nu = 1, 2$)
  - depend on a parameter $n \in \mathbb{N}$ and satisfy

$$\max_{i=0,1,\dots,n_\nu+1} (t^\nu_i - t^\nu_{i-1}) \to^p 0 \qquad (n \to \infty),$$

where $t^\nu_{-1} := 0$ and $t^\nu_{n_\nu+1} := T + \delta$

- The observation data $(Y^1_i)_{i=0}^{n_1}$, $(Y^2_j)_{j=0}^{n_2}$ of X^1 and X^2 are given by:

$$\begin{cases} Y^1_i = X^1_{t^1_i}, & Y^2_j = X^2_{t^2_j-\vartheta} \quad \text{if } \vartheta \ge 0, \\ Y^1_i = X^1_{t^1_i-|\vartheta|}, & Y^2_j = X^2_{t^2_j} \quad \text{if } \vartheta < 0. \end{cases}$$

Here, $\vartheta \in (-\delta, \delta)$ is the unknown lead-lag parameter

Background: HRY model

- Hoffmann et al. (2013) have introduced the following contrast function $U_n(\theta)$ to estimate the parameter ϑ:

$$U_n(\theta) = \begin{cases} \sum_{i,j:t_i^1 \le T} \Delta_i Y^1 \Delta_j Y^2 1_{\{(t_{i-1}^1, t_i^1] \cap (t_{j-1}^2 - \theta, t_j^2 - \theta] \neq \emptyset\}}, & \text{if } \theta \ge 0, \\ \sum_{i,j:t_j^2 \le T} \Delta_i Y^1 \Delta_j Y^2 1_{\{(t_{i-1}^1 + \theta, t_i^1 + \theta] \cap (t_{j-1}^2, t_j^2] \neq \emptyset\}}, & \text{if } \theta < 0, \end{cases}$$

  where $\Delta_i Y^\nu = Y_i^\nu - Y_{i-1}^\nu$ for $\nu = 1, 2$

- $U_n(\theta)$ is the (empirical) cross-covariance function between the returns of Y^1 and Y^2 at the lag θ computed by Hayashi and Yoshida (2005)'s method

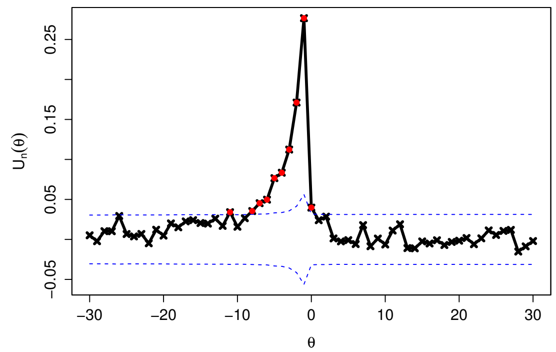
- Hoffmann et al. (2013) have shown that

$$\widehat{\theta}^{HRY} = \arg\max_{\theta \in \mathcal{G}_n} |U_n(\theta)|$$

  is a consistent estimator for ϑ under some regularity conditions while one appropriately takes the finite set $\mathcal{G}_n \subset (-\delta, \delta)$

Motivating example

S&P 500 cash index vs E-mini futures: HRY's $U_n(\theta)$



We depict $U_n(\theta)$ between the S&P500 index and the E-mini S&P500 futures on April 1, 2016 (black line). The horizontal line is in seconds. $\theta > 0$ means that the S&P500 index leads the E-mini futures. The blue dash lines are (pointwise) 99.99% confidence intervals under the null hypothesis $U_n(\theta) = 0$ for each θ, and the red points denote the values outside the confidence intervals (see the function `mllag` in R package **yuima** for details)

Issues of the HRY model

- In the setting of Hoffmann et al. (2013), there is only one lead-lag parameter and $U_n(\theta)$ should be like a scaled Dirac measure
  - However, in the above figure $U_n(\theta)$ is evidently asymmetric around $\widehat{\theta}^{HRY} = -1$ (sec.)
  - There are multiple θ's at which $U_n(\theta)$'s exceed the confidence bands
- To explain such a phenomenon, we propose a model which can "naturally" describe multiple lead-lag relationships

- Why do multiple lead-lag relationships occur?
  - "Heterogenous market hypothesis" (Müller et al., 1997): Market participants act with different time scales
  - Assets are correlated through various (latent) factors

Our approach

- We propose a model taking account of "heterogeneity" of the market
  - Modeling with multiple time scales $\Longrightarrow$ Wavelets!! (cf. Gençay et al., 2002)
- The existing literature on applications of wavelet to lead-lag analysis is based on *discrete-time* modeling (mainly established in Whitcher et al. (1999, 2000) and Serroukh and Walden (2000a,b))

- **Aim of this research**
  - Providing a modeling framework validating wavelet analysis for investigating lead-lag relationships with multiple time scales in a *continuos-time* setting
  - Proposing an estimation procedure for the lead-lag parameters

Lévy's construction revisited

- Lévy(-Ciesielski)'s construction of Brownian motion on $[0,1]$ (cf. Karatzas and Shreve, 1998):

$$B_t^\nu = \xi_0^\nu t + \sum_{j=0}^{\infty} \underbrace{\sum_{k=0}^{2^j-1} \xi_{jk}^\nu \int_0^t \psi_{jk}(s)ds,}_{B^\nu(j)_t} \qquad \nu = 1,2 \qquad (1)$$

  - $\xi_0^\nu, (\xi_{jk}^\nu)_{j,k=0}^\infty$: i.i.d. standard normal variables
  - $\psi_{jk}(s)$: the Haar functions

$$\psi_{jk}(s) = 2^{j/2}\psi(2^j s - k), \qquad \psi(s) = \begin{cases} 1 & 0 \le s < \frac{1}{2}, \\ -1 & \frac{1}{2} \le s < 1, \\ 0 & \text{otherwise.} \end{cases}$$

The construction turns out too restrictive for out purpose.
We generalize it as follows:

Proposition 1 (Dyadic wavelet decomposition of Brownian motion)

Let $B = (B_t)_{t\in\mathbb{R}}$ be a two-sided Brownian motion. Suppose that functions $\widetilde{\phi}, \widetilde{\psi} \in L^2(\mathbb{R})$ satisfy

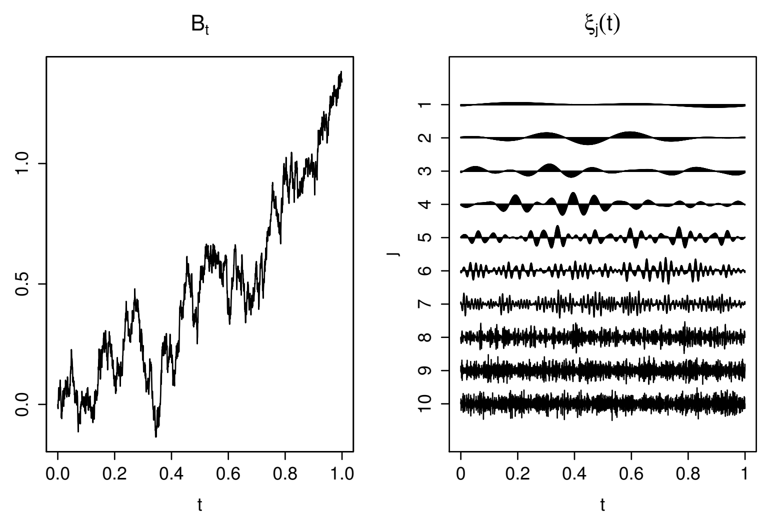
$$|(\mathcal{F}\widetilde{\phi})(\lambda)|^2 + \sum_{j=0}^{\infty} |(\mathcal{F}\widetilde{\psi})(2^{-j}\lambda)|^2 = 1$$

for any $\lambda \in \mathbb{R}$, where $\mathcal{F}$ denotes the Fourier transform. Then we have

$$B_t = \int_0^t \widetilde{\xi} * \widetilde{\phi}(s)ds + \sum_{j=0}^{\infty} 2^j \int_0^t \widetilde{\xi}_j * \widetilde{\psi}_j(s)ds \qquad \text{in } L^2(\mathbb{R})$$

for any $t \in \mathbb{R}$, where $*$ denotes the convolution, $\widetilde{\psi}_j := 2^{j/2}\widetilde{\psi}(2^j\cdot)$ and

$$\widetilde{\xi}(s) = \int_{-\infty}^{\infty} \widetilde{\phi}(s-u)dB_u, \qquad \widetilde{\xi}_j(s) = \int_{-\infty}^{\infty} \widetilde{\psi}_j(s-u)dB_u$$



- We use the Littlewood-Paley wavelets $\widetilde{\phi} = \phi^{LP}$ and $\widetilde{\psi} = \psi^{LP}$ whose Fourier transforms are given by

$$(\mathcal{F}\phi^{LP})(\lambda) = 1_{[-\pi,\pi]}(\lambda),$$
$$(\mathcal{F}\psi^{LP})(\lambda) = 1_{[-2\pi,\pi)\cup(\pi,2\pi]}(\lambda),$$

which make $\widetilde{\xi}_j(s)$'s independent between different levels (consequence of the Parseval identity)

Lévy's construction with dyadic wavelet transform

- For each $\nu = 1, 2$, let $B^\nu = (B^\nu_t)_{t\in\mathbb{R}}$ be a two-sided Brownian motion and set
$$\widetilde{\xi}^\nu_j(s) = \int_{-\infty}^{\infty} \psi^{LP}_j(s-u)dB^\nu_u, \qquad s \in \mathbb{R}$$
- Now we evaluate the cross-covariances between $\widetilde{\xi}^1_j(s)$ and $\widetilde{\xi}^2_j(s)$:
$$\rho_j(\theta) = \mathrm{Cov}\left[\widetilde{\xi}^1_j(s), \widetilde{\xi}^2_j(s+\theta)\right], \qquad \theta \in \mathbb{R},$$
provided their joint stationarity

Lévy's construction with dyadic wavelet transform

Proposition 2

Suppose that $R_j \in [-1, 1]$ and $\theta_j \in \mathbb{R}$ for $j = 0, 1, \dots$. Then, there exists a bivariate Gaussian process $B_t = (B^1_t, B^2_t)$ $(t \in \mathbb{R})$ with stationary increments such that

(i) *both B^1 and B^2 are two-sided Brownian motions,*

(ii) *the cross-spectral density of B is given by*
$$f(\lambda) = \sum_{j=0}^{\infty} R_j e^{-\sqrt{-1}\theta_j\lambda} 1_{\Lambda_j}(\lambda), \qquad \lambda \in \mathbb{R}$$

Statistical model

- For each $\nu = 1, 2$, the log price process $X^\nu = (X^\nu_t)_{t\geq 0}$ of the ν-th asset is given by
$$X^\nu_t = X^0_t + \int_0^t \sigma^\nu_s dB^\nu_s, \qquad t \geq 0,$$
where $(\sigma^\nu_t)_{t\geq 0}$ is a càdlàg process adapted to the filtration $(\mathcal{F}^\nu_t)$ such that the process (B^ν_t) is an $(\mathcal{F}^\nu_t)$-Brownian motion.
- X is observed at discrete times over the interval $[0, T+\delta]$
- $0 \leq t^\nu_1 < t^\nu_2 < \cdots < t^\nu_{n_\nu} \leq T + \delta$: observation times for X^ν
  - t^ν_i's are random variables independent of X
  - $r_N := \max_{\nu=1,2} \max_{i=0,1,\dots,n_\nu+1}(t^\nu_i - t^\nu_{i-1}) \to^p 0$
- Observations $(X^1_{t^1_i})^{n_1}_{i=0}$ and $(X^2_{t^2_i})^{n_2}_{i=0}$ are generally *non-synchronous*

Statistical model

- Given the finest resolution level N, we model the cross-spectral density of $B_t = (B^1_t, B^2_t)$ as
$$f_N(\lambda) = \sum_{j=0}^{N} R_{(j)} e^{-\sqrt{-1}\theta_{(j)}\lambda} 1_{\Lambda_j}(\lambda) = \sum_{j=1}^{N+1} R_j e^{-\sqrt{-1}\theta_j\lambda} 1_{\Lambda_{(j)}}(\lambda),$$
where we set $(j) = N - j + 1$
  - We omit components finer than τ_N from the model because they are not identifiable
  - We relabel indices of the parameters R_j and θ_j so that the finest resolution τ_N corresponds to $j = 1$
    - Convenient for the formulation of asymptotic results when $N \to \infty$

Estimation

- **Our aim**: To construct estimators for θ_j's based on discrete observation data
- To explain the idea behind the construction of our estimator, we focus on the case of $\sigma_s^\nu \equiv 1$ for $\nu = 1, 2$
- Note that θ_j is the unique maximizer of $|\rho_{(j)}(\theta)|$ as long as $R_j \neq 0$
- Motivated by this,
  - we begin by constructing a sensible estimator $\widehat{\rho}_{(j)}(\theta)$ for $\rho_{(j)}(\theta)$, and
  - we construct the estimator for θ_j as a maximizer of $|\widehat{\rho}_{(j)}(\theta)|$

  as in Hoffmann et al. (2013)

Estimation

- Let $U^N(\theta)$ be the inverse Fourier transform of the cross-spectral density $f_N(\lambda)$:
$$U^N(\theta) = (\mathcal{F}^{-1} f_N)(\theta)$$
  - $U^N(\theta)$ is the cross-covariance function
- Then we have
$$\rho_{(j)}(\theta) = 2^{-\frac{(j)}{2}} (U^N * \psi_{(j)}^{LP})(\theta) = \int_{-\infty}^{\infty} U^N(\theta - s)\psi^{LP}(2^{(j)}s)ds$$
- The last form leads to the following cross-covariance estimator at level j:

Estimation

Proposed Estimator

$$\widehat{\rho}_{(j)}(\theta) = \sum_{l=-L_j+1}^{L_j-1} \widehat{U}^N(\theta - l\tau_N)\Psi_j(l), \qquad j \geq 1$$

- $\widehat{U}^N(\theta)$ is the HRY cross-covariance estimator
- $\Psi_j(l)$ is the *autocorrelation wavelets* (cf. Nason et al., 2000)
$$\Psi_j(l) = \sum_{p=0}^{L_j-1-|l|} h_{j,p}h_{j,p+|l|}, \qquad l = 0, \pm 1, \dots, \pm(L_j - 1) \qquad (2)$$
  where $h_{j,0}, h_{j,1}, \dots, h_{j,L_j-1}$ be Daubechies' wavelet filters with length L at the level j $(L_j = (2^j - 1)(L - 1) + 1)$

The lag parameter θ_j is estimated by :

$$\widehat{\theta}_j = \arg\max_{\theta \in \mathcal{G}_j^N} |\widehat{\rho}_{(j)}(\theta)|, \qquad j \geq 1$$

where $\mathcal{G}_j^N = \{l\tau_N : l \in \mathbb{Z}, |l\tau_N| < \delta - L_j\tau_N\}$

Asymptotic theory

Theorem 1

Suppose that $L \to \infty$ and $L\tau_N^\kappa \to 0$ as $N \to \infty$ for any $\kappa > 0$. Under certain regularity conditions, the following statements hold true:

(a) *If a sequence $v_N > 0$ satisfies $L^{-\frac{1}{2}}\tau_N^{-1}v_N \to \infty$ as $N \to \infty$, then*
$$\max_{\theta \in \mathcal{G}_j^N : |\theta - \theta_j| \geq v_N} |\widehat{\rho}_{(j)}(\theta)| \to^p 0$$

(b) *Let (ϑ_N) be a sequence of real numbers such that $\vartheta_N \in \mathcal{G}_j^N$ and $\tau_N^{-1}(\vartheta_N - \theta_j) \to b$ as $N \to \infty$ for some $b \in \mathbb{R}$. Then*
$$\widehat{\rho}_{(j)}(\vartheta_N) \to^p \frac{R_j}{2\pi} \int_{\Lambda_{-j}} F(\lambda)\cos(b\lambda)d\lambda$$

Asymptotic theory

- From Theorem 1, we can derive *consistency* of the estimator $\widehat{\theta}_j$:

Theorem 2

Suppose that $L \to \infty$ and $L^2 \tau_N^{\kappa} \to 0$ as $N \to \infty$ for any $\kappa > 0$. Under the conditions of Theorem 1, if a sequence $v_N > 0$ satisfies $L^{-\frac{1}{2}} \tau_N^{-1} v_N \to \infty$ as $N \to \infty$, then

$$v_N^{-1}(\widehat{\theta}_j - \theta_j) \to^p 0$$

as $N \to \infty$, provided $R_j \neq 0$. In particular, we have $\widehat{\theta}_j \to^p \theta_j$ as $N \to \infty$.

Empirical application

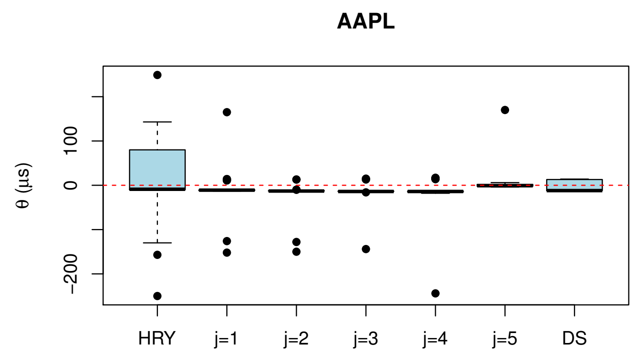
US stock market: quote data

- Cross-market, single-asset analysis
- Time resolution: 1 micro-second ($\tau_N = 1\mu$s)
- Venues: NASDAQ and BATS
- Micro price
- Stocks: AAPL, CSCO, INTC, MSFT
- Source: Daily TAQ Database
- Period: All the trading days in August, 2015
- Between 9:45 and 15:45 (the first and the last 15 min are discarded)
- Search grid: $\mathcal{G}_j^N = \{-250\mu s, -249\mu s, \ldots, 249\mu s, 250\mu s\} (=: \mathcal{G}^N)$
- Wavelet filter: D(20)

Table 1: The average daily numbers of observations

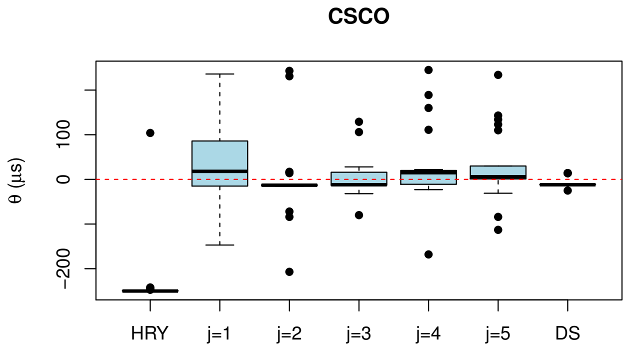
	AAPL	CSCO	INTC	MSFT
NASDAQ (Q)	731,331	331,566	342,239	592,191
BATS (Z)	662,161	169,867	199,156	332,787

Figure 1: Boxplots of the estimates for AAPL: NASDAQ vs BATS



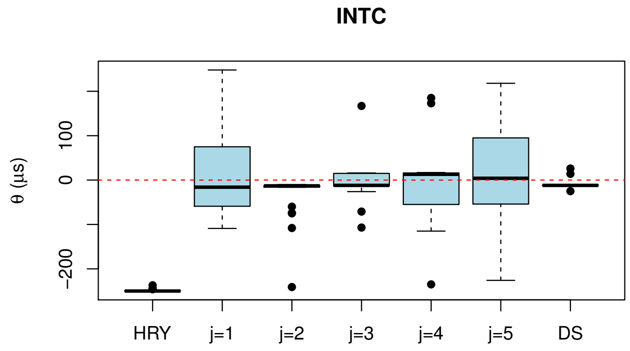
The positive value indicates that the NASDAQ leads the BATS.

Figure 2: Boxplots of the estimates for CSCO: NASDAQ vs BATS



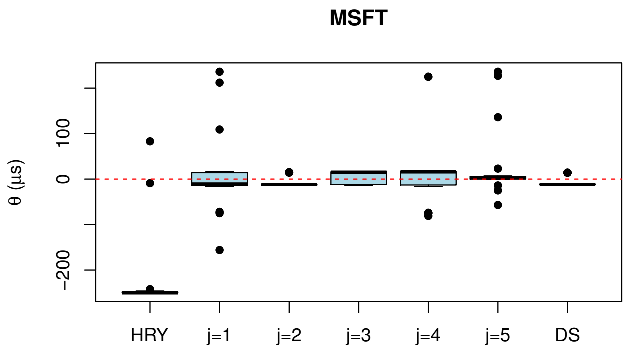
The positive value indicates that the NASDAQ leads the BATS.

Figure 3: Boxplots of the estimates for INTC: NASDAQ vs BATS



The positive value indicates that the NASDAQ leads the BATS.

Figure 4: Boxplots of the estimates for MSFT: NASDAQ vs BATS



The positive value indicates that the NASDAQ leads the BATS.

Table 2: Median and IQR of the estimates

	HRY	$j=1$	$j=2$	$j=3$	$j=4$	$j=5$	DS
AAPL	−9	−11	−13	−14	−14	−1	−12
	(90)	(2)	(1)	(1)	(2)	(4)	(25)
CSCO	−250	18	−13	−12	15	6	−12
	(1)	(101)	(2)	(29)	(31)	(29)	(1)
INTC	−250	−16	−13	−12	13	4	−12
	(1)	(134)	(4)	(28)	(71)	(149)	(1)
MSFT	−250	−11	−12	15	16	3	−12
	(2)	(28)	(1)	(28)	(30)	(6)	(1)

Note. This table reports the sample medians and the IQRs (in parentheses) of the estimates over the whole sample period. The reported values are in micro-seconds.

Conclusions and Future work

- We have established an explicit connection btw wavelet and Lévy-Ciesielski type construction of Brownian motion from a statistical viewpoint
- Based on the model, we have proposed a new estimation method for multi-scale analysis of lead-lag relationships btw two high-frequency time series
- Working papers
  - "Wavelet-based methods for high-frequency lead-lag analysis," arXiv:1612.01232
  - "Multi-scale analysis of lead-lag relationships in high-frequency financial markets," arXiv:1708.03992
- Future work
  - More general model (e.g., time-varying/ stochastic correlation)
  - Asymptotic distribution theory
  - Microstructure noise contamination

Acknowledgement

- Grant-in-Aid for Scientific Research (C), Japan Society for the Promotion of Science (JSPS), 2016–.
- CREST Project, Graduate School of Mathematical Sciences, University of Tokyo, 2014–.

Multi-scale analysis of lead-lag relationships in high-frequency financial markets

Takaki Hayashi[*†‡] Yuta Koike[†§‡]

July 29, 2017

Abstract

We propose a novel estimation procedure for scale-by-scale lead-lag relationships of financial assets observed at a high-frequency in a non-synchronous manner. The proposed estimation procedure does not require any interpolation processing of the original data and is applicable to quite fine resolution data. The validity of the proposed estimators is shown under the continuous-time framework developed in our previous work [15]. An empirical application shows promising results of the proposed approach.

Keywords: High-frequency data; Lead-lag relationship; Multi-scale analysis; Non-synchronous data; Stochastic volatility; Wavelet.

1 Introduction

A financial market consists of various participants. They have different perspectives on the markets and risks with different constraints and different amount of information. In Müller *et al.* [25] it is argued that such various differences are spelled out to a sensitivity to different time scales. Hence they cause a multi-scale structure of the financial market.

The aim of this paper is to investigate such a multi-scale structure in high-frequency financial markets. In this paper we especially focus on lead-lag relationships between financial assets, which is known as a prominent stylized fact of high-frequency financial data (see e.g. [3, 8, 21, 29]). Multi-scale analysis of high-frequency financial data has been performed in a number of articles such as [2, 10, 14, 24, 30]. These articles mainly focus on volatilities of assets. There are little work which conducts multi-scale analysis of lead-lag relationships for high-frequency financial data. One exception is Hafner [11] which has examined multi-scale structures of the lead-lag relationships between the returns, durations and volumes of high-frequency transaction data of the IBM stock. In the meantime, these existing studies mainly focus on empirical applications and are theoretically based on classical discrete time series observed in a long time horizon. However, analysis of high-frequency financial data typically focuses on a short time horizon such as

[*]Keio University, Graduate School of Business Administration, 4-1-1 Hiyoshi, Yokohama 223-8526, Japan

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

[‡]CREST, Japan Science and Technology Agency

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

one day, so it is unclear whether one may apply such a theory to high-frequency financial data. On the other hand, it is nowadays well-recognized that continuous-time modeling provides a powerful tool to analyze high-frequency data observed in a short horizon (cf. Aït-Sahalia and Jacod [1]). Motivated by these reasons, in [15] the authors have developed a relevant continuous-time framework for multi-scale analysis of lead-lag relationships in high-frequency data by introducing two Brownian motions B^1 and B^2 with a scale-by-scale correlation structure. More precisely, they have shown that, for any $R_j \in [-1,1]$ and $\theta_j \in \mathbb{R}$ ($j = 0, 1, \dots$), there exists a bivariate Gaussian process $B_t = (B_t^1, B_t^2)$ ($t \in \mathbb{R}$) with stationary increments such that

(I) both B^1 and B^2 are two-sided Brownian motions,

(II) the cross-spectral density of B is given by

$$f(\lambda) = \sum_{j=0}^{\infty} R_j e^{-\sqrt{-1}\theta_j \lambda} 1_{\Lambda_j}(\lambda), \qquad \lambda \in \mathbb{R}, \tag{1}$$

where $\Lambda_j = [-2^j\pi, -2^{j-1}\pi) \cup (2^{j-1}\pi, 2^j\pi]$ for every $j \in \mathbb{Z}$.

The frequency band Λ_j corresponds to the time scale between 2^{-j} and 2^{-j+1} in the time domain. Also, note that, if $W_t = (W_t^1, W_t^2)$ ($t \in \mathbb{R}$) is a two-sided bivariate Brownian motion with correlation R, for $\theta \in \mathbb{R}$ the process $(W_t^1, W_{t-\theta}^2)$ ($t \in \mathbb{R}$) has the cross-spectral density $Re^{-\sqrt{-1}\theta\lambda}$ ($\lambda \in \mathbb{R}$). Therefore, we can consider that B^1 and B^2 have a lead-lag relationship with the time-lag θ_j in the time scale between 2^{-j} and 2^{-j+1}. Hence, under this model we can understand the multi-scale structure of the lead-lag relationships by estimating the parameters θ_j from observation data.

The main contribution of this paper is to develop a novel estimation procedure for the parameters θ_j based on non-synchronous observations of (volatility-modulated versions of) B^1 and B^2. Although such a procedure has already been proposed in [15], their procedure contains data interpolation onto the grid with the finest observable resolution and is computationally challenging if we are interested in data recorded with sub-second precision. In fact, in a situation where the finest observable resolution is one micro-second we need to store one million observations per one second. Our novel procedure is free from any interpolation processing of the original data and is applicable to data with record times of sub-second precision. A numerical experiment also shows that our new estimators have a superior performance to the interpolation-based estimator when the sampling times exhibit high degree of the non-synchronicity.

The rest of the paper is organized as follows. In Section 2 we present the theoretical setting considered in this paper in details. Our new estimation procedure is described in Section 3. We develop an asymptotic theory associated with the proposed estimators in Section 4. In Section 5 we assess the practical performance of the proposed estimators by a Monte Carlo study, and in Section 6 we apply our procedure to a set of market data. Section 7 concludes the paper. All the proofs are collected in Section 8.

2 Setting

Any high-frequency financial data have the finest observable resolution. We let it correspond to $\tau_N := 2^{-N-1}$ for some $N \in \mathbb{N}$. To derive our theoretical results, we consider the asymptotic theory such that N tends to infinity. Namely, we focus on situations where the finest observable resolution is very fine.

As presented in the Introduction, our theoretical framework is based on a bivariate Gaussian process $B_t = (B_t^1, B_t^2)$ ($t \in \mathbb{R}$) with stationary increments satisfying properties (I)–(II). Since we are mainly interested in the lead-lag relationships at scales close to the finest observation resolution, it is convenient to "relabel" indices of the parameters R_j and θ_j in (1) so that the finest resolution τ_N corresponds to the level $j = 1$ while we consider the asymptotic theory such that N tends to infinity. For this reason, as in [15] we replace property (II) with the following one: The cross-spectral density of B is given by

$$f_N(\lambda) = \sum_{j=1}^{N+1} R_j e^{-\sqrt{-1}\theta_j \lambda} 1_{\Lambda_{(j)}}(\lambda), \qquad \lambda \in \mathbb{R}, \tag{2}$$

where $(j) = N - j + 1$. We also assume that $\theta_j \in (-\delta, \delta)$ for every j with some $\delta > 0$.

Now, for each $\nu = 1, 2$, we consider the log price process $X^\nu = (X_t^\nu)_{t \geq 0}$ of the ν-th asset given by

$$X_t^\nu = X_0^\nu + \int_0^t \sigma_s^\nu dB_s^\nu, \qquad t \geq 0, \tag{3}$$

where $(\sigma_t^\nu)_{t\geq 0}$ is a càdlàg process adapted to the filtration $(\mathcal{F}_t^\nu)$ such that the process (B_t^ν) is an $(\mathcal{F}_t^\nu)$-Brownian motion. We observe the process X^ν on the interval $[0, T+\delta]$ at the sampling times $0 \leq t_0^\nu < t_1^\nu < \cdots < t_{n_\nu}^\nu \leq T + \delta$. The sampling times $(t_i^1)_{i=0}^{n_1}$ and $(t_i^2)_{i=0}^{n_2}$ are random variables which are independent of (X^1, X^2) and implicitly depend on N such that

$$r_N := \max_{\nu=1,2} \max_{i=0,1,\dots,n_\nu+1} (t_i^\nu - t_{i-1}^\nu) \to^p 0$$

as $N \to \infty$, where we set $t_{-1}^\nu := 0$ and $t_{n_\nu+1}^\nu := T + \delta$ for each $\nu = 1, 2$.

3 Construction of the estimators

Our aim is to estimate the parameters θ_j for each j based on discrete observation data $(X_{t_i^1}^1)_{i=0}^{n_1}$ and $(X_{t_i^2}^2)_{i=0}^{n_2}$. We begin by introducing some notation. For each $\nu = 1, 2$, we associate the observation times $(t_i^\nu)_{i=0}^{n_\nu}$ with the collection of intervals $\Pi_N^\nu = \{(t_{i-1}^\nu, t_i^\nu] : i = 1, \dots, n_\nu\}$. We will systematically employ the notation I (resp. J) for an element of Π_N^1 (resp. Π_N^2).

For an interval $H \subset [0, \infty)$, we set $\overline{H} = \sup H$, $\underline{H} = \inf H$, $|H| = \overline{H} - \underline{H}$. In addition, we set $V(H) = V_{\overline{H}} - V_{\underline{H}}$ for a a stochastic process $(V_t)_{t\geq 0}$, and $H_\theta = H + \theta$ for a real number θ.

Now we explain how to construct our estimators. To explain the idea behind the construction, we focus on the case of $\sigma_s^\nu \equiv 1$ for $\nu = 1, 2$. The parameter θ_j is the unique maximizer of the scale-by-scale cross-covariance function $\rho_{(j)}(\theta)$ between B^1 and B^2, which is defined by

$$\rho_{(j)}(\theta) = E\left[\left(\int_{-\infty}^{\infty} \psi_{(j)}^{LP}(s-u) dB_s^1\right)\left(\int_{-\infty}^{\infty} \psi_{(j)}^{LP}(s-u-\theta) dB_s^2\right)\right], \qquad \theta \in \mathbb{R},$$

where $\psi_{(j)}^{LP}(s) = 2^{(j)/2}\psi^{LP}(2^{(j)}s)$ and ψ^{LP} denotes the Littlewood-Paley wavelet:

$$\psi^{LP}(s) = (\pi s)^{-1}(\sin(2\pi s) - \sin(\pi s))$$

(see Sections 2.2–2.3 of [15] for details). Motivated by this fact, we first construct a sensible estimator for $\rho_{(j)}(\theta)$, and then construct the estimator for θ_j as a maximizer of $|\widehat{\rho}_{(j)}(\theta)|$ as in [19].

The idea behind the construction of the estimator for $\rho_{(j)}(\theta)$ is as follows. Let $U^N(\theta)$ be the inverse Fourier transform of $f_N(\lambda)$. Then we have

$$\rho_{(j)}(\theta) = 2^{-\frac{(j)}{2}}(U^N * \psi^{LP}_{(j)})(\theta) = \int_{-\infty}^{\infty} U^N(\theta - s)\psi^{LP}(2^{(j)}s)ds$$

by the convolution theorem. This suggests us to consider the following estimator for $\rho_{(j)}(\theta)$:

$$\widehat{\rho}_{(j)}(\theta) = \sum_{l=-L_j+1}^{L_j-1} \widehat{U}^N(\theta - l\tau_J)\Psi_j(l),$$

where $\widehat{U}^N(\theta)$ is an estimator for $U^N(\theta)$ and $\Psi_j(l)$ is an approximation of $2^{(j)} \cdot \psi^{LP}(2^{(j)}l\tau_N)\tau_N$ (it turns out that the scaling $2^{(j)}$ is necessary due to discretization), which are explicitly defined in the following. Since $U^N(\theta)$ may be regarded as the "cross-covariance function between dB^1 and dB^2", we adopt the following estimator introduced in Hoffmann *et al.* [19] as $\widehat{U}^N(\theta)$:

$$\widehat{U}^N(\theta) = \begin{cases} \sum_{I\in\Pi^1_N, J\in\Pi^2_N:\overline{I}\leq T} X^1(I)X^2(J)K(I, J_{-\theta}) & \text{if } \theta \geq 0, \\ \sum_{I\in\Pi^1_N, J\in\Pi^2_N:\overline{J}\leq T} X^1(I)X^2(J)K(I_\theta, J) & \text{if } \theta < 0, \end{cases}$$

where we set $K(I, J) = 1_{\{I\cap J\neq\emptyset\}}$ for two intervals I and J. This $\widehat{U}^N(\theta)$ can be regarded as the empirical cross-covariance estimator between the returns of X^1 and X^2 at the lag θ computed by Hayashi and Yoshida [16]'s method to handle the non-synchronous sampling times. In the meantime, the Fourier inversion formula yields

$$2^{(j)} \cdot \psi^{LP}(2^{(j)}l\tau_N)\tau_N = \tau_N \int_{-\infty}^{\infty} e^{\sqrt{-1}l\tau_N\lambda} 1_{\Lambda_{(j)}}(\lambda)d\lambda = \int_{-\pi}^{\pi} e^{\sqrt{-1}l\lambda} 1_{\Lambda_{-j}}(\lambda)d\lambda,$$

hence the transfer function of $(2^{(j)} \cdot \psi^{LP}(2^{(j)}l\tau_N)\tau_N)_{l\in\mathbb{Z}}$ is $1_{\Lambda_{-j}}(\lambda)$. In particular, $\Psi_j(l)$ well approximates $2^{(j)} \cdot \psi^{LP}(2^{(j)}l\tau_N)\tau_N$ if the transfer function of $(\Psi_j(l))_{l=-L_j+1}^{L_j-1}$ well approximates $1_{\Lambda_{-j}}(\lambda)$. We construct such a sequence $(\Psi_j(l))_{l=-L_j+1}^{L_j-1}$ from Daubechies' wavelet filter as follows (we refer to Chapters 6–8 of [6], Section 4.8 of [28] and Section 3.4.5 of [32] for details about Daubechies' wavelet filter). Let $(h_p)_{p=0}^{L-1}$ be Daubechies' wavelet filter of (even) length L whose power transfer function $H_L(\lambda) = |\sum_{p=0}^{L-1} h_p e^{-\sqrt{-1}\lambda p}|^2$ is given by

$$H_L(\lambda) = 2\sin^L(\lambda/2) \sum_{p=0}^{L/2-1} \binom{L/2-1+p}{p} \cos^{2p}(\lambda/2), \qquad \lambda \in \mathbb{R}.$$

The associated scaling filter[1] $(g_p)_{p=0}^{L-1}$ is defined via the quadrature mirror relationship as $g_p = (-1)^{p+1}h_{L-p-1}$, $p = 0, 1, \dots, L-1$, hence its power transfer function $G_L(\lambda) = |\sum_{p=0}^{L-1} g_p e^{-\sqrt{-1}\lambda p}|^2$ satisfies $G_L(\lambda) = H_L(\lambda - \pi)$. Then, for every j we construct the associated level j wavelet filter $(h_{j,p})_{p=0}^{L_j-1}$ recursively by $h_{1,p} = h_p$ for $p = 0, 1, \dots, L_1 - 1$ and $h_{j,p} = \sum_{q=0}^{L_{j-1}-1} g_{p-2q}h_{j-1,q}$ for $p = 0, 1, \dots, L_j - 1$, where $L_j = (2^j - 1)(L - 1) + 1$ and $g_p = 0$ for $p \notin \{0, 1, \dots, L-1\}$. Now we define the sequence $(\Psi_j(l))_{l=-L_j+1}^{L_j-1}$

[1]We use the notation that (h_p) denotes the wavelet filter and (g_p) denotes the scaling filter following [28]. Note that the reverse notation is often used in the literature.

by

$$\Psi_j(l) = \sum_{p=0}^{L_j-1-|l|} h_{j,p}h_{j,p+|l|}, \qquad l = 0, \pm 1, \dots, \pm(L_j - 1).$$

These quantities are identical to the *autocorrelation wavelets* from Nason *et al.* [26] (see Definition 3 from [26]). The transfer function $H_{j,L}(\lambda) = \sum_{l=-L_j+1}^{L_j-1} \Psi_j(l)e^{-\sqrt{-1}l\lambda}$ of $(\Psi_j(l))_{l=-L_j+1}^{L_j-1}$ is given by

$$H_{j,L}(\lambda) = H_L(2^{j-1}\lambda)\prod_{i=0}^{j-2} G_L(2^i\lambda), \qquad \lambda \in \mathbb{R}$$

(see Eq.(28) from [26]). Therefore, $H_{j,L}(\lambda)$ well approximates $1_{\Lambda_{-j}}(\lambda)$ as $L \to \infty$ by Theorem 1 from Lai [22] and thus $\Psi_j(l)$ may be used an approximation of $2^{(j)} \cdot \psi^{LP}(2^{(j)}l\tau_N)\tau_N$.

Finally, for every $j \in \mathbb{N}$ we define the estimator $\widehat{\theta}_j$ for θ_j as a solution of the following equation:

$$\left|\widehat{\rho}_{(j)}(\widehat{\theta}_j)\right| = \max_{\theta \in \mathcal{G}^N} \left|\widehat{\rho}_{(j)}(\theta)\right|.$$

Here, we maximize the function $\widehat{\rho}_{(j)}(\theta)$ regarding θ over the finite grid

$$\mathcal{G}^N = \{l\tau_N : l \in \mathbb{Z}, |l| \le \Gamma_N\}$$

with some positive integer Γ_N as in [19].

Remark 1. Given the length L of Daubechies' wavelet filter, we still have several options of $(h_p)_{p=0}^{L-1}$ such as the *external phase wavelet* and the *least asymmetric wavelet* (cf. Section 4.8 of [28]). However, all of them have the same power transfer function $H_L(\lambda)$ by definition, hence $(\Psi_j(l))_{l=-L_j+1}^{L_j-1}$ only depends on the length L of Daubechies' wavelet filters.

4 Asymptotic theory

For a function $f \in L^1(\mathbb{R})$, we denote by $\mathcal{F}f$ the Fourier transform of f:

$$(\mathcal{F}f)(\lambda) = \int_{-\infty}^{\infty} f(t)e^{-\sqrt{-1}\lambda t}dt, \qquad \lambda \in \mathbb{R}.$$

We impose the following conditions to derive our asymptotic results.

Assumption 1. For every $\nu = 1, 2$, the paths of σ^ν are almost surely γ-Hölder continuous for some $\gamma > 0$

Assumption 2. (i) $r_N = O_p(\tau_N^\xi)$ as $N \to \infty$ for any $\xi \in (0, 1)$.
(ii) There are constants $\alpha > 1$, $\beta \in (0, 1)$, $Q > 1$ and an absolutely continuous real-valued function D on $[-\pi, \pi]$ such that

$$\tau_m \sum_{k=0}^{\lceil T\tau_m^{-1}\rceil-1} \int_{-\pi}^{\pi} E\left[\left|D_k^N(\lambda, \theta_N) - D(\lambda)\right|^Q\right] d\lambda = O(\tau_N^\alpha)$$

as $N \to \infty$ for any sequence (θ_N) of real numbers satisfying $\theta_N \in \mathcal{G}^N$ for every N, where $m = \lceil \beta N \rceil$,

$$D_k^N(\lambda, \theta) = \begin{cases} \frac{1}{2\pi\tau_m\tau_N} \sum_{I,J:\underline{I} \in I_m(k)} (\mathcal{F}1_I)(\lambda/\tau_N)(\mathcal{F}1_{J_{-\theta}})(-\lambda/\tau_N)K(I, J_{-\theta}) & \text{if } \theta \ge 0, \\ \frac{1}{2\pi\tau_m\tau_N} \sum_{I,J:\underline{J} \in I_m(k)} (\mathcal{F}1_{I_\theta})(\lambda/\tau_N)(\mathcal{F}1_J)(-\lambda/\tau_N)K(I_\theta, J) & \text{if } \theta < 0 \end{cases}$$

and $I_m(k) = [kT\tau_m, (k+1)T\tau_m)$. Moreover, $D(\lambda) > 0$ for almost all $\lambda \in [-\pi, \pi]$ and $D' \in L^\infty(-\pi, \pi)$.

The simplest situation where Assumption 2 is satisfied is the equidistant and synchronous sampling case such that $t_i^1 = t_i^2 = i\tau_N$ for every i. In this case one can easily see that

$$D_k^N(\lambda, \theta) = \frac{1}{2\pi} \left| \frac{e^{-\sqrt{-1}\lambda} - 1}{\lambda} \right|^2$$

for any $\theta \in \mathcal{G}^N$, hence Assumption 2 is satisfied with $D(\lambda)$ being the quantity in the right side of the above equation. Another example is Lo and MacKinlay [23]'s sampling scheme as described by the following proposition:

Proposition 1. *Suppose that, for each $\nu = 1, 2$, the observation times $(t_i^\nu)_{i=0}^{n_\nu}$ are randomly chosen from $\{i\tau_N : i = 0, 1, \dots, \lfloor (T + \delta)\tau_N^{-1} \rfloor\}$ using Bernoulli trials with success probability $1 - \pi_\nu$. Then, Assumption 2 is satisfied with*

$$\begin{aligned} D(\lambda) &= \frac{1}{\pi\lambda^2} \Re \left[\frac{(1-\pi_1)(1-\pi_2)(1-e^{-\sqrt{-1}\lambda})}{(1-\pi_1 e^{-\sqrt{-1}\lambda})(1-\pi_2 e^{-\sqrt{-1}\lambda})} \right] \\ &= \frac{1-\cos\lambda}{\pi\lambda^2} \frac{(1-\pi_1)(1-\pi_2)(1+\pi_1+\pi_2-\pi_1\pi_2(2\cos\lambda+1))}{|1-\pi_1 e^{-\sqrt{-1}\lambda}|^2 |1-\pi_2 e^{-\sqrt{-1}\lambda}|^2}. \end{aligned}$$

Now we state our asymptotic results. The first result concerns the asymptotic behavior of the estimators $\widehat{\rho}_{(j)}(\theta)$ and can be considered as a counterpart of Propositions 3–4 from [19]:

Theorem 1. *Suppose that $L \to \infty$ and $L\tau_N^\kappa \to 0$ as $N \to \infty$ for any $\kappa > 0$. Suppose also that $(\Gamma_N + L_j)\tau_N < \delta$ for every N. Under Assumptions 1–2, the following statements hold true:*

(a) *If a sequence $v_N > 0$ satisfies $L^{-\frac{1}{2}}\tau_N^{-1} v_N \to \infty$ as $N \to \infty$, then*

$$\max_{\theta \in \mathcal{G}^N : |\theta - \theta_j| \geq v_N} \left| \widehat{\rho}_{(j)}(\theta) \right| \to^p 0$$

as $N \to \infty$.

(b) *Let (ϑ_N) be a sequence of real numbers such that $\vartheta_N \in \mathcal{G}^N$ and $\tau_N^{-1}(\vartheta_N - \theta_j) \to b$ as $N \to \infty$ for some $b \in \mathbb{R}$. Then*

$$\widehat{\rho}_{(j)}(\vartheta_N) \to^p \Sigma_T(\theta_j) R_j \int_{\Lambda_{-j}} D(\lambda) \cos(b\lambda) d\lambda$$

as $N \to \infty$, where

$$\Sigma_T(\theta) = \begin{cases} \int_0^T \sigma_s^1 \sigma_{s+\theta}^2 ds & \text{if } \theta \geq 0, \\ \int_0^T \sigma_{s-\theta}^1 \sigma_s^2 ds & \text{otherwise.} \end{cases}$$

The next theorem concerns the consistency of the estimators $\widehat{\theta}_j$ and can be considered as a counterpart of Theorem 1 from [19]:

Theorem 2. *Suppose that $L \to \infty$ and $L\tau_N^\kappa \to 0$ as $N \to \infty$ for any $\kappa > 0$. Suppose also that $(\Gamma_N + L_j)\tau_N < \delta$ for every N and $\Gamma_N \tau_N \to \delta$ as $N \to \infty$. Under Assumptions 1–2, if a sequence $v_N > 0$ satisfies $L^{-\frac{1}{2}}\tau_N^{-1} v_N \to \infty$ as $N \to \infty$, then*

$$v_N^{-1}(\widehat{\theta}_j - \theta_j) \to^p 0$$

as $N \to \infty$, provided that $R_j \neq 0$ and $\Sigma_T(\theta_j) \neq 0$ a.s. In particular, we have $\widehat{\theta}_j \to^p \theta_j$ as $N \to \infty$.

Remark 2. Theorem 2 shows that our new estimator $\widehat{\theta}_j$ enjoys a similar asymptotic property to that of the estimator proposed in our previous work [15]. As stated in the Introduction, our new estimator has a computational advantage in applications to high-frequency data with sub-second time stamps. In the next section we see that the new one has another advantage in terms of finite sample performance.

5 Simulation study

In this section we assess the finite sample accuracy of our novel estimators $\widehat{\theta}_j$ by a Monte Carlo study. We set $N = 14$, $T = n\tau_N$ with $n = 30,000$.

We simulate model (3) with the following two scenarios of the volatility processes:

Scenario 1 Constant volatilities. $\sigma^\nu \equiv 1$ for $\nu = 1, 2$.

Scenario 2 Stochastic volatilities with a leverage effect. The Heston model is adopted to generate the volatility process σ_t^ν for each $\nu = 1, 2$: The process $v_t^\nu = (\sigma_t^\nu)^2$ is the solution of the following stochastic differential equation:

$$dv_t^\nu = \kappa(\eta - v_t^\nu)dt + \xi\sqrt{v_t^\nu}(\rho dB_t^\nu + \sqrt{1-\rho^2}dW_t^\nu),$$

where W^ν is a standard Wiener process and the initial value v_0^ν is randomly drawn from the stationary distribution of the process v_t^ν in each iteration, i.e. $v_0^\nu \sim \mathrm{Gamma}(2\kappa\eta/\xi^2, 2\kappa/\xi^2)$. We assume that the processes B, W^1 and W^2 are mutually independent. The parameters κ, η, ξ and ρ are chosen as in [4]: $\kappa = 5$, $\eta = 0.04$, $\xi = 0.5$ and $\rho = -0.5$.

The parameters for the spectral density (2) are chosen as in Table 1. Simulation of the paths of the process B is performed in the same way as in [15].

Table 1: Parameters for the spectral density (2)

j	1	2	3	4	5	6	7	8	9–14
R_j	0.3	0.5	0.7	0.5	0.5	0.5	0.5	0.5	0
θ_j/τ_N	−1	−1	−2	−2	−3	−5	−7	−10	0

We use the Lo-MacKinlay sampling scheme presented in Section 4 to generate the sampling times $(t_i^1)_{i=0}^{n_1}$ and $(t_i^2)_{i=0}^{n_2}$. We fix π_1 as $\pi_1 = 1/4$ and vary π_2 as $\pi_2 \in \{1/4, 1/2, 3/4\}$. Recall that π_ν is the occurrence probability of missing observations for the ν-th asset X^ν. Therefore, X^ν is observed less frequently as the value of π_2 increases, hence the degree of the non-synchronicity becomes higher.

We use $L = 20$ as the length of Daubechies' wavelet filter and set $\mathcal{G}^N = \{l\tau_N : l \in \mathbb{Z}, |l| \leq 100\}$. For comparison we also compute the estimator for θ_j proposed in [15], which is defined as a maximizer of the so-called wavelet cross-covariance estimators based on data synchronized by interpolation (we refer to it as WCCF). Here, computation of the wavelet cross-covariance estimators requires specification of Daubechies' wavelet and we use the least asymmetric wavelet with length 20.

We run 1,000 Monte Carlo iterations for each experiment. Table 2 reports the sample median and the median absolute deviation (MAD) of the estimates for each experiment in Scenario 1. We see from Table 2 that both the estimators exhibit good accuracy in the case of $\pi_2 = 1/4$ for the levels $j \leq 7$. It is theoretically natural that the accuracy of the estimators declines as j increases because the contrast function $|\widehat{\rho}_{(j)}(\theta)|$, $\theta \in \mathcal{G}_N$ gets flatter as $L_j = (2^j - 1)(L - 1) + 1$ increases. In the cases of $\pi_2 = 1/2$ and $\pi_2 = 3/4$, the WCCF estimators are strongly biased for the levels $j \geq 3$, while the estimators $\widehat{\theta}_j$ still keep the good precision. Hence our new estimators can handle high-frequency data with rather high degree of the non-synchronicity.

Table 3 shows the simulation results in Scenario 2. As the table reveals, the presence of a time variation and a leverage effect in the volatilities does not affect the performance of our estimators, which is in line with our asymptotic theory.

Table 2: Simulation results in Scenario 1

j	1	2	3	4	5	6	7	8
True	-1	-1	-2	-2	-3	-5	-7	-10
				$\pi_2 = 1/4$				
$\widehat{\theta}_j$	-1 (0)	-1 (0)	-2 (0)	-2 (0)	-3 (0)	-5 (1)	-7 (3)	-9 (9)
WCCF	-1 (0)	-1 (0)	-2 (0)	-2 (0)	-3 (0)	-5 (1)	-7 (3)	-9 (9)
				$\pi_2 = 1/2$				
$\widehat{\theta}_j$	-1 (0)	-1 (0)	-2 (0)	-2 (0)	-3 (0)	-5 (1)	-7 (3)	-9 (9)
WCCF	-1 (0)	-1 (0)	-1 (0)	-1 (0)	-2 (0)	-4 (1)	-6 (3)	-8 (9)
				$\pi_2 = 3/4$				
$\widehat{\theta}_j$	-1 (0)	-1 (0)	-2 (0)	-2 (0)	-3 (0)	-5 (1)	-7 (3)	-9 (9)
WCCF	-1 (0)	-1 (0)	-1 (0)	0 (0)	0 (0)	-2 (1)	-4 (3)	-7 (9)

This table reports the median and the median absolute deviation (in parentheses) of the estimates in Scenario 1 (divided by τ_N).

6 Empirical application

In this section we apply our new method to a set of market data consisting of high-frequency transactions of 4 assets. The 4 assets we will focus on are Apple (AAPL), Cisco Systems (CSCO), Intel (INTC) and Microsoft (MSFT). They are chosen from the stocks which are listed on the NASDAQ exchange and components of the 30 Dow Jones Industrial Average (DJIA) stocks in August 2016. We use intraday transaction data recorded between 9:30 am and 16:00, which are taken from the Daily TAQ database with the accuracy of the timestamp values being one micro-second. The sample period is the whole of August 2016, containing 21 trading days. We investigate the lead-lag relationships between transactions of a single asset executed on different exchanges.[2] The exchanges we will focus on are the NASDAQ, NYSE Arca and BATS exchanges. We report in Table 4 the average daily numbers of transactions of each asset on each exchange.

[2] This subject addresses the issue of determining in which exchange price discovery of the asset occurs. Such an issue is one of the fundamental problems in financial econometrics and has been widely studied in the literature; see e.g. [12, 13, 27, 31].

Table 3: Simulation results in Scenario 2

j	1	2	3	4	5	6	7	8
True	-1	-1	-2	-2	-3	-5	-7	-10
				$\pi_2 = 1/4$				
$\widehat{\theta}_j$	-1 (0)	-1 (0)	-2 (0)	-2 (0)	-3 (0)	-5 (1)	-7 (3)	-9 (9)
WCCF	-1 (0)	-1 (0)	-2 (0)	-2 (0)	-3 (0)	-5 (1)	-7 (3)	-9 (9)
				$\pi_2 = 1/2$				
$\widehat{\theta}_j$	-1 (0)	-1 (0)	-2 (0)	-2 (0)	-3 (0)	-5 (1)	-7 (3)	-9 (9)
WCCF	-1 (0)	-1 (0)	-1 (0)	-1 (0)	-2 (0)	-4 (1)	-6 (3)	-8 (9)
				$\pi_2 = 3/4$				
$\widehat{\theta}_j$	-1 (0)	-1 (0)	-2 (0)	-2 (0)	-3 (0)	-5 (1)	-7 (3)	-9 (9)
WCCF	-1 (0)	-1 (0)	-1 (0)	0 (0)	0 (0)	-2 (1)	-4 (3)	-6 (9)

This table reports the median and the median absolute deviation (in parentheses) of the estimates in Scenario 2 (divided by τ_N).

We use $L = 20$ as the length of Daubechies' wavelet filter and set

$$\mathcal{G}^N = \{-2\text{ms}, -1.999\text{ms}, \dots, 1.999\text{ms}, 2\text{ms}\}$$

as the search grid.

For comparison we also compute the following two estimators for lead-lag relationships.

- Hoffmann-Rosenbaum-Yoshida (HRY) estimator [19]: This estimator is defined as a maximizer of $\widehat{U}^N(\theta)$ over the grid $\theta \in \mathcal{G}^N$:
$$\widehat{\theta}^{HRY} = \arg\max_{\theta \in \mathcal{G}^N} |\widehat{U}^N(\theta)|.$$
- Dobrev-Schaumburg (DS) estimator [9]: This estimator is constructed as follows. For each $\nu = 1, 2$ and each $t \geq 0$, we set $I_t^\nu = 1$ if $t \in \{t_i^\nu : i = 0, 1, \dots, n_\nu\}$ and $I_t^\nu = 0$ otherwise. Then we define
$$A(\theta) = \frac{1}{\min\{n_1, n_2\}} \sum_{k=1}^{\infty} I^1_{k\tau_N} I^2_{k\tau_N + \theta}$$
for each $\theta \in \mathbb{R}$. Now, the DS estimator $\widehat{\theta}^{DS}$ is defined as a maximizer of $A(\theta)$ over the grid $\mathcal{G}^N$:
$$\widehat{\theta}^{DS} = \arg\max_{\theta \in \mathcal{G}^N} A(\theta).$$

Figures 1–3 show the time series of the estimates $\widehat{\theta}^{HRY}$, $\widehat{\theta}_j$ $(1 \leq j \leq 10)$ and $\widehat{\theta}^{DS}$ evaluated every trading day when we focus on the multi-market trading of INTC. We find that the estimates of $\widehat{\theta}_j$ are fairly stable for the level $j = 10$ which corresponds to the time scale between 1.024ms and 2.048ms. On the other hand, the estimates of $\widehat{\theta}^{HRY}$ are scattered and seem to exhibit no regular pattern, while the estimates of $\widehat{\theta}^{DS}$

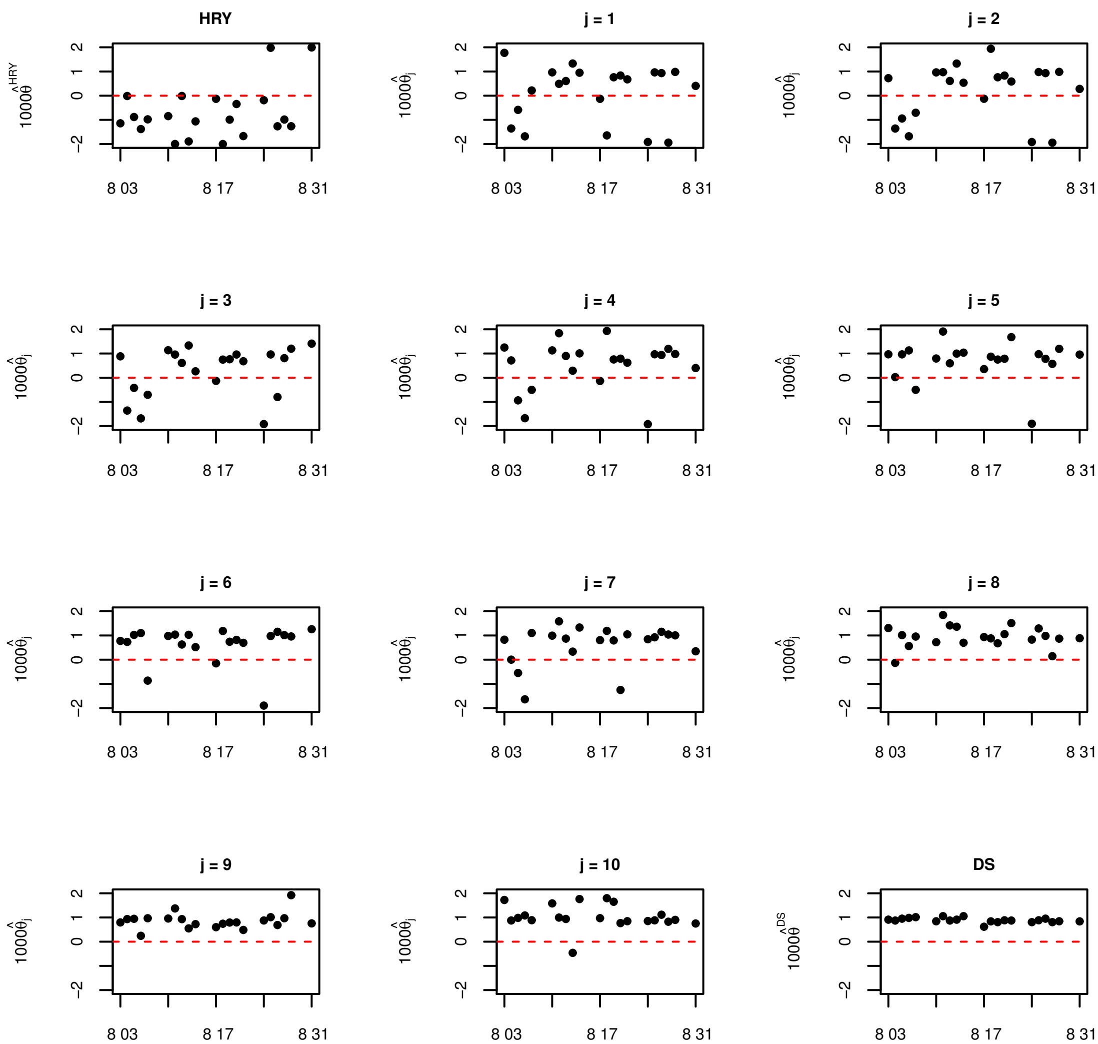
are quite stable and suggest the presence of consistent lead-lag relationships between the trading activity in the three markets.

Tables 5–7 report the sample medians and the sample median absolute deviations (in parentheses) of the estimates $\widehat{\theta}^{HRY}$, $\widehat{\theta}_j$ $(1 \leq j \leq 10)$ and $\widehat{\theta}^{DS}$ over the whole sample period for three pairs of the exchanges. We find that in most cases the estimators $\widehat{\theta}_j$ give more stable estimates than $\widehat{\theta}^{HRY}$. In particular, the estimates of the $\widehat{\theta}_j$ with the levels $j = 9, 10$ are fairly stable for many cases. We also find that the estimates of the $\widehat{\theta}_j$ with the levels $j = 9, 10$ for the different assets are close together. These findings perhaps suggest that there are consistent market participants who are common for these four assets and act with the time scale between 0.512ms and 2.048ms and they cause these systematic lead-lag relationships. Interestingly, we further find that for the pairs NASDAQ-NYSE Arca and NYSE Arca-BATS the estimates of the $\widehat{\theta}_j$ with the levels $j = 9, 10$ are comparatively close to those of $\widehat{\theta}^{DS}$. Note that the estimators $\widehat{\theta}_j$ measure the lead-lag relationships between the asset returns, while the estimator $\widehat{\theta}^{DS}$ measure the lead-lag relationships between the transaction times of the assets, so they measure the essentially different lead-lag relationships. However, if the above consistent market participants cause lead-lag relationships between the three exchanges, it likely occurs that these two kinds of lead-lag relationship are linked each other. In contrast, the estimates of $\widehat{\theta}^{DS}$ for the pair NASDAQ-BATS seem to capture a "deterministic" lead-lag relationship. In order to focus on "stochastic" lead-lag relationships captured by the DS estimator, we re-compute $\widehat{\theta}^{DS}$ for this pair using the grid $\mathcal{G}^N = \{-2\text{ms}, -1.999\text{ms}, \dots, -0.101\text{ms}, 0.101\text{ms}, \dots, 1.999\text{ms}, 2\text{ms}\}$ (i.e. we remove the lags between -0.1ms and 0.1ms from the original search grid). The results are reported in Table 8. We find that the DS estimates in this case are rather close to those of $\widehat{\theta}_j$ with $j = 10$, hence the above remark is indeed applicable to the pair NASDAQ-BATS as well. Finally, the estimates of our estimators $\widehat{\theta}_j$ as well as the DS estimator $\widehat{\theta}^{DS}$ imply the following lead-lag relationships between the three exchanges: The NASDAQ exchange is the fastest one, the BATS exchange is the second fastest one, and the NYSE Arca exchange is the slowest one. It is not surprising that the NASDAQ exchange is the fastest one because all the four assets considered in this study are listed on the NASDAQ exchange; the primary listing exchange typically dominates the price discovery (cf. [12, 13, 27, 31]).

Table 4: The average daily numbers of transactions

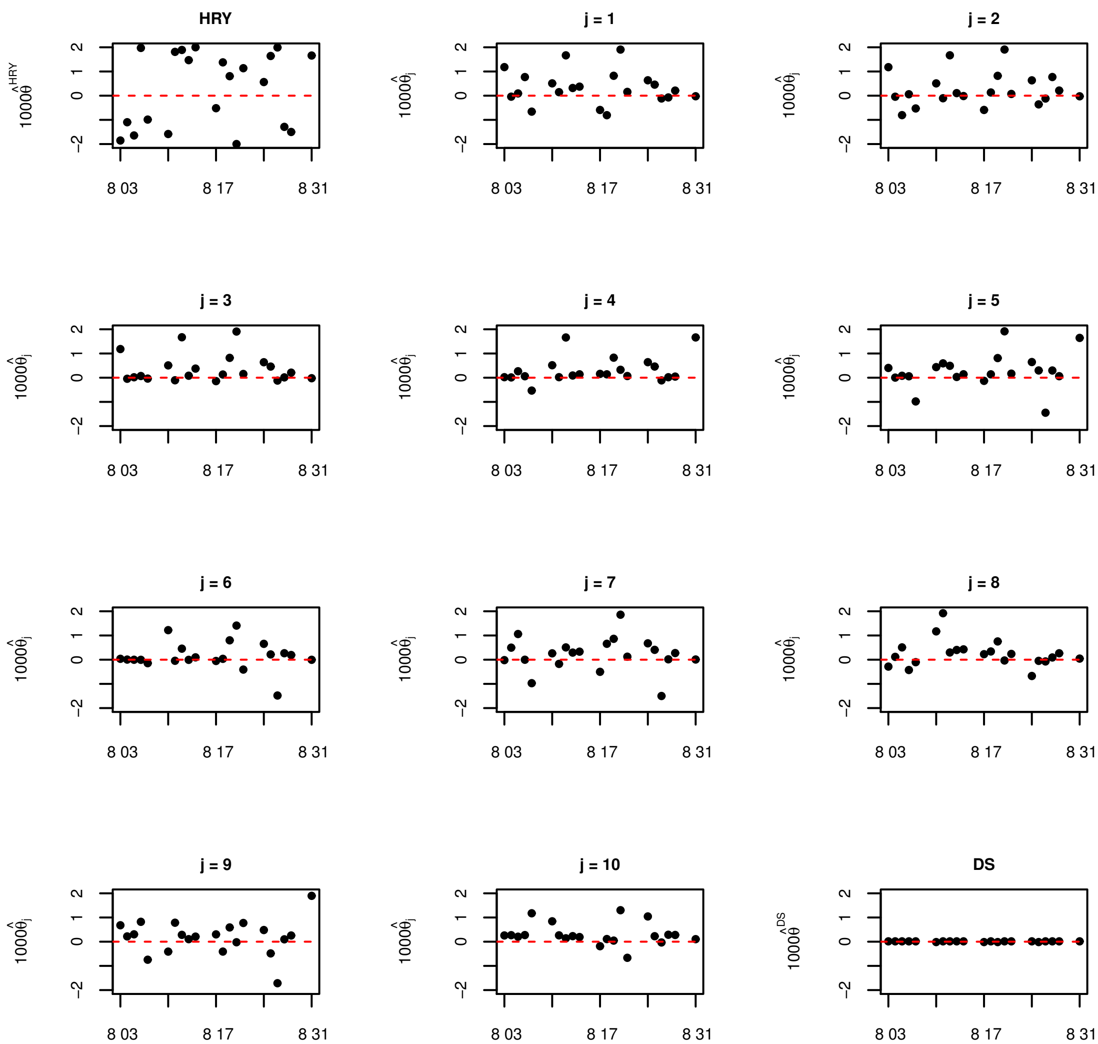
	AAPL	CSCO	INTC	MSFT
NASDAQ	84476	23710	27224	45072
NYSE Arca	58730	13722	18386	24984
BATS	50988	14959	18485	25150

Figure 1: The time series of the estimates $\widehat{\theta}^{HRY}$, $\widehat{\theta}_j$ ($1 \leq j \leq 10$) and $\widehat{\theta}^{DS}$ for INTC: NASDAQ vs Arca



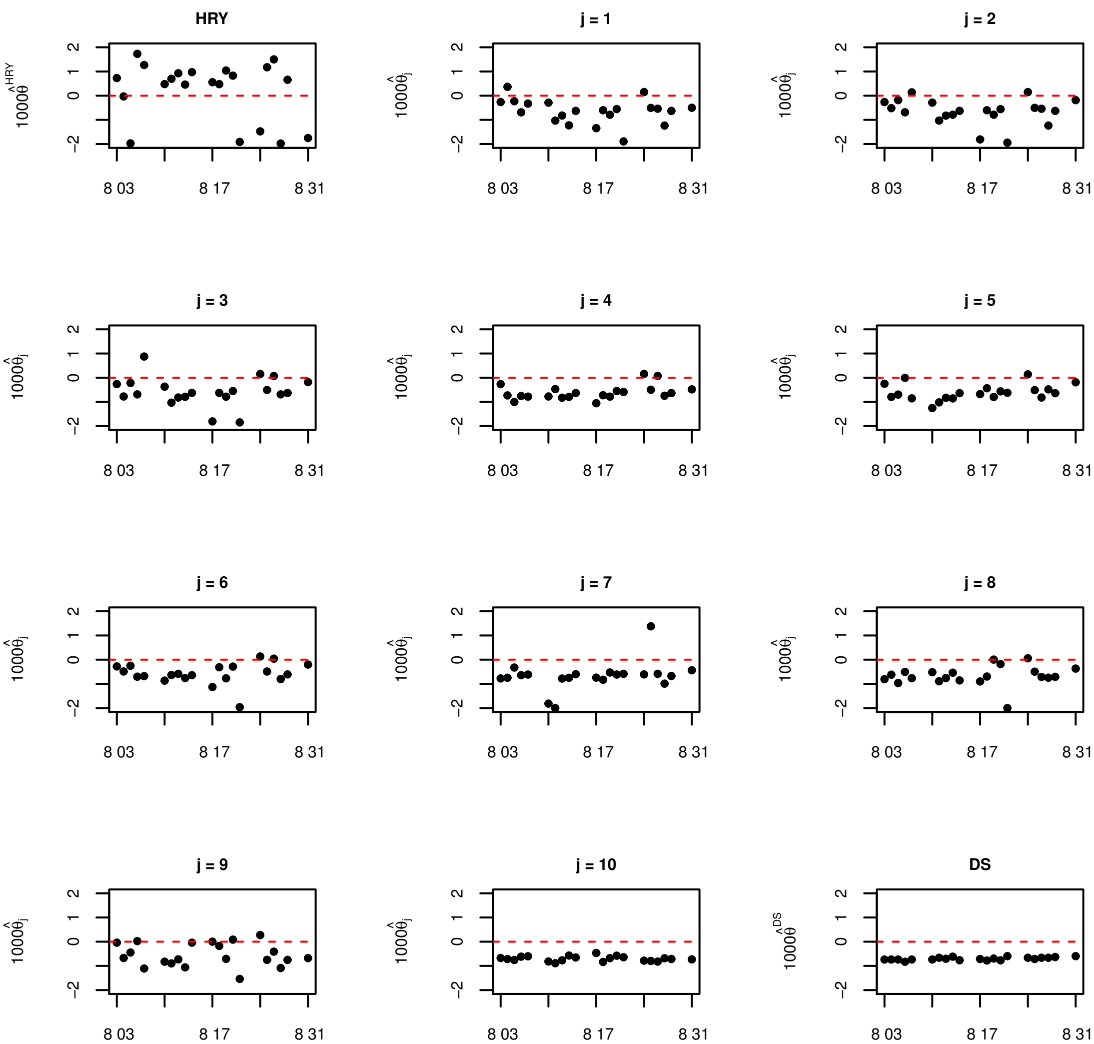
In this figure we depict the time series of the estimates $\widehat{\theta}^{HRY}$, $\widehat{\theta}_j$ ($1 \leq j \leq 10$) and $\widehat{\theta}^{DS}$ for the INTC transaction prices between the NASDAQ and the NYSE Arca exchanges. The upper-left figure corresponds to $\widehat{\theta}^{HRY}$, while the lower-right figure corresponds $\widehat{\theta}^{DS}$. The remaining figures correspond to $\widehat{\theta}_j$ for $j = 1, \ldots, 10$. The horizontal line is labeled by dates. The vertical line is in mili-seconds. The red dash line denotes 0ms. The positive value imply that the NASDAQ leads the NYSE Arca and vice versa.

Figure 2: The time series of the estimates $\widehat{\theta}^{HRY}$, $\widehat{\theta}_j$ $(1 \leq j \leq 10)$ and $\widehat{\theta}^{DS}$ for INTC: NASDAQ vs BATS



In this figure we depict the time series of the estimates $\widehat{\theta}^{HRY}$, $\widehat{\theta}_j$ $(1 \leq j \leq 10)$ and $\widehat{\theta}^{DS}$ for the INTC transaction prices between the NASDAQ and the BATS exchanges. The upper-left figure corresponds to $\widehat{\theta}^{HRY}$, while the lower-right figure corresponds $\widehat{\theta}^{DS}$. The remaining figures correspond to $\widehat{\theta}_j$ for $j = 1, \ldots, 10$. The horizontal line is labeled by dates. The vertical line is in mili-seconds. The red dash line denotes 0ms. The positive value imply that the NASDAQ leads the BATS and vice versa.

Figure 3: The time series of the estimates $\widehat{\theta}^{HRY}$, $\widehat{\theta}_j$ $(1 \leq j \leq 10)$ and $\widehat{\theta}^{DS}$ for INTC: Arca vs BATS



In this figure we depict the time series of the estimates $\widehat{\theta}^{HRY}$, $\widehat{\theta}_j$ $(1 \leq j \leq 10)$ and $\widehat{\theta}^{DS}$ for the INTC transaction prices between the NASDAQ and the NYSE Arca exchanges. The upper-left figure corresponds to $\widehat{\theta}^{HRY}$, while the lower-right figure corresponds $\widehat{\theta}^{DS}$. The remaining figures correspond to $\widehat{\theta}_j$ for $j = 1, \ldots, 10$. The horizontal line is labeled by dates. The vertical line is in mili-seconds. The red dash line denotes 0ms. The positive value imply that the NYSE Arca leads the BATS and vice versa.

Table 5: The medians and MADs of the estimates: NASDAQ vs Arca

	AAPL	CSCO	INTC	MSFT
HRY	0.041 (0.200)	-0.745 (0.933)	-0.986 (0.950)	-0.628 (0.636)
j = 1	0.536 (1.029)	0.539 (1.082)	0.605 (0.557)	0.940 (0.691)
j = 2	0.872 (0.469)	0.900 (0.683)	0.607 (0.549)	0.992 (0.710)
j = 3	0.906 (0.337)	0.880 (0.572)	0.748 (0.669)	0.929 (0.239)
j = 4	0.862 (0.603)	0.910 (0.412)	0.786 (0.575)	0.847 (0.565)
j = 5	0.926 (0.291)	0.842 (0.476)	0.868 (0.246)	0.793 (0.498)
j = 6	0.899 (0.394)	0.894 (0.461)	0.960 (0.286)	0.980 (0.222)
j = 7	0.960 (0.405)	0.903 (0.715)	0.871 (0.344)	0.878 (0.297)
j = 8	0.855 (0.409)	0.725 (0.614)	0.936 (0.351)	1.080 (0.302)
j = 9	0.874 (0.381)	1.043 (0.307)	0.797 (0.219)	0.987 (0.199)
j = 10	0.799 (0.654)	0.924 (0.215)	0.940 (0.170)	0.944 (0.154)
DS	0.839 (0.007)	0.885 (0.053)	0.875 (0.053)	0.839 (0.053)

This table reports the sample medians and the sample median absolute deviations (in parentheses) of the estimates $\widehat{\theta}^{HRY}$ (HRY), $\widehat{\theta}_j$ ($j = 1, \dots, 10$) and $\widehat{\theta}^{DS}$ (DS) over the whole sample period when X^1 is the transaction price process executed on the NASDAQ exchange and X^2 is the one executed on the NYSE Arca exchange.

Table 6: The medians and MADs of the estimates: NASDAQ vs BATS

	AAPL	CSCO	INTC	MSFT
HRY	0.388 (0.645)	-0.615 (2.033)	0.806 (1.760)	-0.381 (2.159)
j = 1	0.123 (0.519)	0.134 (0.519)	0.210 (0.443)	0.016 (0.329)
j = 2	0.080 (0.322)	0.256 (0.789)	0.071 (0.639)	0.236 (0.331)
j = 3	0.034 (0.391)	0.244 (0.338)	0.131 (0.353)	0.059 (0.342)
j = 4	0.108 (0.203)	0.139 (0.470)	0.139 (0.188)	0.050 (0.270)
j = 5	0.173 (0.224)	0.150 (0.289)	0.166 (0.350)	0.088 (0.308)
j = 6	0.061 (0.222)	0.166 (0.279)	0.035 (0.234)	0.102 (0.294)
j = 7	0.075 (0.328)	0.085 (0.388)	0.274 (0.409)	0.066 (0.230)
j = 8	0.088 (0.295)	0.340 (0.403)	0.227 (0.385)	0.154 (0.322)
j = 9	0.167 (0.083)	0.239 (0.248)	0.256 (0.494)	0.249 (0.136)
j = 10	0.139 (0.085)	0.235 (0.142)	0.230 (0.130)	0.190 (0.188)
DS	0.015 (0.000)	0.015 (0.000)	0.015 (0.000)	0.015 (0.000)

This table reports the sample medians and the sample median absolute deviations (in parentheses) of the estimates $\widehat{\theta}^{HRY}$ (HRY), $\widehat{\theta}_j$ ($j = 1, \dots, 10$) and $\widehat{\theta}^{DS}$ (DS) over the whole sample period when X^1 is the transaction price process executed on the NASDAQ exchange and X^2 is the one executed on the BATS exchange.

Table 7: The medians and MADs of the estimates: Arca vs BATS

	AAPL	CSCO	INTC	MSFT
HRY	-0.104 (0.307)	0.771 (0.930)	0.657 (0.566)	1.109 (0.491)
j = 1	-0.747 (0.224)	-0.736 (0.771)	-0.602 (0.406)	-0.673 (0.715)
j = 2	-0.744 (0.154)	-0.740 (0.403)	-0.602 (0.329)	-0.620 (0.704)
j = 3	-0.746 (0.206)	-0.719 (0.350)	-0.627 (0.286)	-0.685 (0.715)
j = 4	-0.753 (0.212)	-0.718 (0.455)	-0.726 (0.151)	-0.778 (0.221)
j = 5	-0.788 (0.077)	-0.713 (0.363)	-0.640 (0.265)	-0.676 (0.383)
j = 6	-0.848 (0.252)	-0.638 (0.283)	-0.609 (0.283)	-0.697 (0.277)
j = 7	-0.733 (0.203)	-0.658 (0.400)	-0.641 (0.160)	-0.721 (0.292)
j = 8	-0.752 (0.289)	-0.678 (0.208)	-0.709 (0.270)	-0.711 (0.423)
j = 9	-0.740 (0.132)	-0.645 (0.199)	-0.679 (0.568)	-0.699 (0.165)
j = 10	-0.713 (0.173)	-0.753 (0.297)	-0.714 (0.105)	-0.787 (0.193)
DS	-0.710 (0.022)	-0.700 (0.099)	-0.715 (0.076)	-0.669 (0.053)

This table reports the sample medians and the sample median absolute deviations (in parentheses) of the estimates $\widehat{\theta}^{HRY}$ (HRY), $\widehat{\theta}_j$ ($j = 1, \dots, 10$) and $\widehat{\theta}^{DS}$ (DS) over the whole sample period when X^1 is the transaction price process executed on the NYSE Arca exchange and X^2 is the one executed on the BATS exchange.

Table 8: The medians and MADs of the DS estimates (different grid): NASDAQ vs BATS

	AAPL	CSCO	INTC	MSFT
DS	0.139 (0.046)	0.247 (0.039)	0.247 (0.053)	0.170 (0.061)

This table reports the sample medians and the sample median absolute deviations (in parentheses) of the estimates $\widehat{\theta}^{DS}$ (DS) with the search grid $\mathcal{G}^N = \{-2\text{ms}, -1.999\text{ms}, \dots, -0.101\text{ms}, 0.101\text{ms}, \dots, 1.999\text{ms}, 2\text{ms}\}$ over the whole sample period when X^1 is the transaction price process executed on the NASDAQ exchange and X^2 is the one executed on the BATS exchange.

7 Conclusion

In this paper we have proposed a new estimation method for multi-scale analysis of lead-lag relationships between two assets based on their high-frequency observation data. The new method is based on the novel estimator for the scale-by-scale cross-covariance functions $\rho_{(j)}(\theta)$ which is constructed as a kind of wavelet transform of the empirical cross-covariance function used in [19]. We have also developed an associated asymptotic theory to ensure the validity of the proposed estimators in the framework established in our previous work [15]. Compared with the estimation method proposed in [15] which is essentially the same as the traditional method used in the wavelet literature, our new estimation method is more appropriate in applications to high-frequency financial data from the computational point of view. Our simulation study has shown that our estimators are also much more suitable to non-synchronously observed data than the previous one. The empirical results have demonstrated that our new method has the ability to provide a deep insight into lead-lag relationships in high-frequency financial markets.

8 Proofs

Throughout the discussions, for sequences (x_N) and (y_N), $x_N \lesssim y_N$ means that there exists a constant $C \in [0,\infty)$ such that $x_N \leq Cy_N$ for large N. Also, for $r > 0$ $\|\cdot\|_r$ denotes the L^r-norm of random variables, i.e. $\|Z\|_r = (E[|Z|^r])^{1/r}$ for a random variable Z.

8.1 Proof of Proposition 1

We begin by proving some lemmas. Let us set $\tilde{\Pi}^\nu_N = \Pi^\nu_N \cup \{(0, t^\nu_0], (t^\nu_{n_1}, T+\delta]\}$ for $\nu = 1,2$ and $\tilde{r}_N = (\sup_{I\in\tilde{\Pi}^1_N} |I|) \vee (\sup_{J\in\tilde{\Pi}^2_N} |J|)$. We denote by P^{Π^1} (resp. E^{Π^1}) the conditional probability (resp. conditional expectation) given $(t^1_i)^{n_1}_{i=0}$.

Lemma 1. $P(\tau_N^{-1}\tilde{r}_N > x) \leq C\tau_N^{-1}e^{-x/C}$

This lemma can be shown in a similar manner to the proof of Lemma 4 from [5], so we omit the proof.

Lemma 2. *Let $\varpi \in (0,1)$ and set $q = \lceil \varpi N \rceil$. Suppose that $\theta_N \geq 0$ for all N. For any $M > 0$, there is a constant $C_M > 0$ such that*

$$\left| E^{\Pi^1}\left[\frac{1}{2\pi\tau_m\tau_N}\sum_{J\in\tilde{\Pi}^2_N}(\mathcal{F}1_I)(\lambda/\tau_N)(\mathcal{F}1_{J_{-\theta_N}})(-\lambda/\tau_N)K(I,J_{-\theta_N})\right] - \frac{\tau_N}{\pi\tau_m\lambda^2}\Re\left[\left(1-e^{-\sqrt{-1}\lambda\tau_N^{-1}|I|}\right)\frac{1-\pi_2}{1-\pi_2 e^{-\sqrt{-1}\lambda}}\right]\right| \leq C_M\tau_N^M$$

uniformly for $\lambda \in \mathbb{R}$ and $I \in \tilde{\Pi}^1_N$ such that $T\tau_q \leq \underline{I} < \overline{I} \leq T(1-\tau_q)$ for every $M > 0$.

Proof. Set

$$J_I = \bigcup_{J\in\tilde{\Pi}^2_N : K(I,J_{-\theta_N})=1} J_{-\theta_N}.$$

Then we have

$$\begin{aligned}
&\frac{1}{2\pi\tau_m\tau_N}\sum_{J\in\tilde{\Pi}_N^2}(\mathcal{F}1_I)(\lambda/\tau_N)(\mathcal{F}1_{J_{-\theta}})(-\lambda/\tau_N)K(I,J_{-\theta_N})\\
&=\frac{1}{2\pi\tau_m\tau_N}(\mathcal{F}1_I)(\lambda/\tau_N)\left\{(\mathcal{F}1_{[\underline{J_I},\underline{I})})(-\lambda/\tau_N)+(\mathcal{F}1_I)(-\lambda/\tau_N)+(\mathcal{F}1_{[\overline{I},\overline{J_I})})(-\lambda/\tau_N)\right\}\\
&=:\mathbf{I}+\mathbf{II}+\mathbf{III}.
\end{aligned}$$

First we consider **I**. We can rewrite it as

$$\mathbf{I}=\frac{\tau_N}{2\pi\tau_m\lambda^2}\left(e^{-\sqrt{-1}\lambda\tau_N^{-1}|I|}-1\right)\left(1-e^{-\sqrt{-1}\lambda\tau_N^{-1}(\underline{I}-\underline{J_I})}\right).$$

Conditionally on $(t_i^1)_{i=0}^{n_1}$, $\tau_N^{-1}(\underline{I}-\underline{J_I})$ follows the geometric distribution with success probability $1-\pi_2$ truncated from above by $\tau_N^{-1}\underline{I}$. More precisely, we have

$$P^{\Pi^1}(\tau_N^{-1}(\underline{I}-\underline{J_I})=k)\begin{cases}\pi_2^k(1-\pi_2) & \text{if } 0\leq k<\tau_N^{-1}\underline{I},\\ \pi_2^{\tau_N^{-1}\underline{I}} & \text{if } k=\tau_N^{-1}\underline{I}.\end{cases}$$

Therefore, we obtain

$$\begin{aligned}
&E^{\Pi^1}\left[1-e^{-\sqrt{-1}\lambda\tau_N^{-1}(\underline{I}-\underline{J_I})}\right]\\
&=1-(1-\pi_2)\sum_{k=0}^{\tau_N^{-1}\underline{I}-1}\pi_2^k e^{-\sqrt{-1}\lambda k}-\pi_2^{\tau_N^{-1}\underline{I}}e^{-\sqrt{-1}\lambda\tau_N^{-1}\underline{I}}\\
&=1-\pi_2^{\tau_N^{-1}\underline{I}}-\frac{(1-\pi_2)(1-\pi_2^{\tau_N^{-1}\underline{I}}e^{-\sqrt{-1}\lambda\tau_N^{-1}\underline{I}})}{1-\pi_2e^{-\sqrt{-1}\lambda}}+\pi_2^{\tau_N^{-1}\underline{I}}(1-e^{-\sqrt{-1}\lambda\tau_N^{-1}\underline{I}})\\
&=\frac{\pi_2(1-e^{\sqrt{-1}\lambda})}{1-\pi_2e^{-\sqrt{-1}\lambda}}-\pi_2^{\tau_N^{-1}\underline{I}}\frac{1-e^{-\sqrt{-1}\lambda\tau_N^{-1}\underline{I}}+\pi_2(1-e^{-\sqrt{-1}\lambda})}{1-\pi_2e^{-\sqrt{-1}\lambda}}+\pi_2^{\tau_N^{-1}\underline{I}}(1-e^{-\sqrt{-1}\lambda\tau_N^{-1}\underline{I}}).
\end{aligned}$$

Consequently, we have

$$\left|E^{\Pi^1}[\mathbf{I}]-\frac{\tau_N}{2\pi\tau_m\lambda^2}\left(e^{-\sqrt{-1}\lambda\tau_N^{-1}|I|}-1\right)\frac{\pi_2(1-e^{-\sqrt{-1}\lambda})}{1-\pi_2e^{-\sqrt{-1}\lambda}}\right|\lesssim\tau_N^M$$

uniformly in $\lambda\in\mathbb{R}$ and $I\in\tilde{\Pi}_N^1$ such that $T\tau_q\leq\underline{I}<\overline{I}\leq T(1-\tau_q)$. Here, we use the inequality $|1-e^{-\sqrt{-1}x}|\leq|x|$ holding for all $x\in\mathbb{R}$ and Lemma

Next we consider **III**. We can rewrite it as

$$\mathbf{III}=\frac{\tau_N}{2\pi\tau_m\lambda^2}\left(1-e^{\sqrt{-1}\lambda\tau_N^{-1}|I|}\right)\left(e^{\sqrt{-1}\lambda\tau_N^{-1}(\overline{J_I}-\overline{I})}-1\right).$$

Now, an analogous argument to the above yields

$$\left|E^{\Pi^1}[\mathbf{III}]-\frac{\tau_N}{2\pi\tau_m\lambda^2}\left(1-e^{\sqrt{-1}\lambda\tau_N^{-1}|I|}\right)\frac{\pi_2(e^{\sqrt{-1}\lambda}-1)}{1-\pi_2e^{\sqrt{-1}\lambda}}\right|\lesssim\tau_N^M$$

uniformly in $\lambda \in \mathbb{R}$ and $I \in \tilde{\Pi}^1_N$ such that $T\tau_q \leq \underline{I} < \overline{I} \leq T(1-\tau_q)$. Hence we have

$$\left| E^{\Pi^1}[\mathbf{III}] - \overline{E^{\Pi^1}[\mathbf{I}]} \right| \lesssim \tau_N^M$$

uniformly for $\lambda \in \mathbb{R}$ and $I \in \tilde{\Pi}^1_N$ such that $T\tau_q \leq \underline{I} < \overline{I} \leq T(1-\tau_q)$.

Finally, we have

$$E^{\Pi^1_N}[\mathbf{II}] = \mathbf{II} = \frac{\tau_N}{\pi\tau_m\lambda^2}\left(1 - \Re\left[e^{-\sqrt{-1}\lambda\tau_N^{-1}|I|}\right]\right)$$

Therefore, a simple computation yields the desired result. □

Proof of Proposition 1. Assumption 2(i) follows from Lemma 1.

Take a constant $\beta \in (0,1)$ arbitrarily. We prove with this β that there are constants $\alpha, Q > 1$ such that Assumption 2(ii) holds true. For the simplicity of exposition, we assume $\theta_N \geq 0$ for all N (this assumption can be easily removed).

Set

$$\tilde{D}^N_k(\lambda,\theta) = \frac{1}{2\pi\tau_m\tau_N} \sum_{I\in\tilde{\Pi}^1_N, J\in\tilde{\Pi}^2_N : \underline{I}\in I_m(k)} (\mathcal{F}1_I)(\lambda/\tau_N)(\mathcal{F}1_{J_{-\theta}})(-\lambda/\tau_N)K(I, J_{-\theta}).$$

It is obvious that

$$\tau_m \sum_{k=0}^{\lceil T\tau_m^{-1}\rceil - 1} \int_{-\pi}^{\pi} E\left[\left|D^N_k(\lambda,\theta_N) - \tilde{D}^N_k(\lambda,\theta_N)\right|^p\right] d\lambda = O(\tau_N^p \tau_m^{p-1})$$

as $N \to \infty$ for any $p > 1$. Therefore, it suffices to show that there are constants $\alpha, Q > 1$ such that

$$\tau_m \sum_{k=0}^{\lceil T\tau_m^{-1}\rceil - 1} \int_{-\pi}^{\pi} E\left[\left|\tilde{D}^N_k(\lambda,\theta_N) - D^N_k(\lambda,\theta_N)\right|^Q\right] d\lambda = O(\tau_N^\alpha) \tag{4}$$

as $N \to \infty$. For the proof we adopt a similar strategy to the proof of Proposition 6 from [5]. Let ϖ be a number such that $\beta < \varpi < 1$ and set $q = \lceil N\varpi \rceil$. Let $\mathcal{E}$ be the event on which the interval $I_{q+2}(u)$ contains at least one point from $\{t^1_i : i \in \mathbb{Z}_+\}$ and one point from $\{t^2_i : i \in \mathbb{Z}_+\}$ for every $u = k\tau_m\tau_{q+2}^{-1}, \dots, (k+1)\tau_m\tau_{q+2}^{-1} - 1$. We have

$$P(\mathcal{E}^c) \leq \tau_m\tau_{q+2}^{-1}(\pi_1^{\tau_N^{-1}\tau_{q+2}} + \pi_2^{\tau_N^{-1}\tau_{q+2}}). \tag{5}$$

In the following we denote by E^A the conditional expectation given an event A.

For $u \in \mathbb{Z}_+$ and $\lambda \in \mathbb{R}$, we set

$$\eta^N_u(\lambda) = \frac{1}{2\pi\tau_m\tau_N} \sum_{I\in\tilde{\Pi}^1_N, J\in\tilde{\Pi}^2_N : \underline{I}\in I_q(u)} (\mathcal{F}1_I)(\lambda/\tau_N)(\mathcal{F}1_{J_{-\theta}})(-\lambda/\tau_N)K(I, J_{-\theta}).$$

Then, we decompose $\tilde{D}^N_k(\lambda,\theta_N) - D(\lambda)$ as

$$\tilde{D}^N_k(\lambda,\theta_N) - D(\lambda) = \left\{ E^{\mathcal{E}}\left[\sum_{u=k\tau_m\tau_q^{-1}}^{(k+1)\tau_m\tau_q^{-1}-1} \eta^N_u(\lambda)\right] - D(\lambda) \right\}$$

$$
+\sum_{\substack{u=k\tau_m\tau_q^{-1}\\ u \text{ is odd}}}^{(k+1)\tau_m\tau_q^{-1}-1}\left(\eta_u^N(\lambda)-E^{\mathcal{E}}[\eta_u^N(\lambda)]\right)+\sum_{\substack{u=k\tau_m\tau_q^{-1}\\ u \text{ is even}}}^{(k+1)\tau_m\tau_q^{-1}-1}\left(\eta_u^N(\lambda)-E^{\mathcal{E}}[\eta_u^N(\lambda)]\right)
$$

$$
=: \mathbb{I}_k^N(\lambda)+\mathbb{II}_k^N(\lambda)+\mathbb{III}_k^N(\lambda).
$$

First we consider $\mathbb{I}_k^N(\lambda)$. Using the inequality $|e^{\sqrt{-1}x}-1|\leq|x|$ holding for $x\in\mathbb{R}$, we have

$$
|\eta_u^N(\lambda)|\lesssim\tau_m^{-1}\tau_N(\tau_N^{-1}r_N)^2\#\{I\in\tilde{\Pi}_N^1:\underline{I}\in I_q(u)\}
$$

uniformly in $\lambda\in\mathbb{R}$ and $u\in\mathbb{Z}_+$. Therefore, noting that $t_i^1-t_{i-1}^1\geq\tau_N$ for every $i=1,\dots,n_1-1$, we obtain

$$
|\eta_u^N(\lambda)|\lesssim\tau_m^{-1}\tau_q(\tau_N^{-1}r_N)^2\lesssim\tau_N^{\varpi-\beta}(\tau_N^{-1}r_N)^2 \tag{6}
$$

uniformly in $\lambda\in\mathbb{R}$ and $u\in\mathbb{Z}_+$. Now, similarly to the proof of Eq.(36) form [5], we can prove

$$
E[r_N^p]=O(\tau_N^p|\log\tau_N|^p) \tag{7}
$$

for any $p>0$. Hence we have

$$
\begin{aligned}
&\left|E^{\mathcal{E}}\left[\sum_{u=k\tau_m\tau_q^{-1}}^{(k+1)\tau_m\tau_q^{-1}-1}\eta_u^N(\lambda)\right]-E\left[\sum_{u=k\tau_m\tau_q^{-1}}^{(k+1)\tau_m\tau_q^{-1}-1}\eta_u^N(\lambda)\right]\right|\\
&\lesssim\tau_m\tau_q^{-1}\left\{E[\tau_m^{-1}\tau_q(\tau_N^{-1}r_N)^2]P(\mathcal{E}^c)+E[\tau_m^{-1}\tau_q(\tau_N^{-1}r_N)^2 1_{\mathcal{E}^c}]\right\}\\
&\lesssim\tau_N^2\left\{E[r_N^2]P(\mathcal{E}^c)+\sqrt{E[r_N^4]}\sqrt{P(\mathcal{E}^c)}\right\}\\
&\lesssim|\log\tau_N|^2\tau_m\tau_{q+2}^{-1}\sqrt{\pi_1^{\tau_N^{-1}\tau_{q+2}}+\pi_2^{\tau_N^{-1}\tau_{q+2}}}
\end{aligned}
$$

uniformly in $\lambda\in\mathbb{R}$ and $k\in\mathbb{Z}_+$. Moreover, Lemma 2 and (6)–(7) imply that

$$
\begin{aligned}
&E\left[\sum_{u=k\tau_m\tau_q^{-1}}^{(k+1)\tau_m\tau_q^{-1}-1}\eta_u^N(\lambda)\right]\\
&=\frac{\tau_N}{\pi\tau_m\lambda^2}\Re\left[E\left[\sum_{I\in\tilde{\Pi}_N^1:\underline{I}\in I_m(k)}\left(1-e^{-\sqrt{-1}\lambda\tau_N^{-1}|I|}\right)\right]\frac{1-\pi_2}{1-\pi_2e^{-\sqrt{-1}\lambda}}\right]+O(\tau_N^{\varpi-\beta}|\log\tau_N|^2)
\end{aligned}
$$

uniformly in λ and k. Since $(t_i^1-t_{i-1}^1)_{i=1}^{n_1}$ is a sequence of i.i.d. variables whose distributions are the geometric distribution with success probability $1-\pi_1$, the Wald identity yields

$$
\begin{aligned}
&E\left[\sum_{u=k\tau_m\tau_q^{-1}}^{(k+1)\tau_m\tau_q^{-1}-1}\eta_u^N(\lambda)\right]\\
&=\frac{\tau_N}{\pi\tau_m\lambda^2}\Re\left[E\left[\#\{I\in\tilde{\Pi}_N^1:\underline{I}\in I_m(k)\}\right]\frac{(1-\pi_2)(1-e^{-\sqrt{-1}\lambda})}{(1-\pi_1e^{-\sqrt{-1}\lambda})(1-\pi_2e^{-\sqrt{-1}\lambda})}\right]+O(\tau_N^{\varpi-\beta}|\log\tau_N|^2)
\end{aligned}
$$

$$= D(\lambda) + O(\tau_N^{\varpi-\beta}|\log \tau_N|^2)$$

uniformly in λ and k. Consequently, we obtain

$$E\left[|\mathbb{I}_k^N(\lambda)|^p\right] = |\mathbb{I}_k^N(\lambda)|^p = O(\tau_N^{(\varpi-\beta)p}|\log \tau_N|^{2p})$$

uniformly in λ and k for any $p > 1$.

Next we consider $\mathbb{III}_k^n(\lambda)$. By construction $(\eta_u^N(\lambda))_{u:\text{ odd}}$ is independent conditionally to $\mathcal{E}$. Therefore, the BDG inequality and (6)–(7) yield

$$E^{\mathcal{E}}\left[\left|\mathbb{III}_k^N(\lambda)\right|^p\right] \lesssim (\tau_m \tau_q^{-1})^{p/2} E\left[\left|\tau_m^{-1}\tau_q(\tau_N^{-1} r_N)^2\right|^p\right] \lesssim (\tau_m^{-1}\tau_q)^{p/2}|\log \tau_N|^{2p}$$

uniformly in λ and k for any $p > 1$. Moreover, (5)–(7) imply that

$$E^{\mathcal{E}^c}\left[\left|\mathbb{III}_k^N(\lambda)\right|^p\right] = O((\tau_m^{-1}\tau_q)^{p/2}|\log \tau_N|^{2p})$$

uniformly in λ and k for any $p > 1$. Consequently, we obtain

$$E\left[\left|\mathbb{III}_k^N(\lambda)\right|^p\right] = O((\tau_m^{-1}\tau_q)^{p/2}|\log \tau_N|^{2p})$$

uniformly in λ and k for any $p > 1$. An analogous argument yields

$$E\left[\left|\mathbb{IIII}_k^N(\lambda)\right|^p\right] = O((\tau_m^{-1}\tau_q)^{p/2}|\log \tau_N|^{2p})$$

uniformly in λ and k for any $p > 1$.

After all, we have

$$\tau_m \sum_{k=0}^{\lceil T\tau_m^{-1}\rceil - 1} \int_{-\pi}^{\pi} E\left[\left|\tilde{D}_k^N(\lambda, \theta_N) - D(\lambda)\right|^p\right] d\lambda = O(\tau_N^{(\varpi-\beta)p/2}|\log \tau_N|^{2p})$$

for any $p > 1$. Now we take $Q > 1$ so that $(\varpi - \beta)Q > 4$ and set $\alpha = (\varpi - \beta)Q/4$. Then (4) holds true. $\square$

8.2 Proof of Theorem 1

First we remark that a standard localization procedure presented e.g. at the beginning of Section 7.3 of [15] allows us to assume that there is a constant $K > 0$ such that

$$|\sigma_t^1| + |\sigma_t^2| \leq K, \qquad |\sigma_t^1 - \sigma_s^1| + |\sigma_t^2 - \sigma_s^2| \leq K|t - s|^\gamma$$

for any $t, s \geq 0$ throughout the proof.

Next we introduce some notation. For each $k \in \mathbb{Z}_+$ and $\theta \in (-\delta, \delta)$, we set

$$\widehat{U}_k^N(\theta) = \begin{cases} \sum_{I,J:\underline{I}\in I_m(k)} X^1(I)X^2(J)K(I, J_{-\theta}) & \text{if } \theta \geq 0, \\ \sum_{I,J:\underline{J}\in I_m(k)} X^1(I)X^2(J)K(I_\theta, J) & \text{if } \theta < 0 \end{cases}$$

and

$$c_k^N(\theta) = \begin{cases} \sigma^1_{k\tau_m}\sigma^2_{(k\tau_m+\theta-r_N)_+} & \text{if } \theta \geq 0, \\ \sigma^1_{(k\tau_m-\theta-r_N)_+}\sigma^2_{k\tau_m} & \text{if } \theta < 0 \end{cases}$$

and

$$\widetilde{U}_k^N(\theta) = \begin{cases} \sum_{I,J:\underline{I}\in I_m(k)} B^1(I)B^2(J)K(I,J_{-\theta}) & \text{if } \theta \geq 0, \\ \sum_{I,J:\underline{J}\in I_m(k)} B^1(I)B^2(J)K(I_\theta,J) & \text{if } \theta < 0 \end{cases}$$

and

$$\overline{U}_k^N(\theta) = \begin{cases} \sum_{I,J:\underline{I}\in I_m(k)} E^\Pi\left[B^1(I)B^2(J)\right]K(I,J_{-\theta}) & \text{if } \theta \geq 0, \\ \sum_{I,J:\underline{J}\in I_m(k)} E^\Pi\left[B^1(I)B^2(J)\right]K(I_\theta,J) & \text{if } \theta < 0. \end{cases}$$

In the following we denote by E^Π the conditional expectation given $(t_i^1)_{i=0}^{n_1}$ and $(t_j^2)_{j=0}^{n_2}$.

Lemma 3. *For any $p > 1$, there is a constant $C_p > 0$ such that*

$$E^\Pi\left[\left|\widehat{U}_k^N(\theta) - c_k^N(\theta)\widetilde{U}_k^N(\theta)\right|^p\right] \leq C_p\tau_m^{(1+\gamma)p}$$

for any $N \in \mathbb{N}$, $k \in \mathbb{Z}_+$ and $\theta \in (-\delta,\delta)$.

Proof. By symmetry it is enough to consider the case of $\theta \geq 0$.

First we apply the so-called reduction procedures used in [17, 18] to every realization of $(I)_{I\in\Pi_N^1}$ and $(J_{-\theta})_{J\in\Pi_N^2}$ (see also the proof of Lemma 2 from [5]). We define a new partition $\tilde{\Pi}_N^1$ as follows: $I \in \tilde{\Pi}_N^1$ if and only if either $I \in \Pi_N^1$ and it has non-empty intersection with two distinct intervals from Π_N^2 or there is $J \in \Pi_N^2$ such that I is the union of all intervals from Π_N^1 included in J. We also define a new partition $\tilde{\Pi}_N^2$ as follows: $J \in \tilde{\Pi}_N^2$ if and only if either $J \in \Pi_N^2$ and $J_{-\theta}$ has non-empty intersection with two distinct intervals from Π_N^1 or there is $I \in \Pi_N^1$ such that J is the union of all intervals from $J' \in \Pi_N^2$ such that $J'_{-\theta}$ is included in I. Due to bilinearity both $\widehat{U}_k^N(\theta)$ and $\widetilde{U}_k^N(\theta)$ are invariant under this procedure. r_N is also unchanged by this application because of its definition. Moreover, by construction we have

$$\max_{J\in\tilde{\Pi}_N^2}\sum_{I\in\tilde{\Pi}_N^1} K(I,J_{-\theta}) \leq 3, \qquad \max_{I\in\tilde{\Pi}_N^1}\sum_{J\in\tilde{\Pi}_N^2} K(I,J_{-\theta}) \leq 3.$$

Consequently, for the proof we may replace (Π_N^1,Π_N^2) by $(\tilde{\Pi}_N^1,\tilde{\Pi}_N^2)$. This allows us to assume that

$$\max_{J\in\Pi_N^2}\sum_{I\in\Pi_N^1} K(I,J_{-\theta}) \leq 3, \qquad \max_{I\in\Pi_N^1}\sum_{J\in\Pi_N^2} K(I,J_{-\theta}) \leq 3. \tag{8}$$

throughout the proof without loss of generality.

We turn to the main body of the proof. We decompose the target quantity as

$$\begin{aligned}
&\widehat{U}_k^N(\theta) - c_k^N(\theta)\widetilde{U}_k^N(\theta) \\
&= \sum_{I,J:\underline{I}\in I_m(k)}\left\{\int_I(\sigma_s^1 - \sigma_{k\tau_m}^1)dB_s^1X^2(J) + \sigma_{k\tau_m}^1B^1(I)\int_J(\sigma_s^2 - \sigma^2_{(k\tau_m+\theta-r_N)_+})dB_s^2\right\}K(I,J_{-\theta}) \\
&=: \mathbf{A}_N + \mathbf{B}_N.
\end{aligned}$$

Let us consider $\mathbf{A}_N$. The Minkovski, Schwarz and BDG inequalities yield

$$
\begin{aligned}
E^{\Pi}\left[|\mathbf{A}_N|^p\right] &\leq \left\{\sum_{I,J:\underline{I}\in I_m(k)}\left(E^{\Pi}\left[\left|\int_I(\sigma_s^1-\sigma_{k\tau_m}^1)dB_s^1X^2(J)\right|^p\right]\right)^{1/p}K(I,J_{-\theta})\right\}^p \\
&\leq \left\{\sum_{I,J:\underline{I}\in I_m(k)}\left(E^{\Pi}\left[\left|\int_I(\sigma_s^1-\sigma_{k\tau_m}^1)dB_s^1\right|^{2p}\right]\right)^{1/2p}\left(E^{\Pi}\left[\left|X^2(J)\right|^{2p}\right]\right)^{1/2p}K(I,J_{-\theta})\right\}^p \\
&\lesssim \left\{\sum_{I,J:\underline{I}\in I_m(k)}\left\|\sup_{s\in I_m(k)}|\sigma_s^1-\sigma_{k\tau_m}^1|\right\|_{2p}\sqrt{|I||J|}K(I,J_{-\theta})\right\}^p,
\end{aligned}
$$

hence by assumption and (8) we obtain

$$
E^{\Pi}\left[|\mathbf{A}_N|^p\right]\lesssim\tau_m^{p\gamma}\left\{\sum_{I,J:\underline{I}\in I_m(k)}(|I|+|J|)K(I,J_{-\theta})\right\}^p\lesssim\tau_m^{p(\gamma+1)}.
$$

By symmetry we also have $E^{\Pi}\left[|\mathbf{B}_N|^p\right]\lesssim\tau_m^{p(\gamma+1)}$. This completes the proof. $\square$

Let us take a number $\xi\in(\frac{\beta+4}{5},1)$ and set $u_N=(\tau_N^{(5\xi-4+\beta)/2}|\log\tau_N|)^{-1}$.

Lemma 4. *There is a constant C such that*

$$
E^{\Pi}\left[\exp\left(\varsigma u_N\left\{\widetilde{U}_k^N(\theta)-\overline{U}_k^N(\theta)\right\}\right)\right]\leq C
$$

for any $N\in\mathbb{N}$, $k\in\mathbb{Z}_+$ and $\theta\in(-\delta,\delta)$.

Proof. Again, by symmetry it suffices to consider the case of $\theta\geq 0$. Moreover, as in the proof of Lemma 3, we may assume (8) without loss of generality.

Let Σ_N be the covariance matrix of $(B^1(I))_{I\in\Pi_N^1:\underline{I}\in I_m(k)},B^2(J))_{J\in\Pi_N^2:\underline{J}\in\tilde{I}_m(k)})^{\top}$, where $\tilde{I}_m(k)=I_m(k)+\theta-r_N$, and set $C_N=\Sigma_N^{1/2}A_N\Sigma_N^{1/2}$, where

$$
A_N=\begin{pmatrix}0 & K_N\\ K_N^{\top} & 0\end{pmatrix},\qquad K_N=(K(I,J_{-\theta})/2)_{(I,J)\in\Pi_N^1\times\Pi_N^2:\underline{I}\in I_m(k),\underline{J}\in\tilde{I}_m(k)}.
$$

We first prove the following equations:

$$
\|C_N\|_{\mathrm{sp}}=O(\tau_N^{-2+3\xi}|\log\tau_N|^2),\qquad\|C_N\|_F^2=O(\tau_N^{-4+5\xi+\beta}|\log\tau_N|^2),\tag{9}
$$

where $\|\cdot\|_{\mathrm{sp}}$ denotes the spectral norm of matrices. By Theorem 5.6.9 from [20] and (8), we have $\|A_N\|_{\mathrm{sp}}\leq\frac{3}{2}$. Therefore, Appendix II(ii)–(iii) from [7] yield

$$
\begin{aligned}
\|C_N\|_F^2\leq\frac{9}{4}\|\Sigma_N\|_F^2=\frac{9}{4}&\sum_{I\in\Pi_N^1:\underline{I}\in I_m(k)}E^{\Pi}\left[B^1(I)^2\right]^2+\frac{9}{4}\sum_{J\in\Pi_N^2:\underline{J}\in\tilde{I}_m(k)}E^{\Pi}\left[B^2(J)^2\right]^2\\
&+\frac{9}{2}\sum_{(I,J)\in\Pi_N^1\times\Pi_N^2:\underline{I}\in I_m(k),\underline{J}\in\tilde{I}_m(k)}E^{\Pi}\left[B^1(I)B^2(J)\right]^2,
\end{aligned}
$$

$$\lesssim r_N\tau_m + \sum_{(I,J)\in\Pi_N^1\times\Pi_N^2:\underline{I}\in I_m(k),\underline{J}\in\tilde{I}_m(k)} E^{\Pi}\left[B^1(I)B^2(J)\right]^2$$

while Corollary 4.5.11 and Theorem 5.6.9 from [20] imply that

$$\begin{aligned}\|C_N\|_{\mathrm{sp}} \le \frac{3}{2}\|\Sigma_N\|_{\mathrm{sp}} &\le \frac{3}{2}\max\left\{\max_{I\in\Pi_N^1:\underline{I}\in I_m(k)} E^{\Pi}\left[B^1(I)^2\right], \max_{J\in\Pi_N^2:\underline{J}\in\tilde{I}_m(k)} E^{\Pi}\left[B^2(J)^2\right]\right\}\\ &\quad + \frac{3}{2}\max_{I\in\Pi_N^1:\underline{I}\in I_m(k)}\sum_{J\in\Pi_N^2:\underline{J}\in\tilde{I}_m(k)}\left|E^{\Pi}\left[B^1(I)B^2(J)\right]\right|\\ &\lesssim r_N + \max_{I\in\Pi_N^1:\underline{I}\in I_m(k)}\sum_{J\in\Pi_N^2:\underline{J}\in\tilde{I}_m(k)}\left|E^{\Pi}\left[B^1(I)B^2(J)\right]\right|.\end{aligned}$$

It holds that

$$\begin{aligned}&E^{\Pi}\left[B^1(I)B^2(J)\right]\\ &= \frac{1}{2\pi\tau_N}\int_{-\infty}^{\infty}(\mathcal{F}1_I)(\tau_N^{-1}\lambda)\overline{(\mathcal{F}1_J)(\tau_N^{-1}\lambda)}f_J(\tau_N^{-1}\lambda)d\lambda\\ &= \frac{1}{2\pi\tau_N}\sum_{i=1}^{N+1}R_i\int_{\Lambda_{-i}}(\mathcal{F}1_I)(\tau_N^{-1}\lambda)\overline{(\mathcal{F}1_J)(\tau_N^{-1}\lambda)}e^{-\sqrt{-1}\tau_N^{-1}\lambda\theta_i}d\lambda\\ &= \frac{1}{2\pi\tau_N}\sum_{i=1}^{N+1}R_i\int_{\Lambda_{-i}}\left(\int_0^{|I|}e^{-\sqrt{-1}\tau_N^{-1}\lambda s}ds\right)\left(\int_0^{|J|}e^{\sqrt{-1}\tau_N^{-1}\lambda s}ds\right)e^{\sqrt{-1}\tau_N^{-1}\lambda(\underline{J}-\underline{I}-\theta_i)}d\lambda.\end{aligned}$$

Since we have

$$\begin{aligned}&\frac{d}{d\lambda}\left\{\left(\int_0^{|I|}e^{-\sqrt{-1}\tau_N^{-1}\lambda s}ds\right)\left(\int_0^{|J|}e^{\sqrt{-1}\tau_N^{-1}\lambda s}ds\right)\right\}\\ &= -\sqrt{-1}\tau_N^{-1}\left(\int_0^{|I|}se^{-\sqrt{-1}\tau_N^{-1}\lambda s}ds\right)\left(\int_0^{|J|}e^{\sqrt{-1}\tau_N^{-1}\lambda s}ds\right)\\ &\quad + \sqrt{-1}\tau_N^{-1}\left(\int_0^{|I|}e^{-\sqrt{-1}\tau_N^{-1}\lambda s}ds\right)\left(\int_0^{|J|}se^{\sqrt{-1}\tau_N^{-1}\lambda s}ds\right),\end{aligned}$$

we obtain

$$\left|\frac{d}{d\lambda}\left\{\left(\int_0^{|I|}e^{-\sqrt{-1}\tau_N^{-1}\lambda s}ds\right)\left(\int_0^{|J|}e^{\sqrt{-1}\tau_N^{-1}\lambda s}ds\right)\right\}\right| \le \frac{\tau_N^{-1}}{2}(|I|^2|J| + |I||J|^2).$$

Therefore, integration by parts implies that

$$\begin{aligned}&\left|\int_{\Lambda_{-i}}\left(\int_0^{|I|}e^{-\sqrt{-1}\tau_N^{-1}\lambda s}ds\right)\left(\int_0^{|J|}e^{\sqrt{-1}\tau_N^{-1}\lambda s}ds\right)e^{\sqrt{-1}\tau_N^{-1}\lambda(\underline{J}-\underline{I}-\theta_i)}d\lambda\right|\\ &\le \left\{|I||J| + \frac{\tau_N^{-1}}{2}(|I|^2|J| + |I||J|^2)\right\}\frac{2^{-i+1}\pi}{\tau_N^{-1}|\underline{J}-\underline{I}-\theta_i|}\end{aligned}$$

as long as $\underline{J} - \underline{I} \neq \theta_i$. Hence we obtain

$$\sum_{J \in \Pi_N^2 : \underline{J} \in \tilde{I}_m(k)} \left| E^{\Pi} \left[B^1(I) B^2(J) \right] \right| \lesssim \tau_N^{-1} r_N^2 + \tau_N^{-2} r_N^3 \sum_{i=1}^{N+1} \sum_{\substack{J \in \Pi_N^2 : \underline{J} \in \tilde{I}_m(k) \\ \underline{J} - \underline{I} - \theta_i \neq 0}} \frac{1}{\tau_N^{-1} |\underline{J} - \underline{I} - \theta_i|}$$
$$= O(\tau_N^{-2+3\xi} |\log \tau_N|^2)$$

by Assumption 2(i). Consequently, we obtain the first equation of (9). Moreover, it holds that

$$\begin{aligned}
&\sum_{(I,J) : \underline{I} \in I_m(k), \underline{J} \in \tilde{I}_m(k)} \left| E^{\Pi} \left[B^1(I) B^2(J) \right] \right|^2 \\
&\leq \frac{N+1}{2\pi \tau_N} \sum_{i=1}^{N+1} \sum_{(I,J) : \underline{I} \in I_m(k), \underline{J} \in \tilde{I}_m(k)} \left| \int_{\Lambda_{-i}} \left(\int_0^{|I|} e^{-\sqrt{-1} \tau_N^{-1} \lambda s} ds \right) \left(\int_0^{|J|} e^{\sqrt{-1} \tau_N^{-1} \lambda s} ds \right) e^{\sqrt{-1} \tau_N^{-1} \lambda (\underline{J} - \underline{I} - \theta_i)} d\lambda \right|^2 \\
&\lesssim N \tau_N^{-2} r_N^3 \tau_m + N \tau_N^{-2} \sum_{i=1}^{N+1} \sum_{\substack{(I,J) : \underline{I} \in I_m(k), \underline{J} \in \tilde{I}_m(k) \\ \underline{J} - \underline{I} - \theta_i \neq 0}} \left\{ |I||J| + \frac{\tau_N^{-1}}{2} (|I|^2 |J| + |I||J|^2) \right\}^2 \frac{1}{(\tau_N^{-1} |\underline{J} - \underline{I} - \theta_i|)^2} \\
&\lesssim N \tau_N^{-2} r_N^3 \tau_m + N^2 \tau_N^{-4} r_N^5 \tau_m = O(\tau_N^{-4+5\xi+\beta} |\log \tau_N|^2),
\end{aligned}$$

hence we obtain the second equation of (9).

Now, noting that $-2 + 3\xi - (5\xi - 4 + \beta)/2 = (\xi - \beta)/2 > 2(1 - \beta)/5 > 0$, from the discussion in Section 3.2.1 of [5], we have

$$\log E^{\Pi} \left[\exp \left(\varsigma u_N \left\{ \widetilde{U}_k^N(\theta) - \overline{U}_k^N(\theta) \right\} \right) \right] = -\frac{1}{2} \log \det(E - 2\varsigma u_N C_N) - \mathrm{tr}[\varsigma u_N C_N]$$

for sufficiently large N due to the first equation of (9). Therefore, by Appendix II-(v) from [7] we obtain

$$\log E^{\Pi} \left[\exp \left(\varsigma u_N \left\{ \widetilde{U}_k^N(\theta) - \overline{U}_k^N(\theta) \right\} \right) \right] \leq \frac{u_N^2}{2} \|C_N\|_F^2 + \frac{|u_N|^3}{3} \frac{\|C_N\|_{\mathrm{sp}} \|C_N\|_F^2}{(1 - \|C_N\|_{\mathrm{sp}})^3}$$

for sufficiently large N due to the first equation of (9). Consequently, (9) yields the desires result. □

Lemma 5. *We have* $\left| \overline{U}_k^N(\theta) \right| \leq 6(\tau_m + r_N)$ *for any* $N \in \mathbb{N}$, $k \in \mathbb{Z}_+$ *and* $\theta \in (-\delta, \delta)$.

Proof. Similarly to the above proofs, we may assume $\theta \geq 0$ and that (8) holds true without loss of generality. Then we have

$$\begin{aligned}
\left| \overline{U}_k^N(\theta) \right| &= \left| \sum_{I,J : \underline{I} \in I_m(k), \underline{J} \in \tilde{I}_m(k)} E^{\Pi} \left[B^1(I) B^2(J) \right] K(I, J_{-\theta}) \right| \\
&\leq \sum_{I,J : \underline{I} \in I_m(k), \underline{J} \in \tilde{I}_m(k)} \left\{ E^{\Pi} \left[B^1(I)^2 \right] + E \left[B^2(J)^2 \right] \right\} K(I, J_{-\theta}) \\
&\leq 3 \left\{ \sum_{I : \underline{I} \in I_m(k)} |I| + \sum_{I : \underline{J} \in \tilde{I}_m(k)} |J| \right\} \leq 3 \cdot 2(\tau_m + \bar{r}_N).
\end{aligned}$$

This completes the proof. □

Lemma 6. *We have*

$$\max_{\theta\in\mathcal{G}^N}\left|\widehat{\rho}_{(j)}(\theta)-\tau_m\sum_{k=0}^{M_N-1}c_k^N(\theta)\int_{-\pi}^{\pi}D(\lambda)H_{j,L}(\lambda)e^{\sqrt{-1}\lambda\theta\tau_N^{-1}}f_N(\lambda/\tau_N)d\lambda\right|\to^p 0$$

as $N\in\mathbb{N}$.

Proof. We decompose the target quantity as

$$\begin{aligned}
&\widehat{\rho}_{(j)}(\theta)-\tau_m\sum_{k=0}^{M_N-1}c_k^N(\theta)\int_{-\pi}^{\pi}D(\lambda)H_{j,L}(\lambda)e^{\sqrt{-1}\lambda\theta\tau_N^{-1}}f_N(\lambda/\tau_N)d\lambda\\
&=\left(\widehat{\rho}_{(j)}(\theta)-\sum_{l=-L_j-1}^{L_j-1}\Psi_j(l)\sum_{k=0}^{M_N-1}c_k^N(\theta-l\tau_N)\widetilde{U}_k^N(\theta-l\tau_N)\right)\\
&\quad+\sum_{l=-L_j+1}^{L_j-1}\Psi_j(l)\sum_{k=0}^{M_N-1}c_k^N(\theta-l\tau_N)\left\{\widetilde{U}_k^N(\theta-l\tau_N)-\overline{U}_k^N(\theta-l\tau_N)\right\}\\
&\quad+\sum_{l=-L_j+1}^{L_j-1}\Psi_j(l)\sum_{k=0}^{M_N-1}\left\{c_k^N(\theta-l\tau_N)-c_k^N(\theta)\right\}\overline{U}_k^N(\theta-l\tau_N)\\
&\quad+\sum_{k=0}^{M_N-1}c_k^N(\theta)\left(\sum_{l=-L_j+1}^{L_j-1}\Psi_j(l)\overline{U}_k^N(\theta-l\tau_N)-\tau_m\int_{-\pi}^{\pi}D(\lambda)H_{j,L}(\lambda)e^{\sqrt{-1}\lambda\theta\tau_N^{-1}}f_N(\lambda/\tau_N)d\lambda\right)\\
&=:\mathbf{I}_N(\theta)+\mathbf{II}_N(\theta)+\mathbf{III}_N(\theta)+\mathbf{IV}_N(\theta).
\end{aligned}$$

First, since we have $\sum_{p=0}^{L_j-1}h_{j,p}^2=1$, it holds that $|\Psi_j(l)|\leq 1$ for every l by the Schwarz inequality. Therefore, we have

$$|\mathbf{I}_N(\theta)|\leq\sum_{l=-L_j-1}^{L_j-1}\sum_{k=0}^{M_N-1}\left|\widehat{U}_k^N(\theta-l\tau_N)-c_k^N(\theta-l\tau_N)\widetilde{U}_k^N(\theta-l\tau_N)\right|,$$

for any $\varepsilon>0$ and $p>1$ we obtain

$$\begin{aligned}
&P\left(\max_{\theta\in\mathcal{G}^N}|\mathbf{I}_N(\theta)|>\varepsilon\right)\\
&\leq\left(\frac{(2L_j-1)M_N}{\varepsilon}\right)^p\sum_{\theta\mathcal{G}^N}\sum_{l=-L_j-1}^{L_j-1}\sum_{k=0}^{M_N-1}E\left[\left|\widehat{U}_k^N(\theta-l\tau_N)-c_k^N(\theta-l\tau_N)\widetilde{U}_k^N(\theta-l\tau_N)\right|^p\right]\\
&\lesssim\tau_N^{-1}LM_N\cdot(LM_N)^p\tau_m^{(\gamma+1)p}=O(\tau_N^{-1}L^{1+p}\tau_m^{p\gamma-1})
\end{aligned}$$

by the Markov inequality and Lemma 3. Since we can take the number p large enough such that $\tau_N^{-1}L^{1+p}\tau_m^{p\gamma-1}\to 0$ as $N\to\infty$ by assumption, we obtain $\max_{\theta\in\mathcal{G}^N}|\mathbf{I}_N(\theta)|\to^p 0$.

Next, for any $\varepsilon>0$ we have

$$P\left(\max_{\theta\in\mathcal{G}^N}|\mathbf{II}_N(\theta)|>\varepsilon\right)$$

$$\leq \sum_{\theta \in \mathcal{G}^N} \sum_{l=-L_j+1}^{L_j-1} \sum_{k=0}^{M_N-1} P\left(\left| \widetilde{U}_k^N(\theta - l\tau_N) - \overline{U}_k^N(\theta - l\tau_N) \right| > \frac{\varepsilon}{KLM_N} \right)$$

with some constant $K > 0$. Therefore, the Markov inequality and Lemma 4 yield

$$P\left(\max_{\theta \in \mathcal{G}^N} |\mathbf{II}_N(\theta)| > \varepsilon \right) \lesssim \tau_N^{-1} LM_N \exp\left(-\frac{\varepsilon u_N}{KLM_N} \right).$$

Since $\tau_N^c u_N / M_N \to \infty$ as $N \to \infty$ for some $c > 0$, by assumption we conclude $\max_{\theta \in \mathcal{G}^N} |\mathbf{II}_N(\theta)| \to^p 0$.

Now we prove $\max_{\theta \in \mathcal{G}^N} |\mathbf{III}_N(\theta)| \to^p 0$. Since $\sum_{l=-L_j+1}^{L_j-1} |\Psi_j(l)| \leq 2L_j - 1$, we have

$$\begin{aligned}
&\max_{\theta \in \mathcal{G}^N} |\mathbf{III}_N(\theta)| \\
&\leq \left((2L_j - 1) \max_{\theta \in \mathcal{G}^N} \max_{l \in \mathbb{Z}: |l| < L_j} \max_{k=0,1,\dots,M_N-1} \left| c_k^N(\theta - l\tau_N) - c_k^N(\theta) \right| \right) \sum_{k=0}^{M_N-1} \max_{\theta \in \mathcal{G}^N} \left| \overline{U}_k^N(\theta) \right|.
\end{aligned}$$

By the Hölder continuity of σ^1, σ^2 and the assumption on L, we have

$$(2L_j - 1) \max_{\theta \in \mathcal{G}^N} \max_{l \in \mathbb{Z}: |l| < L_j} \max_{k=0,1,\dots,M_N-1} \left| c_k^N(\theta - l\tau_N) - c_k^N(\theta) \right| \to^p 0$$

as $N \to \infty$, while Lemma 5 yields $\sum_{k=0}^{M_N-1} \max_{\theta \in \mathcal{G}^N} \left| \overline{U}_k^N(\theta) \right| = O_p(1)$. Hence we obtain the desired result.

Finally we prove $\max_{\theta \in \mathcal{G}^N} |\mathbf{IV}_N(\theta)| \to^p 0$. Noting that

$$\int_{-\pi}^{\pi} D(\lambda) H_{j,L}(\lambda) e^{\sqrt{-1}\lambda\theta\tau_N^{-1}} f_N(\lambda/\tau_N) d\lambda = \sum_{l=-L_j-1}^{L_j-1} \Psi_j(l) \int_{-\pi}^{\pi} D(\lambda) e^{\sqrt{-1}\lambda(\theta - l\tau_N)\tau_N^{-1}} f_N(\lambda/\tau_N) d\lambda,$$

for any $\varepsilon > 0$ we have

$$\begin{aligned}
&P\left(\max_{\theta \in \mathcal{G}^N} |\mathbf{IV}_N(\theta)| > \varepsilon \right) \\
&\leq \sum_{\theta \in \mathcal{G}^N} \sum_{l=-L_j+1}^{L_j-1} P\left(\tau_m \sum_{k=0}^{M_N-1} \left| \tau_m^{-1} \overline{U}_k^N(\theta) - \int_{-\pi}^{\pi} D(\lambda) e^{\sqrt{-1}\lambda(\theta - l\tau_N)\tau_N^{-1}} f_N(\lambda/\tau_N) d\lambda \right| > \frac{\varepsilon}{KL} \right)
\end{aligned}$$

with some constant $K > 0$. Since we have

$$\overline{U}_k^N(\theta - l\tau_N) = \tau_m \int_{-\pi}^{\pi} D_k^N(\lambda, \theta - l\tau_N) e^{\sqrt{-1}\lambda(\theta - l\tau_N)\tau_N^{-1}} f_N(\lambda/\tau_N) d\lambda,$$

it holds that

$$\begin{aligned}
&E\left[\left\{ \tau_m \sum_{k=0}^{M_N-1} \left| \tau_m^{-1} \overline{U}_k^N(\theta) - \int_{-\pi}^{\pi} D(\lambda) e^{\sqrt{-1}\lambda(\theta - l\tau_N)} f_N(\lambda/\tau_N) d\lambda \right| \right\}^Q \right] \\
&\leq (2\pi)^{Q-1} \tau_m \sum_{k=0}^{M_N-1} \int_{-\pi}^{\pi} E\left[\left| D_k^N(\lambda, \theta - l\tau_N) - D(\lambda) \right|^Q \right] d\lambda
\end{aligned}$$

by the Jensen inequality. Therefore, by the Markov inequality we obtain

$$P\left(\max_{\theta\in\mathcal{G}^N}|\mathbf{IV}_N(\theta)|>\varepsilon\right)\lesssim\tau_N^{-1}L^{1+Q}\tau_m\max_{\theta\in\mathcal{G}^N}\sum_{k=0}^{M_N-1}\int_{-\pi}^{\pi}E\left[\left|D_k^N(\lambda,\theta)-D(\lambda)\right|^Q\right]d\lambda.$$

Consequently, Assumption 2 and the assumption on L imply the desired result. This completes the proof. □

Proof of Theorem 1. (a) From Lemma 6 it is enough to prove

$$\max_{\theta\in\mathcal{G}^N:|\theta-\theta_j|\geq v_N}\left|\tau_m\sum_{k=0}^{M_N-1}c_k^N(\theta)\int_{-\pi}^{\pi}D(\lambda)H_{j,L}(\lambda)e^{\sqrt{-1}\lambda\theta\tau_N^{-1}}f_N(\lambda/\tau_N)d\lambda\right|\to^p 0$$

as $N\to\infty$. The above equation follows once we show the following statements: If $\vartheta_N\in\mathcal{G}^N$ ($N=1,2,\dots$) satisfy $|\vartheta_N-\theta_j|\geq v_N$ for every N, then

$$\tau_m\sum_{k=0}^{M_N-1}c_k^N(\vartheta_N)\int_{-\pi}^{\pi}D(\lambda)H_{j,L}(\lambda)e^{\sqrt{-1}\lambda\vartheta_N\tau_N^{-1}}f_N(\lambda/\tau_N)d\lambda\to^p 0$$

as $N\to\infty$. This can be shown in an analogous manner to the proof of Lemma 6 from [15].

(b) From Lemma 6 it suffices to prove

$$\tau_m\sum_{k=0}^{M_N-1}c_k^N(\theta)\int_{-\pi}^{\pi}D(\lambda)H_{j,L}(\lambda)e^{\sqrt{-1}\lambda\theta\tau_N^{-1}}f_N(\lambda/\tau_N)d\lambda\to^p R_j\int_{\Lambda_{-j}}D(\lambda)\cos(b\lambda)d\lambda$$

as $N\to\infty$, which can be shown in an analogous manner to the proof of Lemma 7 from [15]. □

8.3 Proof of Theorem 2

Noting that $\int_{\Lambda_{-j}}D(\lambda)\cos(b\lambda)d\lambda>0$ for any $b\in[-\frac{1}{2},\frac{1}{2}]$ by assumption, the theorem can be shown in an analogous manner to the proof of Theorem 2 from [15] (using Theorem 1 instead of Lemmas 7–8 from [15]). □

Acknowledgments

Takaki Hayashi's research was supported by JSPS Grant-in-Aid for Scientific Research (C) Grant Number JP16K03601. Yuta Koike's research was supported by JST, CREST and JSPS Grant-in-Aid for Young Scientists (B) Grant Number JP16K17105.

References

[1] Aït-Sahalia, Y. and Jacod, J. (2014). *High-frequency financial econometrics*. Princeton University Press.

[2] Barunik, J. and Vacha, L. (2015). Realized wavelet-based estimation of integrated variance and jumps in the presence of noise. *Quant. Finance* **15**, 1347–1364.

[3] Chan, K. (1992). A further analysis of the lead–lag relationship between the cash market and stock index futures market. *Review of Financial Studies* **5**, 123–152.

[4] Christensen, K., Podolskij, M., Thamrongrat, N. and Veliyev, B. (2017). Inference from high-frequency data: A subsampling approach. *J. Econometrics* **197**, 245–272.

[5] Dalalyan, A. and Yoshida, N. (2011). Second-order asymptotic expansion for a non-synchronous covariation estimator. *Ann. Inst. Henri Poincaré Probab. Stat.* **47**, 748–789.

[6] Daubechies, I. (1992). *Ten lectures on wavelets*. SIAM.

[7] Davies, R. B. (1973). Asymptotic inference in stationary Gaussian time-series. *Adv. in Appl. Probab.* **5**, 469–497.

[8] de Jong, F. and Nijman, T. (1997). High frequency analysis of lead-lag relationships between financial markets. *Journal of Empirical Finance* **4**, 259–277.

[9] Dobrev, D. and Schaumburg, E. (2016). High-frequency cross-market trading: Model free measurement and applications. Working paper.

[10] Gençay, R., Gradojevic, N., Selçuk, F. and Whitcher, B. (2010). Asymmetry of information flow between volatilities across time scales. *Quant. Finance* **10**, 895–915.

[11] Hafner, C. M. (2012). Cross-correlating wavelet coefficients with applications to high-frequency financial time series. *J. Appl. Stat.* **39**, 1363–1379.

[12] Harris, F. H. d., McInish, T. H. and Wood, R. A. (2002). Security price adjustment across exchanges: an investigation of common factor components for Dow stocks. *Journal of Financial Markets* **5**, 277–308.

[13] Hasbrouck, J. (1995). One security, many markets: Determining the contributions to price discovery. *Journal of Finance* **50**, 1175–1199.

[14] Hasbrouck, J. (2015). High frequency quoting: Short-term volatility in bids and offers. Working paper, available at SSRN: http://ssrn.com/abstract=2237499.

[15] Hayashi, T. and Koike, Y. (2016). Wavelet-based methods for high-frequency lead-lag analysis. Working paper. Available at arXiv: https://arxiv.org/abs/1612.01232.

[16] Hayashi, T. and Yoshida, N. (2005). On covariance estimation of non-synchronously observed diffusion processes. *Bernoulli* **11**, 359–379.

[17] Hayashi, T. and Yoshida, N. (2008). Asymptotic normality of a covariance estimator for nonsynchronously observed diffusion processes. *Ann. Inst. Statist. Math.* **60**, 357–396.

[18] Hayashi, T. and Yoshida, N. (2011). Nonsynchronous covariation process and limit theorems. *Stochastic Process. Appl.* **121**, 2416–2454.

[19] Hoffmann, M., Rosenbaum, M. and Yoshida, N. (2013). Estimation of the lead-lag parameter from non-synchronous data. *Bernoulli* **19**, 426–461.

[20] Horn, R. A. and Johnson, C. R. (2013). *Matrix analysis*. Cambridge University Press, 2nd edn.

[21] Huth, N. and Abergel, F. (2014). High frequency lead/lag relationships — empirical facts. *Journal of Empirical Finance* **26**, 41–58.

[22] Lai, M.-J. (1995). On the digital filter associated with Daubechies' wavelets. *IEEE Trans. Signal Process.* **43**, 2203–2205.

[23] Lo, A. W. and MacKinlay, A. C. (1990). An econometric analysis of nonsynchronous trading. *J. Econometrics* **45**, 181–211.

[24] Müller, U. A., Dacorogna, M. M., Davé, R. D., Olsen, R. B., Pictet, O. V. and von Weizsäcker, J. E. (1997). Volatilities of different time resolutions — analyzing the dynamics of market components. *Journal of Empirical Finance* **4**, 213–239.

[25] Müller, U. A., Dacorogna, M. M., Davé, R. D., Pictet, O. V., Olsen, R. B. and Ward, J. R. (1993). Fractals and intrinsic time — a challenge to econometricians. Tech. Rep. UAM.1993-08-16, Olsen and Associates.

[26] Nason, G. P., von Sachs, R. and Kroisandt, G. (2000). Wavelet processes and adaptive estimation of the evolutionary wavelet spectrum. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **62**, 271–292.

[27] Ozturk, S. R., van der Wel, M. and van Dijk, D. (2017). Intraday price discovery in fragmented markets. *Journal of Financial Markets* **32**, 28–48.

[28] Percival, D. B. and Walden, A. T. (2000). *Wavelet methods for time series analysis*. Cambridge University Press.

[29] Renò, R. (2003). A closer look at the Epps effect. *Int. J. Theor. Appl. Finance* **6**, 87–102.

[30] Subbotin, A. (2008). A multi-horizon scale for volatility. Unpublished paper. Available at HAL: https://halshs.archives-ouvertes.fr/halshs-00261514.

[31] Subrahmanyam, A. (1997). Multi-market trading and the informativeness of stock trades: An empirical intraday analysis. *Journal of Economics and Business* **49**, 515–531.

[32] Vidakovic, B. (1999). *Statistical modeling by wavelets*. John Wiley & Sons, Inc.

マクロ消費の状態推定問題

国友直人

明治大学政治経済学部

2017 年 8 月

Outline

マクロ消費の動向

この研究は統計局「速報性のある包括的な消費関連指標の在り方に関する研究会」を巡る議論から発展したものであり、佐藤整尚・栗栖大輔・粟屋直の諸氏との共同研究を含んでいる。統計局の研究会の詳細については報告書「消費動向指数 (CTI) に向けて」で HP 上に公開されている。

時間の経過とともに多数の経済時系列が観察されるが、伝統的には多くのマクロ経済時系列データの場合にはデータの収集上・作成上の理由などから月次系列、四半期系列、年次系列などの離散時間単位で計測され、公表されている。しかしながら近年では新しい計測方法、情報処理の利便性が高まり、月次よりも高頻度の観測データも得られるようになっている経済データも存在する。こうした従来よりもより高頻度の時系列データが利用可能になりつつあることにつれてより細かな市場動向の理解も進むと考えられるが、他方、経済時系列ではマクロ経済の大局的な動きにも関心があり、月次、四半期、年次の動きと整合的な統計分析も重要である。

日次データの分解モデル

時刻 t の一次元経済時系列 y_t とする。週次周期 $s=7$ は固定、月次周期 $m=28,29,30,31$ および四半期周期はともに変動することを考慮する必要がある。日次単位からは季節周期の構成日数は変動する。時系列 y_t の加法的分解モデル

$$y_t = x_t + s_t^w + s_t^m + h_t + v_t \quad (t=1,\cdots,T),$$

を考察する。x_t はトレンド成分、s_t^w は週次成分、s_t^m は月次成分、h_t が特別休日成分、v_t は不規則変動である。加法的分解モデルを採用し、まｚｙ簡単化の為に循環成分をゼロとする。不規則成分 v_t は $N(0,\sigma^2)$ にしたがう互いに独立な確率変数、トレンド成分は

$$x_t = x_{t-1} + v_t^x \quad (t=1,\cdots,T),$$

にしたがい、不規則成分 v_t^x は確率分布 $N(0,\sigma_x^2)$ にしたがう互いに独立な確率変数、なお Kitagawa (2010) を参考とする。

週次成分は

$$(1+L+\cdots+L^6)s_t^w = v_t^w \quad (t=1,\cdots,T),$$

とする。L はラグ作用素、不規則成分 v_t^w は確率分布 $N(0,\sigma_w^2)$ にしたがう互いに独立な確率変数とする。月次成分は

$$(1+L+\cdots+L^{11})[\sum_{t\in I_i(t)} s_t^m] = v_{I_i(t)}^m \quad (t=1,\cdots,T),$$

とする。月次時系列に基づく季節調整法とは整合的ではあるが、月次状態成分は退化しているので、結局 $s_{I_i(t)}^m = \sum_{t\in I_i(t)} s_i^m$ を推定すればよい。不規則成分 s_i^m を確率分布 $N(0,\sigma_m^2)$ にしたがう互いに独立な確率変数とすると、$v_{I_i(t)}^m$ の分散は σ_m^2 に比例する。
統計的問題として新しい観点は月次成分の扱いであり、状態空間表現では制約条件はかなり疎であるが、時間に依存、高次元問題となる。

特別休暇部分 $h_t = 0$, 月次効果 $s_t = 0$ とすると加法モデル

$$y_t = x_t + s_t^w + v_t \quad (t = 1, \cdots, T)$$

は decomp モデルにおいて季節周期 $s = 7$ に対応。そこでカルマンフィルタを利用して状態変数 $\hat{x}_t(1), \hat{s}_t^w(1)$ $(t = 1, \cdots, T)$ を推定する。分散の最尤推定値を $\hat{\sigma}^2(1), \hat{\sigma}_x^2(1), \hat{\sigma}_w^2(1)$ として、次に観測値 $Y_i = \sum_{t \in I_i(t)}$ $(i = 1, \cdots, n)$ に対して加法モデル

$$Y_i = \sum_{t \in I_i(t)} x_t + \sum_{t \in I_i(t)} s_t^w + \sum_{t \in I_i(t)} s_t^m + \sum_{t \in I_i(t)} v_t \quad (i = 1, \cdots, n),$$

にカルマンフィルタリングを適用、月次指標 $s_{I_i(t)}^m(1) = \sum_{t \in I_i(t)} s_i^m$ を推定する。$c_i = \{\#t | t \in I_i(t)\}$ とすると右辺の分散は c_i $(i = 1, \cdots, n)$ に比例する不均一分散の時系列分解モデル、推定値 $\hat{\sigma}_m^2(1)$ を得る。

この二段階フィルタリング操作を繰り返えすことより日次データ $\{y_t\}$ から求める尤度関数は $k \to \infty$ のとき

$$L_T(\hat{\sigma}^2(k), \hat{\sigma}_x^2(k), \hat{\sigma}_w^2(k), \hat{\sigma}_m^2(k)) \xrightarrow{p} L_T(\hat{\sigma}^2, \hat{\sigma}_x^2, \hat{\sigma}_w^2, \hat{\sigma}_m^2)$$

として状態変数の推定値を得る。重要な論点は集計されたデータ（すなわち月次データや四半期データ）から推定される時系列分解は曜日効果つきの季節調整モデルと整合的、日次データより季節調整も同時に行うことが可能。

特別休日効果 (special holiday effects)h_t は日本の正月やお盆休みなどはカレンダーによって決定されるが曜日周期 7 には依存しない。回帰項

$$h_t = \sum_{t \in D_j} \beta_{D_j(t)} x_t I(t \in D_j) + v_t^h$$

と表現する。v_t^h は休日効果ノイズ、$I(t \in D_j)$ は j 番目の特定休日を表すダミー変数、係数 $\beta_{D_j(t)}$ は時間とともに変化するなど様々な定式化が可能。

消費の日次データ

利用するデータは 2000 年 1 月から 2016 年 10 月までの約 6000 の日次データである。一般にはあまり知られていないが、総務省統計局では 2000 年から日次データの集計を開始しているが、一般には月次データ、四半期データを様々な用途で基礎的に消費データとして利用されている。日次 2010.1.1 から 500 個のデータ、月次 2000.1-2016.12 のデータ。

月次マクロ指標

マクロ経済データの場合には収集上・作成上の理由から日々の系列、月次系列、四半期系列、年次系列など様々な頻度と異なる時間的タイミングで計測され、公表。消費、投資、政府支出、輸出入など主要なマクロ時系列は調査や作成上の理由から相互に調整されて作成されていない。しかしマクロ経済の動向を理解、政策評価など行う立場からは望ましくない。また直近の状況を理解するためには早めのデータ作成が望ましいが、国全体のデータ作成には多くの情報が必要である。例えば GDP やその主要な項目は最速で四半期、数カ月後以降に公表される。後からより正確と思われるデータが利用可能となり、その結果として、公表した後に過去の公表数値が改訂されることも多い。直近に公表されている四半期データが直近で得られる月次系列などの情報と見かけ上で矛盾する事例もある。また日本の政府統計では担当部署が分かれていることも問題をより複雑化する。

マクロ消費動向は、サンプリングによる家計消費と商業動態統計など生産・販売の動向の乖離が重要な問題。家計調査データは世帯をベースにした標本調査の集計値、家計調査データは近年の世帯数の変化などの動向を考慮する必要がある。他方、生産・販売の調査データには企業消費・政府消費やインバウンド消費など GDP における家計消費概念とは必ずしも整合的でない数値なども算入、マクロ消費を計測する際にはこうした項目の影響を勘案する必要がある。またエコノミストは GDP 最終消費の数値を重視、GDP 速報の推計では家計面と企業面における消費の情報を統合した数値を四半期ベースで作成、GDP 推計の確報は、生産面のより細かな推計値を利用 (内閣府 (2010))。

マクロ指標の四半期データを
y_{1t} $(t = 4(i-1)+3j; i = 1,\cdots,n; j = 1,2,3,4; T = 12n)$
とする。より高頻度な月次データを $p-1$ 次元ベクトル
$\mathbf{y}_{2t}$ $(t = 12(i-1)+j, i = 1,\cdots,n; j = 1,\cdots,12; T = 12n)$
とすると、利用可能な情報の下で観測不能な真のトレンドの状態 x_{1t} $(t = 12(i-1)+j, i = 1,\cdots,n; j = 1,\cdots,12)$ を推定する問題。
なお Kunitomo and Sato (2016) は観測不能な真の非定常トレンド $\mathbf{x}_t = (x_{kt})$ 間に線形関係 $\beta_x'\mathbf{y}_t = O_p(1)$, また観測不能な真の非定常季節性 $\mathbf{s}_t = (s_{kt})$ 間の線形関係 $\beta_s'\mathbf{s}_t = O_p(1)$, が成り立つとき未知母数ベクトルである β_x,β_s を推定する SIML(分離情報最尤法) と呼ばれる方法を提案。トレンド成分間の線形関係に注目し、季節成分ベクトル間の制約を無視して

$$\beta_x'\mathbf{x}_t = \beta_0 + u_t^{(x)}$$

とすると、最終消費系列のトレンド成分が共和分関係 (co-integrated relations) にあることを対応する。

消費系列の場合には最終消費系列 y_{1t} を構成する需要側時系列を y_{2t}、供給側時系列を y_{3t} として、線形関係 (あるいは共和分関係) を利用することで観測できない目的変数の月次系列のトレンドの状態推定の方法を考えるのである。利用可能は四半期データと月次データの情報を有効に利用することで互いに矛盾のない状態推定を実現することが目的である。こうした状態変数表現を利用することでマクロ変数の状態推定が可能。

最適な状態推定

時系列データの一部のみ観測されている場合については良く知られている既存の方法は存在しない。非定常状態変数に関係 (あるいは構造方程式) があるとき $\mathrm{ranl}(\mathbf{x}_t\mathbf{x}_t^{'})$ が確率的に $p-1$ となり、時間的に変動するベクトル $\boldsymbol{\beta}_{xt}$ が存在して

$$\mathbf{x}_t^{'}\boldsymbol{\beta}_{xt} = O_p(1)$$

を考察。統計的フィルターを利用すると係数ベクトル $\boldsymbol{\beta}_{xt}$、あるいは階数条件を用いて状態推定を行う必要がある。いずれの場合にも完全観測の場合には尤度関数は初期条件 $\mathbf{Y}_0$ の下、母数 $\boldsymbol{\theta}$ に対し

$$\begin{aligned} L_T(\boldsymbol{\theta}) &= \prod_{t=1}^{T} f(\mathbf{y}_t|\mathbf{Y}_{t-1},\boldsymbol{\theta}) \\ &= \prod_{i=1}^{n}\prod_{j=1}^{s} f(\mathbf{y}_{(s(i-1)+j}|\mathbf{Y}_{(s(i-1)+j-1},\boldsymbol{\theta}) \end{aligned}$$

$f(\mathbf{y}_t|\mathbf{Y}_{t-1},\boldsymbol{\theta})$ は条件付き密度関数、$\mathbf{Y}_{t-1}$ は時刻 $t-1$ で利用可能な情報。また階数制約条件の下で母数ベクトル $\boldsymbol{\beta}$ を推定する SIML 法を Kunitomo-Sato (2016) が開発。四半期データが完全観測の場合には四半期尤度は

$$L_Q(\boldsymbol{\theta}) = \prod_{i=1}^{n} f(\mathbf{y}_{4i}|\mathbf{Y}_{4(i-1)},\boldsymbol{\theta})$$

尤度関数より原系列 $\mathbf{y}_t$ の同時分布を定める、状態変数および母数 $\boldsymbol{\theta}$ をフィルタリングおよび四半期最尤推定が可能。月次データが不完全観測の場合には、まず四半期最尤推定を行い、階数制約条件を利用して月次 (疑似) 最尤推定を行うのが自然。

SIML

Let y_{ij} be the $i-$th observation of the $j-$th time series at t_i^n for $i = 1, \cdots, n; j = 1, \cdots, p; 0 = t_0^n \leq t_1^n \leq \cdots \leq t_n^n = T$. We ususally set $n = T$, $t_i^n - t_{i-1}^n = 1$, $\mathbf{y}_i = (y_{1i}, \cdots, y_{pi})^{'}$ be a $p \times 1$ vector and $\mathbf{Y}_n = (\mathbf{y}_i^{'}) \; (= (y_{ij}))$ be an $n \times p$ matrix of observations and $\mathbf{y}_0$ is the initial observation vector. We consider the situation when the underlying non-stationary trends $\mathbf{x}_i \; (= (x_{ji}))$ at t_i^n $(i = 1, \cdots, n)$ are not necessarily the same as the observed time series and let $\mathbf{s}_i^{'} = (s_{1i}, \cdots, s_{pi})$ and $\mathbf{v}_i^{'} = (v_{1i}, \cdots, v_{pi})$ be the vectors of the seasonal components and the noise components at t_i^n, respectively, which are independent of $\mathbf{x}_i$.

Then

$$\mathbf{y}_i = \mathbf{x}_i + \mathbf{s}_i + \mathbf{v}_i$$

where $\mathbf{x}_i$ are a sequence of non-stationary trend components, $\mathbf{s}_i$ are a sequence of seasonal components, $\mathbf{v}_i$ are a sequence of independent noise components with $\mathcal{E}(\mathbf{v}_i) = \mathbf{0}$ and $\mathcal{E}(\mathbf{v}_i\mathbf{v}_i^{'}) = \mathbf{\Sigma}_v$. (We assume that $\mathbf{\Sigma}_v$ are positive definite and finite.)
To estimate the structural relationships among the hidden random variables; the trend components and seaonal components when we have stationary and non-stationary errors-in-variables models.

Let $\boldsymbol{\beta}$ be a $p \times 1$ vector and we want to estimate

$$\boldsymbol{\beta}^{'}\mathbf{y}_i = O_p(1) \ \ (i = 1, \cdots, n) \ ,$$

more generally, let $\mathbf{B}$ be a $q \times p$ $(q \leq p)$ matrix and we want to estimate

$$\mathbf{B}\mathbf{y}_i = O_p(1) \ \ (i = 1, \cdots, n)$$

Similarly, some structural relations among seasonal components can be written as

$$\mathbf{B}_s\mathbf{s}_i = O_p(1) \ \ (i = 1, \cdots, n).$$

Consider the situation when $\Delta \mathbf{x}_i$ and $\mathbf{v}_i$ $(i = 1, \cdots, n)$ are mutually independent and each of the component vectors are independently, identically,0 and normally distributed as $N_p(\mathbf{0}, \mathbf{\Sigma}_x)$ and $N_p(\mathbf{0}, \mathbf{\Sigma}_v)$, respectively. We use an $n \times p$ matrix $\mathbf{Y}_n = (\mathbf{y}_i^{'})$ and consider the distribution of $np \times 1$ 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

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

Given the initial condition $\mathbf{y}_0$ the conditional maximum likelihood (ML) estimator can be defined as the solution of maximizing the conditional log-likelihood function except a constant as

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

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

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

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 $\mathbf{C}_n^{-1}\mathbf{C}_n'^{-1} = \mathbf{P}_n\mathbf{D}_n\mathbf{P}_n'$ and $\mathbf{D}_n$ is a diagonal matrix with the k-th element

$$d_k = a_{kn}^* = 2[1-\cos(\pi(\frac{2k-1}{2n+1}))] = 4\sin^2(\pi/2)[(2k-1)/(2n+1)]$$

The conditional likelihood function given the initial condition is proportional to

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

where

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

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 $N_p(\mathbf{0}, \boldsymbol{\Sigma}_x + a_{kn}\boldsymbol{\Sigma}_v)$, and the coefficients a_k is a dense sample of $4\sin^2(x)$ in $(0, \pi/2)$.

It is natural to use $\mathbf{z}_k\mathbf{z}_k^{'}$ to estimate $a_{kn}\boldsymbol{\Sigma}_v + \boldsymbol{\Sigma}_x$ 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, $(1/m_n)\sum_{k=1}^{m_n} a_{kn}$ 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 $\hat{\boldsymbol{\Sigma}}_x$ can be defined by

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

This estimator of the variance-covariance of non-stationary trends is trying to use the information on trends in the frequency domain, which corresponds to only the trend parts without measurement errors from the time series observations. For $\hat{\boldsymbol{\Sigma}}_x$, the number of terms m_n should be dependent on n. Then we need the order requirement that $m_n = O(n^{\alpha})$ and $0 < \alpha < 1$, which is the first property of the macro-SIML estimation.

A Nonstationary Common Trend Case

Let $\mathbf{y}_i = \mathbf{x}_i + \mathbf{v}_i$, $0\mathbf{Y}_n = (\mathbf{y}_i^{'})$, and the vectors $\mathbf{x}_i$ satisfy

$$\mathbf{x}_i = \mathbf{x}_{i-1} + \boldsymbol{\pi}\mu_i \ ,$$

where $\boldsymbol{\pi}$ is a $p \times 1$ vector, μ_i are i.i.d. (one-dimensional) random variables following $N(0, \sigma_\mu^2)$ and $\mathbf{v}_i$ are i.i.d. (p-dimensional) random variables following $N_p(\mathbf{0}, \boldsymbol{\Sigma}_v)$ with the variance-covariance matrix $\boldsymbol{\Sigma}_v$, which is a non-singular matrix.

We take $\mathbf{b} = \sigma_\mu \boldsymbol{\pi}, \mathbf{A} = a_{kn}\boldsymbol{\Sigma}_v$ and apply the matrix formulae such that for a positive definite $\mathbf{A}$ we have

$$|\mathbf{A} + \mathbf{b}\mathbf{b}^{'}| = |\mathbf{A}|[1 + \mathbf{b}^{'}\mathbf{A}^{-1}\mathbf{b}]$$

and

$$[\mathbf{A} + \mathbf{b}\mathbf{b}^{'}]^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{b}[1 + \mathbf{b}^{'}\mathbf{A}^{-1}\mathbf{b}]^{-1}\mathbf{b}^{'}\mathbf{A}^{-1}$$

for $\boldsymbol{\Sigma}_x = \mathbf{b}\mathbf{b}^{'}$.

Then the likelihood function L_n is proportional to $(-1/2)$ times

$$
\begin{aligned}
L_{1n} &= \sum_{k=1}^{n} \Big[\log|a_{kn}\mathbf{\Sigma}_v| + \log(1 + a_{kn}^{-1}\mathbf{b}'\mathbf{\Sigma}_v^{-1}\mathbf{b}) + a_{kn}^{-1}\mathbf{z}_k'\mathbf{\Sigma}_v^{-1}\mathbf{z}_k \\
&\quad - \frac{a_{kn}^{-1}(\mathbf{z}_k'\mathbf{\Sigma}_v^{-1}\mathbf{b})^2}{a_{kn} + \mathbf{b}'\mathbf{\Sigma}_v^{-1}\mathbf{b}} \Big] \\
&= \sum_{k=1}^{n} \log|a_{kn}\mathbf{\Sigma}_v| + \sum_{k=1}^{n} a_{kn}^{-1}\mathbf{z}_k'\mathbf{\Sigma}_v^{-1}\mathbf{z}_k \\
&\quad + \sum_{k=1}^{n} \left[\log(1 + a_{kn}^{-1}c) - \frac{a_{kn}^{-1}(\mathbf{z}_k'\mathbf{\Sigma}_v^{-1}\mathbf{b})^2}{a_{kn} + c} \right] ,
\end{aligned}
$$

where

$$
c = \sigma_\mu^2 \boldsymbol{\pi}'\mathbf{\Sigma}_v^{-1}\boldsymbol{\pi} \ .
$$

We need a normalization for the vector $\boldsymbol{\pi}$. If we take a simple normalization, the maximum likelihood estimator of $\boldsymbol{\pi}$ could be a quite complicated solution of the likelihood equation even when $p = 2$. One possible normalization is to take $\boldsymbol{\beta}' = (1, -\boldsymbol{\beta}_2')$ and then the maximization is not a trivial task.

As an alternative way to solve the present problem is to use the conditions that

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

and

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

The rank of matrix $\mathbf{\Sigma}_x$ is one while the matrix $\mathbf{\Sigma}_v$ has a full rank.

$$\left[\hat{\mathbf{\Sigma}}_{x.SIML} - \lambda \hat{\mathbf{\Sigma}}_{v.SIML}\right] \hat{\beta}_{SIML} = \mathbf{0} \ ,$$

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

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

or

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

where

$$\mathbf{Z}_n = (\mathbf{z}_k^{'}) = \mathbf{P}_n \mathbf{C}_n^{-1} \left(\mathbf{Y}_n - \mathbf{1}_n \bar{\mathbf{y}}_0^{'} \right) \ ,$$

$\hat{\mathbf{\Sigma}}_{v.SIML}$ is a SIML estimator of $\mathbf{\Sigma}_v$, and λ is the (scalar) eigen value.

Because the rank of $\mathbf{\Sigma}_x$ is degenerated and it is one, we need to take the smallest eigenvalue λ_1. We have the $\hat{\beta}_{SIML}$, which is called the SIML estimator of β. A simplified (consistent) estimation may be given by

$$\hat{\mathbf{\Sigma}}_{x.SIML} \times \hat{\beta}_{SIL} = \mathbf{0} \ ,$$

that is,

$$\hat{\mathbf{\Sigma}}_{x.SIML} \times \left[\begin{array}{c} 1 \\ -\hat{\beta}_{2.SIL} \end{array} \right] = \mathbf{0} \ .$$

We can solve as

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

which can be the least squares estimator for the transformed variables and it is called the SILS estimator.

Gaussian Likelihood

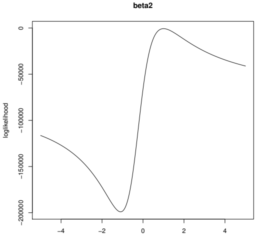


Figure 1 : Likelihood Function of0 β_2 $(n = 1,000)$

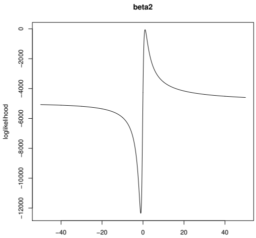


Figure 2 : Likelihood Function of β_2 ($n = 1,000$)

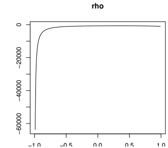


Figure 3 : Likelihood Function of ρ $(n = 1,000)$

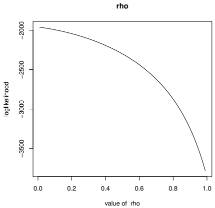


Figure 4 : Wrong Likelihood Function of ρ $(n = 1,000)$

文献

- Akaike, H (1989), "Choosing priors and its applications," in *Bayesian Statistics and Applications*, edited by Suzuki and Kunitomo, Tokyo University Press, in Japanese.
- 北川源四郎 (2005)「時系列解析入門」，岩波書店。
- Anderson, T.W. (1984), "Estimating Linear Statistical Relationships," *Annals of Statistics*, 12, 1-45.
- Johansen, S. (1995), *Likelihood Based Inference in Cointegrated Vector Autoregressive Models*, Oxford UP.
- Kunitomo, N. and S. Sato (2008), "Separating Information Maximum Likelihood Estimation of Realized Volatility and Covariance with Micro-Market Noise," Discussion Paper CIRJE-F-581, Graduate School of Economics, University of Tokyo, (http://www.e.u-tokyo.ac.jp/cirje/research/dp/2008), in *North American Journal of Economics and Finance* (2013).
- Kunitomo, N. and S. Sato (2017), "Trend, Seasonality and Economic Time Series : the Nonstationary Errors-in-variables Models," Discussion Paper CIRJE-F-977, Graduate School of Economics, University of Tokyo, (http://www.e.u-tokyo.ac.jp/cirje/research/dp/2015), revised.

釧路・経済統計キャンプ２０１７

ＧＤＰ統計の見方について

東京大学大学院経済学研究科
佐藤　整尚

概要：

- ＧＤＰのいろいろな見方
  - 歴史的経緯
  - 年率換算伸び率の問題点
- ノイズを除いた伸び率の重要性
  - 安定的な伸び率（平滑化伸び率）の提案
- トレンド・循環成分（ＴＣ）の抽出
  - 安定的な伸び率の推定にはＴＣ成分が必要
- 状態空間モデルとフィルタリング
  - ＴＣ成分は観測されないので統計的な推計が必要
- ＴＣ成分を用いた月次系列の推定
  - ＧＤＰ統計の例
  - 消費統計の例
- 多変量季節調整モデルの試み

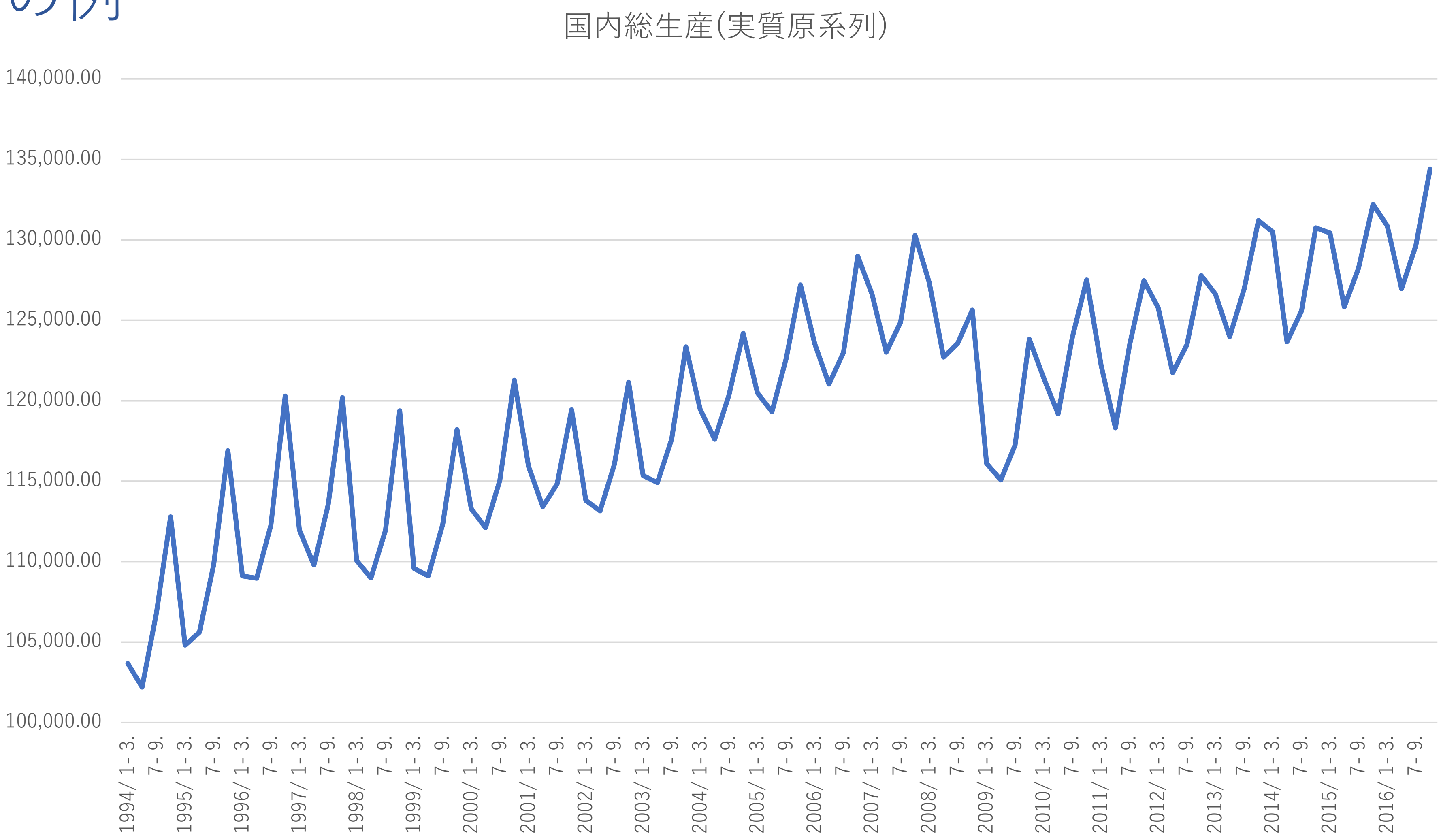
2016年9月2日の経済教室（日経新聞）

ＧＤＰ（ＧＮＰ）の見方のいろいろ

歴史的経緯：新聞での見出し

- 50年代半ば―60年代半ば
  - 始めは実額
  - 四半期系列は4-5か月後に公表、前期比
- 60年代半ばー70年代初め
  - 前年同期比
  - 前期比
- 70年代以降
  - 年率換算の前期比が登場（72年）＜－政府の成長見通しとの比較のため
  - 以後、実質季節調整済み系列の前期比を年率換算（(今期／前期)^4-1）したものが見出しに出てくるようになる。

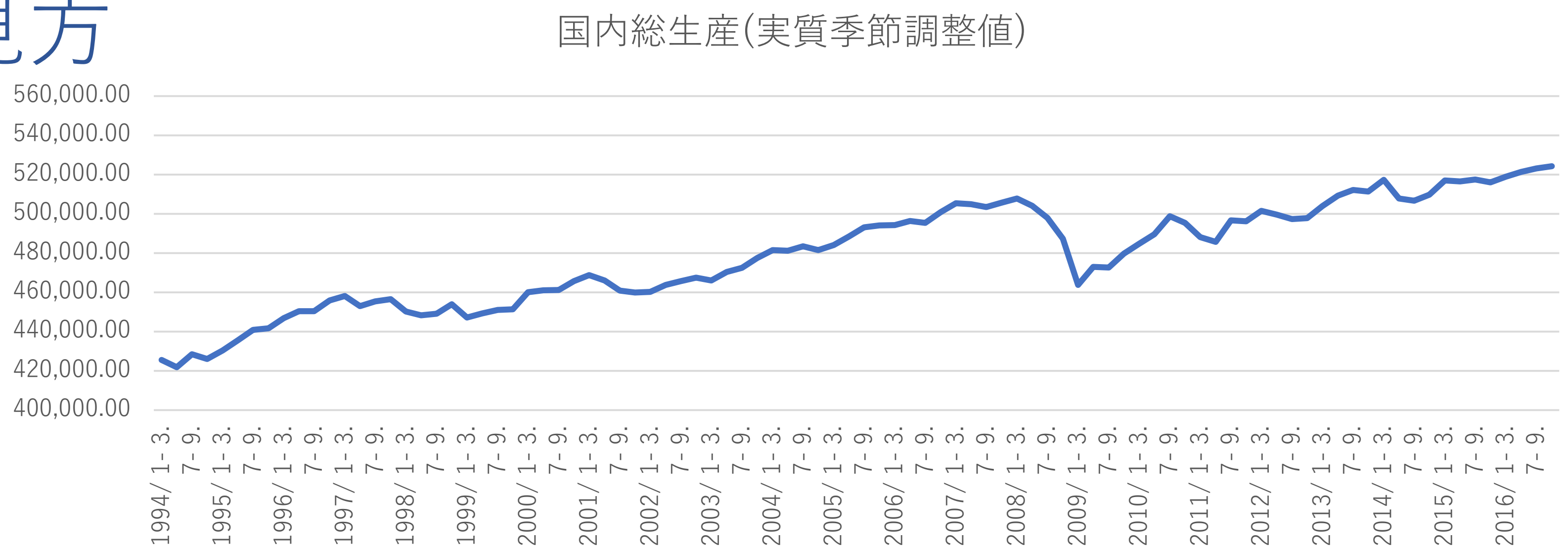
時系列の例



季節性もあり、このままでは、読み取りにくい。

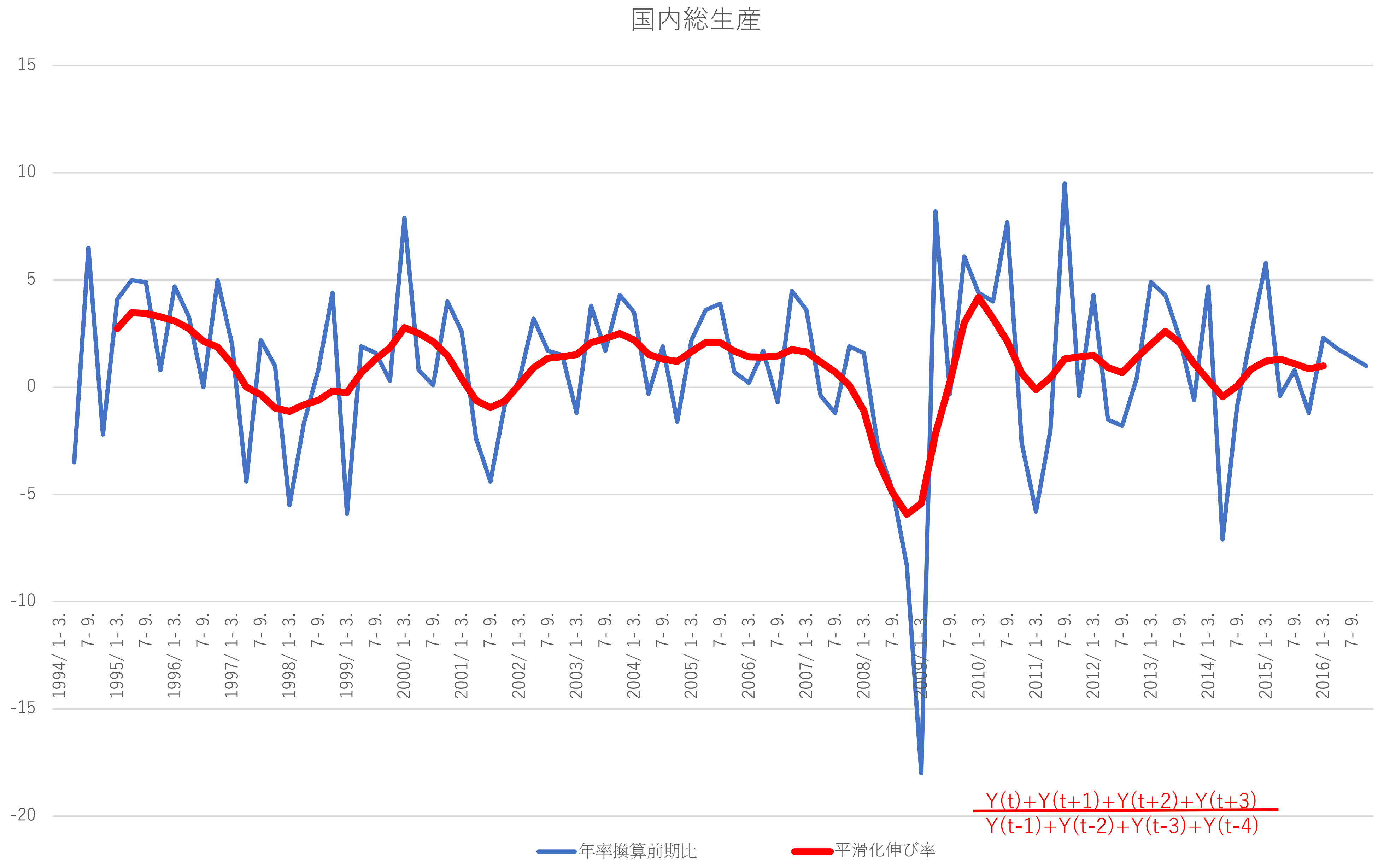
時系列の見方

Ｘ１２－ＡＲＩＭＡによる季節調整

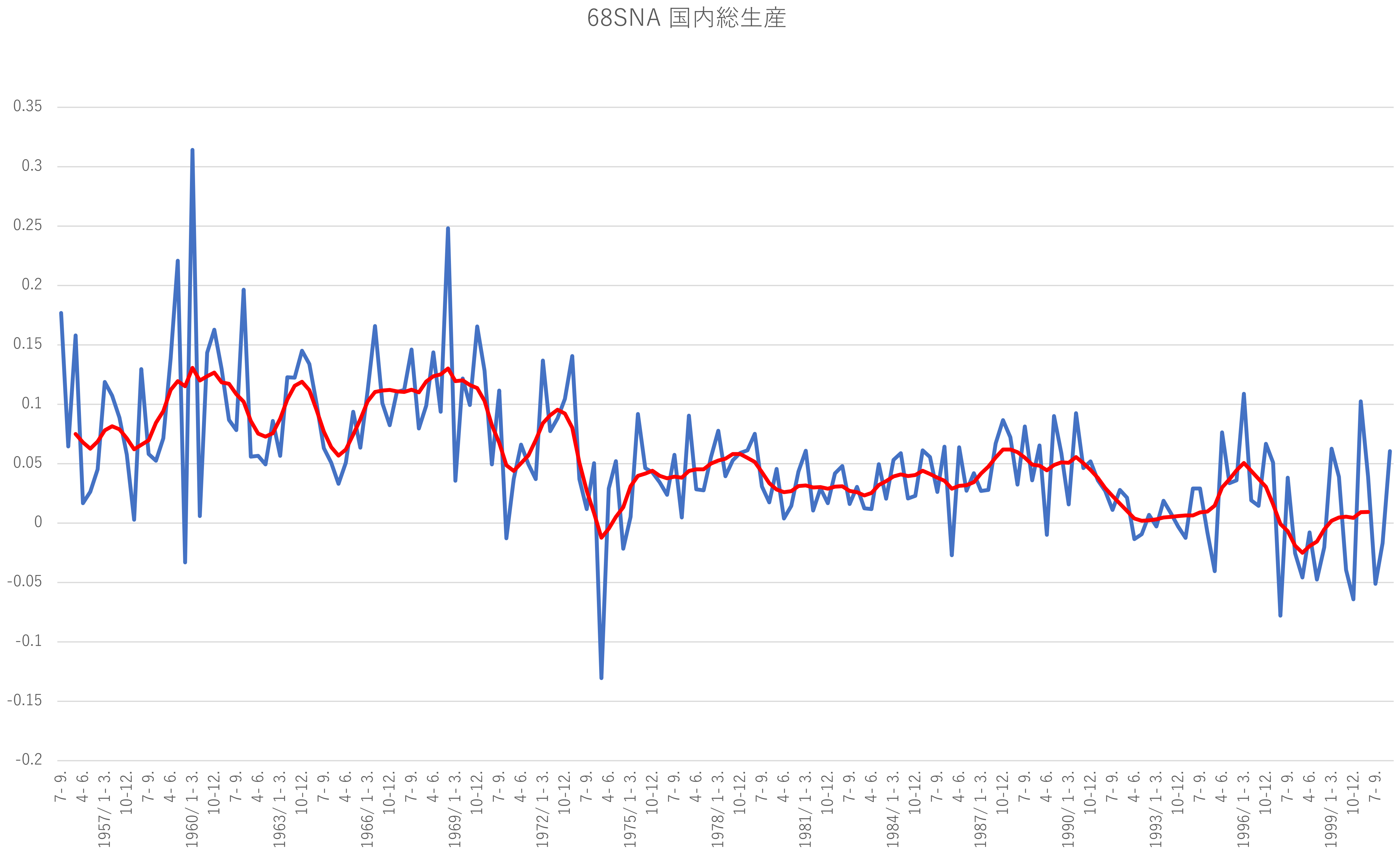


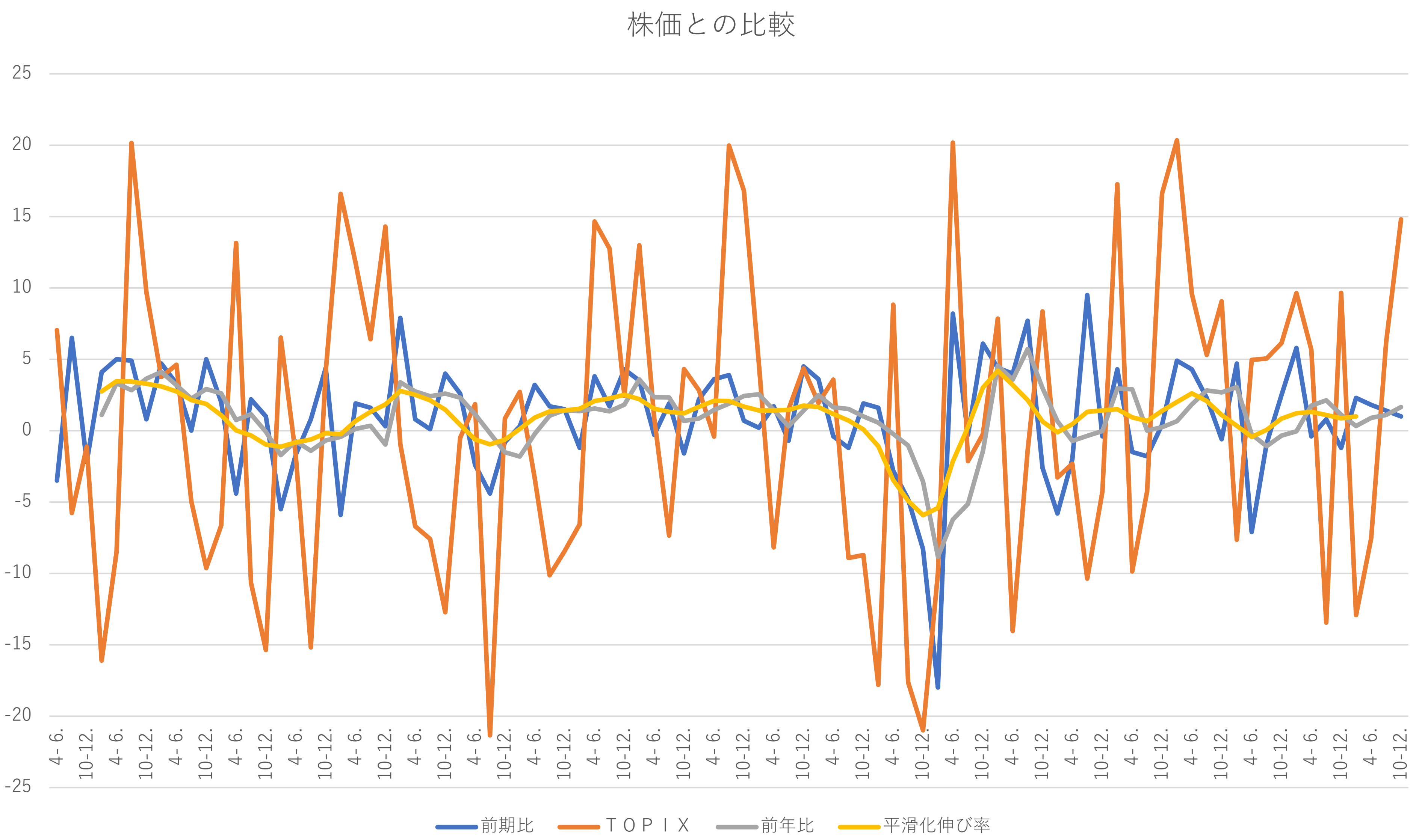
水準よりは
伸び率が重要

やや、遅れる。



国内総生産
15
10
5
0
-5
-10
-15
-20
1994/ 1 - 3.
7 - 9.
1995/ 1 - 3.
1996/ 1 - 3.
1997/ 1 - 3.
1998/ 1 - 3.
1999/ 1 - 3.
2000/ 1 - 3.
2001/ 1 - 3.
2002/ 1 - 3.
2003/ 1 - 3.
2004/ 1 - 3.
2005/ 1 - 3.
2006/ 1 - 3.
2007/ 1 - 3.
2008/ 1 - 3.
2009/ 1 - 3.
2010/ 1 - 3.
2011/ 1 - 3.
2012/ 1 - 3.
2013/ 1 - 3.
2014/ 1 - 3.
2015/ 1 - 3.
2016/ 1 - 3.
$\frac{Y(t)+Y(t+1)+Y(t+2)+Y(t+3)}{Y(t-1)+Y(t-2)+Y(t-3)+Y(t-4)}$
年率換算前期比
平滑化伸び率


68SNA 国内総生産
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
-0.05
-0.1
-0.15
-0.2
7- 9.
4- 6.
1957/ 1- 3.
10-12.
7- 9.
4- 6.
1960/ 1- 3.
10-12.
7- 9.
4- 6.
1963/ 1- 3.
10-12.
7- 9.
4- 6.
1966/ 1- 3.
10-12.
7- 9.
4- 6.
1969/ 1- 3.
10-12.
7- 9.
4- 6.
1972/ 1- 3.
10-12.
7- 9.
4- 6.
1975/ 1- 3.
10-12.
7- 9.
4- 6.
1978/ 1- 3.
10-12.
7- 9.
4- 6.
1981/ 1- 3.
10-12.
7- 9.
4- 6.
1984/ 1- 3.
10-12.
7- 9.
4- 6.
1987/ 1- 3.
10-12.
7- 9.
4- 6.
1990/ 1- 3.
10-12.
7- 9.
4- 6.
1993/ 1- 3.
10-12.
7- 9.
4- 6.
1996/ 1- 3.
10-12.
7- 9.
4- 6.
1999/ 1- 3.
10-12.
7- 9.



相関係数：０．１６（前期比）
０．０７（前年比）
０．２８（平滑化伸び率）

ＧＤＰ成長率と金融データとの相互相関係数

	前期比	前年同期比	平滑化伸び率
TOPIX	0.255	0.159	0.369
JPY	0.110	0.139	0.133
JGB	0.084	0.181	0.192

Robust Year on Year Change (前年同月比平均伸び率)

- remove the effect from the noise of Y_{t-12}

$$r_t = Y_t / Y_{t-12} - 1$$

$$r_t \approx \log(Y_t) - \log(Y_{t-12})$$

We assume the following decomposition.

$$\log(Y_t) = T_t + S_t + C_t + N_t$$

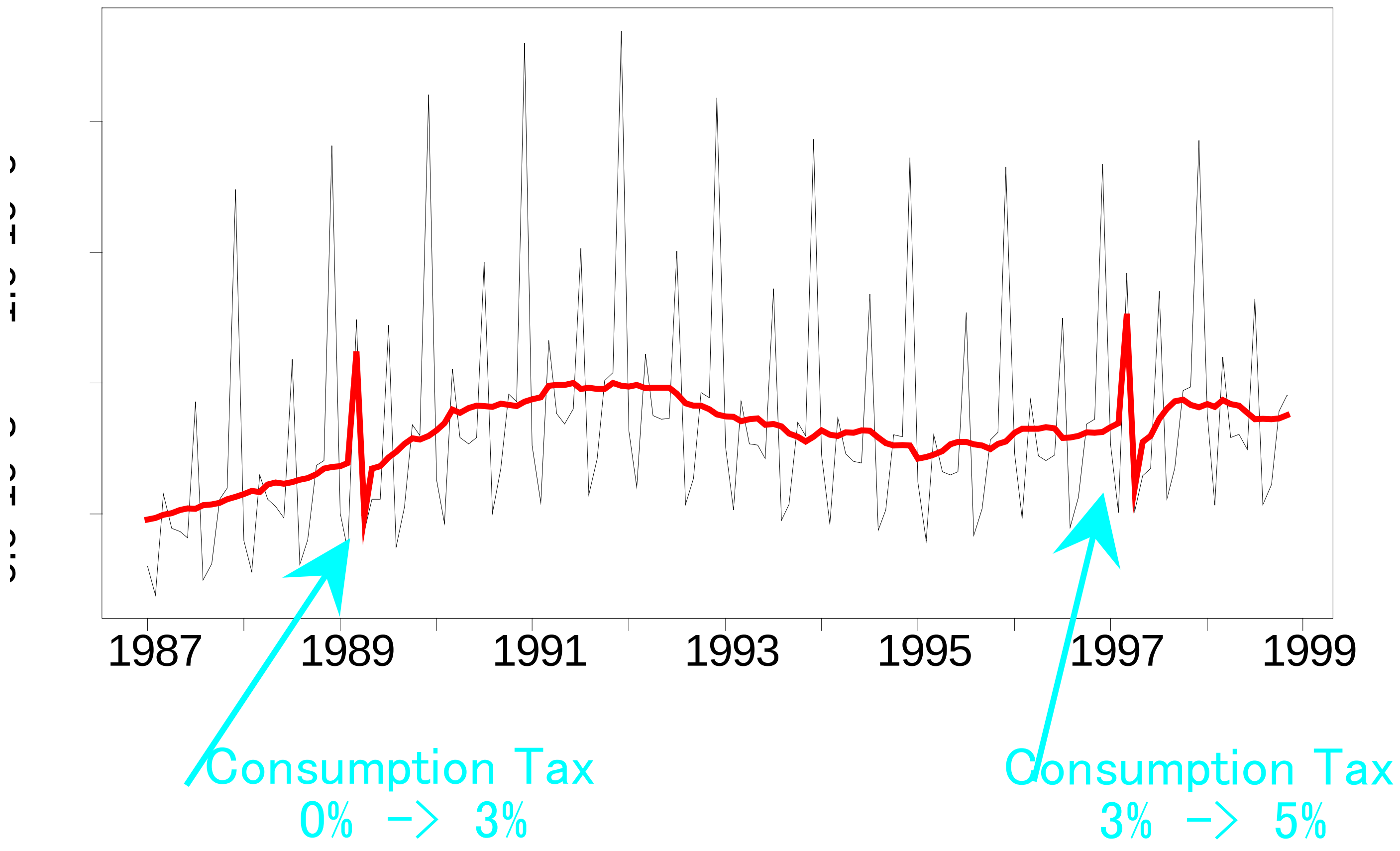
(T: Trend, S: Seasonal, C: Cycle, N: Noise)

<u>definition of “robust year on year change”</u>

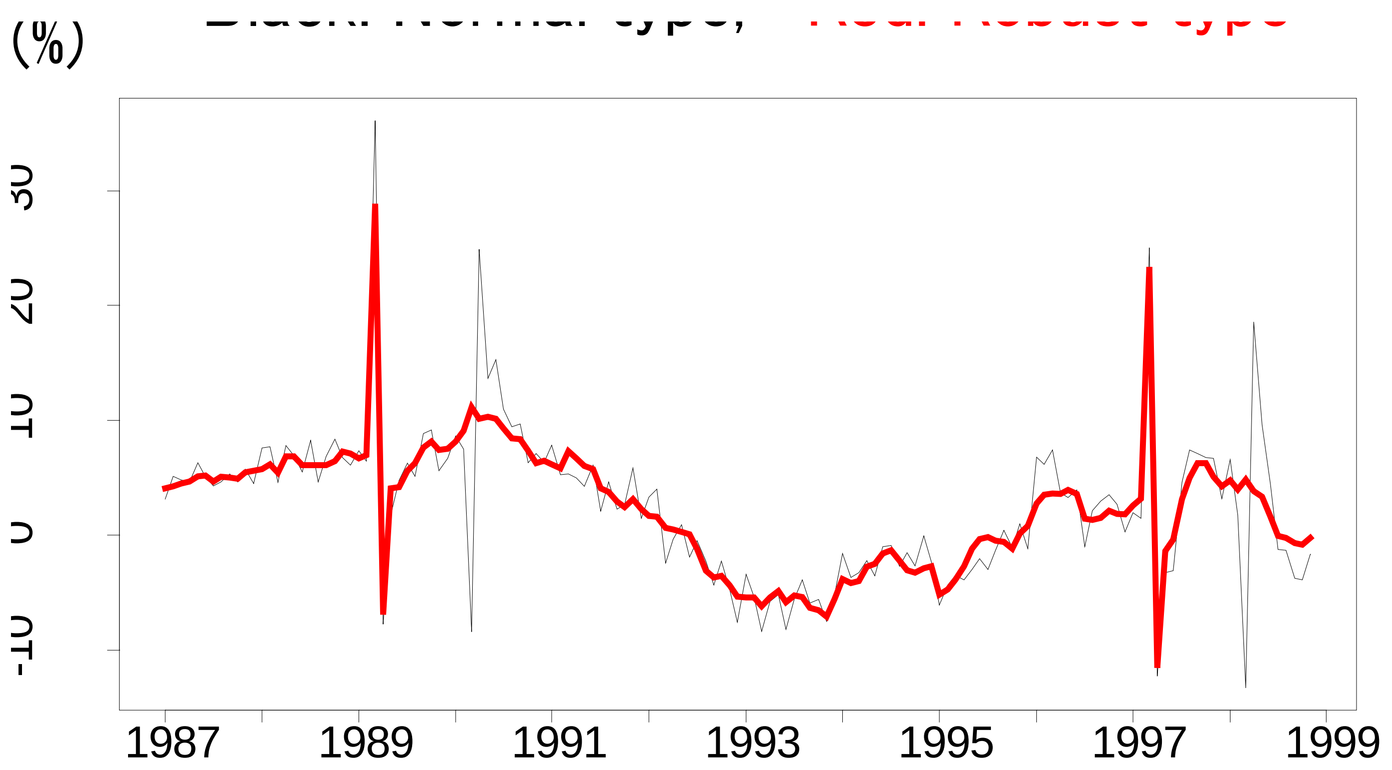
$$r_t^* = T_t + C_t + N_t - T_{t-12}$$

Example

- Data: Monthly sales of department stores in Japan (1987-1998)

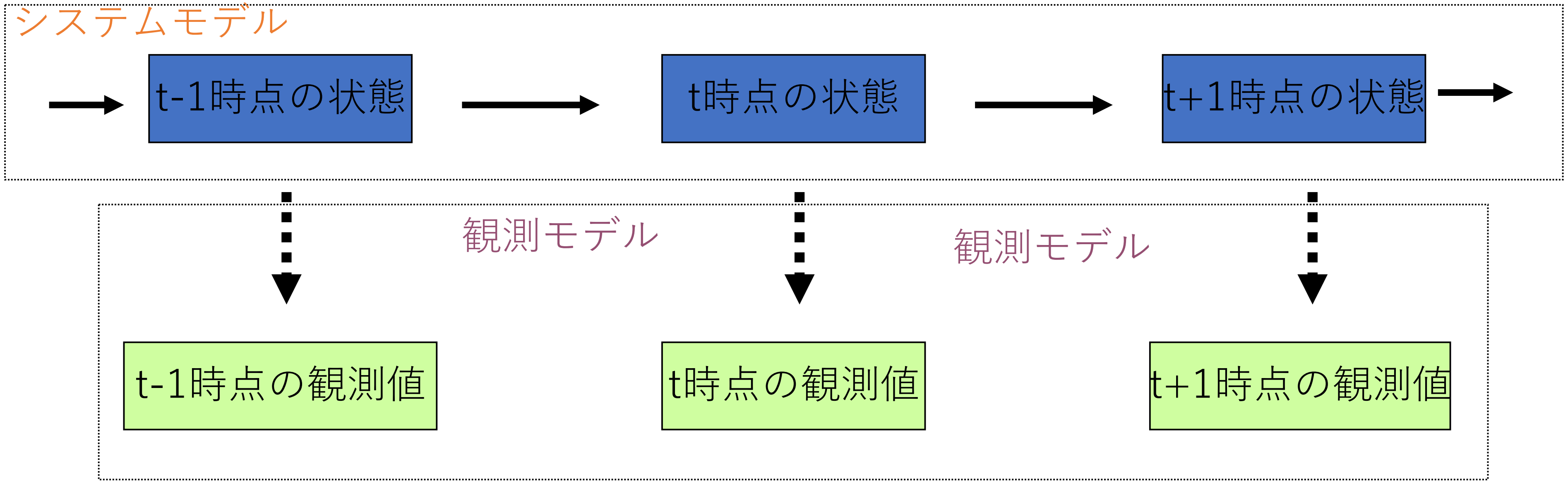


Comparing 2 types of
" Year on Year Change"



状態空間モデルとは

- システムモデル（状態モデル）と観測モデルからなる。
- システムモデルには時系列的性質が仮定される。
（前期の値に依存して、今期が決まる。）



状態空間モデル

- 線形・ガウスの場合（ｘが状態変数）

$$x_t = Fx_{t-1} + Gv_t$$

（システムモデル）

$$y_t = Hx_t + u_t$$

（観測モデル）

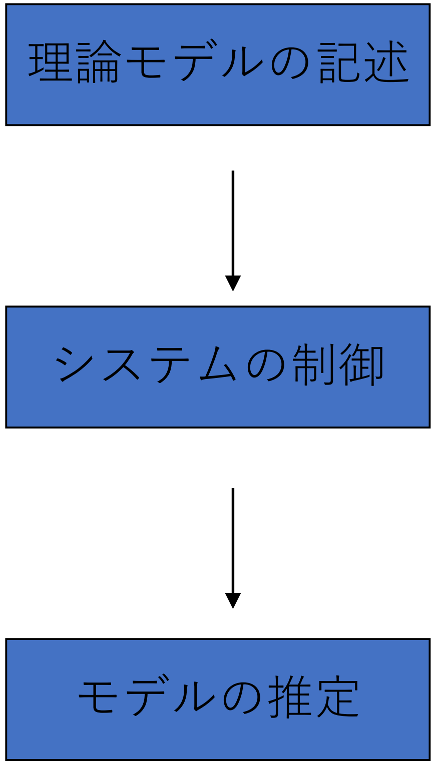
＃観測値Ｙから状態Ｘの推定をフィルタリングという。
（カルマンフィルタ、Ｄｅｃｏｍｐ法など）

状態空間モデルの目的

- モデルの当てはめやパラメータの推定を行う。
- 観測値から状態変数を推定する。
- モデルに基づき予測を行う。
- モデルに基づき平滑化を行う。

状態空間モデルの歴史

- もともとは物理システムの記述に使われていた。
- 1960年代カルマンにより、制御工学での利用が進んだ。（カルマンフィルタ）
- 1970年代赤池により、統計科学への応用が始まる。

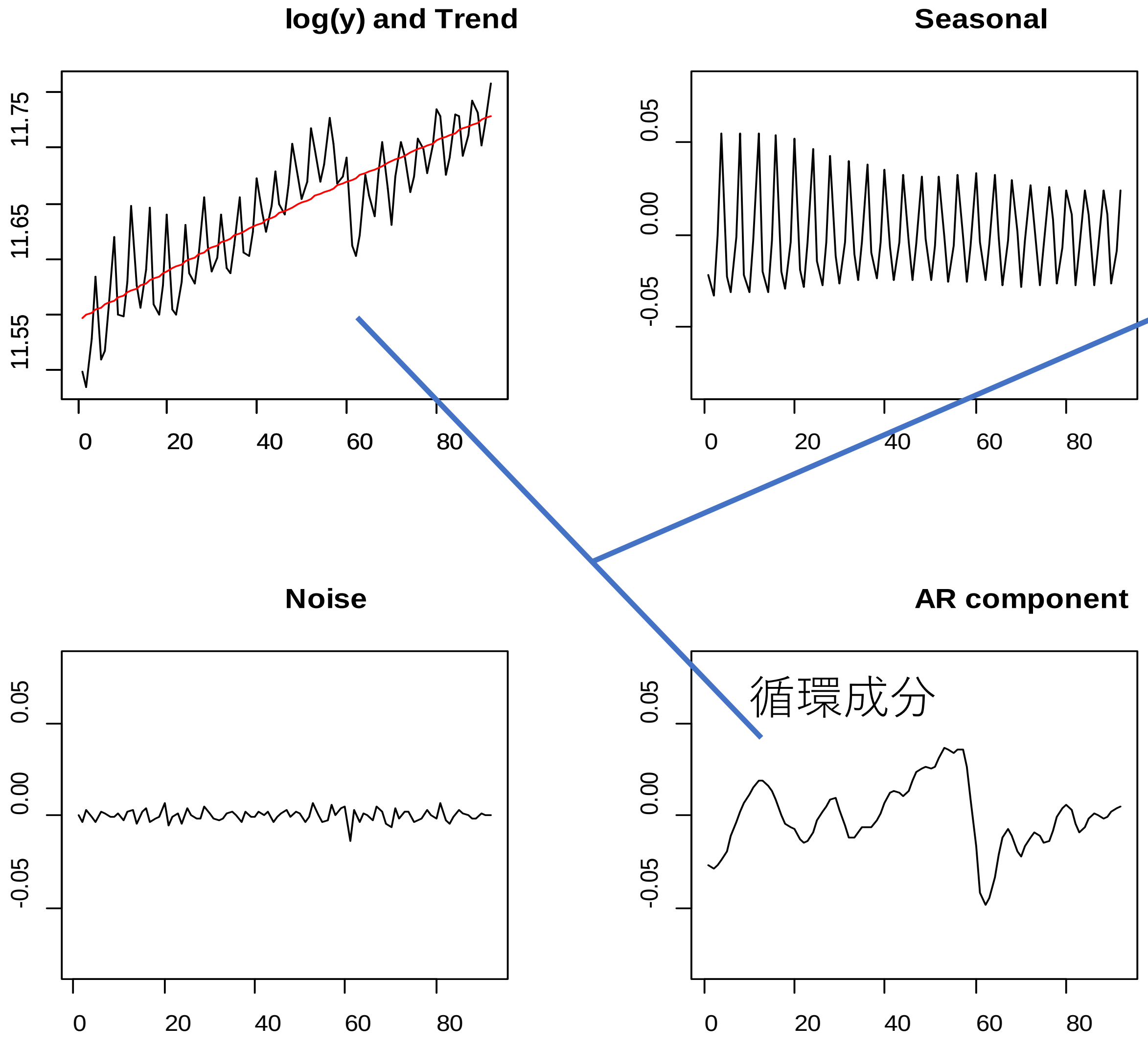


日本から始まったといっても過言ではない。
季節調整法に応用したのは、世界的に見てもっとも早い。
（Ｄｅｃｏｍｐ法）

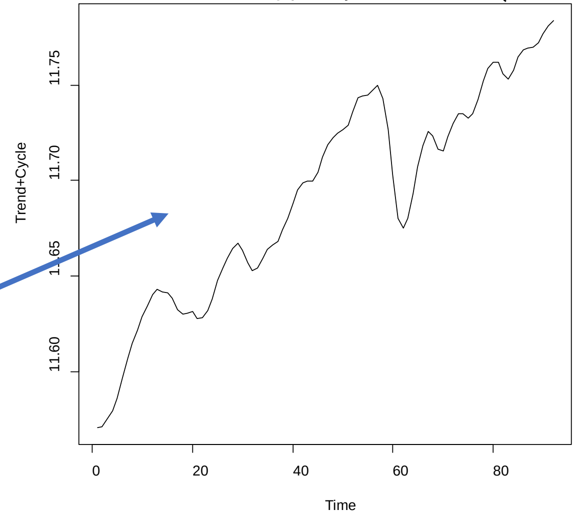
Ｄｅｃｏｍｐ法による

国内総生産に対して状態空間モデルを適用

Log(GDP)＝
トレンド＋
季節性＋
ＡＲ＋
ノイズ



トレンド＋循環成分（ＴＣ）

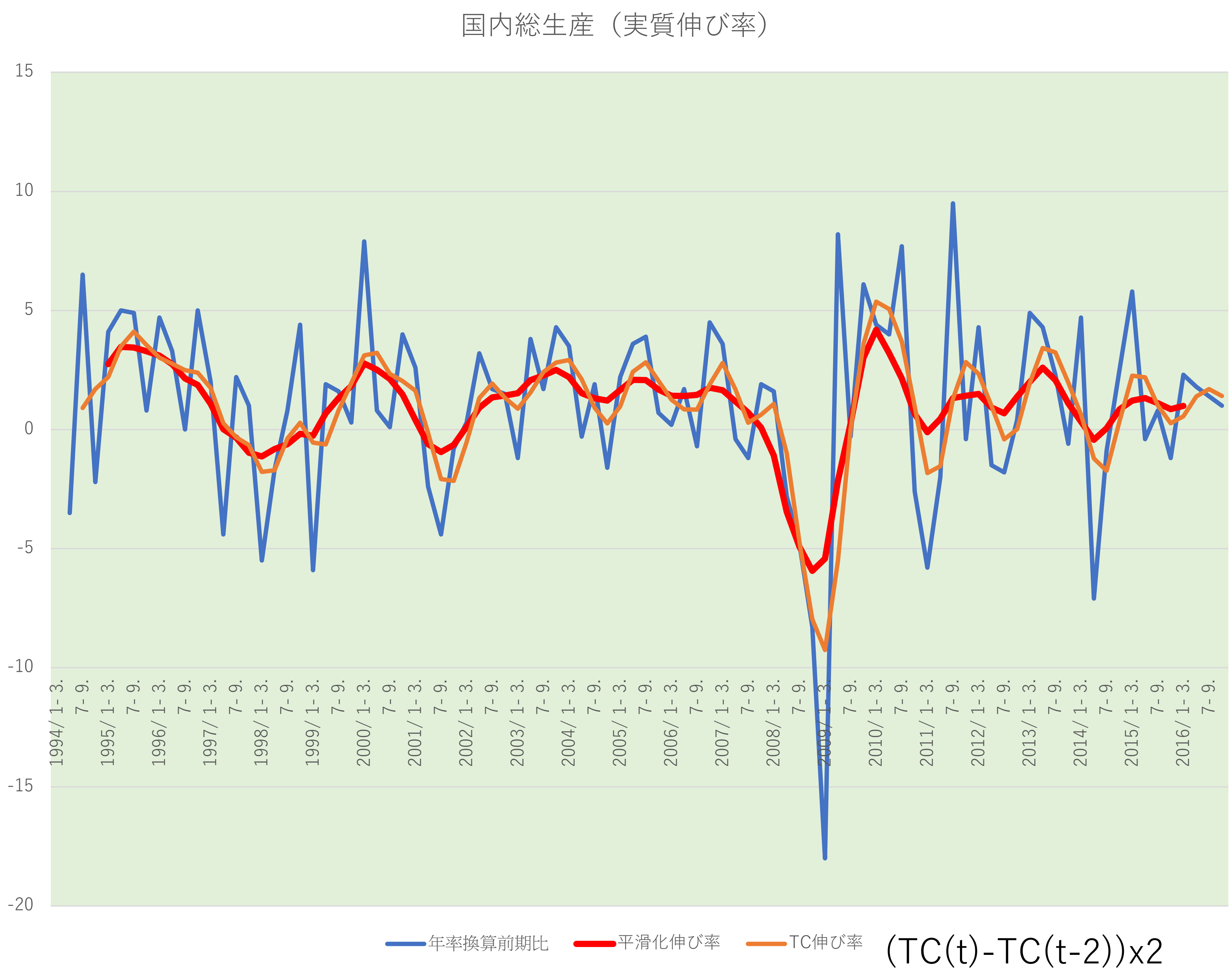


$$\log(Y_t) = T_t + S_t + C_t + I_t$$

$$T_t = 2T_{t-1} - T_{t-2} + u_{1,t}$$

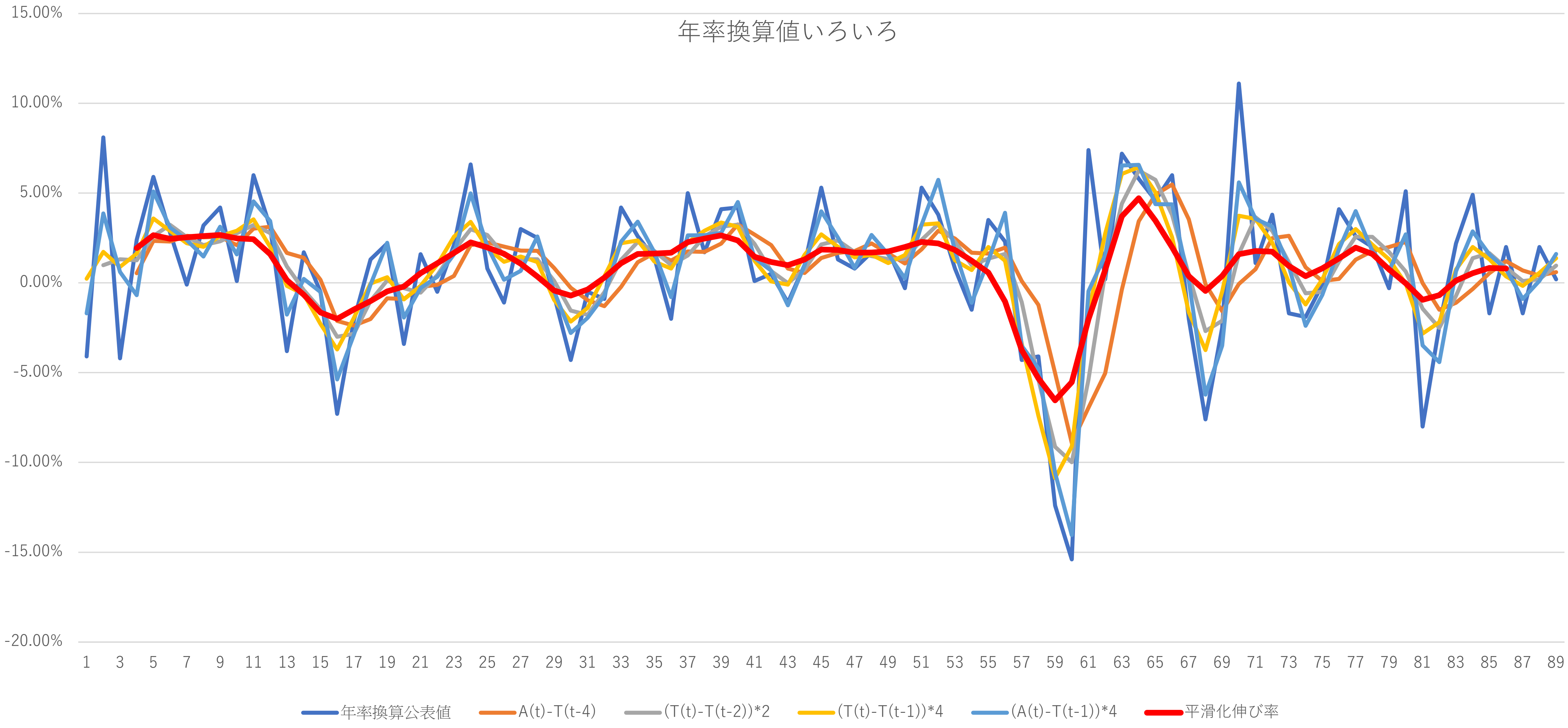
$$S_t = -\sum_{i=1}^{3} S_{t-i} + u_{2,t}$$

$$C_t = a_1 C_{t-1} + a_2 C_{t-2} + u_{3,t}$$



(TC(t)-TC(t-2))x2

参考：
日本のマクロ経済統計の課題



T(t)=Trend(t)+AR(t)

A(t)=T(t)+noise(t)

多変量トレンドモデル

$$Y_{1,t} \cong Y_{2,t} + Y_{3,t} + \cdots + Y_{7,t}$$

（ＧＤＰ）＝（消費）＋（住宅）＋（投資）＋（政府支出）＋（政府投資）＋（その他）

$$Y_{1,t} = T_{2,t} + T_{3,t} + \cdots + T_{7,t} + S_{1,t} + I_{1,t}$$

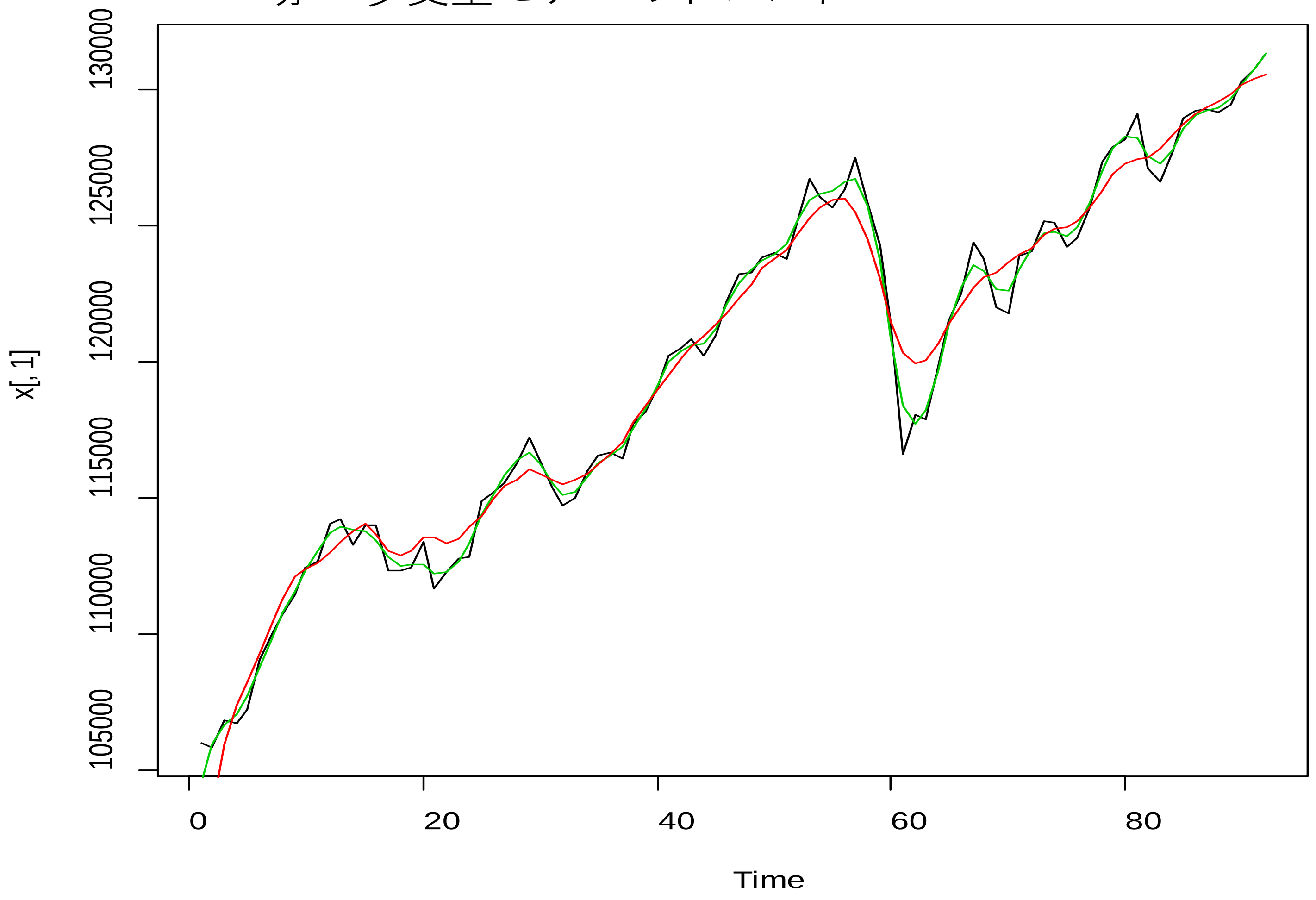
$$Y_{i,t} = T_{i,t} + S_{i,t} + I_{i,t} \qquad (i = 2, \cdots, 7)$$

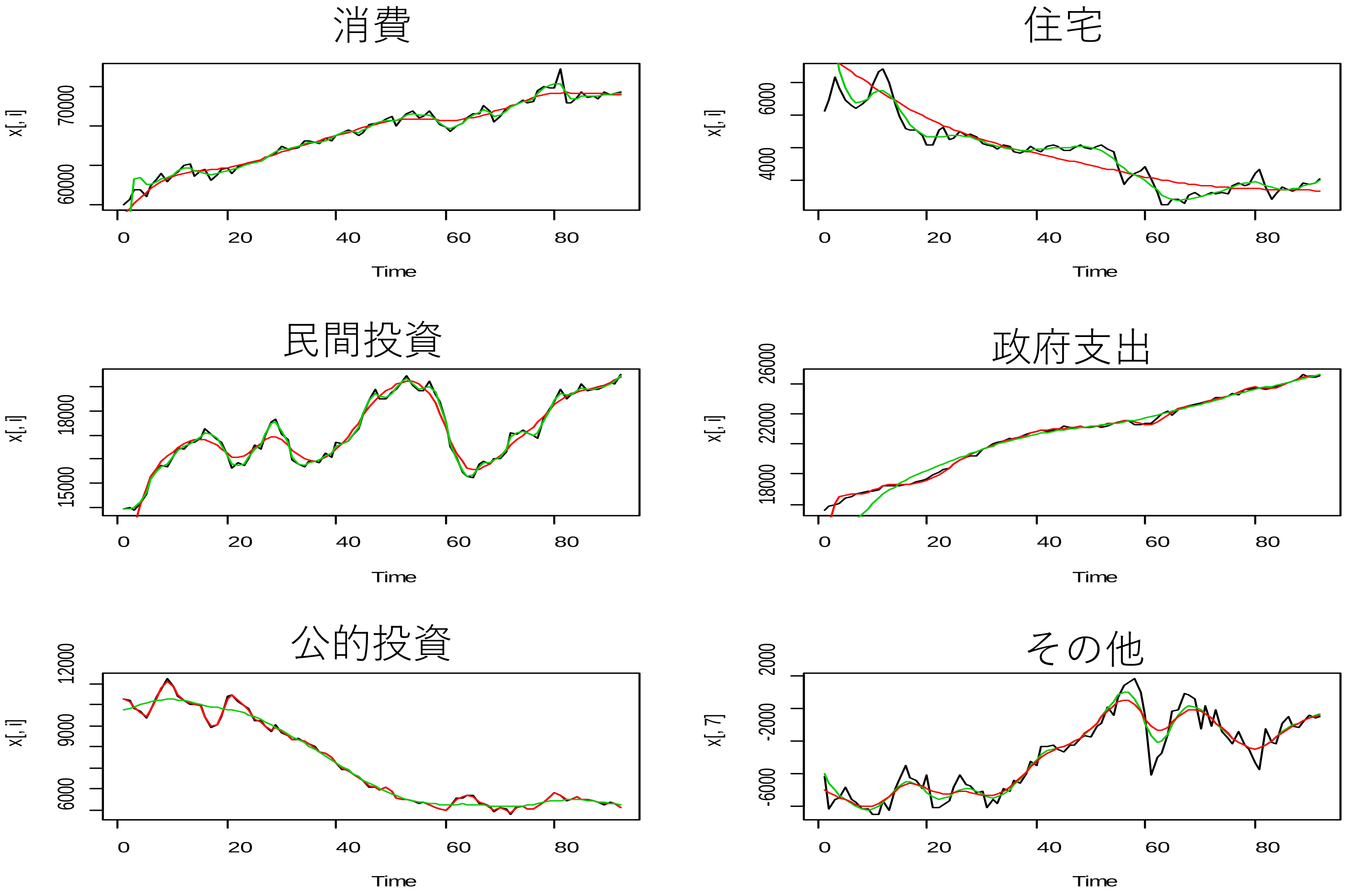
$$S_{i,t} = -\sum_{k=1}^{3} S_{i,t-k} + v_{i,t}$$

$$T_{i,t} = 2T_{i,t-1} - T_{t-2} + u_{i,t} \qquad (i = 2, \cdots, 7)$$

ＧＤＰ

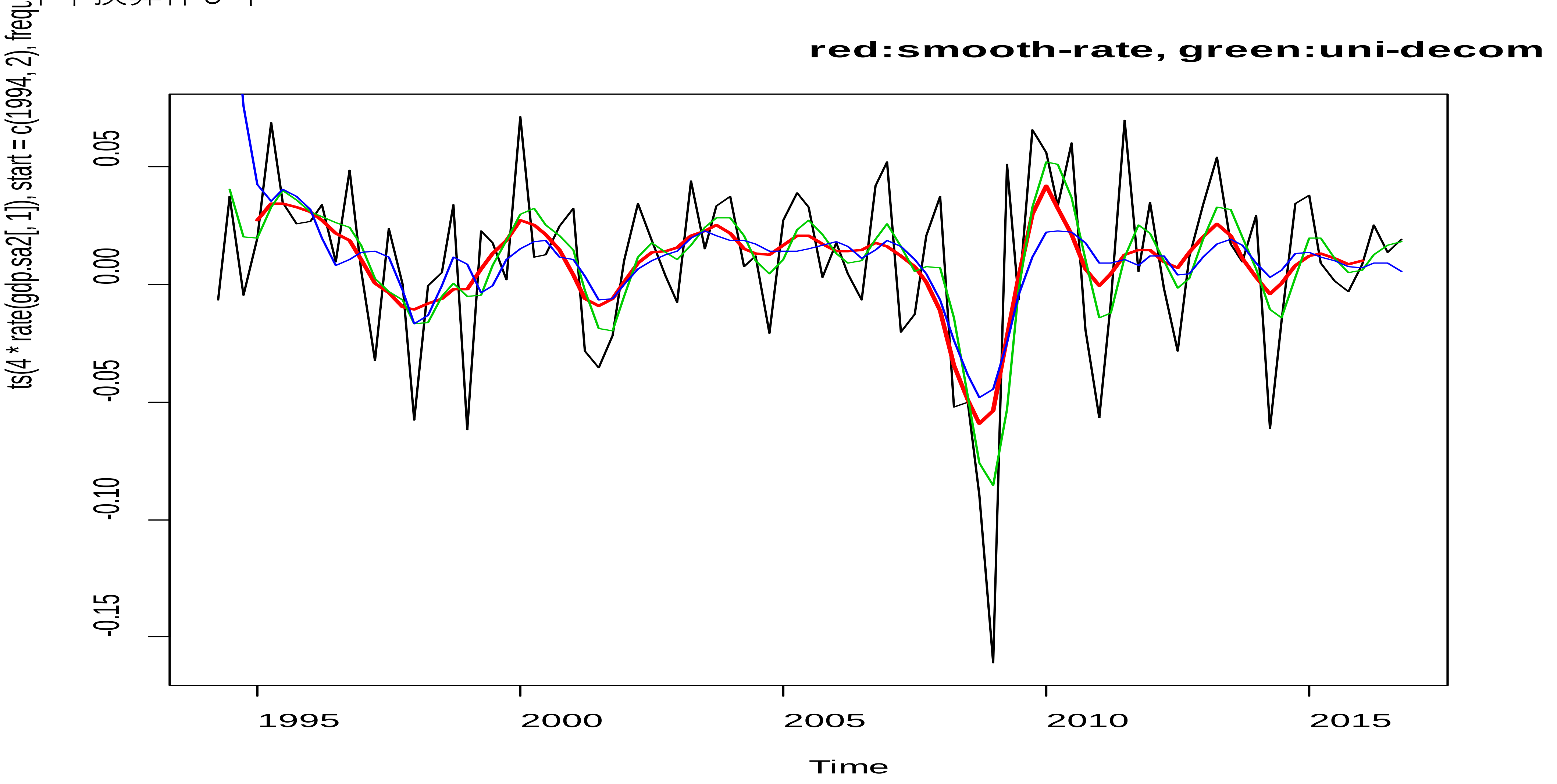
黒：季節調整値、緑：1変量モデルのトレンド
赤：多変量モデルのトレンド





黒：季節調整値、緑：1変量モデルのトレンド
赤：多変量モデルのトレンド

ＧＤＰ年率換算伸び率



red:smooth-rate, green:uni-decom
ts(4 * rate(gdp.sa2[, 1]), start = c(1994, 2), freq
0.05
0.00
-0.05
-0.10
-0.15
1995
2000
2005
2010
2015
Time

より一般的な多変量トレンドモデル

$$Y_{1,t} = T_{1,t} + a_2 T_{2,t} + \cdots + a_m T_{m,t} + S_{1,t} + I_{1,t}$$

$$Y_{i,t} = T_{i,t} + S_{i,t} + I_{i,t} \qquad (i = 2, \cdots, m)$$

$$S_{i,t} = -\sum_{k=1}^{p-1} S_{i,t-k} + v_{i,t}$$

$$T_{i,t} = 2T_{i,t-1} - T_{t-2} + u_{i,t} \qquad (i = 1, \cdots, m)$$

多変量トレンドＡＲモデル

$$Y_{1,t} \cong Y_{2,t} + Y_{3,t} + Y_{4,t}$$

（ＧＤＰ）＝（消費）＋（住宅＋投資＋その他）＋（政府支出＋政府投資）

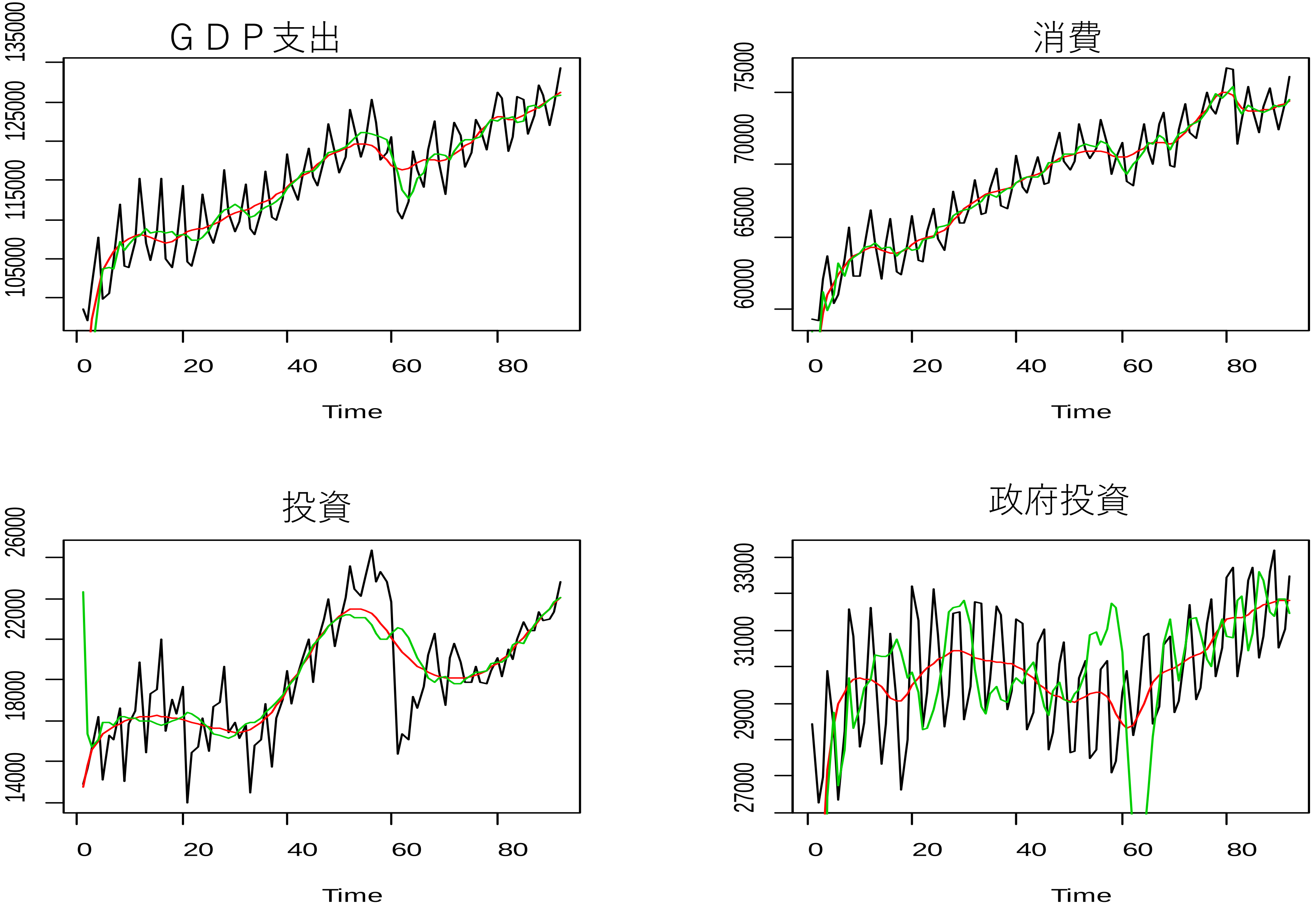
$$Y_{1,t} = T_{2,t} + T_{3,t} + T_{4,t} + A_{2,t} + A_{3,t} + A_{4,t} + S_{1,t} + I_{1,t}$$

（ＧＤＰ）

$$Y_{i,t} = T_{i,t} + A_{i,t} + S_{i,t} + I_{i,t} \qquad (i = 2, \cdots, 4)$$

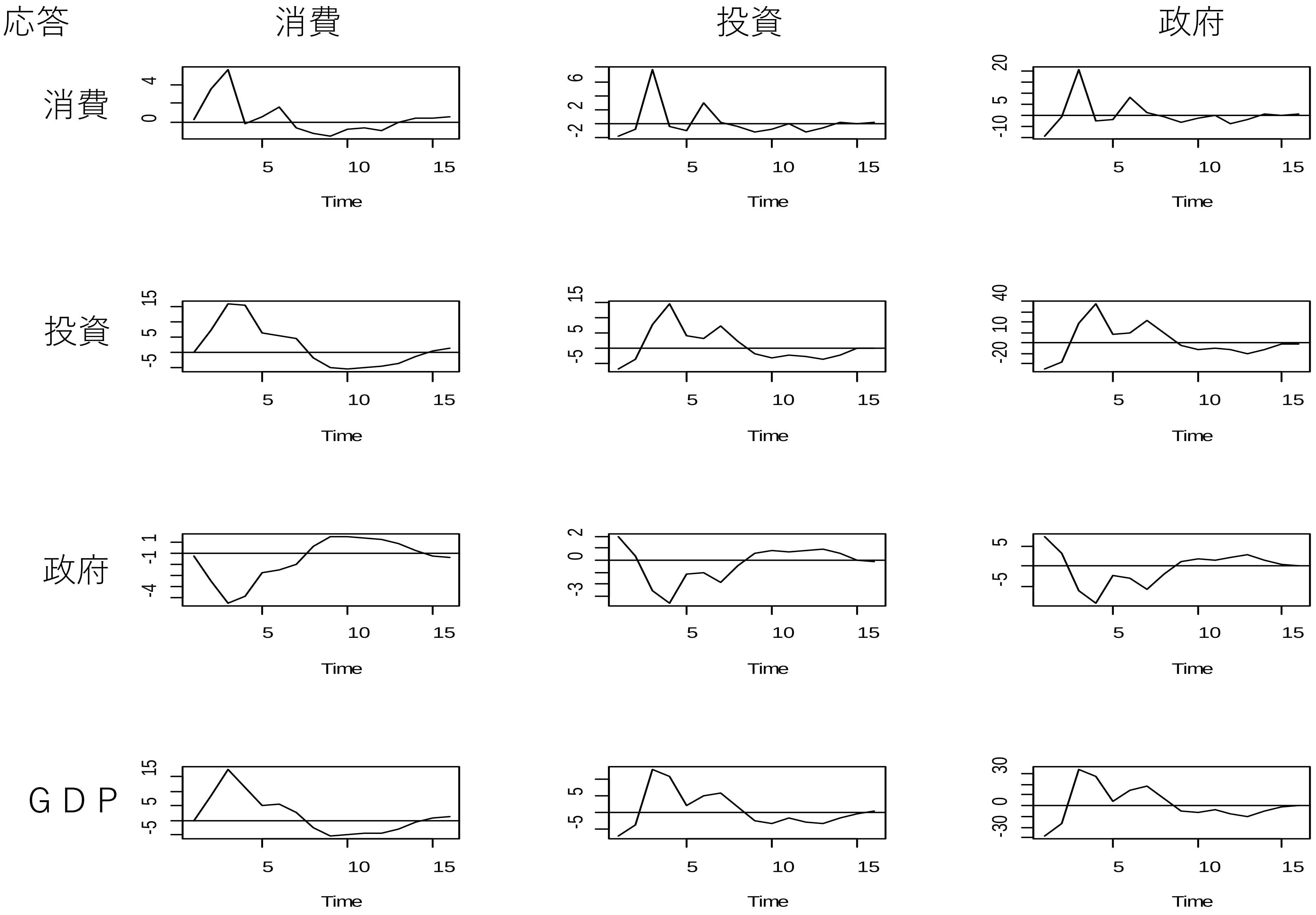
（消費、etc）

$$\begin{pmatrix} A_{2,t} \\ A_{3,t} \\ A_{4,t} \end{pmatrix} = B_1 \begin{pmatrix} A_{2,t-1} \\ A_{3,t-1} \\ A_{4,t-1} \end{pmatrix} + B_2 \begin{pmatrix} A_{2,t-2} \\ A_{3,t-2} \\ A_{4,t-2} \end{pmatrix} + \begin{pmatrix} \eta_{2,t} \\ \eta_{3,t} \\ \eta_{4,t} \end{pmatrix}$$

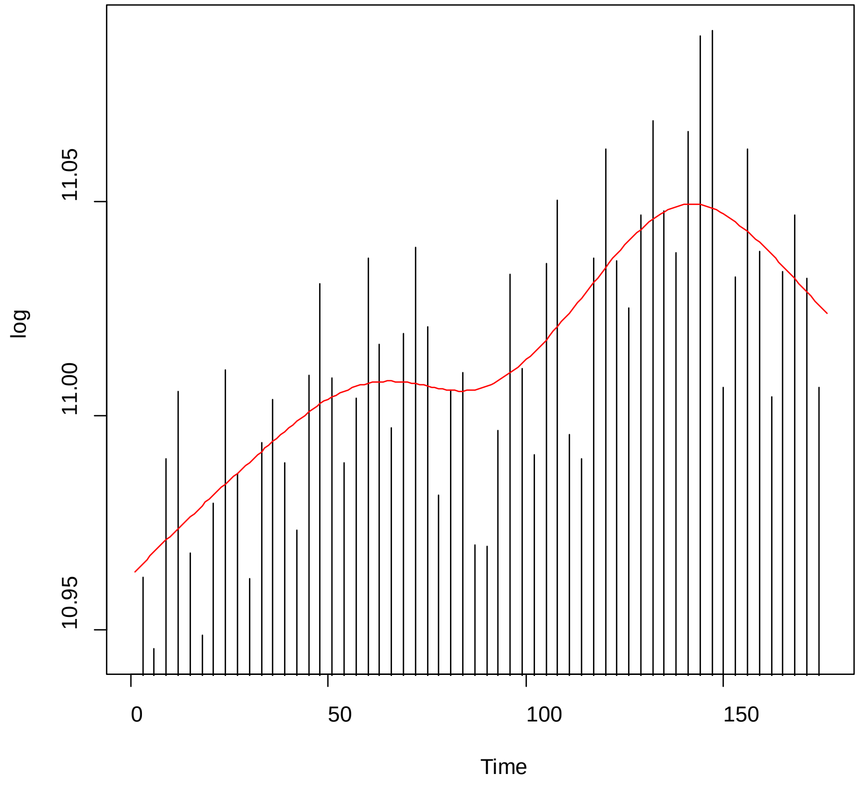


ＧＤＰ支出
135000
125000
115000
105000
0
20
40
60
80
Time
消費
75000
70000
65000
60000
0
20
40
60
80
Time
投資
26000
22000
18000
14000
0
20
40
60
80
Time
政府投資
33000
31000
29000
27000
0
20
40
60
80
Time

インパルス応答



最終消費支出（四半期）



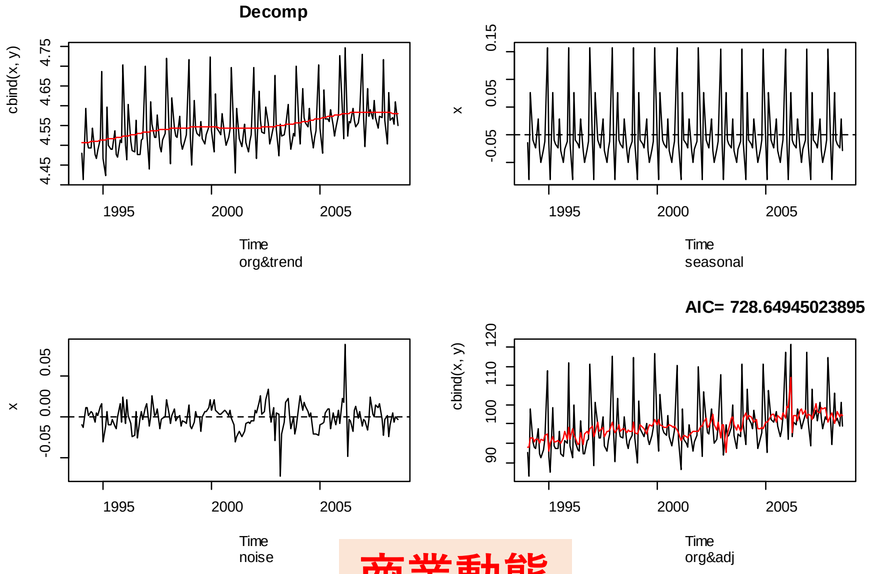
月次トレンドの推定値（ＴＣ）

Log(最終消費）＝トレンド＋季節性＋ノイズ

3，6，9，12月以外は欠損値

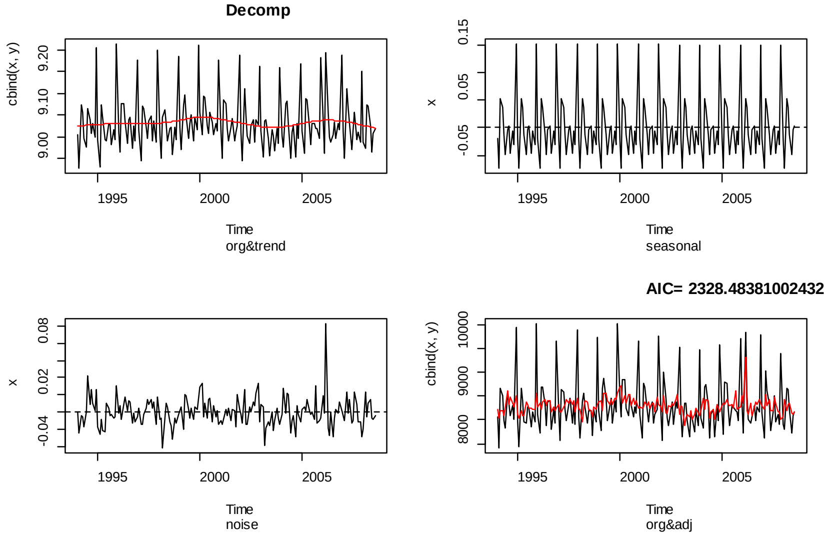
これだと、ただ補間しただけ。月次の情報が入っていない。→　他の消費データの活用

外生変数（説明変数、とりあえず既存の消費関連データを利用）

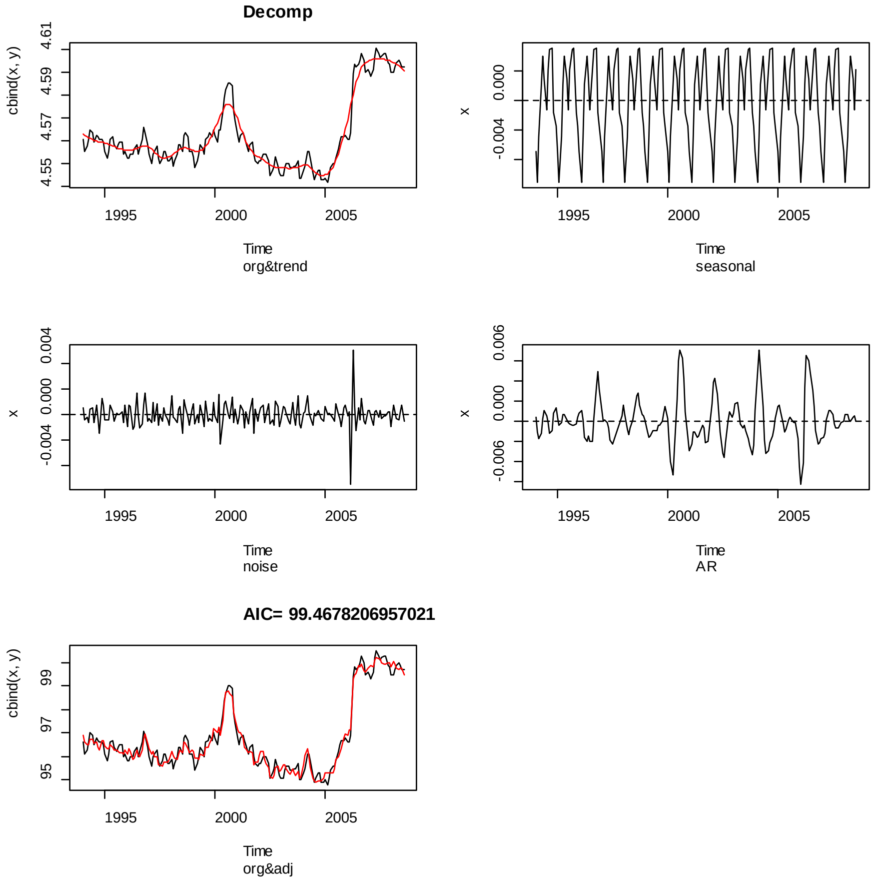


商業動態

すべてのデータに対してＴＣ系列を推定。



家計調査（マクロ、2人以上世帯）



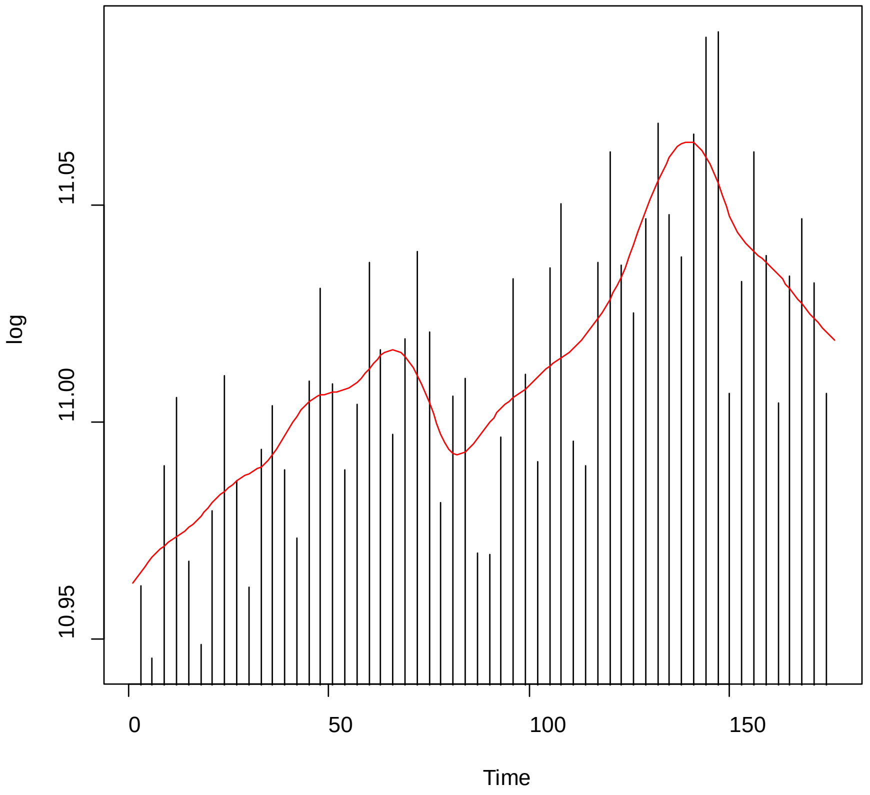
消費者物価

推定はＤｅｃｏｍｐ法にて行った。

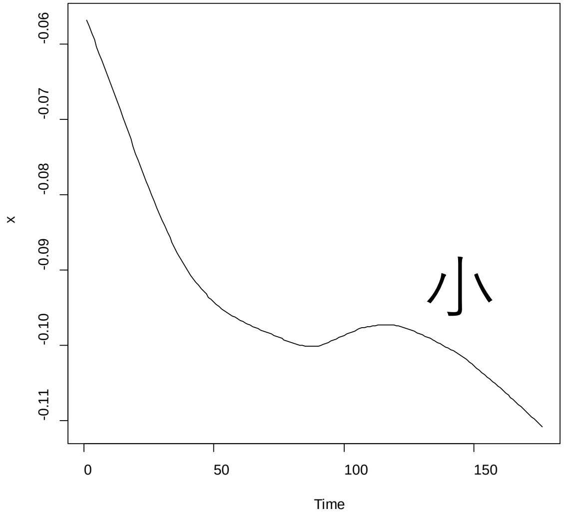
既存の消費データを用いた推計例

Log(最終消費）＝トレンド＋季節性＋回帰部分＋ノイズ

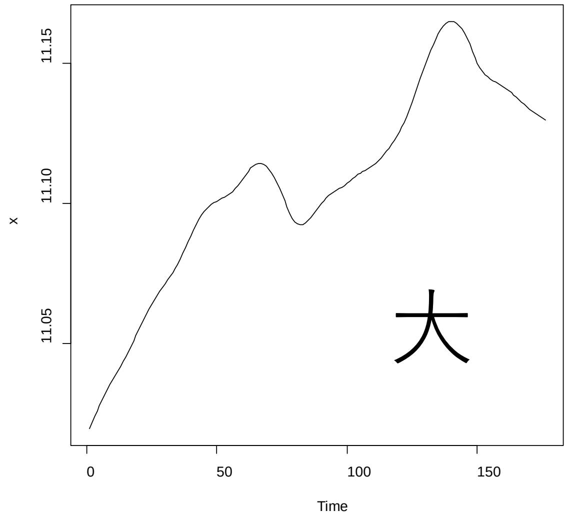
状態空間モデルによる推定（外生変数あり）



月次推計値
（トレンド成分＋回帰成分）



トレンド成分



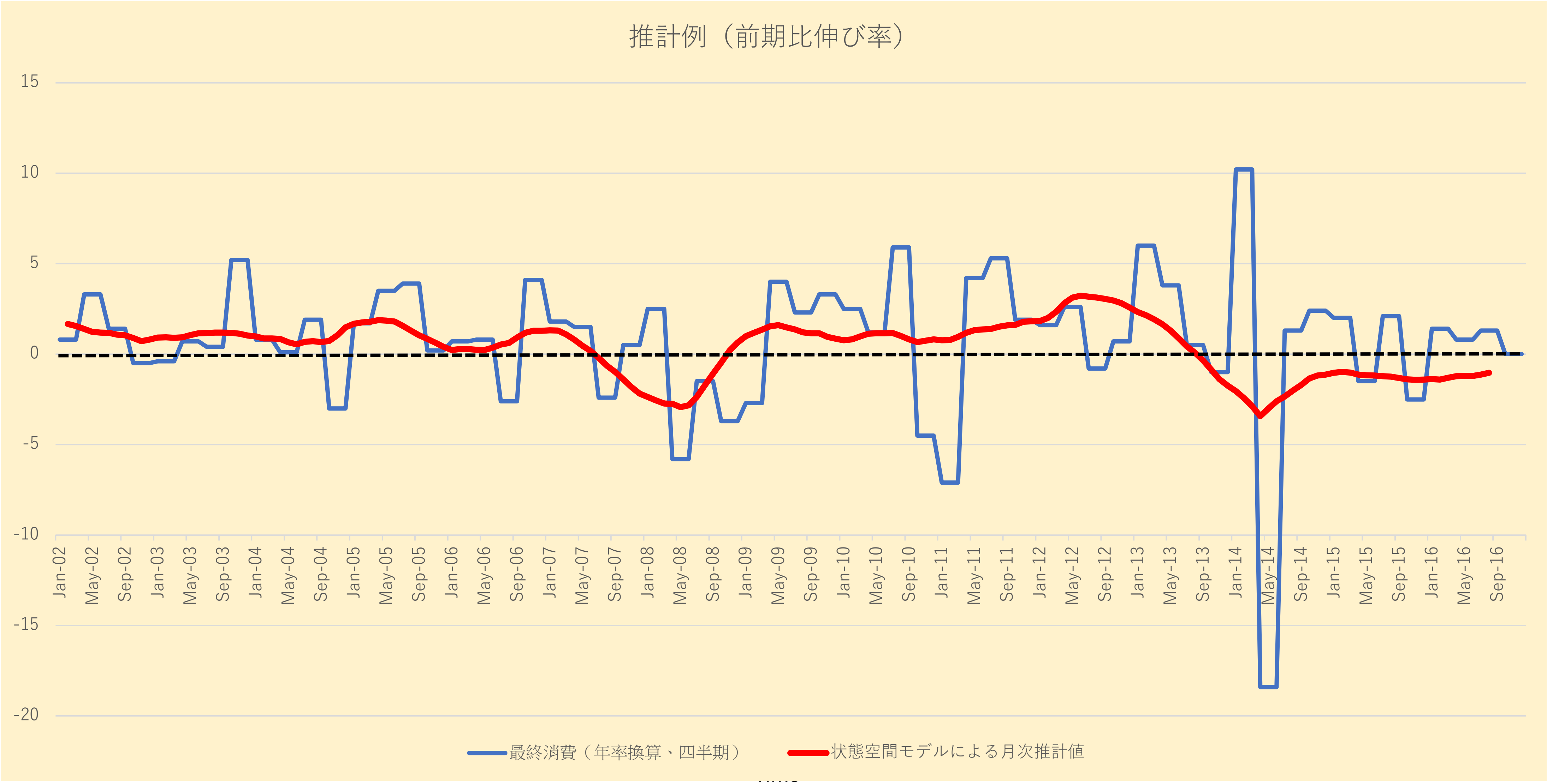
回帰成分

$$\log(Y_t) = T_t + S_t + R_t + I_t$$

$$T_t = 2T_{t-1} - T_{t-2} + u_{1,t}$$

$$S_t = -\sum_{i=1}^{11} S_{t-i} + u_{2,t}$$

$$R_t = a_1 Z_{1,t} + a_2 Z_{t-2} + \cdots + u_{4,t}$$



推計例（前期比伸び率）
15
10
5
0
-5
-10
-15
-20
Jan-02
May-02
Sep-02
Jan-03
May-03
Sep-03
Jan-04
May-04
Sep-04
Jan-05
May-05
Sep-05
Jan-06
May-06
Sep-06
Jan-07
May-07
Sep-07
Jan-08
May-08
Sep-08
Jan-09
May-09
Sep-09
Jan-10
May-10
Sep-10
Jan-11
May-11
Sep-11
Jan-12
May-12
Sep-12
Jan-13
May-13
Sep-13
Jan-14
May-14
Sep-14
Jan-15
May-15
Sep-15
Jan-16
May-16
Sep-16
最終消費（年率換算、四半期）
状態空間モデルによる月次推計値

High-frequency Financial Data and G-Causality Analysis

Kosuke Oya[1] Ryo Kinoshita[2]

[1]Graduate School of Economics, Osaka Univ.

[2]Tokyo Keizai University

2017 年 8 月 5 日
「釧路・経済統計キャンプ 2017」at 釧路公立大学

Outline

1. Review: strength of causality

Time domain: Granger (1963): strength of causality $\{ X_{kt} \} \Rightarrow \{ X_{jt} \}$,

$$1 - \mathrm{var}_j[\,Q\,] \,/\, \mathrm{var}_j[\,Q(k)\,]$$

- Q: set of q stochastic processes $\{ X_{it} \}$, $(i = 1, \ldots, q)$
- $Q(k)$: set Q excluding $\{ X_{kt} \}$
- $\mathrm{var}_j[\,Q\,]$: var. of one step prediction error of the process $\{ X_{jt} \}$ with Q

Frequency domain: spectral decomposition of Granger's measure

- Geweke (1982): for VAR model under specific conditions
- Hosoya (1991) for general stationary process
- Yao and Hosoya (2000): for nonstationary process

Related studies: Relative power contribution (RPC)

- Akaike (1968): under mutulally uncorrelated noise assumption
- Tanokura and Kitagawa (2016): allow possible correlated noises

Causality measures for two series $\{ X(t),\ Y(t) \}$

Geweke (1982)'s measure

$$F_{Y\to X} = \log\left\{ \frac{\mathrm{var}\Big(\text{one step prediction error of } X(t) \text{ given}\{X(s)\}_{-\infty}^{t-1}\Big)}{\mathrm{var}\Big(\text{one step prediction error of } X(t) \text{ given}\{X(s), Y(s)\}_{-\infty}^{t-1}\Big)} \right\}$$

Hosoya (1991)'s measure

$$M_{Y\to X} = \log\left\{ \frac{\mathrm{var}\Big(\text{one step prediction error of } X(t) \text{ given}\{X(s)\}_{-\infty}^{t-1}\Big)}{\mathrm{var}\Big(\text{one step prediction error of } X(t) \text{ given}\{X(s), Y_{0,-1}(s)\}_{-\infty}^{t-1}\Big)} \right\}$$

- $Y_{0,-1}(t)$: residuals of linear projection of $Y(t)$ onto $\{X(s)\}_{-\infty}^{t}$, $\{Y(s)\}_{-\infty}^{t-1}$
  the specific shock of $Y(t)$ which is uncorrelated with $X(t)$
- $M_{Y\to X}$: transform. of Granger's strength of causality for $Y_{0,-1}(t) \to X(t)$

Spectral decomposition: Hosoya (1991)

$$M_{Y\to X} = \int_{-\pi}^{\pi} M_{Y\to X}(\lambda)d\lambda, \quad M_{Y\to X}(\lambda) = \log\left(\frac{\tilde{f}_{11}(\lambda)}{\tilde{f}_{11}(\lambda) - \tilde{f}_{12}(\lambda)\tilde{f}_{22}^{-1}(\lambda)\tilde{f}_{12}^{*}(\lambda)}\right), \quad -\pi < \lambda \le \pi.$$

- $\tilde{f}(\lambda)$: spectral density matrix of $\{ X(t),\ Y_{0,-1}(t) \}$,
- $M_{Y\to X}(\lambda) = F_{Y\to X}(\lambda)$ when $\{ X(t),\ Y(t) \}$ has the stationary AR representation.
- $M_{Y\to X}(\lambda) = -\log\{1 - RPC_{Y\to X}(\lambda)\}$

Assumption

- $\{ X(t), Y(t) \}$ is the second-order stationary and has a spectral density matrix $f(\lambda)$,
- $\int_{-\pi}^{\pi} \log\det f(\lambda)\, d\lambda > -\infty \ \left(\Rightarrow f(\lambda) \text{ has a factorization } f(\lambda) = \frac{1}{2\pi}\Lambda(e^{-i\lambda})\Lambda(e^{-i\lambda})^* \right)$,
- $\Sigma = \Lambda(0)\Lambda(0)^*$ is the cov. matrix of $\underline{\text{one step prediction error of } \{X(t),\ Y(t)\}}$.

$\uparrow$

the resids. of projection $\{X(t),\ Y(t)\}$ onto $\{X(s)\}_{-\infty}^{t-1}$ & $\{Y(s)\}_{-\infty}^{t-1}$

Spectral density function for stationary VARMA model

VARMA model for $Z(t) = (X(t),\ Y(t))'$

$$A(L)Z(t) = B(L)\epsilon(t), \quad \epsilon(t) \sim N(0,\Sigma) \ \text{ and } \ \Sigma = (\sigma_{ij})$$

$$f(\lambda) = \frac{1}{2\pi}\Lambda(e^{-i\lambda})\Lambda^*(e^{-i\lambda}), \quad \Lambda(e^{-i\lambda}) = A^{-1}(e^{-i\lambda})\, B(e^{-i\lambda})\, \Sigma^{1/2}$$

$$\tilde{f}(\lambda) = (\tilde{f}_{ij}),$$

$$\tilde{f}_{11}(\lambda) = f_{11}(\lambda), \quad \tilde{f}_{22}(\lambda) = \tfrac{1}{2\pi}(\sigma_{22} - \sigma_{21}\sigma_{11}^{-1}\sigma_{12}), \quad \tilde{f}_{12}(\lambda) = \tilde{f}_{21}(\lambda)^*,$$

$$\tilde{f}_{21}(\lambda) = \begin{bmatrix} -\Sigma_{21}\Sigma_{11}^{-1} & 1 \end{bmatrix} B(e^{-i\lambda})^{-1}A(e^{-i\lambda}) \begin{bmatrix} f_{11}(\lambda) \\ f_{21}(\lambda) \end{bmatrix}$$

- Let θ and $M(\theta,\ \lambda)$ be the model parameter vector and the causality measure at frequency λ, then the estimator of $M(\theta,\ \lambda)$ is obtained as $M(\hat{\theta},\ \lambda)$

Frequency-wise causality using HF data I

- We provide an alternative approach to obtain the asymp. variance estimate of $M(\hat{\theta}, \lambda)$. This approach does not require numerical differentiation.

Variance estimation via Subsampling

- Let $\{ Z(t) \}$ follow a stationary VARMA process.
- Entire sample $\{ Z(1), \ldots, Z(T) \}$ is divided into non-overlapping subsamples,

$$\{Z(1), \ldots, Z(T_b)\},\ \{Z(T_b+1), \ldots, Z(2T_b)\},\ \ldots,\ \{Z(T-T_b+1), \ldots, Z(T)\},$$

  where T and T_b are the size of the entire sample and the block size of the subsample under the assumption of $T = nT_b$.
- $\hat{\theta}_T$ and $\hat{\theta}_{T_b,i}$ are the estimators using the entire sample and the i-th subsample.
- $\sqrt{T}(\hat{\theta}_T - \theta)$ and $\sqrt{T_b}(\hat{\theta}_{T_b,i} - \theta)$ have the same limiting distribution $N(0, \Psi(\theta))$ under standard regularity conditions with additional assumptions $T_b/T \to 0$ and $T_b \to \infty$ as $T \to \infty$.

Frequency-wise causality using HF data II

- Suppose that $g(\theta, \lambda)$ is a differentiable positive scalar-valued function of θ and λ.

$$\sqrt{T}(g(\hat{\theta}_T, \lambda) - g(\theta, \lambda)) \text{ and } \sqrt{T_b}(g(\hat{\theta}_{T_b,i}, \lambda) - g(\theta, \lambda)) \xrightarrow{d} \mathrm{N}(0,\ V(\theta, \lambda))$$

- Subsampling estimator of $V(\theta, \lambda)$

$$\hat{V}(\lambda) = \frac{1}{n}\sum_{i=1}^{n} T_b(g(\hat{\theta}_{T_b,i}, \lambda) - \bar{g}(\lambda))^2,\ \ \bar{g}(\lambda) = n^{-1}\sum_{i=1}^{n} g(\hat{\theta}_{T_b,i}, \lambda)$$

- Under the conditions that $\{(\sqrt{T}(g(\hat{\theta}_T, \lambda) - g(\theta, \lambda)))^4\}$ are uniformly integrable,

$$\frac{1}{n}\sum_{i=1}^{n} E|\sqrt{T_b}(g(\hat{\theta}_{T_b,i}, \lambda) - g(\theta, \lambda))|^2 \longrightarrow V(\theta, \lambda) \text{ as } n \to \infty$$

  and the proper mixing condition for $\{Z(t)\}$,

$$\hat{V}(\lambda) \to V(\theta, \lambda) \text{ in } L^2 \text{ as } T \to \infty.$$

  See Carlstein (1986) and Fukuchi (1999) for details.

Test for a change in frequency-wise causality using HF data I

- Two subsample periods divided before and after a structural break.
- Period k consists of n_k days, $k = 1$ and 2 stands for before and after a break.
- Assume that a sample size of intra-daily data for the i-th day in period k is

$$T_k^{(i)} = T_k + o(n_k^{-1}).$$

- θ_k and $M(\theta_k, \lambda)$ are the parameter and the measure at freq. λ for period k.
- For i-th day in period k, we have $\hat{\theta}_k^{(i)}$ and $M(\hat{\theta}_k^{(i)}, \lambda)$ using intra-daily data.

Change in measures: $CM(\lambda) = M(\theta_1, \lambda) - M(\theta_2, \lambda)$

Estimates of $CM(\lambda)$: $\widehat{CM}(\lambda) = \bar{M}_1(\lambda) - \bar{M}_2(\lambda), \quad \bar{M}_k(\lambda) = \frac{1}{n_k}\sum_{i=1}^{n_k} M(\hat{\theta}_k^{(i)}, \lambda)$

Test for a change in frequency-wise causality using HF data II

- Because $\sqrt{T_k}(M(\hat{\theta}_k^{(i)}, \lambda) - M(\theta_k, \lambda)) \xrightarrow{d}$ normal with mean 0 and some variance,

$$\sqrt{n_k T_k}(\bar{M}_k(\lambda) - M(\theta_k, \lambda)) \xrightarrow{d} N(0, H_k(\lambda)) \text{ and } \frac{\widehat{CM}(\lambda) - CM(\lambda)}{\sqrt{\frac{H_1(\lambda)}{n_1 T_1} + \frac{H_2(\lambda)}{n_2 T_2}}} \overset{a}{\sim} N(0, 1)$$

- Replacing T_b, n, $\sqrt{T_b}g(\hat{\theta}_{T_b,i}, \lambda)$ and $\sqrt{T_b}\bar{g}(\lambda)$ in $\hat{V}(\lambda)$ with

$$T_k,\ n_k,\ \sqrt{T_k}M(\hat{\theta}_k^{(i)}, \lambda)\ \ \sqrt{T_k}\bar{M}_k(\lambda)$$

provides $\hat{V}(\lambda) = \hat{H}_k(\lambda)$.

Test of no change in measures at frequency λ: $H_0 : CM(\lambda) = 0$ and $H_1 : CM(\lambda) \neq 0$

$$\frac{\widehat{CM}(\lambda)}{\sqrt{\frac{\hat{H}_1(\lambda)}{n_1 T_1} + \frac{\hat{H}_2(\lambda)}{n_2 T_2}}} \xrightarrow{d} \text{N(0, 1) under } H_0.$$

a causal relationship between the Nikkei 225 and the Nikkei 225 mini I

How does Quantitative and Qualitative Monetary Easing (QQE) affect Japanese stock and future market?

- 15 sec. intraday data: index and (future) mini
- sample period: 2013/02/25 - 2013/05/16
- 2013/04/04: break point (BOJ announced the Introduction of "QQE")
- estimate the causalities for each day.
- take averages of the causalities for each pre-event and post-event period.
- test the difference of the two averages of causalities.

pre-event period		event period			post-event period	
2/25	3/25	3/26	4/4	4/15	4/16	5/16

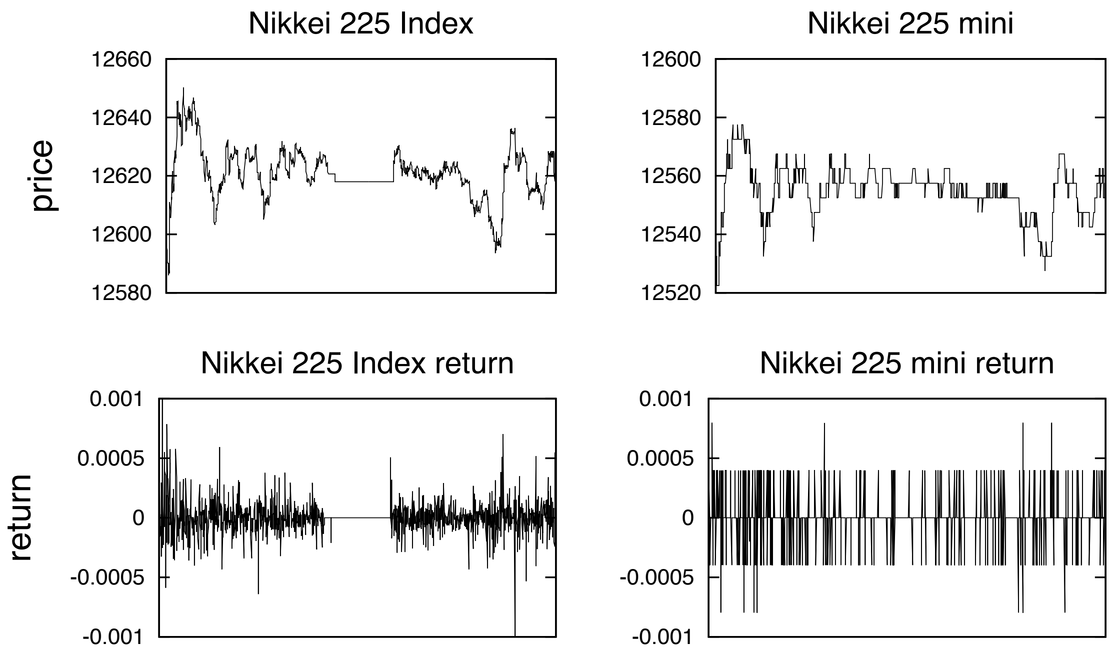
a causal relationship between the Nikkei 225 and the Nikkei 225 mini II

- correspondence between frequencies and cycles when we use the 15-second interval data is shown as follow.

long run	6min cycle	3min cycle	1min cycle	45sec cycle	36sec cycle	30sec cycle
0	$\pi/6$	$2\pi/6$	$3\pi/6$	$4\pi/6$	$5\pi/6$	π

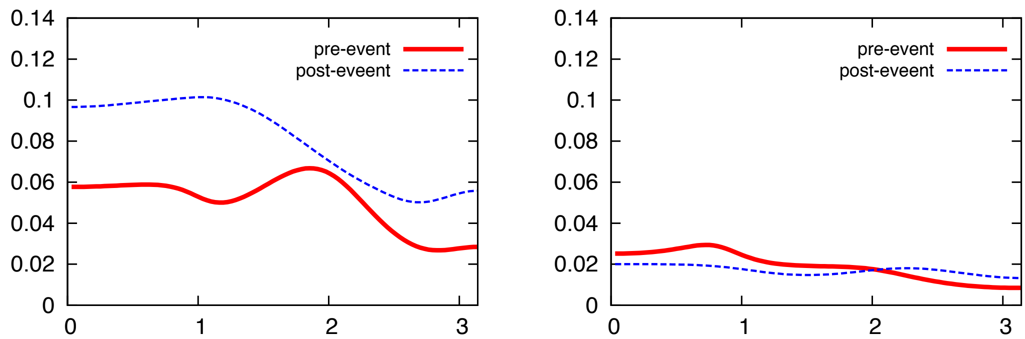
- To compute the log returns, we use mid-quote
- We separately analyze the intra-daily series for the morning session (9:00 to 11:30) and the afternoon session (12:30 to 15:00).
- The intra-daily series of a typical date, March 21, 2013, is shown in next slide.

a causal relationship between the Nikkei 225 and the Nikkei 225 mini III



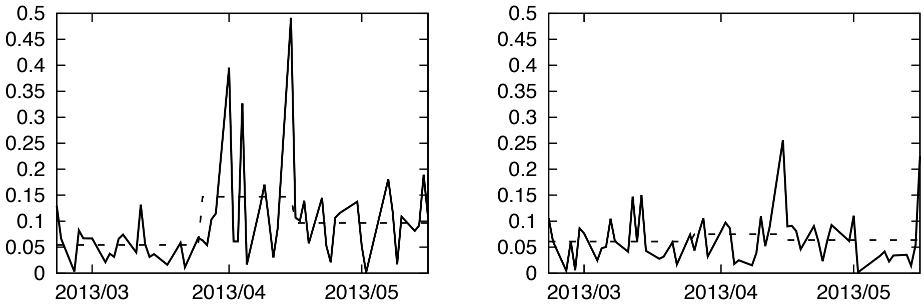
a causal relationship between the Nikkei 225 and the Nikkei 225 mini IV

G-causality in the morning session: futures to index (*left*) and index to f. (*right*)

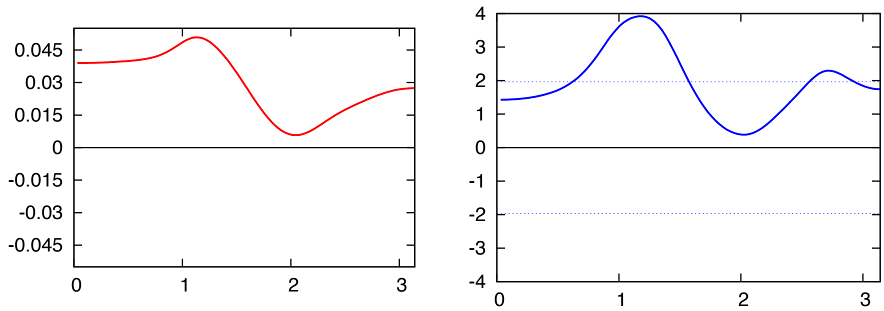


- Causality (futures → index) peaks at a frequency of approximately 1.8 for the pre-event period, and that the peak shifts to near the frequency 1.2 for the post-event period.
- This implies that the component with a 40-second cycle is mainly reflected in the futures in advance of the index for the pre-event period, and the 1.8-minute cycle is a main component for the post-event period.
- We also find that the causality increase for all frequencies after the announcement of QQE.
- In contrast, the causality from the stock index to the futures are small for all frequencies.

a causal relationship between the Nikkei 225 and the Nikkei 225 mini V



Transitions of G-causality estimates at $\pi/3$ (*left*) and at $2\pi/3$ (*right*)



Change in causality (*left*) and test statistics (*right*) in the morning session

a causal relationship between the Nikkei 225 and the Nikkei 225 mini VI

- Causalities are relatively small for the pre-event period and become volatile in the event period at frequency $\pi/3$.
- For the post-event period, the causalities tend to be large compared with those for the pre-event period and the increase is statistically significant.
- In contrast, the small increase at frequency $2\pi/3$ is not significant.
- The changes are significant at frequencies from approximately 0.6 to 1.5, which correspond to one-minute to 2.5-minute cycles.
- The overall causality from the futures to the stock index are 0.051 and 0.081 for the pre- and post-event periods, and this causality change is not significant.

Summarize

- there are some predictabilities from the Nikkei 225 mini to the Nikkei 225 in both morning and afternoon sessions.
- These predictabilities strengthened in most frequencies after the announcement of the QQE.

For details, see chapter 5 in Hosoya et al. (2017).

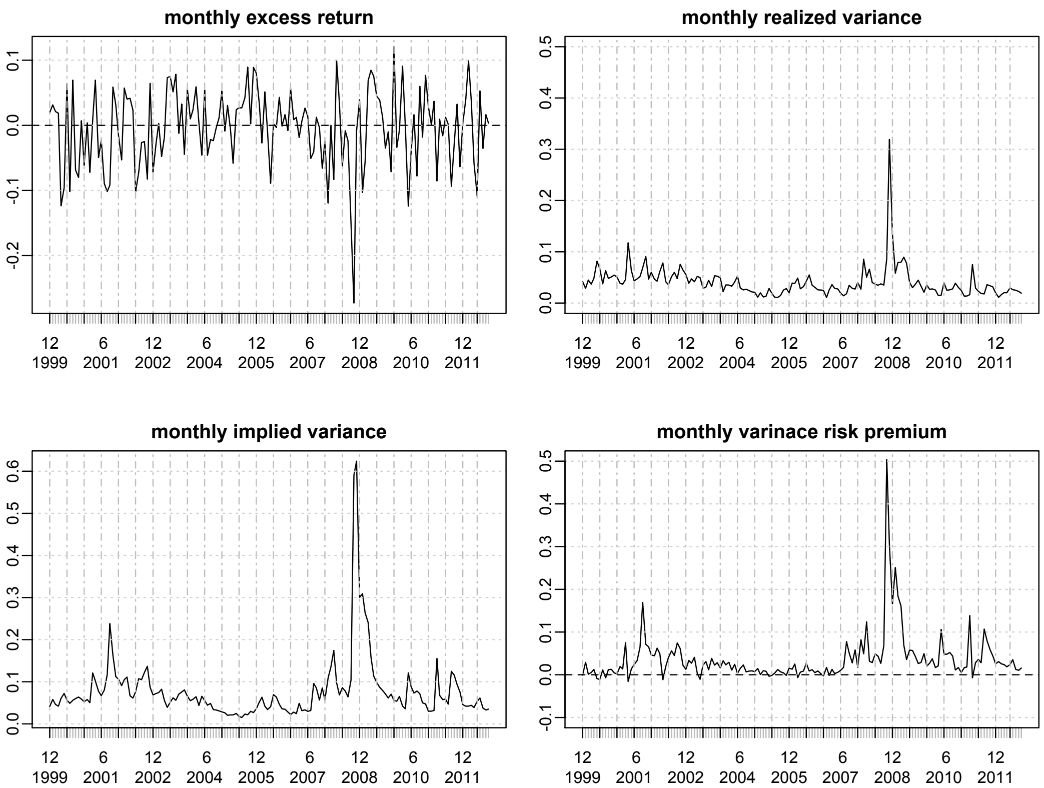
Return predicatiblity of variance risk premium

- the diff. between the risk-neutral and the physical expectations of the future market variation $\langle \ln S \rangle_{t+1}$

$$VRP_t = \underbrace{E_t^{\mathbb{Q}}[\,\langle \ln S \rangle_{t+1}\,]}_{\text{(model free vol. index)}^2} - \underbrace{E_t[\,\langle \ln S \rangle_{t+1}\,]}_{\text{conditional expectation}} = VXJ_t^2 - E_t[\,RV_{t+1}\,],$$

$$\text{where } RV_{t+1} = \sum_{i=1}^{n}(\ln S_{t_i} - \ln S_{t_{i-1}})^2 \xrightarrow{p} \langle \ln S \rangle_{t+1} = \int_t^{t+1} \sigma_s^2 ds$$

- 1 month unit time interval is used in previous studies
- S_{t_i} : i-th transaction price at t_i, $t \le t_0 < \cdots < t_n \le t+1$
- $\ln S_{t_i}$ follows the continuous semi-martingale process
- $\langle \ln S \rangle_t$: quadratic variation
- RV_t and model based $\widehat{RV}_{t+1}$ are used as the proxy for $E_t[\,RV_{t+1}\,]$



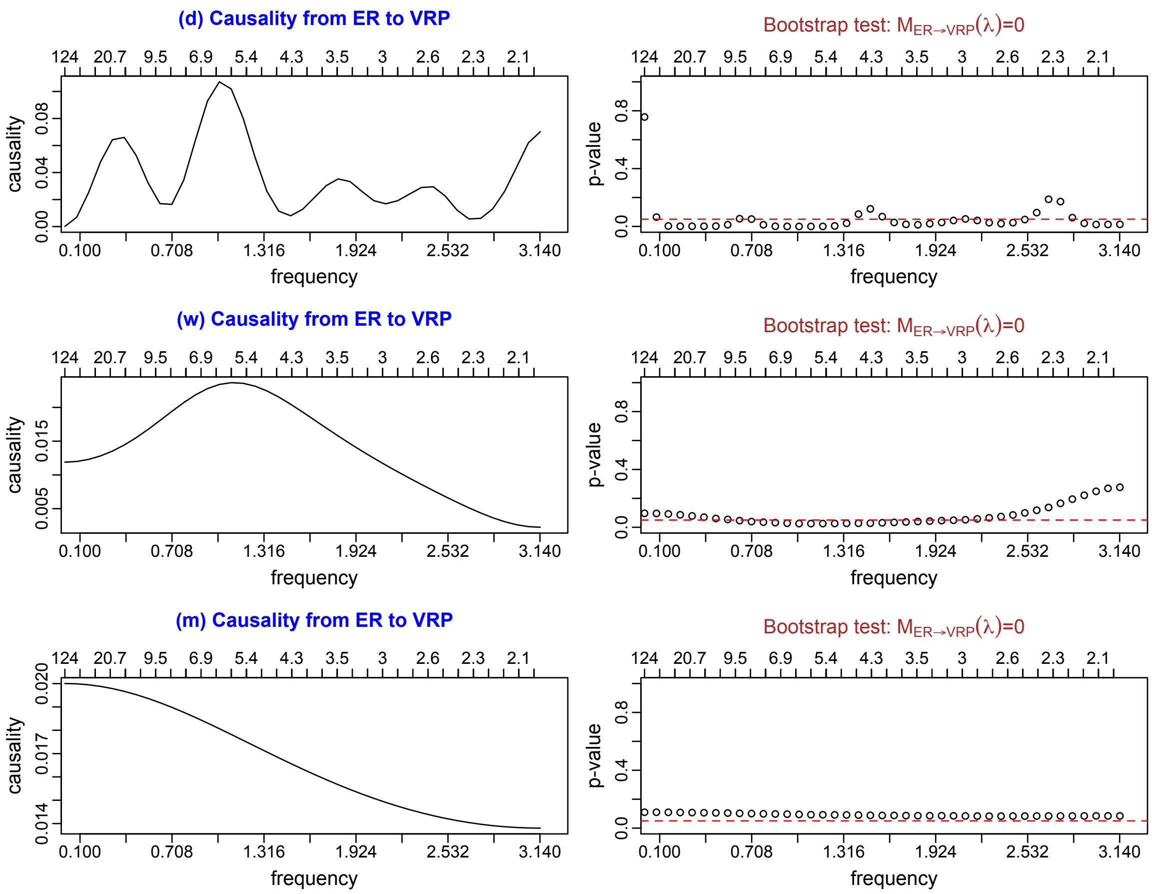
$$E_t[\,RV_{t+1}\,] = \widehat{RV}_{t+1}$$

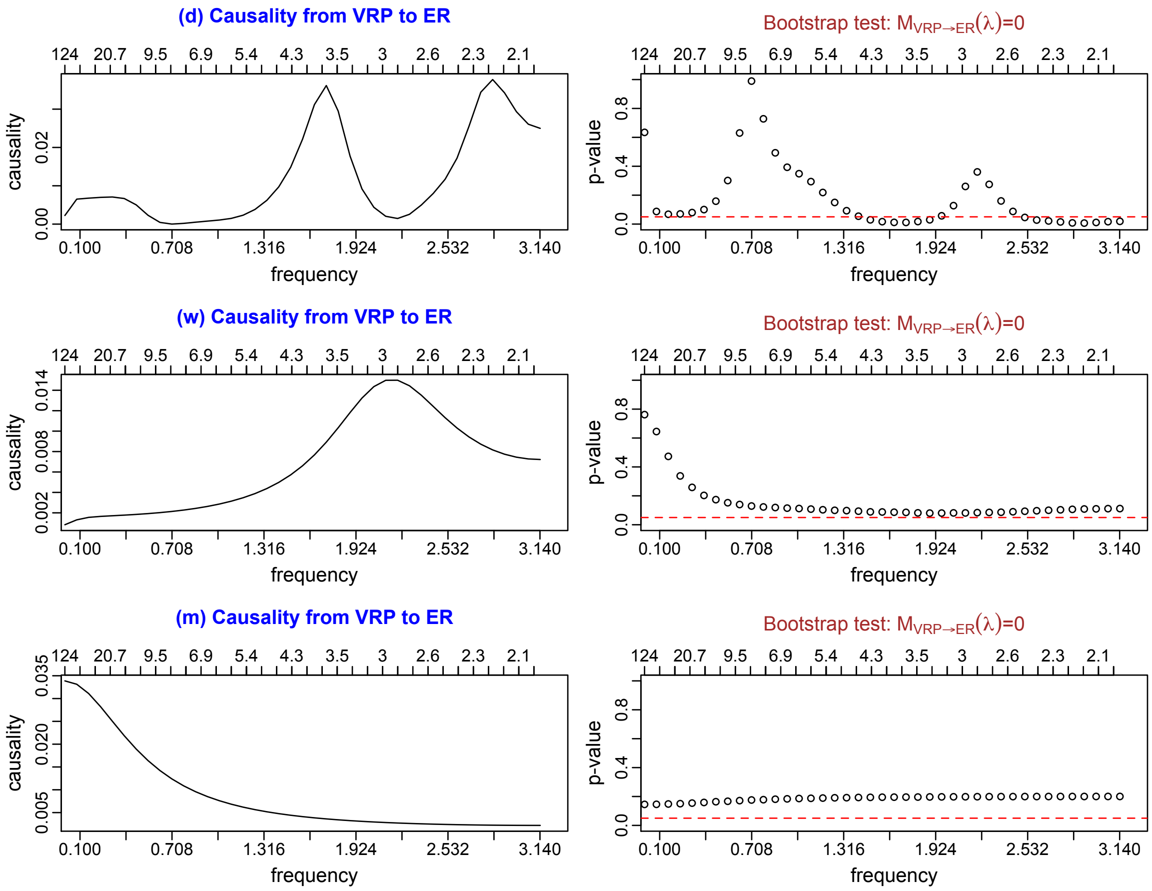
Causality from VRP to ER

VAR(p) model for $z(t) = [\, VRP_t,\ ER_t \,]'$

$$z_t = \Phi_1 z_{t-1} + \cdots + \Phi_p z_{t-p} + \varepsilon_t$$

- $VRP_t = VXJ_t^2 - \widehat{RV}_{t+1}$ where $\widehat{RV}_{t+1}$ is from HAR model by Corsi (2009)
- Daily data is from March 19, 1998 to August 15, 2014 (4035 obs.), weekly data is from July 10, 1998 to April 18, 2014 (824 obs.), monthly data is from December 1999 to September 2012 (154 obs.).
- Lag order of VAR models for monthly, weekly and daily data are selected by Hannan-Quinn information criteria (2, 3 and 11).
- Bootstrap test for $H_0 : M_{VRP\to ER}(\lambda) = 0$ and $H_0 : M_{ER\to VRP}(\lambda) = 0$





Summary for empirical analysis

Causality from ER to VRP

- (d) significant peaks are found at 2 days, 5 days and 21 days cycles.
- (w) unimodal, causality at 5–6 weeks cycle is strongest.
- (m) causality at low frequency is strongest and then declines as $\lambda \uparrow$,
  but insignificant for almost whole freq.

Causality from VRP to ER

- (d) two significant peaks at 4 days and 2.2 days cycles, but insignificant at 5–10 days and 3 days cycles.
- (w) unimodal, peak at mid freq (2.9–3.1 weeks cycles)
- (m) causality at high freq.(2 months) is smallest, then increases towards low freq, but insignificant for whole frequency (2 months or longer cycles).

References

AKAIKE, H. (1968): "On the use of a linear model for the identification of feedback systems," *Annals of the Institute of statistical mathematics*, 20, 425–439.

CARLSTEIN, E. (1986): "The use of subseries values for estimating the variance of a general statistic from a stationary sequence," *The Annals of Statistics*, 1171–1179.

CORSI, F. (2009): "A simple approximate long-memory model of realized volatility," *Journal of Financial Econometrics*, nbp001.

FUKUCHI, J.-I. (1999): "Subsampling and model selection in time series analysis," *Biometrika*, 86, 591–604.

GEWEKE, J. (1982): "Measurement of linear dependence and feedback between multiple time series," *Journal of the American statistical association*, 77, 304–313.

GRANGER, C. W. J. (1963): "Economic processes involving feedback," *Information and control*, 6, 28–48.

HOSOYA, Y. (1991): "The decomposition and measurement of the interdependency between second-order stationary processes," *Probability theory and related fields*, 88, 429–444.

HOSOYA, Y., K. OYA, T. TAKIMOTO, AND R. KINOSHITA (2017): *Characterizing Interdependencies of Multiple Time Series: Theory and Applications (in press)*, Springer.

TANOKURA, Y. AND G. KITAGAWA (2016): *Indexation and Causation of Financial Markets*, Springer.

YAO, F. AND Y. HOSOYA (2000): "Inference on one-way effect and evidence in Japanese macroeconomic data," *Journal of Econometrics*, 98, 225–255.

Simultaneous multivariate point process models and G-causality analysis of international financial markets

国友直人・栗栖大輔・粟屋直

明大 政経, 東大 経済 D2, D1

2017 年 8 月 5 日
研究集会
「新しい時系列計量分析の理論と応用」

Outline

- Introduction
- SMPP models
- SHPP models
- Causality analysis of SHPP models
- Asymptotic theory
- Simulation
- Empirical study
- Conclusion

Introduction

- In economic and financial time series, we sometimes observe abrupt and large movements which often have significant influence not only on a single financial market but also several different markets. (e.g. the financial crisis of 2007-08).
- Such events are observed nearly at the same time if trading hours of some market overlap with others.
- We model these “simultaneous events” as co-jumps of a multivariate marked point process (simultaneous multivariate point process, SMPP).

Introduction

In our paper,

- New marked Hawkes processes which allows co-jumps (simulteneous Hawkes point process, SHPP) are proposed.
- A Granger non-causality test for SHPP is developed.
- A causality measure in the frequency domain is developed.
- We applied proposed methods to empirical studies.

SMPP models

- We use the idea of Solo('07) for the construction of SMPP.
- Consider d-dimensional point process $N = (N_1(t), \dots, N_d(t))_{t\in[0,T]}$ which may have co-jumps. Since $dN_j(t) \in \{0,1\}$, $0 \le \forall j \le d$ and $\forall t \in [0, T]$, there are can be $2^d - 1$ outcomes of N.
- We consider a $(2^d - 1)$-dimensional auxiliary point process N^* which counts these events.

$$N^*(t) = (N_1^*(t), \dots, N_{2^d-1}^*(t)).$$

- N^* do not have co-jumps.

SMPP models

- We consider the case when $d = 3$. In this case N^* is a 7-dimensional point process $N^* = (N_1^*, \ldots, N_7^*)$.
- For example,

  N_j^*, $j = 1, 2, 3$ counts the events that only j-th component jumps.

  N_j^*, $j = 4, 5, 6$ counts the events that two components simultaneously jump.

  N_7^* counts the events that every components jump.

$$
\begin{aligned}
N_1 &= N_1^* + N_4^* + N_5^* + N_7^*, \\
N_2 &= N_2^* + N_4^* + N_6^* + N_7^*, \\
N_3 &= N_3^* + N_5^* + N_6^* + N_7^*.
\end{aligned}
$$

- There is a one-to-one map between N and N^*.
- We use this relationships for the construction of SMPP.

SHPP models

We generalize Hawkes process by using this idea. (simultaneous Hawkes point process (SHPP)).

- Devide the observation period $[0, T]$ into n intervals: $I_i = (t_{i-1}, t_i]$, $0 = t_0 < t_1 < \cdots < t_{n-1} < t_n = T$.
- $Y(s) = (Y_1(s), \ldots, Y_d(s))$: d-dimensional stochastic process (ex. log-return).
- Set threshold values $u = (u_1, \ldots, u_d)$ for each component.
- We count events that at least one component of $Y = (Y_1, \ldots, Y_d)$ exceeds threshold at t_i. We define these events as jumps of a point process N and model N by an auxiliary point process N^*.
- Jump sizes $X_j(t_i) = Y_j(t_i) - u_j$, $(Y_j(t_i) \geq u_i)$ are marks of N^*.

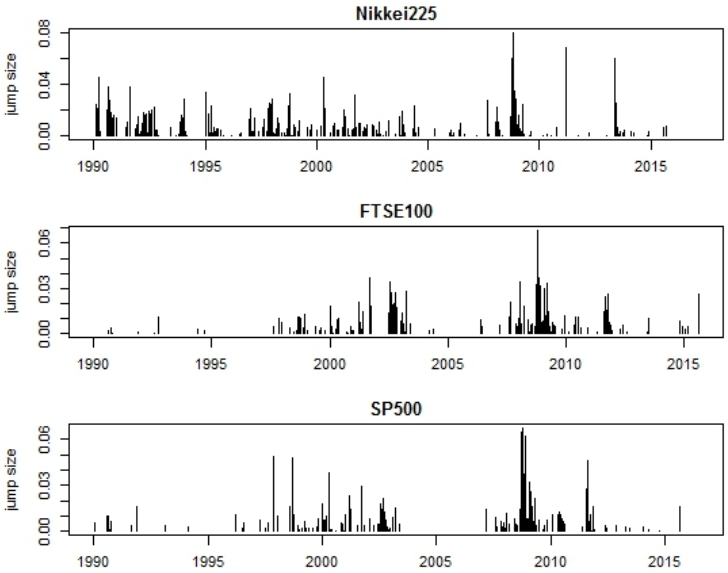
SHPP models

Let $\{t^*_{1,i}\}$, $\{t^*_{2,i}\}$ and b$\{t^*_{3,i}\}$ be jump times of N^*_1, N^*_2 and N^*_3 respectively. We consider the following intensity functions (rate of exceedance, $d = 2$):

$$\begin{aligned}\lambda^*_j(t|\mathcal{H}^*_t) = \lambda^*_{j,0} &+ \sum_{i:t^*_{1,i}\le t} c^*_{j1}(X_1(t^*_{1,i}))g^*_{j1}(t - t^*_{1,i}) \\ &+ \sum_{i:t^*_{2,i}\le t} c^*_{j2}(X_2(t^*_{2,i}))g^*_{j2}(t - t^*_{2,i}) \\ &+ \sum_{i:t^*_{3,i}\le t} c^*_{j3}(X_1(t^*_{3,i}), X_2(t^*_{3,i}))g^*_{j3}(t - t^*_{3,i}),\ j = 1,2,3\end{aligned}$$

$\lambda^*_{j,0}$: positive constant,
$\mathcal{H}^*$: history of N^*,
$c^*_{ji}(\cdot) : \mathbb{R} \to \mathbb{R}_+$: impact function,
$g^*_{ji}(\cdot) : \mathbb{R} \to \mathbb{R}_+$: decay function.

SHPP models


Nikkei225
jump size
0.00
0.04
0.08
1990
1995
2000
2005
2010
2015
FTSE100
jump size
0.00
0.03
0.06
1990
1995
2000
2005
2010
2015
SP500
jump size
0.00
0.03
0.06
1990
1995
2000
2005
2010
2015

SHPP models

Point process models have been used in the literature of financial time series analysis as models of “regular” time series.

- Grothe et al.('14, JoE)

$$c_{ji}(x) = 1 + G_{ji}^{\leftarrow}(F_{i,t}(x_i)), \quad g_i(t) = e^{-\gamma_i t}, \quad X_j(t) \sim \mathrm{GPD}(\xi_j, \sigma_j(t)),$$

$G^{\leftarrow}$: inverse dist. func. of the exp. dist. with mean δ_{ji}, $\gamma_i > 0$.

- Kunitomo, Ehara and Kurisu('17, JJSS(in Japanese))

$$c_{ji}(x) = A_{ji} x_i^{\delta_{ji}} + B_{ji}, \quad g_i(t) = e^{-\gamma_i t}, \quad X_j(t) \sim \mathrm{GPD}(\xi_j, \sigma_j),$$

$A_{ji}, B_{ji}, \gamma_i > 0$, $0 \le \delta_{ji} \le 1$.

SHPP models

In our study we consider the SHPP with the following intensity function:

$$\begin{aligned}\lambda_j^*(t|\mathcal{H}_t^*) = \lambda_{j,0}^* &+ \sum_{i:t_{1,i}^* \le t} c_{j1}^*(X_1(t_{1,i}^*))e^{-\gamma_1^*(t-t_{1,i}^*)} \\ &+ \sum_{i:t_{2,i}^* \le t} c_{j2}^*(X_2(t_{2,i}^*))e^{-\gamma_2^*(t-t_{2,i}^*)} \\ &+ \sum_{i:t_{3,i}^* \le t} c_{j3}^*(X_1(t_{3,i}^*), X_2(t_{3,i}^*))e^{-\gamma_3^*(t-t_{3,i}^*)}, \ j = 1, 2, 3.\end{aligned}$$

$X_i(s) \overset{i.i.d.}{\sim} \mathrm{GPD}(\sigma_i, \xi_i)$, $\forall s \in \mathbb{R}$, $X_i \perp\!\!\!\perp X_j$, $i \neq j$.

Stationarity of SHPP model

SHPP model is stationary $\Leftrightarrow \mathrm{spr}(C^*(\Gamma^*)^{-1}) < 1$,
where $C^* = (C_{ji}^*)_{1 \le i,j \le 3}$, $C_{ji}^* = E[c_{ji}^*(X_1)]$,
$\Gamma^* = diag(\gamma_1^*, \ldots, \gamma_3^*)$.

Causality analysis of SHPP models

Florens and Fougére('96, Ecta) generalized the concept of Granger causality (Granger('69, Ecta)) to continuous time stochastic processes. The IGNC can be seen as a completion of their analysis. We consider the case when $d = 2$.

- Granger-non-causality(GNC) ($N = N^*$, no co-jumps)

  N_2 is G-non-causal to $N_1 \Leftrightarrow c_{12}(x) = 0 \Leftrightarrow \alpha_{12} = 0$.

- Instantaneous GNC(IGNC) ($N \neq N^*$, co-jumps)

  N_2 is instantaneously G-non-causal to N_1
  $\Leftrightarrow c^*_{12}(x) = 0, c^*_{13}(x) \neq 0$.

  N_2 is IGNC to N_1
  $\Leftrightarrow c^*_{12}(x) \neq 0, c^*_{13}(x) = 0$.

  N_2 is IGNC to N_1
  $\Leftrightarrow c^*_{12}(x) = c^*_{13}(x) = 0$.

Hawkes('71, Biometrika) introduced the spectral density matrix of a multivariate Hawkes process with the following intensity function:

$$\lambda(t) = \lambda_0 + \int_{-\infty}^{t} \Gamma(t-s)N(ds),$$

where $\Gamma(t) = (\gamma_{ij}(t))$ is a $d \times d$ matrix and $\gamma_{ij}(t) = 0$ for $t < 0$. Let Γ^* be the Fourier transformation of Γ. Then the spectral density matrix is given by

$$f(\omega) = \frac{1}{2\pi}[I_d - \Gamma^*(\omega)]^{-1}\Sigma[I_d - (\Gamma^*(\omega))^T]^{-1},$$

where $\Sigma = diag(\sigma_{ii})$ is the variance matrix of N.

Causality analysis of SHPP models

When there can be co-jumps, the spectral density matrix of N^* is given by

$$f(\omega) = \frac{1}{2\pi}[I_d, O][D - D\Gamma^*(\omega)]^{-1}\Sigma^*[D^\top - (\Gamma^*(\omega))^\top D^\top]^{-1}[I_d, O]^\top,$$

where $\Sigma^* = diag(\sigma^*_{ii})$ is the variance matrix of N, D is a transformation matrix of N into N^*. Since the (i,i)-component of $f(\omega)$ can be represented as

$$f_{ii}(\omega) = \sum_{k=1}^{2^d-1} |a_{ik}(\omega)|^2 \sigma^*_{ii}, \;\; i,j = 1,\ldots,2^d-1$$

where a_{ik} are functions of complex variables, we define RPC(relative power contribution) and IRPC(instantaneous RPC) by

$$RPC_{k\to i}(\omega) = \frac{|a_{ik}(\omega)|^2\sigma_{kk}}{f_{ii}(\omega)}, \;\; i = 1,\ldots,2^d-1, k = 1,\ldots,d,$$

$$IRPC_{k\to i}(\omega) = \frac{|a_{ik}(\omega)|^2\sigma_{kk}}{f_{ii}(\omega)}, \;\; i = 1,\ldots,2^d-1, k = d+1,\ldots,2^d-1.$$

Asymptotic theory

We give theoretical results on GNC and IGNC test.

- $N = (N_i)_{1 \leq i \leq d}$: d-dimensional $(\mathcal{F}_t)_{t \in \mathbb{R}_+}$ adapted simple point process on $\mathbb{R}_+$.
- $A = (A_i)_{1 \leq i \leq d}$: continuous compensator of N.
- $g_T(t) = (g_T^{(i,j)}(t))$: $p \times d$ predictable process.
- $\Sigma = (\sigma_{i,j})_{1 \leq i,j \leq p}$: $\mathcal{F}_0$-measurable non negative definite random matrix.

Asymptotic theory

Assume the following conditions: When $T \to \infty$,

$$\max_{1\le i,j\le p}\max_{1\le k\le d}\left(E\left[\frac{1}{T}\int_0^T |g_T^{(i,k)}(t)g_T^{(j,k)}(t)|dA_k(t)\right]\right) < \infty, \tag{1}$$

$$\frac{1}{T^{1+\delta}}\max_{1\le k\le d} A_k(T) \xrightarrow{P} 0, \quad \forall\delta > 0, \tag{2}$$

$$\frac{1}{T}\int_0^T \sum_{k=1}^d g_T^{(i,k)}(t)g_T^{(j,k)}(t)dA_k(t) \xrightarrow{d} \sigma_{i,j}, \quad 1 \le i \le d, \tag{3}$$

$$\max_{1\le k\le d} E\left[\frac{1}{T}\int_0^T \|g_T^{(\cdot,k)}(t)\|^2 I(\|g_T^{(\cdot,k)}(t)\| > c)dA_k(t)|\mathcal{F}_0\right] \xrightarrow{P} 0, \forall c > 0. \tag{4}$$

$g_T^{(\cdot,k)}(t) = (g_T^{(1,k)}(t), \ldots, g_T^{(p,k)}(t))^\top$.

Asymptotic theory

Theorem 1

Suppose conditions (1), (2), (3) and (4) are satisfied. Then, we have that

$$X_T = \frac{1}{\sqrt{T}} \int_0^T \sum_{k=1}^{d} g_T^{(\cdot,k)}(t)[dN_k(t) - dA_k(t)] \xrightarrow{\mathcal{F}_0-stably} N_p(0, \Sigma).$$

- For MLEs of multivariate point processes, $X_T = \frac{1}{\sqrt{T}} \frac{\partial L_T(\theta)}{\partial \theta}$.
- The asymptotic variance can be random.
- In this case, the Wilks' property (the asymptotic properties of LR test) holds.

Asymptotic theory

- $L_T(\theta)$: log-likelihood of N
- $L_T^*(\theta)$: log-likelihood of N^*
- θ_0: true parameters of N^*, $\theta_0 \in \Theta(\subset \mathbb{R}^p)$: compact.
- $\hat{\theta}_{ML}$: MLE of N
- $\hat{\theta}_{ML}^*$: $(p-r) \times 1$ MLE of N^*, $0 \leq r \leq p$. $\hat{\theta}_{ML}^* \subset \Theta_1(\subset \Theta)$.

Asymptotic theory

We assume the following conditions for the asymptotic properties of LR test:

$I(\theta_0)$: $\mathcal{F}_0$-measurable p.d. matrix, When $T \to \infty$,

$$\frac{1}{T}\int_0^T \sum_{k=1}^{d} \left(\frac{\partial \log \lambda_k(t,\theta)}{\partial \theta}\right)\left(\frac{\partial \log \lambda_k(t,\theta)}{\partial \theta}\right)^{\top} \lambda_k(t,\theta)dt \xrightarrow{P} I(\theta_0), \qquad (5)$$

$$\frac{1}{T}\int_0^T \sum_{k=1}^{d} \frac{1}{\lambda_k(t,\theta)} \frac{\partial^2 \lambda_k(t,\theta)}{\partial\theta\partial\theta^{\top}}[dN_k(t) - \lambda_k(t,\theta)dt] \xrightarrow{P} 0. \qquad (6)$$

$$\frac{1}{\sqrt{T}}\int_0^T \sum_{k=1}^{d} \frac{\partial \log \lambda_k(t,\theta)}{\partial \theta}[dN_k(t) - \lambda_k(t,\theta)dt] \xrightarrow{\mathcal{F}_0-stably} N_p(0, I(\theta_0)), \qquad (7)$$

Asymptotic theory

Theorem 2

(i) When there exist co-jumps, $\hat{\theta}_{ML}$ is not consistent under the likelihood function $L_T(\theta)$.

(ii) Suppose conditions (5), (6), (7) are satisfied. Then we have that

$$2(L_T^*(\theta_0) - L_T^*(\hat{\theta}_{ML}^*)) \xrightarrow{d} \chi(r),$$

where r is the number of restrictions of θ_0 and $\chi(r)$ is the χ^2 random variable with r degree of freedoms.

- We apply this result for the Granger non-causality test.

Simulation

DGP: $N^* = (N_1^*, N_2^*, N_3^*)$,

$$\lambda_j^*(t|\mathcal{H}_t^*) = \lambda_{j,0}^* + \sum_{i:t_{1,i}^* \le t} \alpha_{j,1}^* X_1(t_{1,i}^*) e^{-\gamma^*(t-t_{1,i}^*)}$$
$$+ \sum_{i:t_{2,i}^* \le t} \alpha_{j,2}^* X_1(t_{2,i}^*) e^{-\gamma^*(t-t_{2,i}^*)}$$
$$+ \sum_{i:t_{3,i}^* \le t} \alpha_{j,3}^* \max(X_1(t_{3,i}^*), X_2(t_{3,i}^*)) e^{-\gamma^*(t-t_{3,i}^*)}, \quad j = 1, 2, 3,$$

$X_j(t_{k,l}^*) \overset{i.i.d.}{\sim} \mathrm{GPD}(\sigma_j, \xi_j)$, $j = 1, 2$,
We used Gaussian copula for the dependence of $(X_1(t_{3,l}^*), X_2(t_{3,l}^*))$.
$(\sigma_1, \xi_1) = (0.007, 0.22)$,
$(\sigma_2, \xi_2) = (0.008, 0.15)$.

Simulation

	α^*_{11}	α^*_{12}	α^*_{13}	α^*_{21}	α^*_{22}	α^*_{23}	
True	0.57000	0.00000	0.19000	0.00010	0.71000	0.09500	
Mean	0.63641	0.00259	0.12387	0.03994	0.76318	0.07905	
RMSE	0.01045	0.00426	0.00913	0.00568	0.01004	0.00557	
	α^*_{31}	α^*_{32}	α^*_{33}	γ^*	$\lambda^*_{1,0}$	$\lambda^*_{2,0}$	$\lambda^*_{3,0}$
True	0.05900	0.12000	0.20000	0.02700	0.00930	0.00530	0.00084
Mean	0.06748	0.13922	0.11315	0.02859	0.00853	0.00427	0.00107
RMSE	0.00272	0.00380	0.00963	0.00033	0.00019	0.00017	0.00007

Table: Summary of numerical experiments. Simulation size $N = 100$. For GPD(σ_j, ξ_j), we set $(\sigma_1, \xi_1) = (0.007, 0.22)$, $(\sigma_2, \xi_2) = (0.008, 0.15)$.

Empirical study

- We test GNC and IGNC among financial markets, Tokyo, New York, London and Hong Kong by using SHPP model.
- Kunitomo et al.('17, JJSS) investigates the Granger causality between Tokyo and New York or London by using multivariate marked Hawkes process (special case of SHPP).
- In that case, since the trading hours of Tokyo do not overlap with those of other markets, there exist no co-jumps.
- Example : Tokyo, Hong Kong

  $\rightarrow$Trading hours overlap with each other.

  $\rightarrow$ We cannot use the model used in Kunitomo et al.('17, JJSS) since there can be co-jumps (the simultaneous jumps)

  $\rightarrow$ We need to use SHPP models which allows co-jumps.

Empirical study

- Market index: Nikkei225, S&P500, FTSE100, HSI.
- Period: 1990/1/2 $\sim$ 2015/8/25.
- Model: Bivariate SHPP model.
- Threshold: $u = -2\%$ (about 5% of the data).
- Method: MLE $\rightarrow$ LR test.

Empirical study（Jump Size）

Markets	σ_i	ξ_i
Tokyo	0.00806(0.00065)	0.16874(0.06431)
New York	0.00765(0.00076)	0.21538(0.08082)
London	0.00850(0.00084)	0.10799(0.07717)
Hong Kong	0.00861(0.00055)	0.15773(0.05076)

Table: Estimated GPD parameters.

Empirical study（Model Selection）

As a first step, we choose a model which used in the causality analysis based on the minimum AIC principle.

- Model 1: $c_{ji}^*(x) = 1$,
- Model 2: $c_{ji}^*(x) = a_{ji}^* \max(x_{(j)})$,
- Model 3: $c_{ji}^*(x) = a_{ji}^* \max(x_{(j)})^{\delta^*}$.

Empirical study（Causality Test）

Null	T-NY	T-L
$c_{21}(x) = 0$	accept	accept
$c_{12}(x) = 0$	reject	reject

Table: GNC test at the 5% significant level. T: Tokyo (N_1), NY: New York (N_2), L: London (N_2).

Null		T-HK
Type 1	$c^*_{12}(x) = 0$	accept
	$c^*_{21}(x) = 0$	accept
Type 2	$c^*_{13}(x) = 0$	reject
	$c^*_{23}(x) = 0$	accept
Type 3	$c^*_{12}(x) = 0,\ c^*_{13}(x) = 0$	reject
	$c^*_{21}(x) = 0,\ c^*_{23}(x) = 0$	accept

Table: IGNC test at the 5% significant level, T: Tokyo(N_1), HK: Hong Kong(N_2).

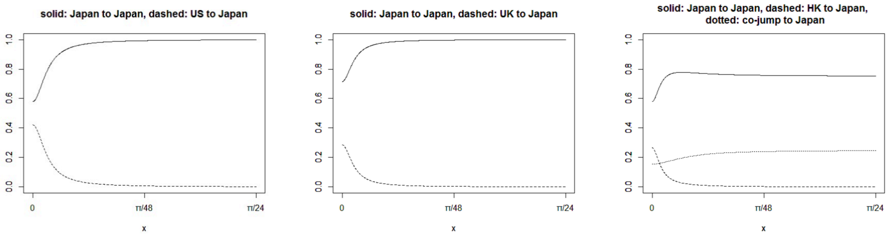


Figure: RPC and IRPC from New York, London and Hong Kong to Tokyo.

- For the relationships between Tokyo-NY, and Tokyo-London, the self contribution play major contribution while there are some contribution from NY or London to Tokyo.
- For the relationship between Tokyo-HK, the instantaneous contribution plays a major contribution in all frequencies as well as the self contribution.

Conclusion

- Point process models with co-jumps are introduced.
- We proposed SHPP models.
- Asymptotic properties of MLE and LR test of SMPP models are investigated.
- We checked finite sample properties of MLEs of SHPP by simulation.
- GNC, IGNC, RPC and IRPC are defined and applied to the Granger causality analysis among financial markets.

Reference

- Florens, J.P. and Fougére, D. (1996), Noncausality in continuous time. *Econometrica*, 64, 1195-1212.
- Granger, C.W. (1969), Investigating Causal Relations by Econometric Models and Cross-Spectral Methods. *Econometrica*, 37, 161-194.
- Grothe, O., Korniichuk, V. and Manner, H. (2014), Modeling multivariate extreme events using self-exciting point processes. *J. Econometrics* 182, 269-289.
- Hawkes, A. G. (1971), Spectra of Some Self-Exciting and Mutually Exciting Point Processes. *Biometrika*, 58-1, 83-90.

- Kunitomo, N., Ehara, A. and Kurisu, D. (2017), Causality analysis of financial markets by using the multivariate Hawkes type models. *J. Japan Statist. Society* 46, 137-171.
- Solo, V. (2007), Likelihood function for multivariate point processes with coincidences. *Proceedings of the 46th IEEE Conference on Decision and Control New Orleans, LA, USA, Dec. 12-14, 2007*