

経済統計・政府統計の理論と応用 2019¹

国友直人² & 山本拓³ & 久保川達也⁴

(共編)

2019年2月

¹文部科学省・科学研究費プロジェクト「経済統計・政府統計の応用からの提言」(2015年度~2018年度)が、2019年1月31日に東京大学経済学研究科において開催した講演をまとめたものである。

²明治大学政治経済学部

³一橋大学

⁴東京大学経済学研究科

前書き

本報告書は、日本学術振興会・科学研究費プロジェクト「経済統計・政府統計の理論と応用からの提言」（2015年度－2018年度、研究代表者：山本 拓）が、2019年1月31日（木）に東京大学小島ホールにおいて開催した2018年度の研究集会における講演内容をまとめたものである。

本プロジェクトの目的は、経済統計・政府統計における主要な課題の技術的および制度的問題を、統計学的な立場から理論的・学術的に検討すること、ならびに経済統計・政府統計の応用の際の問題点を明らかにし、それらの解決案を模索・提言することである。

本研究集会は上記の目的のもとに、プロジェクトの過半をしめる統計学の研究者と実際に経済統計・政府統計の作成者または直接的な利用者として携わっている方々との直接的な交流の場を提供しようとするものである。

研究集会は9件の報告からなり、第1セッションでは、公的統計に関わる課題・問題が扱われた。まず標本調査における複数回答における過少回答の傾向の分析とそれに対する対策が議論された。次に高齢化社会に向けて、主として「全国消費実態調査」に基づいて、高齢無業者のみを対象とした貯蓄・消費行動の実証分析が紹介された。最後は少子高齢化に伴う人口構成の予測、ならびに人的資産の質すなわち人口資質についてのさまざまな側面からの評価・予測について議論が展開された。

第2セッションでは、経済統計における課題への理論的なアプローチが取り上げられた。まず個人情報秘匿に関するランダムネスの評価について理論的な検討がされた。次に、多変量時系列データについて、統一的に長期トレンド、季節性、短期の要因に分解する統計的方法が提案された、最後に、テキスト・データから情報を取り出し予測を行う試みが紹介された。

第3セッションではさまざまな経済統計について、対応するモデル分析の新しい統計的手法が取り上げられた。まずは非定常な経済データの回帰モデルにおいて、非定常性の困難さを回避する簡単な推定法に試みが紹介された。次に、パネル・データの推定において covariance structure analysis を用いた有用な推定法が提案された。最後に、ファイナンス・データにおける多変量の確率的ボラティリティ・モデルがベイシアン的視点から議論された。

上記のように多岐にわたるトピックを扱う研究集会が、経済統計・政府統計に様々な形で関わる人々の刺激となり、今後の分析の発展ならびに各種統計自体の改善つながることを期待する次第である。

2019年2月

編者

科学研究プロジェクト「経済統計・政府統計の理論と応用」研究集会

日程： 2019年1月31日(木)

会場： 東京大学経済学研究科 小島ホール2階 コンファレンスルーム

オーガナイザー： 国友 直人・山本 拓

プログラム

<挨拶>

13:00~13:05 山本拓 (一橋大学)

<セッション I> 公的統計の課題解決

Chair: 山本拓 (一橋大学)

13:05~13:35 「複数回答方式における過少回答傾向とその対策」 土屋隆裕 (横浜市立大学)

13:35~14:05 「高齢無業世帯の貯蓄と消費行動について」 川崎茂 (日本大学)

14:05~14:35 「人口減少・超高齢社会と人的資本— 政府統計における人口資質の視点 —」
金子隆一 (明治大学)

<休憩>

<セッション II> 経済統計の課題解決

Chair: 川崎茂 (日本大学)

14:50~15:20 「ランダムネスによるプライバシー保護の評価」 星野伸明 (金沢大学)

15:20~15:50 「経済時系列分析の新展開：多次元・トレンド・季節性への SIML アプローチ」
国友直人 (明治大学)

15:50~16:20 「テキストデータからの情報抽出を利用した時系列予測」 川崎能典 (統計数理研究所)

<休憩>

<セッション III> 経済統計データ分析の展開

Chair: 国友直人 (明治大学)

16:35~17:05 「Statistical Inference in Regression Models with Possibly
Integrated Processes」 * 千木良弘朗 (東北大学)・山本拓 (一橋大学)

17:05~17:35 「Covariance Structure Analysis of Panel Regression Models」 早川和彦 (広島大学)

17:35~18:05 「Multivariate Stochastic Volatility Model with Realized Volatilities and Pairwise
Realized Correlations」 * 大森裕浩 (東京大学)・山内雄太 (東京大学)

*は共著の場合の報告者

Theory and Applications of Economic and Government Statistics 2019

Abstract

This report is a summary of the unpublished talks and papers presented at the fourth and the final conference held on January 31, 2019 under the title "Theory and Applications of Economic Statistics and Government Statistics." The topics of the talks are on (i) recent issues of government statistics, (ii) recent issues of economic statistics, and (iii) statistical methodologies for economic statistics. The papers presented at the fourth conference have been written in Japanese or English.

複数回答方式における過少回答傾向とその対策

土屋 隆裕（横浜市立大学）

1

あなたは、ここ1ヶ月の間に次のようなことがありましたか？

複数回答方式

問1 あなたは、ここ1ヶ月の間に次のようなことがありましたか？(MA)

- 気が散りやすい、ひとつのことに集中しにくい
- よく寝れない、もしくは睡眠薬が欠かせない
- 日付や曜日を思い出すのに時間がかかる
- 調子が良い時と悪い時の違いが大きい
- ここ半年間で2kg以上体重が増った
- わけもなく疲れる、もしくは何もはじめる気がしない
- 目が回る、もしくはふわふわする
- 熱っぽい、もしくは熱が一時的に上がる
- よくつまずく、もしくは転びやすい
- 見えにくい、もしくは見えない
- 聞こえにくい、もしくは聞こえない
- 動作がおそい、ぎこちない
- 腰痛、もしくは肩・膝・その他の関節の痛み
- 背骨あるいは関節の変形
- 皮膚の荒れがある
- トイレに近い、もしくは間に合わない
- 3日以上便が出ないか排便できない不快感あり、もしくは下剤を使っている
- 噛むのに躊躇する、もしくは食事中か食後にむせる

強制選択方式

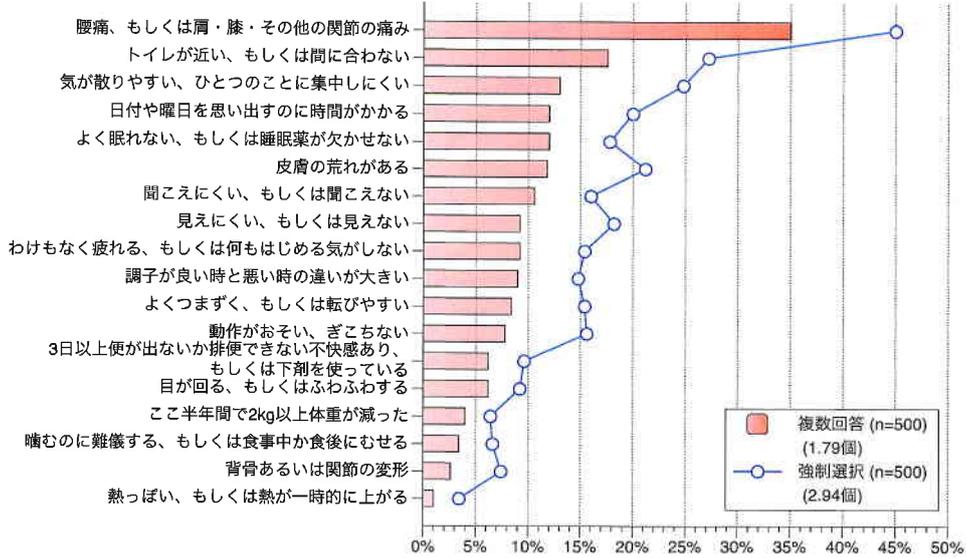
問1 あなたは、ここ1ヶ月の間に次のようなことがありましたか？(SA)

- | | あった | なかった |
|--------------------------------------|-----------------------|-----------------------|
| A 気が散りやすい、ひとつのことに集中しにくい | <input type="radio"/> | <input type="radio"/> |
| B よく寝れない、もしくは睡眠薬が欠かせない | <input type="radio"/> | <input type="radio"/> |
| C 日付や曜日を思い出すのに時間がかかる | <input type="radio"/> | <input type="radio"/> |
| D 調子が良い時と悪い時の違いが大きい | <input type="radio"/> | <input type="radio"/> |
| E ここ半年間で2kg以上体重が増った | <input type="radio"/> | <input type="radio"/> |
| F わけもなく疲れる、もしくは何もはじめる気がしない | <input type="radio"/> | <input type="radio"/> |
| G 目が回る、もしくはふわふわする | <input type="radio"/> | <input type="radio"/> |
| H 熱っぽい、もしくは熱が一時的に上がる | <input type="radio"/> | <input type="radio"/> |
| I よくつまずく、もしくは転びやすい | <input type="radio"/> | <input type="radio"/> |
| J 見えにくい、もしくは見えない | <input type="radio"/> | <input type="radio"/> |
| K 聞こえにくい、もしくは聞こえない | <input type="radio"/> | <input type="radio"/> |
| L 動作がおそい、ぎこちない | <input type="radio"/> | <input type="radio"/> |
| M 腰痛、もしくは肩・膝・その他の関節の痛み | <input type="radio"/> | <input type="radio"/> |
| N 背骨あるいは関節の変形 | <input type="radio"/> | <input type="radio"/> |
| O 皮膚の荒れがある | <input type="radio"/> | <input type="radio"/> |
| P トイレに近い、もしくは間に合わない | <input type="radio"/> | <input type="radio"/> |
| Q 3日以上便が出ないか排便できない不快感あり、もしくは下剤を使っている | <input type="radio"/> | <input type="radio"/> |
| R 噛むのに躊躇する、もしくは食事中か食後にむせる | <input type="radio"/> | <input type="radio"/> |

2

あなたは、ここ1ヶ月の間に次のようなことがありましたか？

(65歳以上を対象としたインターネット調査)



定期的な使用薬

定期的に使っている薬は？

複数選択

- 飲み薬
- ぬり薬
- 目薬
- 座薬
- 貼り薬
- 吸入薬
- 注射薬
- 点鼻・点耳薬

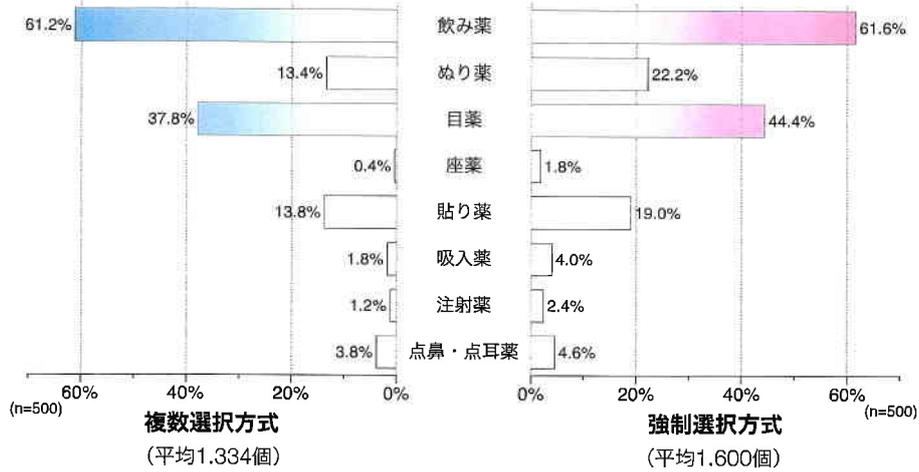
強制選択

- 使用している 使用していない
- 飲み薬
 - ぬり薬
 - 目薬
 - 座薬
 - 貼り薬
 - 吸入薬
 - 注射薬
 - 点鼻・点耳薬

定期的な使用薬

定期的に使用している薬は？

(65歳以上を対象としたインターネット調査)

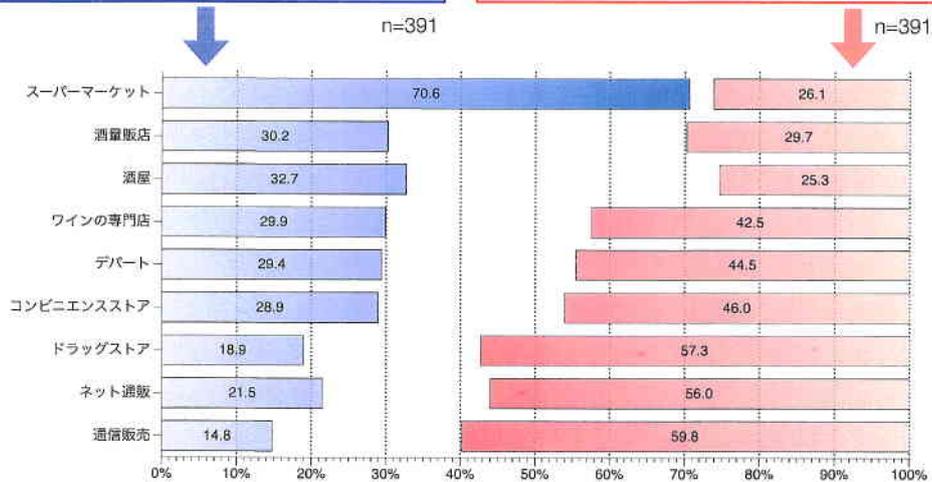


5

ワインの購入先

あなたがワインを購入したことがあるのは、次のうちどこですか。

あなたがワインを購入したことがないのは、次のうちどこですか。



6

複数回答方式で過少回答傾向が生じる原因

- ・ 少数の主な「該当」項目のみを選択するため
- ・ 非該当のときに「非該当」選択肢を明示的に選ぶ必要がないため

7

次のうち、あなたがそう思うものをいくつでも選んでください。

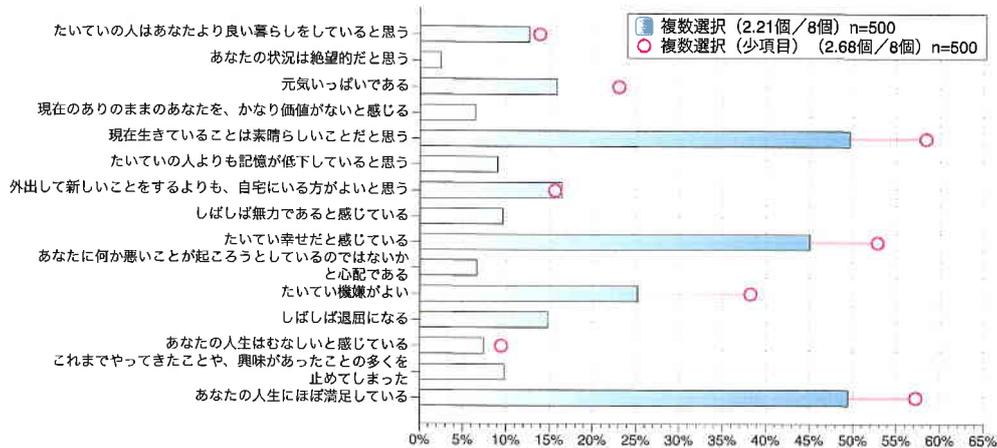
- | | |
|---|---|
| <input type="checkbox"/> たいていの人はあなたより良い暮らしをしていると思う | <input type="checkbox"/> たいていの人はあなたより良い暮らしをしていると思う |
| <input type="checkbox"/> あなたの状況は絶望的だと思う | <input type="checkbox"/> 元気いっぱいである |
| <input type="checkbox"/> 元気いっぱいである | <input type="checkbox"/> 現在生きていることは素晴らしいことだと思う |
| <input type="checkbox"/> 現在のありのままのあなたを、かなり価値がないと感じる | <input type="checkbox"/> 外出して新しいことをするよりも、自宅にいる方がよいと思う |
| <input type="checkbox"/> 現在生きていることは素晴らしいことだと思う | <input type="checkbox"/> たいてい幸せだと感じている |
| <input type="checkbox"/> たいていの人も記憶が低下していると思う | <input type="checkbox"/> たいてい機嫌がよい |
| <input type="checkbox"/> 外出して新しいことをするよりも、自宅にいる方がよいと思う | <input type="checkbox"/> あなたの人生はむなしと感じている |
| <input type="checkbox"/> しばしば無力であると感じている | <input type="checkbox"/> あなたの人生にほぼ満足している |
| <input type="checkbox"/> たいてい幸せだと感じている | |
| <input type="checkbox"/> あなたに何か悪いことが起ころうとしているのではないかと心配である | |
| <input type="checkbox"/> たいてい機嫌がよい | |
| <input type="checkbox"/> しばしば退屈になる | |
| <input type="checkbox"/> あなたの人生はむなしと感じている | |
| <input type="checkbox"/> これまでやってきたことや、興味があったことの多くを止めてしまった | |
| <input type="checkbox"/> あなたの人生にほぼ満足している | |

8

うつ尺度

つぎのうち、あなたがそう思うものをいくつか選んでください。

(65歳以上を対象としたインターネット調査)



9

うつ尺度

つぎのうち、あなたがそう思うものをいくつか選んでください。

(65歳以上を対象としたインターネット調査)

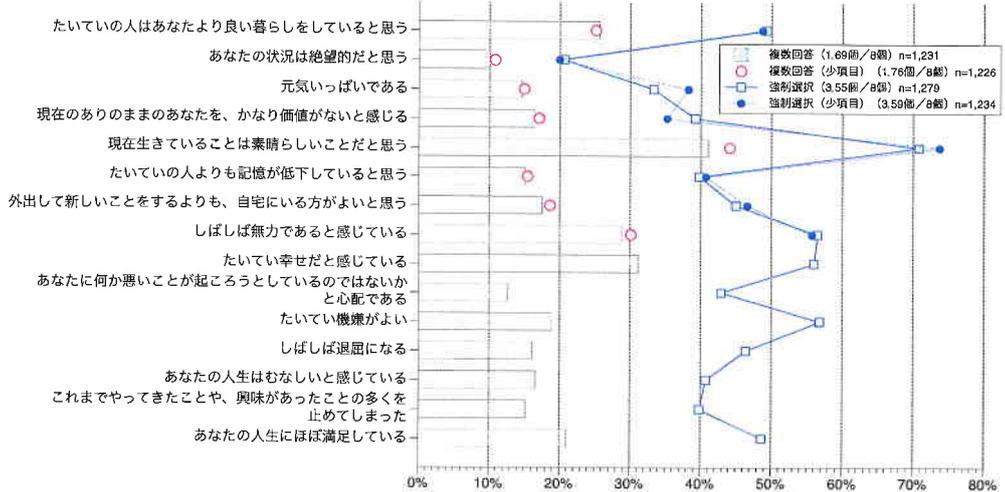
- ・ 対象者の年齢層を広げたらどうなるか？
- ・ 初頭効果ではないか？

10

うつ尺度

つぎのうち、あなたがそう思うものをいくつか選んでください。

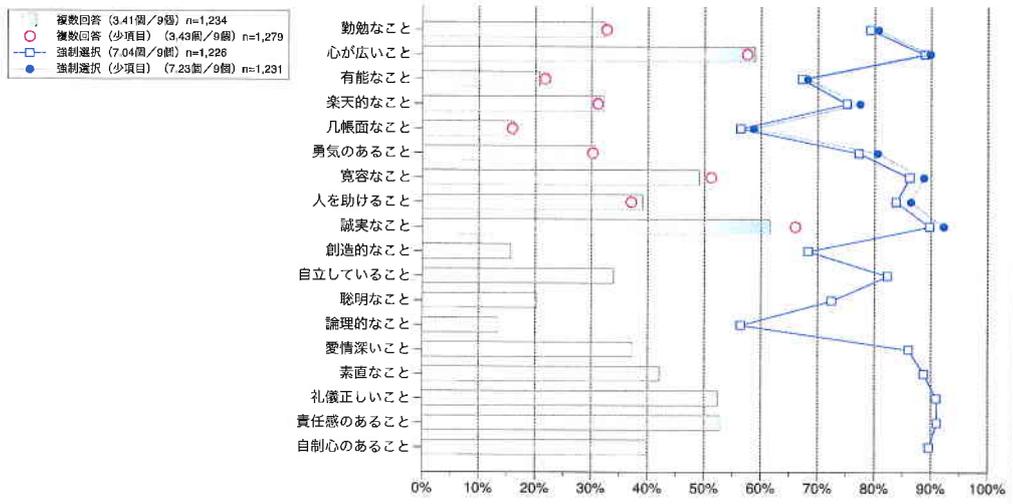
(20歳以上を対象としたインターネット調査)



Rokeachの価値

次のうち、あなたが人として重要だと思う性質をいくつか選んでください。

(20歳以上を対象としたインターネット調査)



プレチェック方式

健康状態に関する調査票

問1 あなたは、この1か月間に以下の症状がありましたか。
(あったものは「あった」にチェックを付け直してください)

熱がある	<input type="radio"/> なかった <input checked="" type="radio"/> あった	腹痛・胃痛	<input type="radio"/> なかった <input checked="" type="radio"/> あった
体がだるい	<input type="radio"/> なかった <input checked="" type="radio"/> あった	痔による痛み・出血など	<input type="radio"/> なかった <input checked="" type="radio"/> あった
眠れない	<input type="radio"/> なかった <input checked="" type="radio"/> あった	歯が痛い	<input type="radio"/> なかった <input checked="" type="radio"/> あった
いらいらしやすい	<input type="radio"/> なかった <input checked="" type="radio"/> あった	歯ぐきのはれ・出血	<input type="radio"/> なかった <input checked="" type="radio"/> あった
もの忘れする	<input type="radio"/> なかった <input checked="" type="radio"/> あった	かみにくい	<input type="radio"/> なかった <input checked="" type="radio"/> あった
頭痛	<input type="radio"/> なかった <input checked="" type="radio"/> あった	発疹(じんま疹・できものなど)	<input type="radio"/> なかった <input checked="" type="radio"/> あった
めまい	<input type="radio"/> なかった <input checked="" type="radio"/> あった	かゆみ(湿疹・水虫など)	<input type="radio"/> なかった <input checked="" type="radio"/> あった
目のかすみ	<input type="radio"/> なかった <input checked="" type="radio"/> あった	肩こり	<input type="radio"/> なかった <input checked="" type="radio"/> あった
物を見づらい	<input type="radio"/> なかった <input checked="" type="radio"/> あった	腰痛	<input type="radio"/> なかった <input checked="" type="radio"/> あった
耳なりがする	<input type="radio"/> なかった <input checked="" type="radio"/> あった	手足の関節が痛む	<input type="radio"/> なかった <input checked="" type="radio"/> あった

15

プレチェック方式

健康状態に関する調査票

問1 あなたは、この1か月間に以下の症状がありましたか。
(あったものは「あった」にチェックを付け直してください)

熱がある	<input type="radio"/> なかった <input checked="" type="radio"/> あった	腹痛・胃痛	<input type="radio"/> なかった <input checked="" type="radio"/> あった
体がだるい	<input type="radio"/> なかった <input checked="" type="radio"/> あった	痔による痛み・出血など	<input type="radio"/> なかった <input checked="" type="radio"/> あった
もの忘れする	<input type="radio"/> なかった <input checked="" type="radio"/> あった	かみにくい	<input type="radio"/> なかった <input checked="" type="radio"/> あった
頭痛	<input type="radio"/> なかった <input checked="" type="radio"/> あった	発疹(じんま疹・できものなど)	<input type="radio"/> なかった <input checked="" type="radio"/> あった
めまい	<input type="radio"/> なかった <input checked="" type="radio"/> あった	かゆみ(湿疹・水虫など)	<input type="radio"/> なかった <input checked="" type="radio"/> あった
目のかすみ	<input type="radio"/> なかった <input checked="" type="radio"/> あった	肩こり	<input type="radio"/> なかった <input checked="" type="radio"/> あった
物を見づらい	<input type="radio"/> なかった <input checked="" type="radio"/> あった	腰痛	<input type="radio"/> なかった <input checked="" type="radio"/> あった
耳なりがする	<input type="radio"/> なかった <input checked="" type="radio"/> あった	手足の関節が痛む	<input type="radio"/> なかった <input checked="" type="radio"/> あった

回答者負担は複数回答方式と同じ(はず)

回答後調査票は強制選択方式と同じ見た目

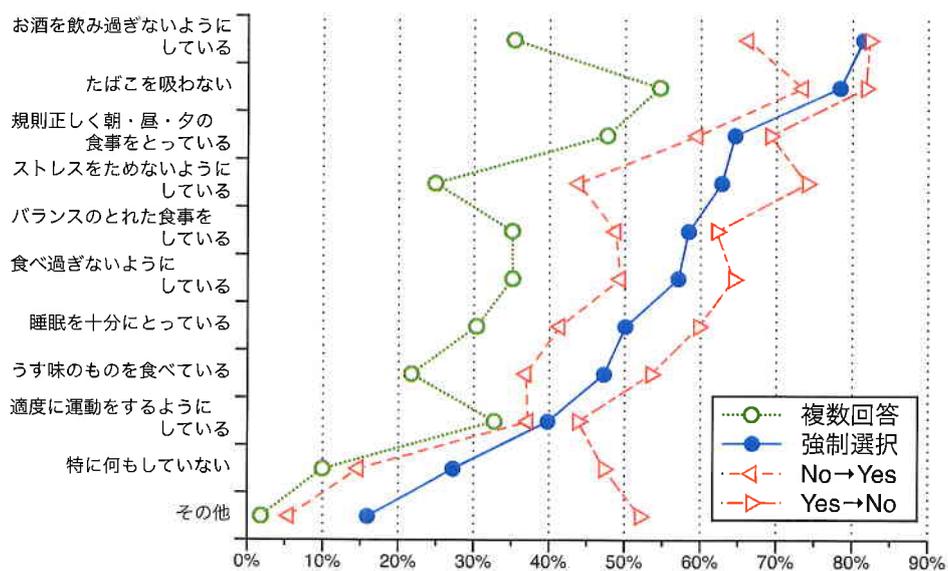
16

インターネットによる比較実験調査の概要

- ・ **調査対象**：20～69歳のネット調査会社登録モニター
- ・ **調査票と回収数**：4種類のPDF × 800人
 - 複数回答方式
 - 強制選択方式
 - No → Yes
 - Yes → No
- ・ **調査内容**：健康状態に関する内容8項目
- ・ **調査時期**：2017年11月～12月

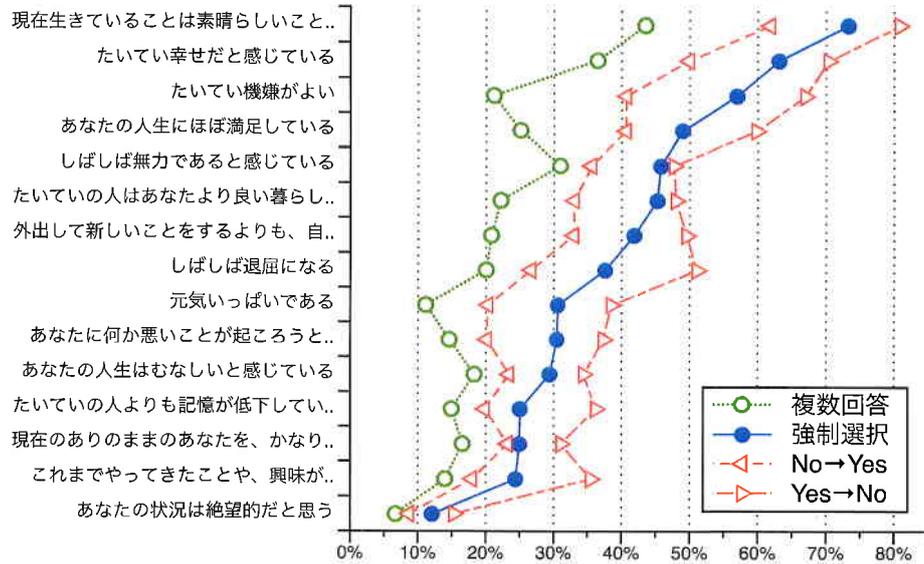
17

健康のためにしていること



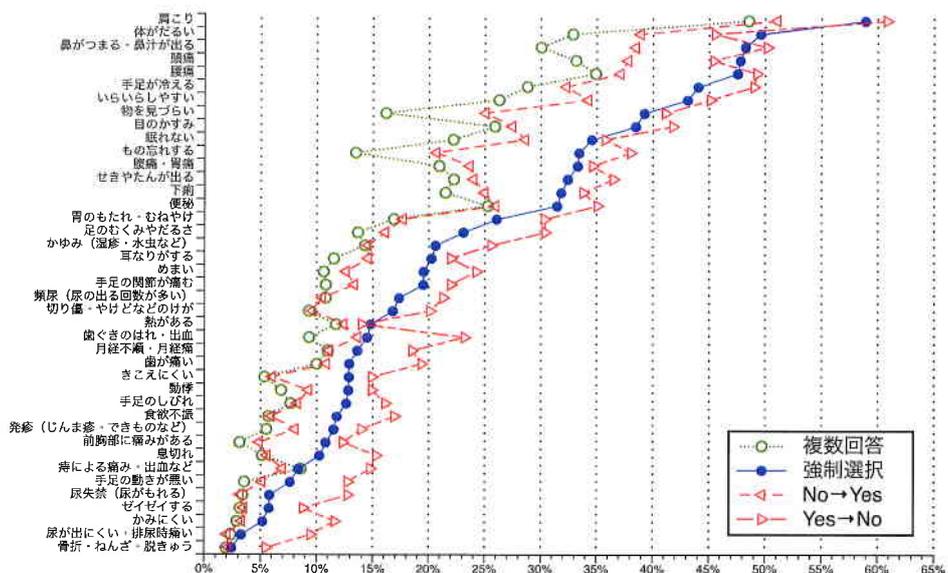
18

うつ尺度



19

この1か月間の症状



20

高齢無業世帯の 貯蓄と消費行動について

日本大学経済学部

川崎 茂

kawasaki.shigeru@nihon-u.ac.jp

1

問題意識

- 長期の景気拡大局面の中、家計消費はなぜ伸び悩んでいるのか？
- 賃金上昇が消費につながらない
 - ➔ 社会保障負担の増加？高齢化による消費の停滞？
- 家計消費に占める高齢者世帯のウェイトは増加
 - ➔ 高齢者の消費動向を分析することが重要
- 有業高齢者の消費行動は、壮年有業者と類似
 - ➔ 消費関数などの研究は多数存在
- 無業高齢者の消費行動は、壮年有業者と相違
 - ➔ 消費関数を推計したものは見当たらない。解明が不十分では？
- 多くの先行研究で、ライフサイクル仮説の観点で高齢化と消費を分析
 - ➔ 高齢者の保有する資産は大きく、その影響も分析する必要がある。

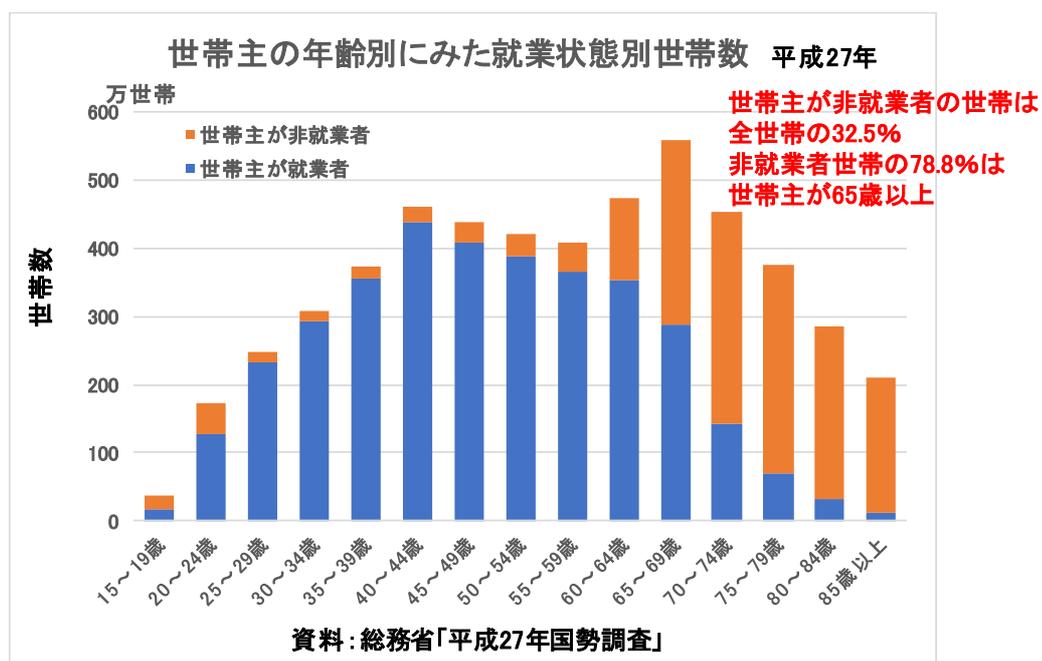
2

分析の概要

- 資産は、所得・消費の結果として生じる一方で、資産が所得・消費（特に消費）に影響を及ぼしうる。
 - ➔ 勤労所得の影響を除くため、リタイアした無業の高齢者世帯に注目
資産の保有が消費にどのような影響を与えるか分析する。
- 主として「平成26年全国消費実態調査」のデータを使用。
 - ➔ 5年に1回、全国の約5万世帯を対象として調査
家計収支については9～11月、資産については11月末残高を調査
(本年の調査から、「全国家計構造統計」に改称。)
- 65歳以上の夫婦のみから成る世帯について、消費関数を推計
 - ➔ 基礎的なファクトファインディングの段階
高齢化の進展に伴う家計消費の変化を予測するための基礎となることを期待

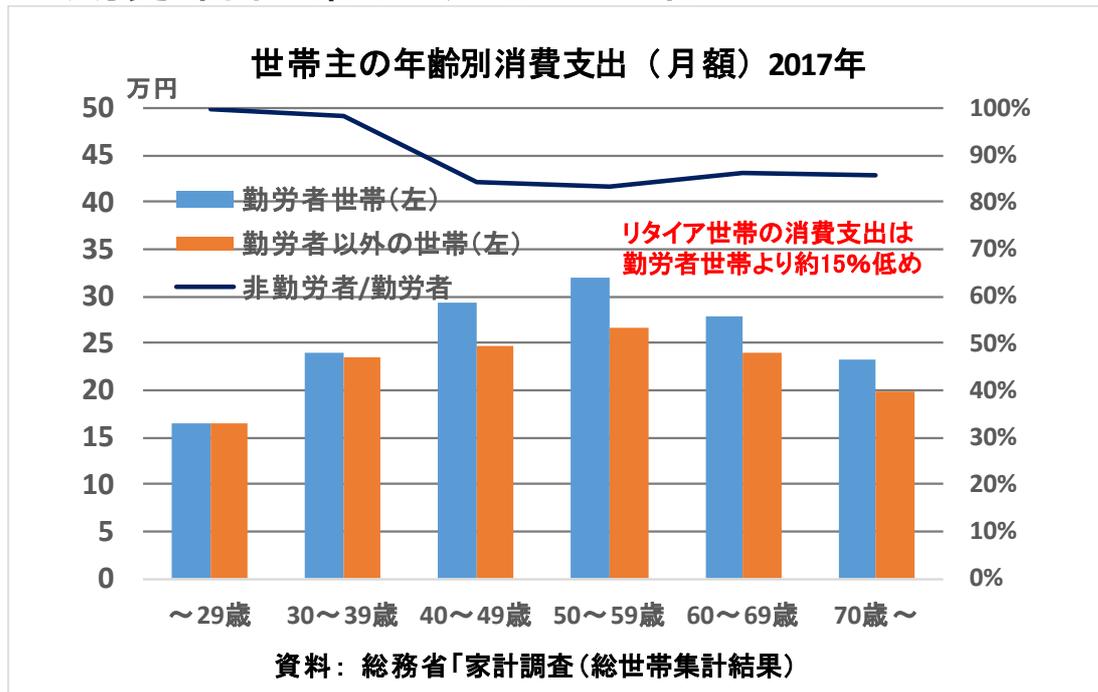
3

無業の高齢者世帯の増加



4

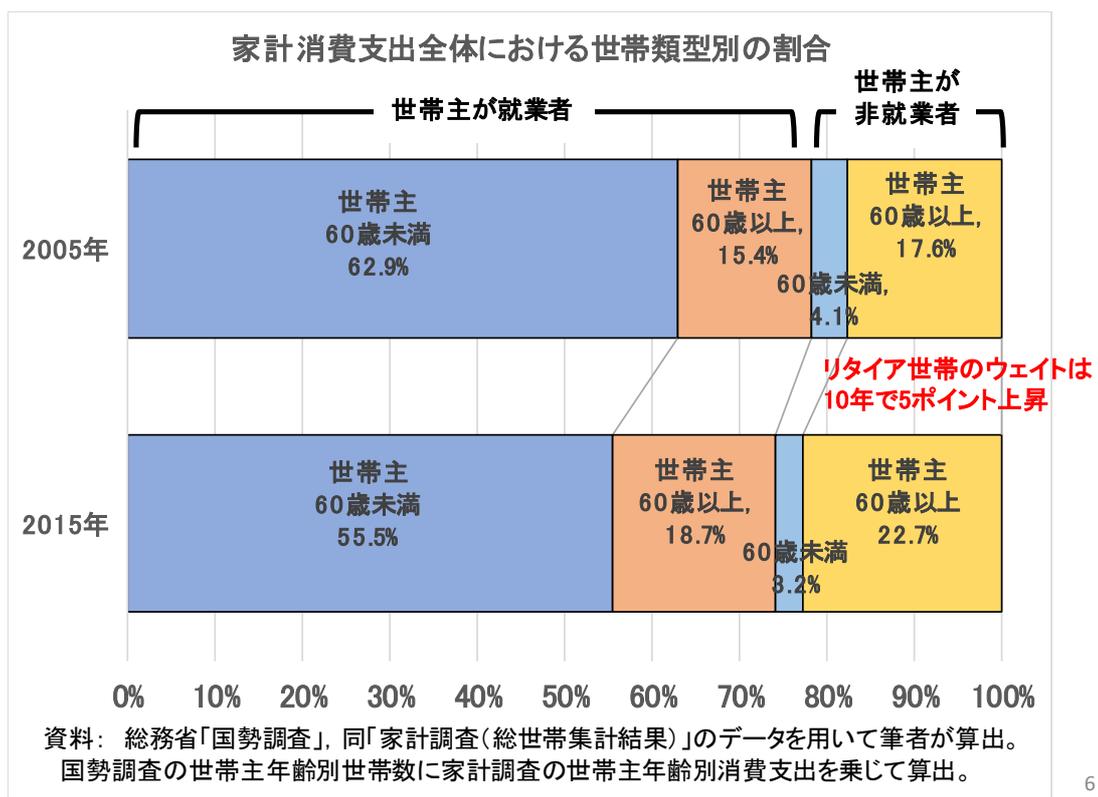
退職者の世帯の消費支出は 勤労者世帯よりも15%低め



注：「勤労者以外の世帯」には、世帯主が自営業主の世帯等も含まれる。

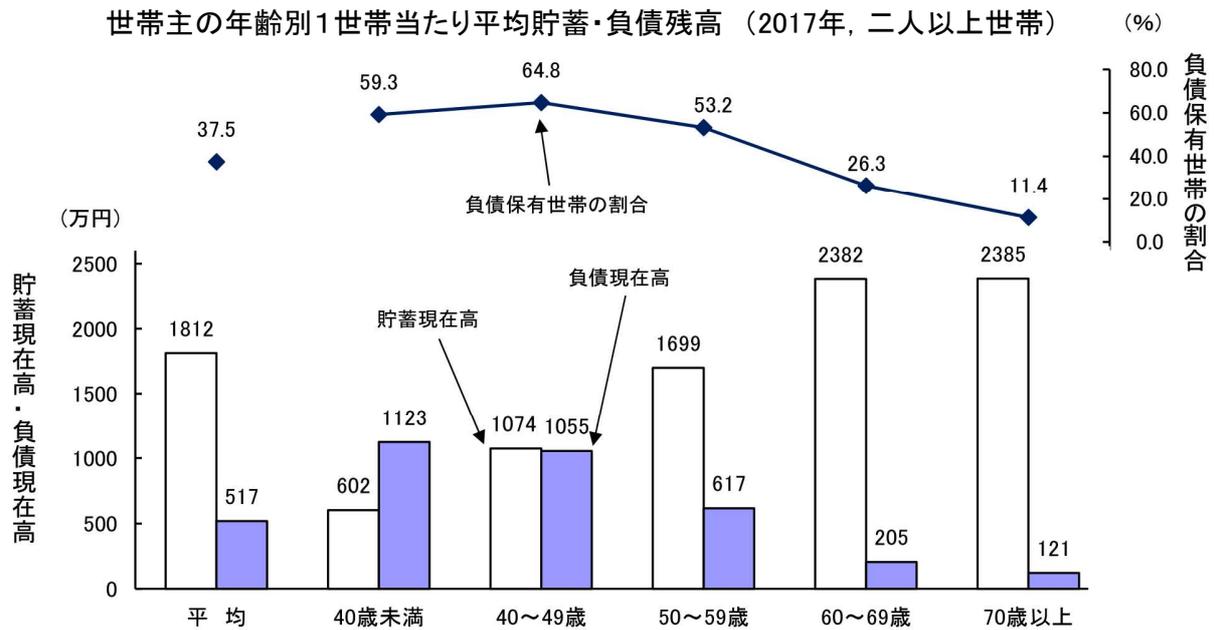
5

退職者の世帯の消費のウェイトが上昇



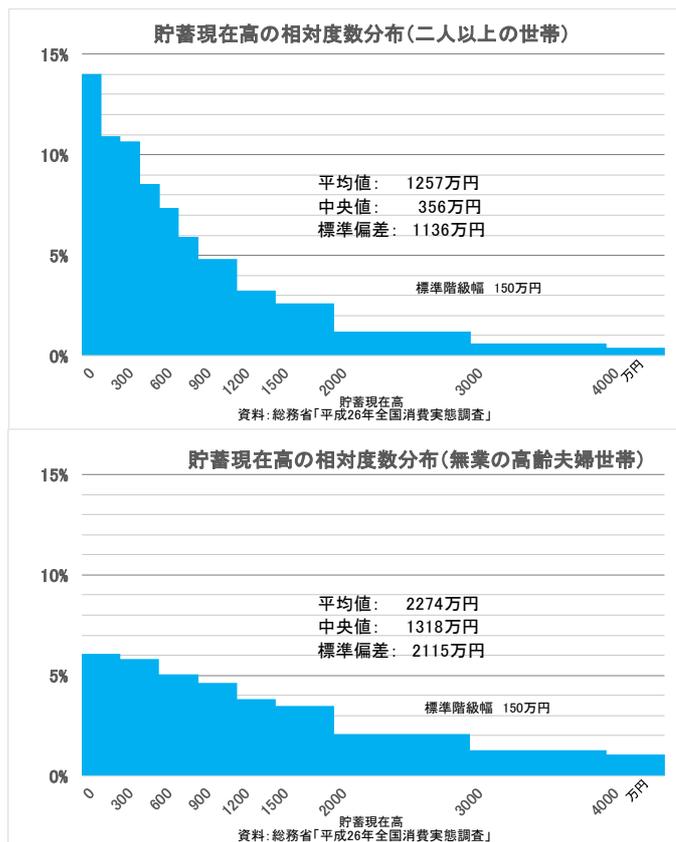
6

高齢者世帯の平均貯蓄高は2000万円超

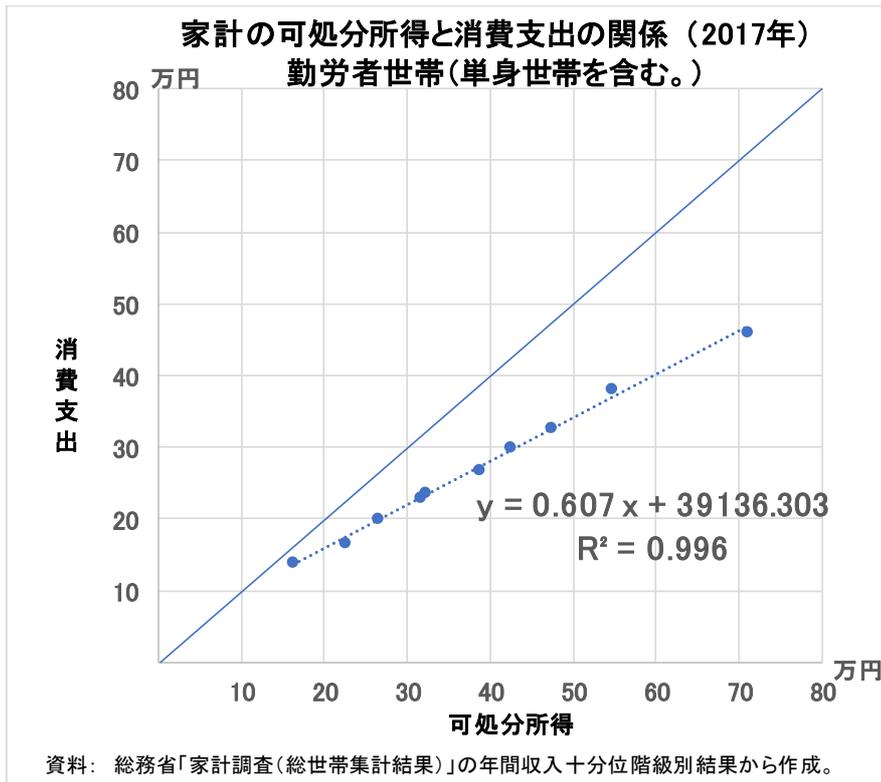


資料：総務省「家計調査報告」(貯蓄・負債編) 平成29年平均結果(二人以上世帯)

退職者世帯の貯蓄保有額は分散が大きい

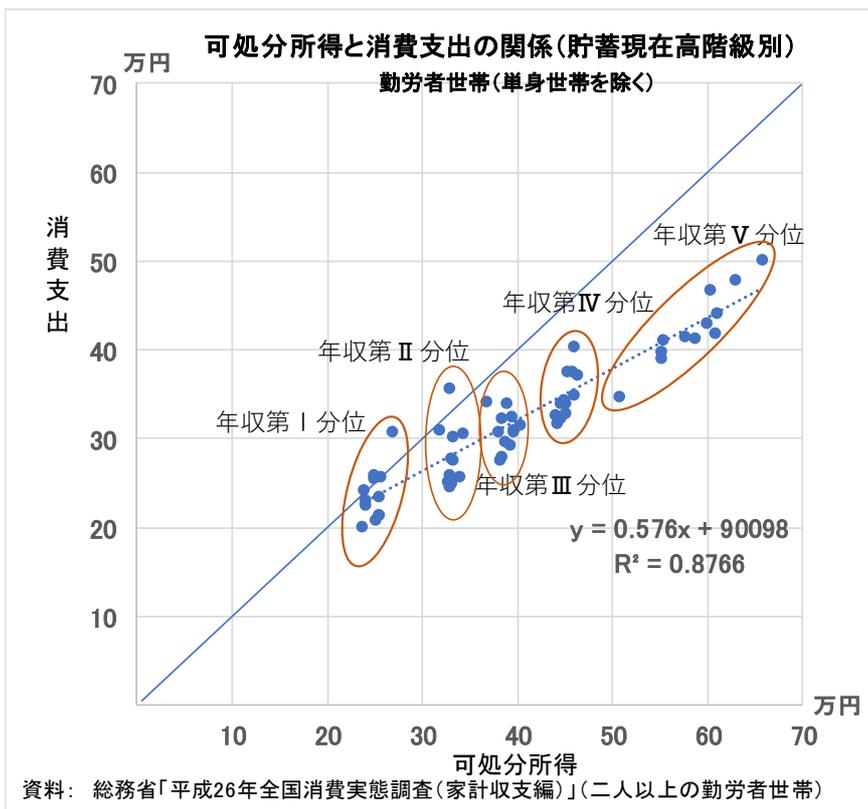


消費関数 - 消費支出と可処分所得の関係



9

消費関数に貯蓄残高を組み入れる



使用したデータは、
年間収入5分位階級別、貯蓄現在高階級別の消費支出および可処分所得の集計値。

10

貯蓄残高を組み入れた消費関数の推計

使用したデータ

平成26年全国消費実態調査 フロー編第5表 年間収入五分位階級, 貯蓄現在高階級別1世帯
 当たり1か月間の収入と支出

年収(5) × 貯蓄(12) = 60区分の消費支出, 可処分所得, 貯蓄現在高の集計値
 N=60

消費支出(Y): 平成26年9月～11月の1世帯1か月当たりの金額(円単位)

可処分所得(X1): 同上

貯蓄現在高(X2): 平成26年末現在の貯蓄高(千円単位)

回帰式と主な統計量

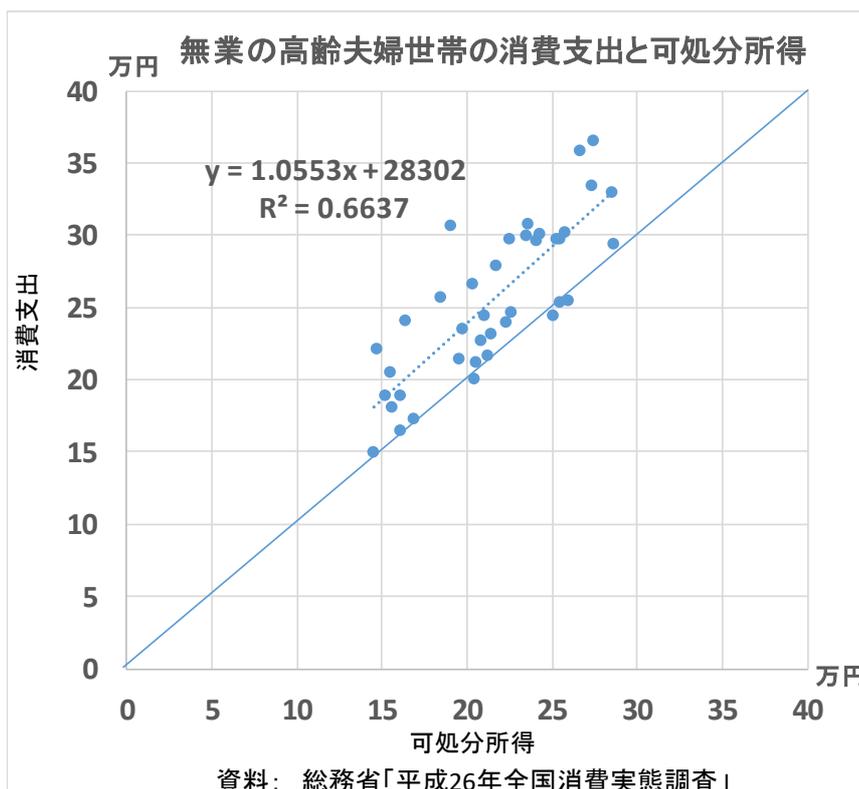
$$Y = 0.5594 X1 + 1.3065 X2 + 76330$$

(39.960) (13.533) (12.935) ()内は t 値

R2 = 0.9707 自由度補正済R2 = 0.9697 標準誤差 = 12466.6

	平均	標準偏差	相関係数	
消費支出	320,776	70,997		
可処分所得	400,507	115,409	消費支出	可処分所得
貯蓄現在高	15,614	16,734	可処分所得	0.9363
			貯蓄現在高	0.0874

無業の高齢夫婦世帯の消費関数



使用したデータは、
 年間収入階級別、
 貯蓄現在高階級別
 の消費支出および
 可処分所得の集計
 値。

無業の高齢世帯の消費関数の推計

使用したデータ

平成26年全国消費実態調査 第104-2表 65歳以上の夫婦のみの世帯(有業者なし)年間収入階級, 貯蓄現在高階級別1世帯当たり1か月間の収入と支出

年収(5) × 貯蓄(10) = 50区分の消費支出, 可処分所得, 貯蓄現在高の集計値
(世帯数の少ないセルは統合。N=44)

消費支出(Y): 平成26年9月～11月の1世帯1か月当たりの金額(円単位)

可処分所得(X1): 同上

貯蓄現在高(X2): 平成26年末現在の貯蓄高(千円単位)

回帰式と主な統計量

$$Y = 0.9551 X1 + 0.6638 X2 + 35485$$

(8.105) (2.906) (1.4323)

R2 = 0.7276 自由度補正済R2 = 0.7124 標準誤差 = 37916.8

	平均	標準偏差
消費支出	255,074	53,131
可処分所得	214,889	41,547
貯蓄現在高	20,992	20,547

相関係数		
	消費支出	可処分所得
可処分所得	0.8147	-
貯蓄現在高	0.4801	0.2926

13

まとめ

- 無業の高齢世帯の消費に対する可処分所得と貯蓄高の影響が存在することが確認された。
- ➔ 無業世帯が増加する中, その消費の決定要因を解明することは重要
- この分析で未解明の課題について, 今後取り組みたい。
 - 住宅など実物資産の影響, 負債の影響
 - 年金, 税, 社会保険料等の影響
 - リタイア後の貯蓄残高の貯蓄取り崩しのペースに影響を与える要因
 - データの精度 (高齢化による資産管理能力の低下)
- ➔ ミクロデータ, パネルデータによる分析が必要。

14

人口減少・超高齢社会と人的資本

— 政府統計における新しい人口資質の視点 —

金子隆一¹

わが国が世界に先駆けて直面している人口減少・少子高齢化は、高齢層の増大・現役世代縮小という人口バランスの崩れによって経済社会システムの機能を阻害し、社会の継承や存続そのものを脅かしている。一方でわが国は世界一の健康長寿と高い教育水準を有しており、とりわけ高齢層は量的増大とともに健康ならびにスキルや教養という質の面でも大きく変化しつつある。こうした人的資本の質の面での充実、すなわち人口資質の向上は、国や地域の生産性向上に寄与するだけでなく、人々がより豊かな人生をおくる前提ともなるものである。ことから人口資質の向上は、人口減少・少子高齢化によるマクロ面（経済システム）での不具合を是正するとともに、ミクロ面（人々の人生や生活）での発展を促す働きを持っている。本論では、こうした人口資質を定量的に捉え、今後必要となる社会システム再構築の際の指針形成に資するため、政府統計のあり方や指標、さらには統計的方法論についての検討を試みたい。

1. はじめに

わが国は現在、明治期以降の人口成長の時代から恒常的な人口減少の時代へと歴史的な転換を経験しつつある。この人口減少は当然ながら消費市場や労働力の規模の縮小をもたらすが、同時に人口高齢化という構造変化によって、社会を支える層と社会から支えられる層の人口バランスを大きく変えることになる。こうした人類未経験の変化は、現行の経済社会システムが機能する範囲を大きく逸脱するものであり、社会の継承や存続を脅かすことになるであろう。

そのような人口変動を制止する方法として、少子化の解消と移民が取り上げられることがある。しかし、現在の日本人口が持つ人口モメンタムを前提とすると、日本人口は少子化を完全に解消しても今後5~60年は減少を止めることはなく、結局2割程度の人口を失うことがわかっている²。また、少子化によって不足する労働力をすべて移民でまかなうとすれば、数十年のうちに日本人口の3~4割が外国人としなくてはならない。こうした変化は、これまで移民受け入れの経験のない日本社会にとって、新たな課題を生み出すことになる。そして受け入れた移民は、たとえば老齢となった後には年金、医療・介護サービスの需要

¹ 明治大学 政治経済学部、rkaneko@meiji.ac.jp

² 人口モメンタムとは、人口の年齢構造に内在する人口増加または減少を促進する特性。日本人口は1997年以降、負の人口モメンタムを擁する年齢構造となっている。また、ここで少子化を完全に解消するとは、出生率が人口置換水準になることを意味している。少子化を解消しても当面人口減少を止めることができないことについては、金子（2018）などを参照のこと。

者となるなど、社会負担も増大することを忘れてはならない。また、そもそも今後世界的に人口高齢化が進み、各国とも若年労働力は不足するなか、外国人労働力の確保は難しくなるだろう。結局、人口減少そのものを速やかに解消する方策はないと考えるのが自然であろう。そうであれば、われわれはこの日本社会の存立のためにいったい何ができるのであろうか。

おそらく唯一の方途は、ある程度の人口規模の縮小を受け入れながらも、人口の「質」を活用することによって、社会全体の生産性を維持することであろう。たとえば、わが国は世界一の健康長寿と非常に高い教育水準を有している³。とりわけ高齢層は量的増大とともに、健康ならびにスキルや教養という質の面でも大きく変化しつつある。こうした人的資本の「質」の面での充実、言い換えれば人口資質の向上は、国や地域の生産性向上に寄与している。しかし、現行制度下においては、世代ごとに改善を続ける健康度や教育水準は、生産性の向上にフルに転換されているわけではない。たとえば、従来からの定年制度によって年齢の高い労働力は一律の年齢での引退を余儀なくされている。これは高齢において一定程度維持されている生産力を制度によって放棄する仕組みに他ならない。重要なことは日本社会においては、これらの引退層においても高い生産意欲が維持されており、適切な形で生産に結び付けることができれば、経済的な自助となる同時に彼ら自身の自己実現にもつながる点である。すなわち、人口資質の向上とこれを適切に生産性に結び付ける仕組みは、人口減少・少子高齢化によるマクロ面、すなわち経済社会システムにおける不具合を是正するとともに、ミクロ面、すなわち人々の人生や生活の発展を促す働きを有しており、今後の日本社会の維持に必須のものであると考えられる。

本論では、こうした人口資質を定量的に捉え、社会システム再構築の際の指針形成に資するための政府統計のあり方や指標、さらには統計的方法論について検討してみたい。

2. 日本の人口動向と人口資質

(1) わが国の近代化と人口資質の変化

幕末から明治維新幕末にかけて動き出したわが国の近代化は、欧米先進国からの新しい産業技術の移転と資本主義経済の導入によって急速な進展をみせ、人々の暮らしや生涯も大きく変貌していった。とりわけ彼らの生存状況の改善は、まず、それまで乳幼児期に大量に失われていた生命を救い、青年層の人口を増大させた。この生活の質の改善にともなう人口構造変化は、豊富で若い労働力の供給を可能とし、わが国のその後の産業の発展、経済の拡大に大いに寄与した。これは一般に多産多死から多産少死を経て少産少死へと向かう「人口転換」と呼ばれる歴史過程のもたらした帰結の一つである。

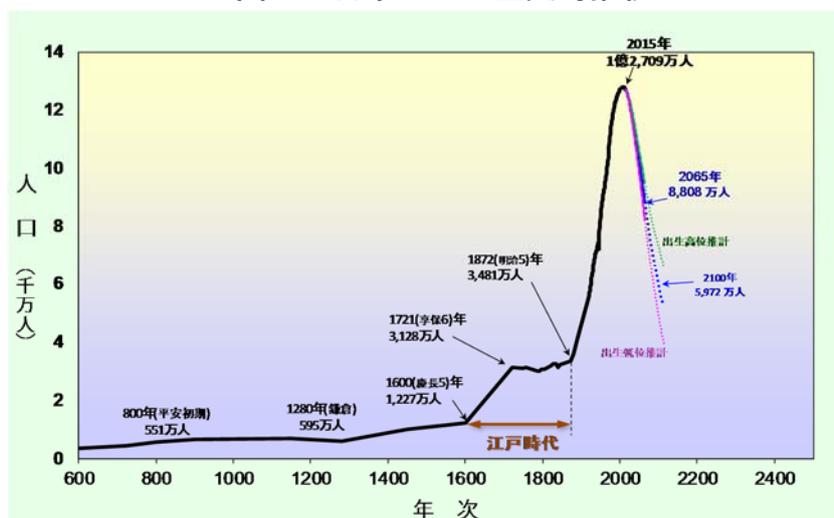
³ WHO(世界保健機関)2018年によると、2016年時点男女総数の平均寿命について日本は世界第1位(84.2年)となっている。2位スイスとの差は0.9年であった。国内統計である簡易生命表2017年による平均寿命は、男性81.09年、女性87.26年である。

その後のわが国の経済成長と社会発展は、労働力の量的増大だけではなく、生存状況の改善の前提となる健康状態の向上や、教育程度の上昇という質的变化によっても支えられたと考えられる。とりわけ戦後になり、第一次産業から第二次、そして第三次へと漸次産業構造の転換が進むと、労働力はその供給量だけでなく、労働者個々の技能や知識によって全体の生産性が大きく左右されるようになった。高度情報化が進む現在において、その傾向はなお一層強まっている。

(2) 日本社会の存続を脅かす過度の人口減少と少子高齢化

前述のとおり、わが国の人口推移は、明治期以降の急速な増加のフェーズから、現在は一転して減少のフェーズに移行している（図1）。政府による現時点における将来見通しを示した「日本の将来推計人口」によれば、2015年国勢調査時点の総人口（外国人を含めた日本領内における常住人口）12,709万人は、50年後の2065年に8,808万人（2015年当初人口の69.3%）に減少し、その後もこの趨勢が続けば、2100年59,718万人（同47.0%）を経て、100年後2115年には50,555万人（39.8%）まで縮小する。すなわち、日本社会は今後50年間で約3割、100年間で約6割の人口を失うことになる（表1）。

図1 日本人口の歴史的推移



資料：国立社会保障・人口問題研究所「人口統計資料集」（1846年までは鬼頭宏「人口から読む日本の歴史」、1847～1870年 森田優三「人口増加の分析」、1872～1920年 内閣統計局「明治五年以降我国の人口」、1920～2015年 総務省統計局「国勢調査」「推計人口」、2016～2115年国立社会保障・人口問題研究所「日本の将来推計人口(平成29年推計)」[死亡中位仮定]。

表1 日本の人口推移（年齢3・4区分）：2015年人口=100とする指数

年次	人口						
	総数	0~14歳	15~64歳	65歳以上			
					65~74歳	75歳以上	
1900 (明治 33)	34.5	91.0	34.9	7.1	13.6	0.0	
1965 (昭和 40)	77.3	157.8	86.6	18.2	24.5	11.5	
2000 (平成 12)	99.9	116.1	111.8	65.1	74.3	55.2	
2010 (22)	100.8	105.6	105.8	87.1	87.1	87.0	
2015 (27)	100.0	100.0	100.0	100.0	100.0	100.0	
2020 (新元号 2)	98.6	94.5	95.8	106.9	99.6	114.7	
2030 (12)	93.7	82.9	89.0	109.7	81.4	140.2	
2050 (32)	80.2	67.5	68.3	113.4	81.1	148.1	
2065 (47)	69.3	56.3	58.6	99.8	64.6	137.7	
2100 (82)	47.0	38.3	39.8	67.5	44.7	92.1	
2115 (97)	39.8	32.6	33.5	57.4	38.3	77.9	

資料：2015年までは国勢調査、2020年以降は「日本の将来推計人口」平成29年推計(出生中位・死亡中位)による。

これまで労働力の主な供給源となってきた生産年齢(15~64歳)の人口に限ってみれば、2015年現在の人口はこの年齢層の人口の最大を記録した1995年8,726万人から、7,728万人へと、すでに11.4%減少しているが、あらためて2015年を基準とすると、50年後の2065年には6割程度(58.6%)、に、100年後の2115年には1/3(33.5%)へと縮小することが見込まれている。

以上のような恒常的な人口減少と未曾有の少子高齢化は、高齢層増大・現役世代縮小という人口バランスの崩れによって、現行の経済社会システムに不全をもたらし、その持続可能性を脅かしている。

3. 将来の日本社会と人口資質

(1) 健康・寿命

前述のとおり明治期以降のわが国における平均寿命の伸長、すなわち長寿化は目覚ましいものであるが、それは高齢まで生存する人口を増加させる働きが有ることから、現在世界一となる人口高齢化をもたらしたとの考えが一般的である。たしかに寿命の伸長は高齢化に寄与するものであるが、人口高齢化の最大の要因は少子化であり、その効果は長寿化のそれをはるかに凌駕する。したがって、国際的に見てほぼ同程度に長寿化が進行する先進諸国のなかでも、出生率が人口置換水準に近い西北欧の国々や米国と、その水準に遠く及ばない独や南欧諸国、ならびに日本をはじめとする東アジアの国々とは⁴、人口高齢化の将来像には大きな違いが見込まれている⁵。

⁴ 出生率の二極化については、Rindfuss,Choe,Brauner-Otto (2016) を参照のこと。

⁵ たとえば、2050年の従属人口指数は、出生率の高いスウェーデン 69.7%、米国 64.8%であるに対して、出生率が低迷する日本では 95.8%、イタリアでも 91.2%に及ぶ(いずれも国連推計 2017 年による。日本政府の将来推計人口(平成 29 年推計)では、93.2%である)。

以上のように、長寿化社会と高齢化社会は本来違う状況を表すことに注意が必要であり、前者は人々が健康で長生きする社会、後者は相対的に高齢者が多くなることによって多くの課題が生ずる社会を示している。すなわち、人口高齢化は人口の量的、構造的変化を指すのに対して、長寿化は健康や生存状況の変化を意味している。人口高齢化に対しては、むしろ人々がより健康で長生きとなる長寿化を推進することによって、たとえば医療、介護等の負担を減らすことが考えられるなど、これらの概念は明確に区別することが重要である。

さて、わが国では速やかな長寿化によって人々は年々健康を増進し、寿命も伸長しているから、上述のようにその成果を社会的にも活用することが考えられるが、それは人口高齢化に対してどのような定量的効果を持つだろうか。とりわけ、高齢者を従来のように固定したイメージで捉えずに、少し見方変えると違った将来像が見えるはずである。

たとえば、「高齢」への境となる 65 歳時の健康度は昔に比べると大幅に改善していると考えられる。しかしながら、実際に健康度を測定することは簡単ではない。そもそも「健康」の概念を客観的に規定することは困難であるし、仮にそれを医学的・生物学的に定義できたとしても、その定量的測定はやはり難しいはずである。しかし、人口統計学においてはこれに代替する指標がある。それは生命表における平均余命である。

平均余命とは、対象期間の死亡の起こり方、すなわち年齢別死亡率を元にして、任意の年齢 (x 歳) の者が、平均してあと何年生きることになるかを算定した数値である。いわば各年齢の集団が死亡からどの程度離れた距離にいるのかを示しているが、この距離を当該年齢集団の平均的な健康度と考えれば、平均余命を健康度として見ることができる。

表 2 には、いくつかの年次における 65 歳時平均余命を男女別表として示した。これによれば平均余命は 1960 年男性 11.6 年、女性 14.1 年であり、その後、年次とともに長くなっていることがわかる。これを健康度と見なせば、たとえば 2010 年において 1960 年 65 歳集団と同じ健康度となる年齢を特定することができる。すなわち 2010 年生命表において平均余命が男性 11.6 年、女性 14.1 年となる年齢を見出せばよい。それらは男性 74.7 歳、女性 76.5 歳であった（これを 1960 年基準 65 歳時余命等価年齢と呼ぶことにする）。つまり、1960 年から 2010 年の間に高齢者の健康度は大幅に改善し、当初 65 歳とした高齢の開始年齢が男性 9.7 年、女性 11.5 年高くなったことを示している。

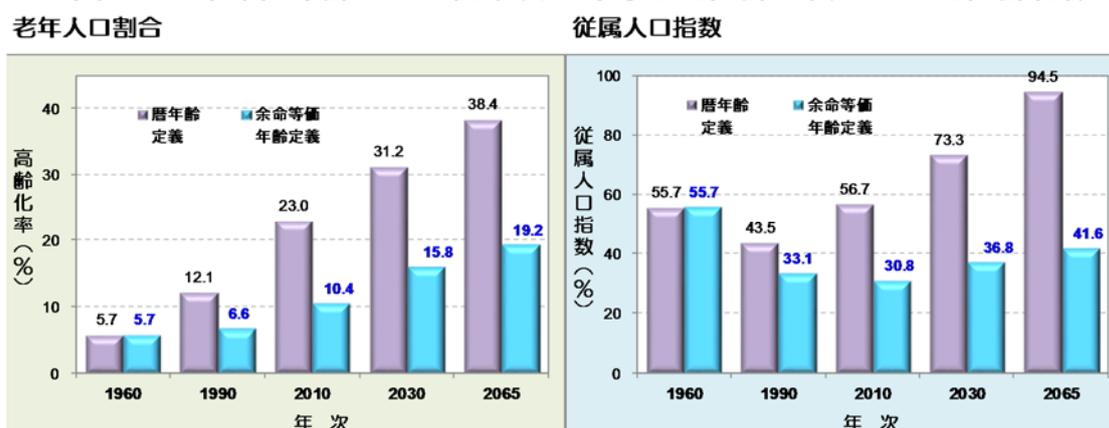
表 2. 男女 65 歳時の平均余命，余命等価年齢(1960 年基準)，老年猶予年数
わが国の高齢層は時代とともに若返っている

男性				女性			
年次	65歳時 平均余命	1960年基準 65歳等価年齢	65歳時 猶予年数	年次	65歳時 平均余命	1960年基準 65歳等価年齢	65歳時 猶予年数
1960	11.6	65.0	0.0	1960	14.1	65.0	0.0
1990	16.2	71.6	6.6	1990	20.0	72.2	7.2
2010	18.7	74.7	9.7	2010	23.7	76.5	11.5
2030	20.7	77.1	12.1	2030	25.7	78.5	13.5
2065	22.6	79.3	14.3	2065	27.9	80.8	15.8

資料：2010 年以前は社人研死亡データベース、2030 年、2065 年は「日本の将来推計人口(平成 29 年推計) 死亡中位仮定を用いて算出。

この同一健康度による高齢開始年齢(1960 年基準 65 歳時余命等価年齢)の上昇傾向は、将来的にも継続することが見込まれており、2065 年には男女ともに 80 歳前後(男性 79.3 歳、女性 80.8 歳)になるとみられる。この動的な高齢開始年齢に従えば、人口高齢化の将来像は大きく変わる。図 2 には、その場合の老年人口割合(高齢化率)と従属人口指数の変化を示した。高齢化率については、2065 年で従来の暦年定義による指標値は 38.4%に達するのに対して、新定義による指標値は 19.2%と半減している。すなわち「高齢者」の頻度は国民 5 人中 2 人と見込まれたものが、5 人中 1 人に納まることになる。また、扶養される人口と扶養する人口のバランスを示した従属人口指数では、従来の定義では 2065 年 94.5%と見込まれたものが、41.6%に大きく縮減しており、これは人口ボーナスの状況が将来も続くという将来像となっている。したがって、暦年齢ではなく、「健康度」によって「高齢」を定義した場合には、扶養される人口と、扶養する人口のバランスが崩れることはないことになる。

図 2. 余命等価年齢による健康度を考慮した高齢定義を用いた高齢化像



資料：表 2 に同じ。

ただし、こうした新たな高齢化の将来像は、高齢者における健康の増進をフルに生産性の向上に変換できたとしたときに実現すると考えられるものであり、そのためには現行の経済社会システムの刷新が必須である。現行のシステムは、定年制度などをはじめとして

暦年齢に強く依存している部分が大きく、健康という人口資質の変化を活用することはほとんどできていないと考えられる。

また、そうした制度変更は、部分的な対応では機能せず、システム全体を体系的に再構築して行く必要があるだろう。たとえば、「高齢」定義年齢の上昇に合わせて年金支給開始年齢を高めて行くことは、もしそれだけ単独で行われれば、むしろ混乱を招くことが必至である。第一にここで捉えた健康度の増進は、あくまで集団の平均的傾向を示すものであり、実際には高齢期においては健康の個人差が大きいのであるから、制度の適用年齢を一律に変えることは、趣旨が適合せず不利益を被る層を多く発生させることになる。

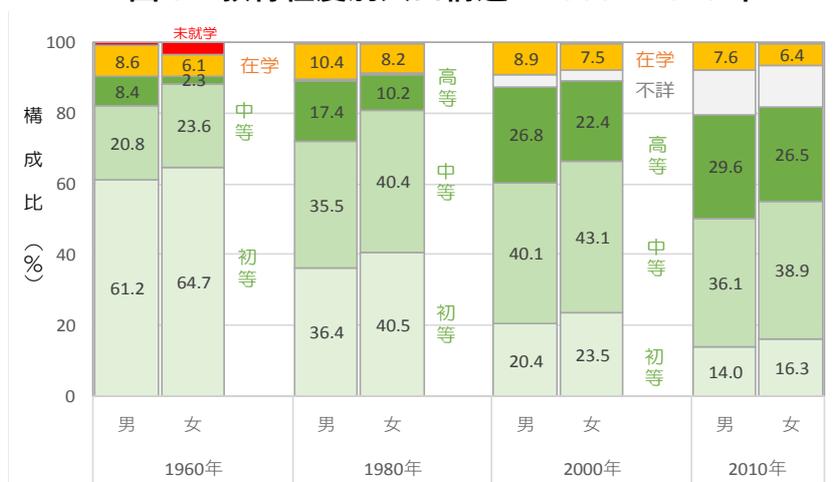
ここで理解すべきことは、現行システムが個人一人ひとりの健康度の違いといった詳細を扱うことがまったくできないという事実である。ひるがえって、新たなシステムではそうした個人の事情を把握し、それに応じたきめ細かな対応ができるものでなくてはならない。これはすでに指摘したように測定の困難がともなうが、これを実現することは、今後の社会の持続にとっては必須であるとする。現行システムの中でも、比較的新しい介護保険において、個人ごとの要介護度を認定し、これに対応したサービスを提供するという仕組みは、今後の社会に必要な新たなタイプのシステムの先駆けとみることができる。

ただし、これはすでに経験されている通り、非常にコストのかかる仕組みであり、仮に安易な運用がなされれば、公平性を確保することが困難となる。したがって、認定の手続きをより効率的に、また同時に客観的に行うためには科学的な視点を導入することが必須である。具体的には、「要介護度」や「健康度」といった概念を医学的、経済学的、社会学的観点から科学的に規定し、それを個人のおける効率的かつ正確な同定に結び付ける手続きを開発することに他ならない。これを実現する前提には、国民全体における介護の量的、質的需要や健康の実態を常にモニタリングする体制が必要となるであろう。そのためには、個人番号（マイナンバー）を基礎として、レセプトや診療群分類包括評価（DPC）等の既存データの活用に加えて、国民「要介護度」「健康度」把握のための統計調査～データベースの運用、連携が必要となるであろう。近年における個人番号（マイナンバー）の導入やビッグデータ活用環境の発達、活用事例の蓄積は、そうした国民に関する政府統計データベース開発の基礎を与えるものと見ることができる。

(2) 教育・教養

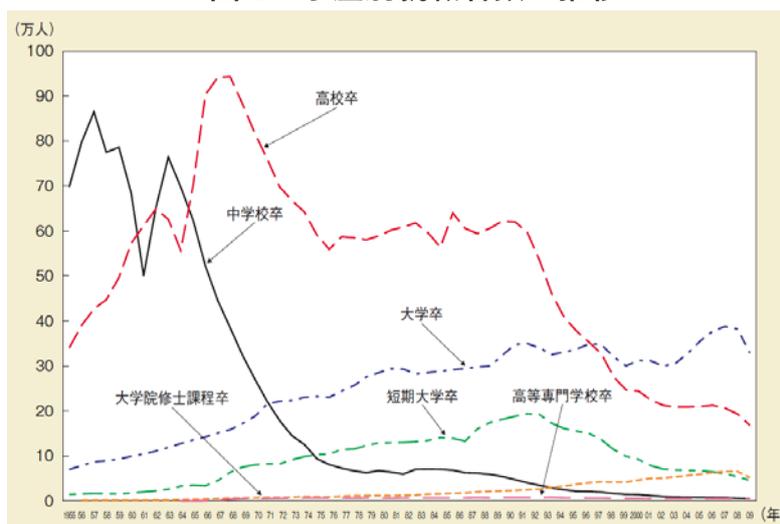
国民の生産性を向上させ、同時に個々の生活や人生に豊かさをもたらす人口資質として健康の次に重視すべき項目は、教育・教養であると考えられる。わが国における国民の教育程度は、戦後急速に高まり、現在に至るまで向上を続けている（図3）。具体的に学歴別の就職者数の推移を見ても、近年少子化により就業者総数が減少しているにも関わらず、大学以上卒業の就職者は増える傾向にあった（図4）。

図3 教育程度別人口構造：1960～2010年



注：総務省統計局『国勢調査報告』による。15歳以上人口。初等教育は小学校・中学校・高小（旧青年学校を含む）を、中等教育は高校・旧中を、高等教育は短大・高専・大学・大学院をそれぞれ卒業した者。

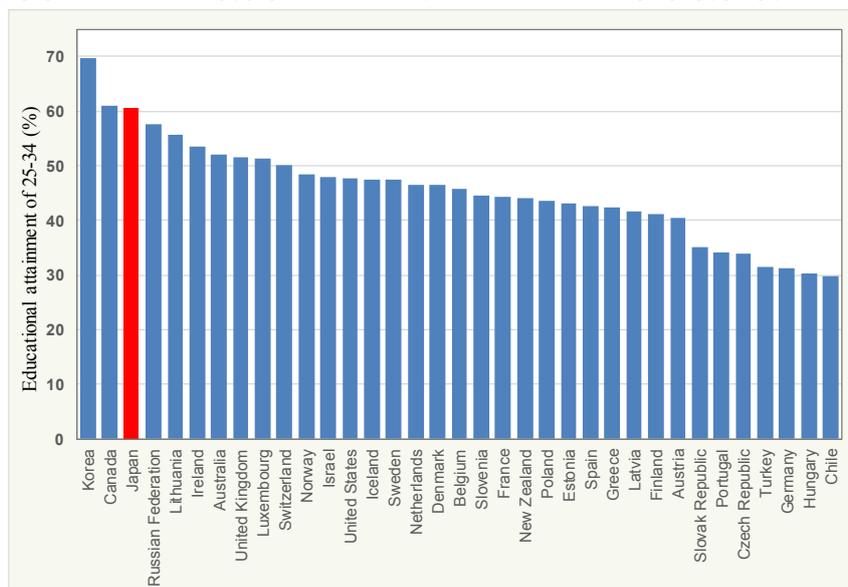
図4 学歴別就職者数の推移



注：数値は、各年の卒業者における就職者数であり、進学しかつ就職した者を含む。
資料：文部科学省「学校基本調査」

また、わが国は現在、国際的に見ても大学卒業割合の多い、教育水準の高い社会として見ることができる（図5）。これらは先進諸国において、産業構造が急速に高度情報化に向けてシフトしている現在、国全体としての生産性を維持し発展させて行く上で、少なくとも潜在的には有利な条件を有していることになる。

図5 OECD各国25～34歳における大学卒業割合の比較



Source: OECD(2018) *Education at a Glance 2018: OECD Indicators*.

しかし、国民の教育程度の領域においても、現行の企業や社会システムにおいて、わが国の有する優位な条件が十分に活用されているとはいえない。まず、教育の内容が現在の急速に変わる産業構造や国際環境に沿った有効なものであるかどうかを十分に検討する必要がある。とりわけ今後のビジネスシーンにおいて要求される外国語能力も含めたコミュニケーション能力や創造力、問題解決能力といった高度な技能を解発するカリキュラムが重要となってくるであろう。経済のグローバル化が進み、あらゆる分野においてAIなどに代表される革新的技術の導入が進むとみられるなか、それらに柔軟に対応して行く人材の育成が社会を維持するためのカギとなるはずである。

また、教育程度の向上はわが国だけでなく、世界的の進展することが見込まれており、とりわけ現在の発展途上諸国においてはその効果が著しいものになると予想される。Lutzらは、そうした教育の普及によって、人口の推移自体が影響を受けるとして、教育程度の進展を内生化した将来人口推計を行っている（図6、Lutz et al. 2018）。その意図は、教育の普及に対する努力の程度によって、世界人口の将来像が大きく変わることを示すことと、教育を介しての人的資本の開発が、例外なく高齢化を経験して行く各国の人々の福祉に資することを示すことである。

21世紀を通して、ほぼ世界一の高齢化国として歩む日本にとっては、教育ならびに人材育成についても世界をリードすることが、自らの社会を存続させるためにも必要なことであるが、それは後に続く他の国々にとっても重要な経験知となる。そうした視点から刻々と変化する情勢に即した教育システムの試行錯誤的な再構築に取り組む必要があるだろう。その際には、国民の詳細な就学状況、学力、技能ならびに教育機関における取組等に関するモニタリング情報が必須となる。教育の分野においても、国民的データベースの構築、運用が望まれる。

図6 教育水準を内生化したモデルによる将来推計人口：
イギリス、フランス、スウェーデン

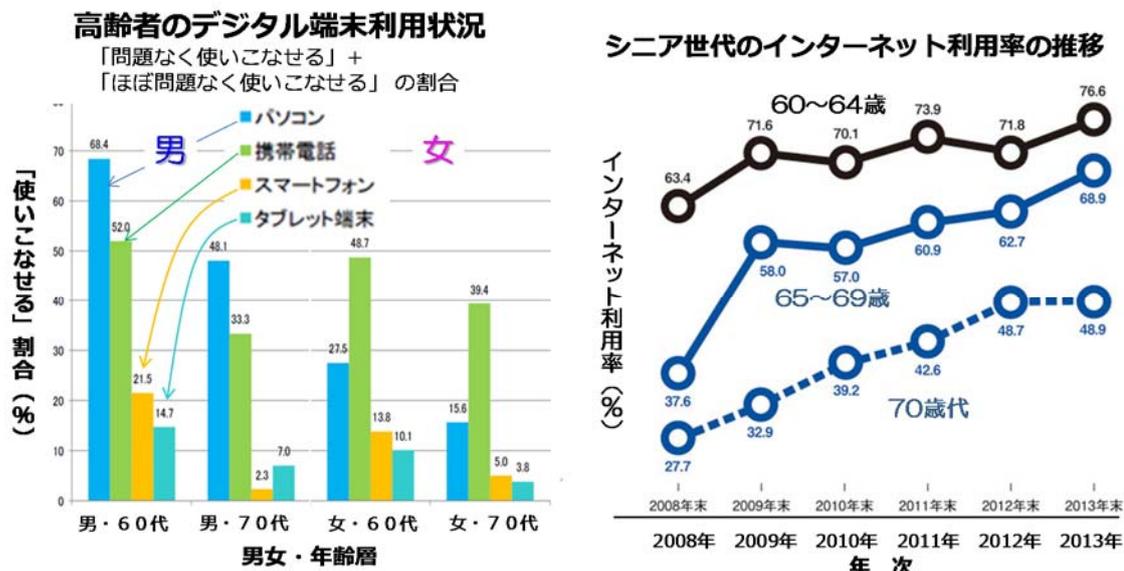


Source: Wolfgang Lutz, et al. (Eds.), 2018, *Demographic and human capital scenarios for the 21st century - 2018 assessment for 201 countries*. Luxembourg: European Union.

(3) 技能・スキル

今後に進展するとみられる多くの技術革新を考えると、そうした現場で働く人々に要請される職業的スキルも年々変化して行くことは必至である。こうした技能の刷新をいかに行うかということも、社会の生産性を維持するうえで重要な要素となる。たとえば、近年の高齢者についてみると、インターネットやデジタル機器の急速な普及に対して、当初はまったく利用できない層が大勢を占めたが、現在では職業上でも生活上でも積極的に活用する層が多数となっている（図7）。彼らの多くは40歳代でこれらの新技術に仕事上で遭遇し、基本的な概念を習得した経験があるものと思われる。このように革新的な技術であっても広範で柔軟な基礎を有する場合は対応が可能になると考えられるから、今後重要となる技術に関しては、早い時期から基本原理の啓蒙に努めることが有効であろう。また、高齢者は新技術を理解できないとする固定概念は排すべきである。

図7 変わる高齢者像：デジタル・シニアの広がり

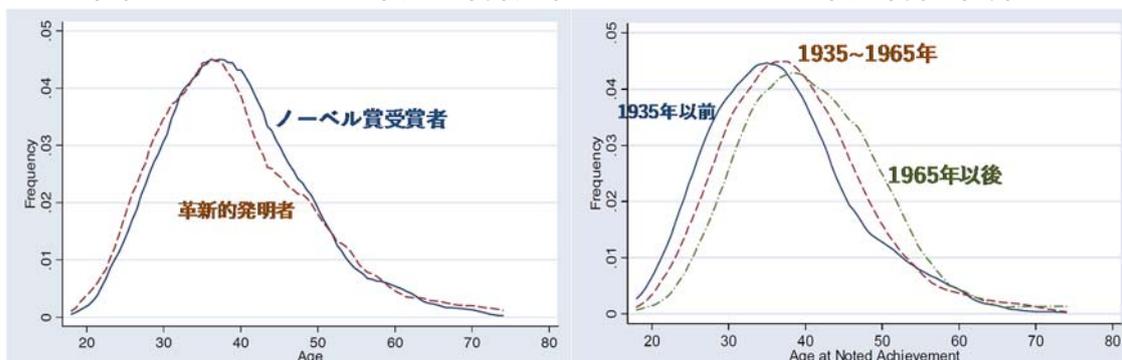


資料：電通中高年調査 2015 年、齊藤徹(電通総研)「高齢者のライフスタイルと消費・働き方」(政府税調提出資料)より。総務省が毎年公表している「通信利用動向調査」小林直樹「2000 万人のデジタルシニア、年平均 19 万円のネット消費を狙え」(日経デジタルマーケティング記事 2014 年 12 月 22 日)より

しかしながら、革新的技術の開発は若い年代にしかできないから、人口高齢化が進行するにしたがって、イノベーションは低調になるとの見方がある。この点に関して、過去のイノベーションを生み出した年齢の分布をみると、確かに 30 歳代後半をピークに頻度は下がる傾向にあるが、たとえば 50 歳付近でもピーク時の半数を維持している(図 8-a)。また、そのピーク年齢は、時代とともに高齢側へシフトが見られ、より高齢で生み出されるイノベーションの割合が増えている(図 8-b)。したがって、革新的技術の開発といえども、必ずしも特定の年齢層に固定して生じているわけではないことがわかる。

図8 イノベーション形成の年齢分布

a. 20 世紀のイノベーション形成の年齢分布 b.イノベーション形成年齢の高齢シフト



Source: Benjamin F. Jones, “Age and Great Invention” (working paper, National Bureau of Economic Research, Cambridge, MA, May 2005), 36, Figure 2, <https://www.nber.org/papers/w11359>.

この例から言えることは、技術革新のための資源配分や機会を、固定的に特定の年齢に

集中させるようなことは必ずしも効率的ではなく、むしろ多様な知識、技能に期待して、広い層に機会を与えることがより創造的であり得るということである。こうした活動についても、インターネット上に広く情報交換の場を設け、公的な情報研究機関によるビッグデータ収集のしくみを持つことが有効であろう。

(4) ライフサイクル統計の必要性

人口減少、少子高齢化がもたらす社会の生産性低下に対して、人口資質向上による対処の可能性と、これに必要となる公的統計情報とその国民的データベースの必要性を見てきた。より一般化して表現するなら、国民生活の質をライフサイクルの視点からモニタリングする統計システムが求められるということができる。ここではその中で最も基礎的な国民の経済活動をライフサイクルの視点からとらえるための国民移転勘定 National Transfer Account (NTA)について簡単に検討したい。

国民移転勘定とは、国民の勤労所得、消費、資産、私的移転、さらに税、社会保障などを介した公的移転など、個人あるいは世代における経済フローを年齢別に把握するための統計システムである。各国のチームから構成される国際的プロジェクトによって運営され、現在では世界 46 カ国で統計が計算されている。算出には、わが国では全国消費実態調査、家計調査などのマイクロデータと国民経済計算における国民所得勘定や年金・医療保険統計といったマクロデータを用いる。

ここでは全年齢にわたる労働所得、消費ならびにその収支（消費－労働所得）について、1965年、90年、2015年男女について推計した結果と、これらを元に2040年、2065年について将来推計した結果を図9に示す。さらにこれらの年次の年齢別数値を元に、各年ごとに補間推計、将来推計を行い、年間の収支を男女別にしめしたグラフを図10に示した。これによれば、今後、男女・年齢別にみた個人の平均生産性が変わらないとする、男女・全年齢について合計した収支（消費－労働所得）、すなわち消費超過（赤字）の幅が人口高齢化によって増大し行くことがわかる。これは国民経済が持続可能でないことを示しており、個人の平均生産性を高める必要があることを示している。これに対しては、少子化の是正や外国人労働力の導入によって対処するにせよ、本稿で検討してきた人口資質の向上によらなければ、実質的に社会の持続可能性を確保することはできないであろう。

図9 労働所得・消費・収支（消費－労働所得）：1965, 90, 2015, 40, 65年

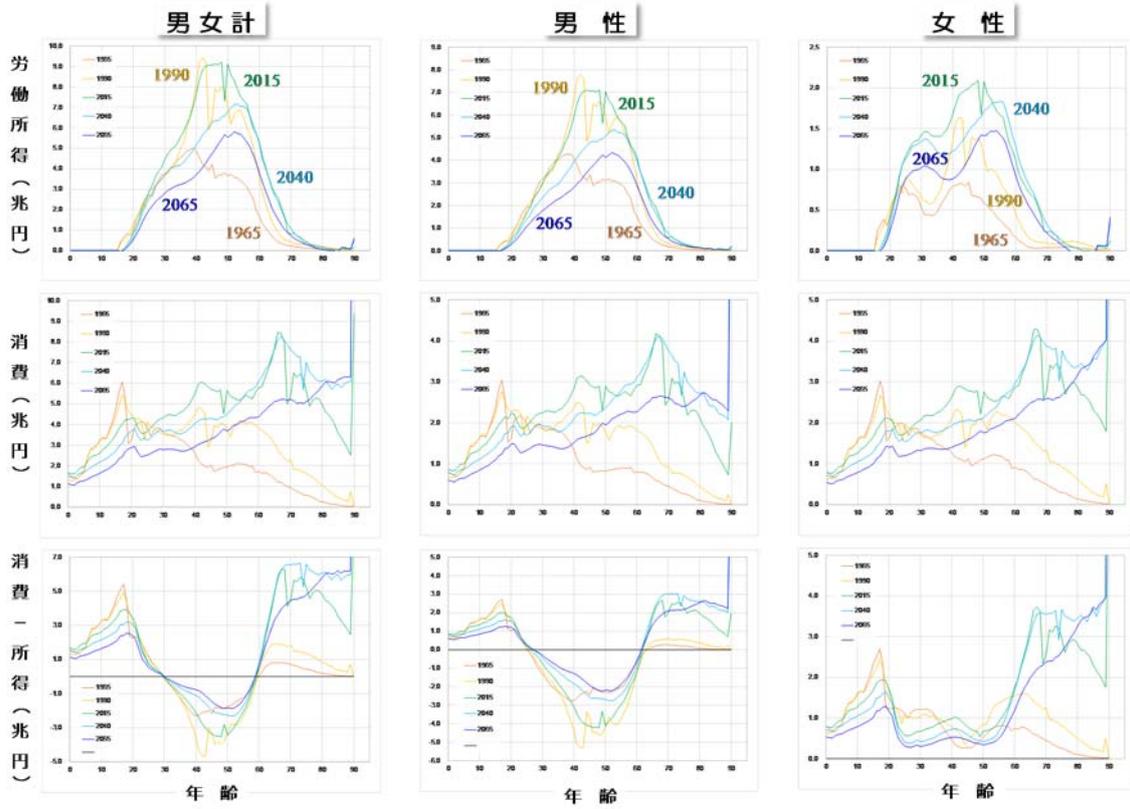
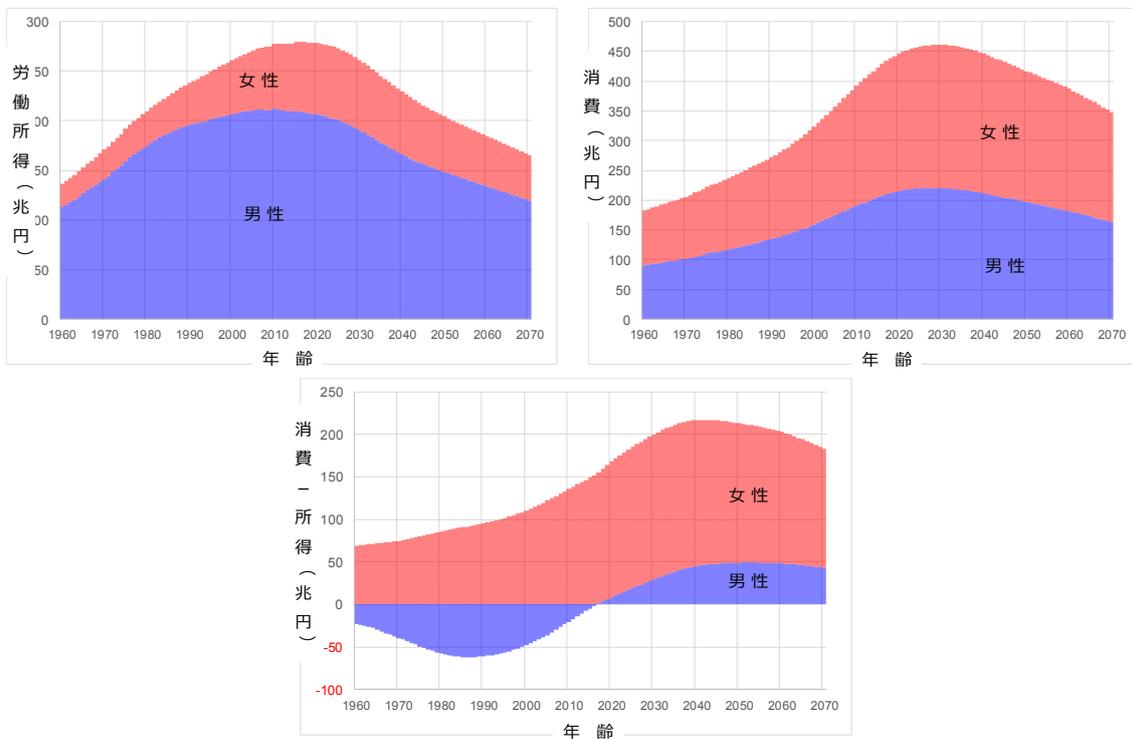


図10 労働所得・消費・収支（消費－労働所得）の年次推移: 1960～2070年



5. 考察とまとめ

わが国は世界に先駆けて人口減少・少子高齢化を進展させ、高齢層の増大と現役世代縮小という人口バランスの崩れによって経済社会システムの機能は阻害され、社会の継承や存続そのものが脅かされている。しかし一方で、わが国は世界一の健康長寿と高い教育水準を有しており、とりわけ高齢層は量的増大とともに健康ならびにスキルや教養という質の面でも大きく変化しつつある。

こうした人的資本の質の面での充実、すなわち人口資質の向上は、国や地域の生産性向上に寄与するだけでなく、人々がより豊かな人生をおくる前提ともなるものである。ことから人口資質の向上は、人口減少・少子高齢化によるマクロ面（経済システム）での不具合を是正するとともに、ミクロ面（人々の人生や生活）での発展を促す働きを持っている。

本論では、こうした人口資質を定量的に捉え、今後必要となる社会システム再構築の際の指針形成に資するため、政府統計のあり方や指標、さらには統計的方法論についての検討を試みた。

まず、生命表から得られる平均余命を健康度と見なして、「高齢」定義に対してこの健康度による動的な年齢を適用すると、今後の人口高齢化について従来とはまったくことなる将来像が得られることがわかった。扶養する人口と扶養される人口とのバランスについては、人口ボーナス期に匹敵する状態を維持することができるのである。

また、教育・教養、さらには技能・スキルという人口資質の要素については、戦後大きく改善が続き、国際的にも高い水準にあるから、これらを効率的に生産性に結び付けることができれば、社会全体の生産性維持に大いに寄与できると考えられる。

ただし、人口資質の向上を生産性向上に結び付けるには、暦年齢に強く依存する現行の社会システムでは不十分であり、システム全体を個人の特性を詳細に把握し、これらの情報により個人の持つ潜在的生産性を最大限に引き出すような新たなシステムを体系的に再構築する必要がある。そのためには、まず健康度や教育程度といった特性を定義し、定量的測定を行うための統計情報と、個人差を詳細に把握するための国民的データベースを構築することが必須となる。

次に、こうした人口資質による高齢化への対処の本質的意味について考察したい。まずは高齢人口における健康度や教育程度の向上が潜在的に生産性向上に寄与し、支えられる立場から部分的にせよ支える立場に移ることによって人口構造のアンバランスの是正に働くということである。これが機能するためには個人差を把握し、適切に対処する新たな仕組みが必要であると述べた。とすると、このしくみの対象は、必ずしも高齢者に限らなくてもよいはずである。たとえば、生産年齢層であっても何らかの障害を持つ者に対しては、制度がその詳細を把握し、これに即した支援サービスや補助金支給によって、その個人の持つ生産性の発揮を保障するのである。すなわち、すべての成人について、何らかの事情によって持てる生産性を十分に発揮できていない個人については、制度が阻害要因を把握し、是正のための支援等を提供するというのが、ここで求めている制度の本質的な働きと

ということになる。この制度システムが最終的に目指すのは、国民すべてがその持てる能力を最大限に発揮できることが、あたかも個人の基本的人権であるように保障することである。これは教育、人材育成においても当てはまる構図である。

最後に、日本社会の持続可能性を本質的に阻害しているしくみについて考えたい。たとえば、人口高齢化は社会理念である民主主義の多数決原理と、資本主義経済の根幹にある市場原理の双方を介して、社会の資源配分を高年齢層に振り向ける性質を持つことが指摘されている (Preston 1984)。それは子育て世代を含む若年層の生活を劣化させることになり、少子化を助長する働きを持つ。少子化は人口高齢化の最大の要因であるから、ここに悪循環が存在する (Lutz, et al. 2006)。したがって、現代社会が拠って立つ政治と経済の基本理念が少子化を助長し、社会の持続可能性を阻害していることになる。はたして、この悪循環のどこを断つべきであろうか？おそらく、資源配分の段階であろう。高齢者達における経済合理性に任せた意思決定にこそ悪循環の起点を見るべきと考えられる。しかしながら、彼らは経済合理的に思考する権利を有している。自らの限られた生存時間の中に、最大の効用を得ようと行動することは、彼らの基本的人権である。ここで人口資質のまだ検討していない第三の項目が重要な役割を持つのではないか。それは倫理・品格である。彼らには自分個人の経済合理的利益を超えて、たとえば社会の継承と存続に価値を置く決定を下すこともできる。というのは、自らの生も先達から継承したこの社会という培地によって生まれたのであるから、自らも先達と同様の行動を取るということには十分な根拠があるからである。ただし、その決定権が当人にある以上、倫理・品格と呼ばれる第三の人口資質の働きに、この社会の持続可能性はかかっていると考えられるのである。

以上のように、人口高齢化がもたらす本質的課題に対して、人口資質の向上によるのみ対応し得ると考えられるのであり、これを有効とする具体的制度、あるいは社会システムの再構築には、健康などの操作的定義を策定し得る統計情報と、個人の特性や個人差を把握することのできる国民的データベースの保有、運営が前提になると考えられるのである。

文献

Lutz, W., Skirbekk V. and M.R. Testa, (2006), "The low fertility trap hypothesis: Forces that may lead to further postponement and fewer births in Europe", Vienna Yearbook of Population Research, pp. 167-192.

Lutz, Wolfgang et al. (Eds.), 2018, *Demographic and human capital scenarios for the 21st century - 2018 assessment for 201 countries*. Luxembourg: European Union.

OECD (2018). *Education at a Glance 2018: OECD Indicators*.
<https://doi.org/10.1787/888933801582> (accessed at 2018.2.1).

Preston Samuel H. (1984). "Children and the elderly: divergent paths for America's dependents." *Demography* 21, 435-457.

Rindfuss, Choe, Brauner-Otto (2016), "The Emergence of Two Distinct Fertility Regimes in Economically Advanced Countries." *Popul.Res.Policy Rev* 35:287-304.

United Nations, (2017), World Population Prospects: The 2017 Revision. United Nations, Department of Economic and Social Affairs, Population Division.

金子隆一・村木厚子・宮本太郎(2018)『新時代からの挑戦状-少親多死社会をどう生きるか』厚生労働統計協会 (ISBN-10: 4875117736)

総務省が毎年公表している「通信利用動向調査」

小林直樹「2000万人のデジタルシニア、年平均19万円のネット消費を狙え」(日経デジタルマーケティング記事 2014年12月22日)より

国立社会保障・人口問題研究所 (2017)『日本の将来推計人口(平成29年推計)』厚生労働統計協会.

電通中高年調査 2015年、斉藤徹(電通総研)「高齢者のライフスタイルと消費・働き方」(政府税調提出資料)より

内閣統計局編 (1930)『明治五年以降我国の人口』(調査資料 第3輯)、内閣統計局.

森田優三(1944)『人口増加の分析』日本評論社.

ランダムネスによる プライバシー保護の評価

星野 伸明

金沢大学・経済学経営学系

2019年1月31日

本研究は科研費、統計数理研究所共同利用研究経費の助成を受けている。

1

概要

1. ランダムな個票データの情報漏洩危険性
2. 差分プライバシーと微分プライバシー
3. 各種ランダム度数メカニズムの危険性評価
4. 有望なメカニズム — 擬似多項分布の紹介

2

ランダムな個票データは安全？

- 情報保護のため、公開する個票データを乱数にすることはよく行われる。
 - 例) サンプルング、ランダムノイズ付加、ランダムスワッピング、模造
- 公開個票データから母集団の母数が推定されると情報漏洩が判明：
 - 既知属性所与で条件付き（母集団）分布のサイズが1と分かれば、識別開示
 - 既知属性所与で条件付き（母集団）分布が退化と分かれば、属性開示
 - 既知属性所与で条件付き（母集団）分布の分散が小と分かれば、推測開示
- ランダムな出力（=標本）から母数（=母集団分布）を推定する際の精度を、情報漏洩の危険性と考える。
- 母数推定精度（推定量の分散）は、対数尤度関数の傾きが決めている。
- 「差分プライバシー」は対数尤度関数の傾きを押さえることで、母数推定の精度を押さえている。

差分プライバシー) Machanavajjhala (2008) の枠組み

- $\vec{m} = (m_1, m_2, \dots, m_J)$ が「標本」。 m_j は第 j セルの観測度数。
 - $m = \sum_{j=1}^J m_j$ は標本サイズ。
- $\vec{n} = (n_1, n_2, \dots, n_J)$ が「母集団」。 n_j は第 j セルの母集団度数。
 - $n = \sum_{j=1}^J n_j$ は母集団サイズ。
- $\vec{n}' = (n'_1, n'_2, \dots, n'_J)$ は \vec{n} と一個体のみ違う度数ベクトル。つまり適当な $j \neq k$ について、 $n'_j = n_j + 1, n'_k = n_k - 1$ となっている。
- ϵ -差分プライベート $\Leftrightarrow \forall (\vec{m}, \vec{n})$

$$\left| P(\vec{m}; \vec{n}) / P(\vec{m}; \vec{n}') \right| \leq \exp(\epsilon) \quad (1)$$

Notes

- \vec{n} と \vec{n}' の差は一個体の移動だが、追加（ないし削除）した場合の尤度比で差分プライバシーを定義してもよいはず。
 - 移動では母集団に含まれていることに変わりはない。
- 「標本」を個票データとして公開するのでなければ、 \vec{m} が非負整数のベクトルである必要はない。
 - \vec{n} の各要素にラプラスノイズを乗せた \vec{m} は差分プライベートである。
 - 差分プライベートな実数のベクトルを事後処理して非負整数のベクトル化しても差分プライベート。しかしこの方法の挙動解析は面倒。
 - * 寺田ほか (2016) は総和条件を満たす非負整数となるように事後処理。
 - 本報告では総和条件を満たす非負整数の多変量分布を考察。

微分プライバシー

- 一個体を追加ないし削除した場合の尤度比で差分プライバシーを定義：

$$\left| \frac{\log P(\vec{m}; \vec{n}) - \log P(\vec{m}; \vec{n} + \Delta)}{|\Delta|} \right| \leq \epsilon, \quad |\Delta| = 1$$

- ここで $|\Delta| \rightarrow 0$ の極限の解釈：

$$|\partial \log P(\vec{m}; \vec{n}) / \partial n_j| \leq \epsilon \tag{2}$$

- 全ての \vec{m} について (2) 式が成立するなら、正則条件の下で

$$V(\hat{n}_j) \geq 1/\epsilon^2$$

- $I(n_j) := E_{\vec{m}}[(\partial \log P(\vec{m}; \vec{n}) / \partial n_j)^2] \leq \epsilon^2$ である。この時クラメールラオの不等式が成立するなら $V(\hat{n}_j) \geq 1/I(n_j)$
- つまり母数推定精度の上限を与えている。

考察) 単純無作為非復元抽出

- リサンプリング機構として単純無作為非復元抽出を採用した場合 (超幾何分布) :

$$P(\vec{m}; \vec{n}) = \frac{\binom{n_1}{m_1} \binom{n_2}{m_2} \dots \binom{n_J}{m_J}}{\binom{n}{m}}$$

- この時 ϵ -差分プライバシーは達成出来ない。
 - $P(\vec{m}; \vec{n})/P(\vec{m}; \vec{n}')$ の分母が 0、分子が正となる場合がある。ex) $n_k = 1, m_k = 1$
- \vec{m} の値域が母数 \vec{n} に依存しているのが原因。
- \vec{m} の値域を \vec{n} に依存しないようにするには、例えば

$$P(\vec{m}; \vec{n}, \vec{l}) = \frac{\binom{n_1+l_1}{m_1} \binom{n_2+l_2}{m_2} \dots \binom{n_J+l_J}{m_J}}{\binom{n+l}{m}} \tag{3}$$

のような方法が考えられる。これは一種の模造機構である。

- $l_j > m$ の時に \vec{m} の値域は \vec{n} に依存しない。

考察) 超幾何分布による模造

- 模造機構として (3) 式を採用。変形した結果 :

$$P(\vec{m}; \vec{n}, \vec{l}) = \binom{m}{\vec{m}} \frac{\Gamma(n - m + l + 1)}{\Gamma(n + l + 1)} \prod_{j=1}^J \frac{\Gamma(n_j + l_j + 1)}{\Gamma(n_j + l_j - m_j + 1)}$$

- この時 ϵ -差分プライベート \Leftrightarrow

$$\left| \frac{1 + \min_j l_j}{1 + \min_j l_j - m} \right| \leq \exp(\epsilon) \tag{4}$$

- 注) $\forall j, l_j > m$

- $n \gg m$ なら実用性があるかもしれないが。
- Machanavajjhala(2008) の例 : $(m = \text{百万}, \epsilon = 7) \Rightarrow l_j \geq 1000911.714$

超幾何分布の微分プライバシー

- 微分プライバシー $\Leftrightarrow \forall(\vec{m}, \vec{n})$

$$\left| \frac{\partial \log P(\vec{m}; \vec{n})}{\partial n_j} \right| = \left| \frac{1}{n_j + l_j} + \frac{1}{n_j + l_j - 1} + \dots + \frac{1}{n_j + l_j - m_j + 1} - \frac{1}{n + l} - \frac{1}{n + l - 1} - \dots - \frac{1}{n + l - m + 1} \right| \leq \epsilon$$

- 上式左辺が最大化されるのは $n_j = 0, m_j = m$ の時なので ($l_j > m$)、微分プライバシー \Leftrightarrow

$$\left| \frac{1}{\min_j l_j} + \frac{1}{\min_j l_j - 1} + \dots + \frac{1}{\min_j l_j - m + 1} - \frac{1}{n + l} - \frac{1}{n + l - 1} - \dots - \frac{1}{n + l - m + 1} \right| \leq \epsilon$$

- $m, n, \min_j l_j (n \gg m)$ が大きければ上式は近似的に

$$\left| \frac{\min_j l_j}{\min_j l_j - m} \right| \approx \left| \frac{(\min_j l_j)(n + l - m)}{(\min_j l_j - m)(n + l)} \right| \leq \exp(\epsilon) \Leftrightarrow (4)$$

考察) 単純無作為復元抽出

- 復元抽出なら $n_j = 0$ のセルのみ m_j の値域が n_j に依存。
- 模造機構として単純無作為復元抽出を採用した場合 (多項分布):

$$P(\vec{m}; \vec{n}, \vec{\beta}) = \binom{m}{\vec{m}} \prod_{j=1}^J \left(\frac{n_j + \beta_j}{n + \beta} \right)^{m_j}$$

- $\beta_j > 0$ なら $n_j = 0$ でも m_j の値域は変わらない。

- この時 ϵ -差分プライベート \Leftrightarrow

$$\left(1 + \frac{1}{\min_j \beta_j} \right)^m \leq \exp(\epsilon) \tag{5}$$

- Machanavajjhala(2008) の例: ($m = \text{百万}, \epsilon = 7$) $\Rightarrow \beta_j \geq 142857$

単純無作為復元抽出の微分プライバシー

- 微分プライバシー $\Leftrightarrow \forall(\vec{m}, \vec{n})$

$$\left| \frac{\partial \log P(\vec{m}; \vec{n})}{\partial n_j} \right| = \left| \frac{m_j}{n_j + \beta_j} - \frac{m}{n + \beta} \right| \leq \epsilon$$

- 上式左辺が最大化されるのは $n_j = 0, m_j = m$ の時なので、微分プライバシー \Leftrightarrow

$$\left| \frac{m}{\min_j \beta_j} - \frac{m}{n + \beta} \right| \leq \epsilon$$

- n が大きいときに上式は近似的に

$$\exp\left(\frac{m}{\min_j \beta_j}\right) \leq \exp(\epsilon)$$

- さらに $\min_j \beta_j$ が大きい時

$$\left(1 + \frac{1}{\min_j \beta_j}\right)^m \approx \exp\left(\frac{m}{\min_j \beta_j}\right) \leq \exp(\epsilon) \Leftrightarrow (5)$$

考察) 負の超幾何分布による模造

- Machanavajjhala (2008) の模造機構 (負の超幾何分布) :

$$P(\vec{m}; \vec{n}, \vec{\alpha}) = \binom{m}{\vec{m}} \frac{\Gamma(n + \alpha_{\cdot})}{\Gamma(n + \alpha_{\cdot} + m)} \prod_{j=1}^J \frac{\Gamma(n_j + \alpha_j + m_j)}{\Gamma(n_j + \alpha_j)}$$

- この時 ϵ -差分プライベート \Leftrightarrow

$$\left| \frac{\min_j \alpha_j + m}{\min_j \alpha_j} \right| \leq \exp(\epsilon) \tag{6}$$

- $\alpha_j = O(m)$ くらいで左辺を bound 出来る。

- Machanavajjhala(2008) の例 : ($m = \text{百万}, \epsilon = 7$) $\Rightarrow \alpha_j \geq 914$

負の超幾何分布の微分プライバシー

- 微分プライバシー $\Leftrightarrow \forall(\vec{m}, \vec{n})$

$$\left| \frac{\partial \log P(\vec{m}; \vec{n})}{\partial n_j} \right| = \left| \frac{1}{n_j + \alpha_j} + \frac{1}{n_j + \alpha_j + 1} + \dots + \frac{1}{n_j + \alpha_j + m_j - 1} - \frac{1}{n + \alpha} - \frac{1}{n + \alpha + 1} - \dots - \frac{1}{n + \alpha + m - 1} \right| \leq \epsilon$$

- 上式左辺が最大化されるのは $n_j = 0, m_j = m$ の時なので、微分プライバシー \Leftrightarrow

$$\left| \frac{1}{\min_j \alpha_j} + \frac{1}{\min_j \alpha_j + 1} + \dots + \frac{1}{\min_j \alpha_j + m - 1} - \frac{1}{n + \alpha} - \frac{1}{n + \alpha + 1} - \dots - \frac{1}{n + \alpha + m - 1} \right| \leq \epsilon$$

- $m, n, \min_j \alpha_j (n \gg m)$ が大きければ上式は近似的に

$$\left| \frac{\min_j \alpha_j + m}{\min_j \alpha_j} \right| \approx \left| \frac{(m + \min_j \alpha_j)(n + \alpha)}{\min_j \alpha_j (n + m + \alpha)} \right| \leq \exp(\epsilon) \Leftrightarrow (6)$$

考察) 擬似多項分布による模造

- 擬似多項分布 (type 2) による模造機構：

$$P(\vec{m}; \vec{n}, \vec{\gamma}) = \binom{m}{\vec{m}} \frac{1}{(n + \gamma)(n + \gamma + m)^{m-1}} \prod_{j=1}^J (n_j + \gamma_j)(n_j + \gamma_j + m_j)^{m_j - 1}$$

- この時 ϵ -差分プライベート \Leftrightarrow

$$\left| \left(1 + \frac{1}{\min_j \gamma_j} \right) \left(1 + \frac{1}{\min_j \gamma_j + m} \right)^{m-1} \right| \leq \exp(\epsilon) \tag{7}$$

- この条件は m が大きいとき近似的に

$$(1 + 1/\min_j \gamma_j) \leq \exp(\epsilon - 1)$$

であり、 γ_j を m 依存で膨らませる必要がない。

- Machanavajhala(2008) の例： $(m = \text{百万}, \epsilon = 7) \Rightarrow \gamma_j \geq 0.00248491$

擬似多項分布の微分プライバシー

- 微分プライバシー $\Leftrightarrow \forall(\vec{m}, \vec{n})$

$$\left| \frac{\partial \log P(\vec{m}; \vec{n})}{\partial n_j} \right| = \left| \frac{m_j - 1}{n_j + \gamma_j + m_j} + \frac{1}{n_j + \gamma_j} - \frac{1}{n + \gamma} - \frac{m - 1}{n + \gamma + m} \right| \leq \epsilon$$

- 上式左辺が最大化されるのは $n_j = 0, m_j = m$ の時なので、微分プライバシー \Leftrightarrow

$$\left| \frac{m - 1}{\min_j \gamma_j + m} + \frac{1}{\min_j \gamma_j} - \frac{1}{n + \gamma} - \frac{m - 1}{n + \gamma + m} \right| \leq \epsilon$$

- n が大きければ上式は近似的に

$$\exp\left(\frac{m - 1}{\min_j \gamma_j + m} + \frac{1}{\min_j \gamma_j}\right) \leq \exp(\epsilon)$$

- さらに $\min_j \gamma_j$ が大きいとき左辺を近似して

$$\left(1 + \frac{1}{\min_j \gamma_j}\right) \left(1 + \frac{1}{\min_j \gamma_j + m}\right)^{m-1} \Leftrightarrow (7)$$

微分プライバシーと差分プライバシー

- Machanavajjhala の枠組みだと、尤度の n に依存する部分が分母と分子でキャンセルして、DP 条件が n に依存しない。
 - 条件がシンプルになる。
 - n に依存しなくていいのか：母集団に占める一個体の重みが効かない？
- 微分プライバシーだと n が $n + \delta$ に変化する影響が n に依存して表れる。
- 以上の例では $n \rightarrow \infty$ として微分プライバシー条件から n を消去すれば差分プライバシー条件と近似的に等しい。

差分プライベートな模造機構の比較

- ($m = n = 100$ 万, $\epsilon = 7$, $J = 100$ 万) かつ全てのセルで $l = 1000911.714$, $\alpha = 914$, $\beta = 142857$, $\gamma = 0.0024849$ が等しいとする。
- これらの条件の下で $n_j = 1$ 万のセル度数の期待値：

超幾何分布	多項分布	負の超幾何分布	擬似多項分布
1.00999	1.06999	11.9279	9975.22

- $n_j = 0$ のセルに期待値を配るので $n_j > 0$ のセルの期待値は過小となる。
- $n_j = 0$ のセルの保護をあきらめて n_j 依存で $(l_j, \alpha_j, \beta_j, \gamma_j)$ を決めれば、 $n_j > 0$ のセルの期待値を不偏に出来る。
- 擬似多項分布はよさそうに見えるが、ほとんど実証していない。
 - Hu and Hoshino (2018, LNCS 11126)

擬似二項分布（青色）と二項分布（燈色）の比較

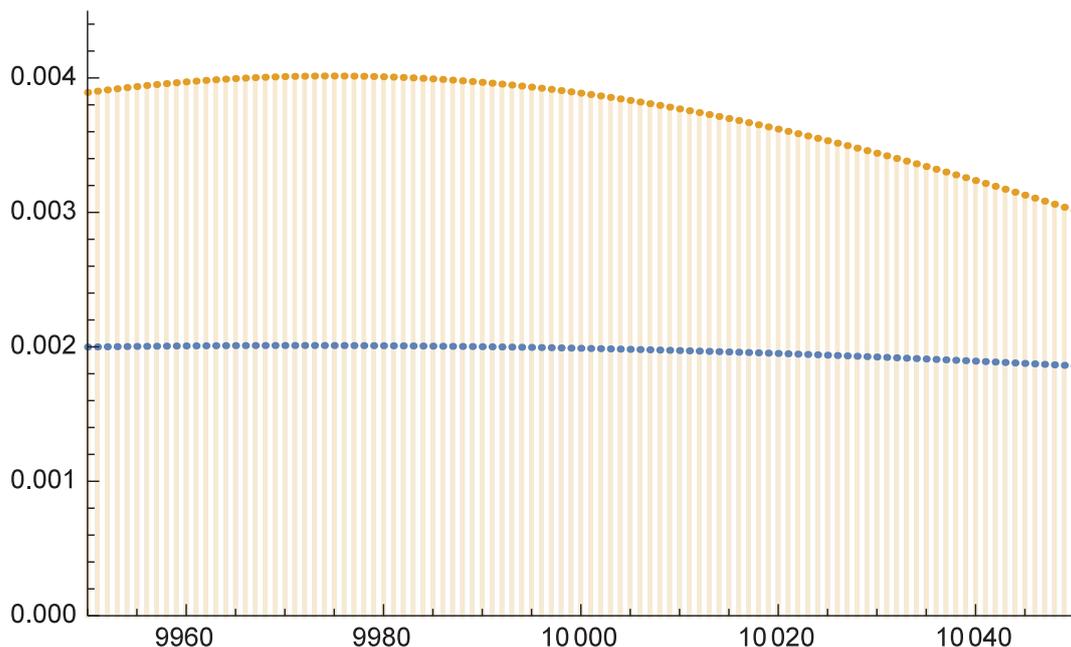


Figure 1: Probability Masses (Average 9975.22) of QB & Bin ($n = J = 1$ million, $\gamma = 0.0024849$)

擬似多項分布の紹介

The probability mass function (pmf) of $QM(\pi_1, \dots, \pi_F; n, \beta)$:

$$p(y_1, \dots, y_F) = \frac{n!}{y_1! \dots y_F!} \frac{1}{(1 + n\beta)^{n-1}} \prod_{f=1}^F \pi_f (\pi_f + y_f \beta)^{y_f - 1}$$

- y_f : the frequency of the f -th cell
- n : the given sum of F frequencies
- π_f : the f -th cell probability ($\pi_f \geq 0, \sum \pi_f = 1$)
- β : this parameter controls the variances of frequencies ($\beta \geq -\min_f \pi_f/n$) *
- $QM(\pi_1, \dots, \pi_F; n, 0) = \text{Multinomial}(\pi_1, \dots, \pi_F; n)$
- Our parameterization is different from the original (Consul and Mittal, 1977).

*Rejection sampling needs $\beta \geq 0$

Closure under the collapse of cells

The probability mass function of $QM(\pi_1, \dots, \pi_F; n, \beta)$ is expressed as

$$p(y_1, \dots, y_F) = \frac{\prod_{f=1}^F A_{y_f}(\pi_f, \beta)}{A_n(1, \beta)},$$

where

$$A_k(x, z) = x(x + kz)^{k-1}/k!, \quad k = 0, 1, 2, \dots,$$

is called “Abel polynomials”.

- Abel polynomials satisfy a convolution property (e.g., Charalambides, 2006, p.206):

$$A_n(x + y, z) = \sum_{k=0}^n A_k(x, z) A_{n-k}(y, z), \quad x \in \mathbb{R}, y \in \mathbb{R}, z \in \mathbb{R}.$$

⇒ QM is closed under the collapse of cells (H, 2009).

Conditional distribution method

Multivariate \mathbf{Y} can be generated by the sequential sampling of univariate margins conditionally:

$$\mathbf{Y} \stackrel{d}{=} (Y_1|Y_2, \dots, Y_F) \dots (Y_{F-2}|Y_{F-1}, Y_F)(Y_{F-1}|Y_F)Y_F$$

Theorem 1 If $\mathbf{Y} \sim \text{QM}(\pi_1, \dots, \pi_F; n, \beta)$ then

$(Y_f|Y_{f+1} = y_{f+1}, \dots, Y_F = y_F) \sim \text{QB}(\pi_f / (1 - \sum_{g=f+1}^F \pi_g); n - \sum_{g=f+1}^F y_g, \beta)$ for $f = 1, \dots, F$.

- $\text{QB}(\pi; n, \beta) = \text{QM}(\pi, 1 - \pi; n, \beta)$: “Quasi-Binomial (type 2)”

- Note that $\beta \geq -\min_{f \in [F]} \frac{\pi_f}{n} \geq$

$$-\min \left\{ \frac{\pi_f}{(1 - \sum_{g=f+1}^F \pi_g)(n - \sum_{g=f+1}^F y_g)}, \frac{\sum_{g=1}^{f-1} \pi_g}{(1 - \sum_{g=f+1}^F \pi_g)(n - \sum_{g=f+1}^F y_g)} \right\}$$

- Theorem holds even after exchanging the indices of cells.

– Faster to sample when $\pi_1 \leq \pi_2 \leq \dots \leq \pi_F$; see Ho et al. (1979)

• We can generate QM samples if we can generate QB samples.

Multi-stage sampling

Example (2-stage sampling) The first stage generates $\sum_{g=1}^f Y_g \sim \text{QM}(\sum_{g=1}^f \pi_g; n, \beta)$. The second stage generates (Y_1, \dots, Y_f) given their sum. The second stage distribution is:

Theorem 2 If $\mathbf{Y} \sim \text{QM}(\pi_1, \dots, \pi_F; n, \beta)$ then for $f = 1, \dots, F$ and $m = 0, \dots, n$,

$$(Y_1, \dots, Y_f | \sum_{g=1}^f Y_g = m) \sim \text{QM}(\pi_1 / (\sum_{g=1}^f \pi_g), \dots, \pi_f / (\sum_{g=1}^f \pi_g); m, \beta).$$

- Note that $\beta \geq -\min_{f \in [F]} \frac{\pi_f}{n} \geq -\min_{h \in [f]} \frac{\pi_h / (\sum_{g=1}^f \pi_g)}{m}$

• A recursive argument validates multi-stage sampling from the QM distribution.

• Multi-stage sampling may be faster:

– See Malefaki and Iliopoulos (2007)

– It lowers the rejection rate of our sampling to be proposed.

Rejection sampling from QB

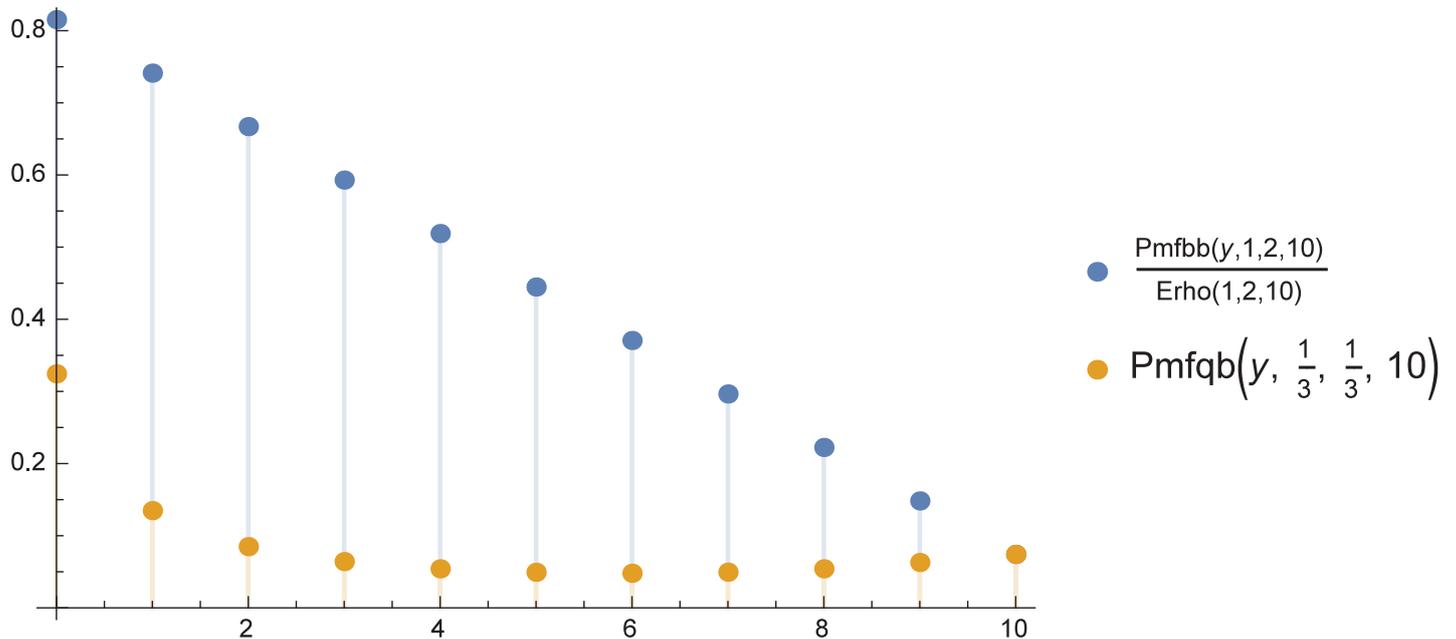


Figure 2: Dominating QB pmf by BB pmf (Average acceptance rate \doteq 20%)

BB distribution

Beta(a_1, a_2) distribution is defined by the following density function:

$$f(p) = \frac{\Gamma(a_1)\Gamma(a_2)}{\Gamma(a_1 + a_2)} p^{a_1-1} (1 - p)^{a_2-1}, \quad 0 \leq p \leq 1,$$

where a_1, a_2 are usually required to be positive,

$Y | p \sim \text{Binomial}(n, \pi)$ then Y is called $\text{BB}(a_1, a_2; n)$ distributed, with pmf

$$p_{BB}(y) = \frac{n!}{y!(n - y)!} \frac{\Gamma(a_1)}{\Gamma(a_1 + n)} \frac{\Gamma(a_1 + y)}{\Gamma(a_1)} \frac{\Gamma(a_2 + n - y)}{\Gamma(a_2)}, \quad y = 0, \dots, n.$$

- Both BB and QB distributions belong to the class of CCP distributions (H, 2009).
- To equalize the “cell probabilities” of BB and QB, it has to be

$$\pi_1 = a_1 / (a_1 + a_2), \quad \beta = 1 / (a_1 + a_2)$$

Largest difference between pmfs

We need to find the minimum ratio of

$$\frac{p_{BB}(y)}{p_{QB}(y)} = \frac{\Gamma(a. + 1)\Gamma(a_1 + y)\Gamma(a_2 + n - y)(a. + n)^{n-1}}{\Gamma(a. + n)\Gamma(a_1 + 1)\Gamma(a_2 + 1)(a_1 + y)^{y-1}(a_2 + (n - y))^{n-y-1}} \quad (8)$$

to decide the multiplication rate of p_{BB} to dominate p_{QB} .

Theorem 3 Suppose that n is a positive integer, and a_1, a_2 are positive real numbers. Denote the value of y that minimizes Equation (8) by y^* . Then $y^* = n$ when $a_1 < a_2$, and $y^* = 0$ when $a_2 < a_1$. When $a_1 = a_2$, (8) is minimized at $y = 0$ and $y = n$.

Efficiency of rejection sampling

By symmetry, we only consider the case of $a_1 < a_2$. Acceptance rates at $y = 0, 1, \dots, n$ are

$$\frac{\Gamma(a_1 + n)\Gamma(a_2 + 1)}{\Gamma(a_1 + y)\Gamma(a_2 + n - y)} \frac{(a_1 + y)^{y-1}(a_2 + n - y)^{n-y-1}}{(a_1 + n)^{n-1}} =: \rho(y).$$

Then

$$E(\rho(Y)) = \min_y \frac{p_{BB}(y)}{p_{QB}(y)} = \frac{p_{BB}(n)}{p_{QB}(n)} = \frac{\Gamma(a. + 1)\Gamma(a_1 + n)}{\Gamma(a. + n)\Gamma(a_1 + 1)} \left(\frac{a. + n}{a_1 + n}\right)^{n-1} =: r(a_1, a_2, n).$$

- $\lim_{a \rightarrow \infty} r(\pi a, (1 - \pi)a, n) = 1$
- $\lim_{n \rightarrow \infty} r(a_1, a_2, n) = O(n^{-a_2})$

Table 1: Average Acceptance Rates: $r(0.1/\beta, 0.9/\beta, n)$

n	β									
	2^{-1}	2^{-2}	2^{-3}	2^{-4}	2^{-5}	2^{-6}	2^{-7}	2^{-8}	2^{-9}	2^{-10}
10^1	0.13	0.06	0.03	0.02	0.03	0.06	0.14	0.29	0.50	0.69
10^2	0.00	0.00	0.01	0.07	0.26	0.51	0.71	0.84	0.92	0.96
10^3	0.12	0.34	0.59	0.77	0.87	0.94	0.97	0.98	0.99	1.00
10^4	0.81	0.90	0.95	0.97	0.99	0.99	1.00	1.00	1.00	1.00
10^5	0.98	0.99	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00
10^6	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
10^7	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
10^8	0.84	0.84	0.84	0.84	0.84	0.84	0.84	0.84	0.84	0.84
10^9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Algorithm 1 (Rejection sampling from the QB distribution) *The following procedure generates a sample from $QB(a_1/(a_1 + a_2); n, 1/(a_1 + a_2))$ for a positive integer n .*

When $0 < a_1 < a_2$,

1. Generate $p \sim \text{Beta}(a_1, a_2)$
2. Generate $y|p \sim \text{Binomial}(n, p)$
3. Generate $u \sim U(0, 1)$
4. If $u > \rho(y)$ then goto 1
5. Output y .

When $0 < a_2 < a_1$,

1. Swap a_2 and a_1 .
2. Generate $p \sim \text{Beta}(a_1, a_2)$
3. Generate $y|p \sim \text{Binomial}(n, p)$
4. Generate $u \sim U(0, 1)$
5. If $u > \rho(y)$ then goto 2
6. Output $n - y$.

統計的推測精度の管理に向けて

- 差分プライバシーというきつい基準の下でも、擬似多項分布は有用なデータを生成できるかもしれない。
- 微分プライバシー概念の方が含意が明瞭。昨年お話したようにベル多項式を用いると見通しが良くなる。
- 差分プライバシーは全ての母数の下での安全性を考えるが、所与の母数（発行したいデータ）の下での安全性でよいはず。
 - n について最大化しない条件付き差分/微分プライバシーを押しさえればよい。
 - 条件付きで議論すれば、不偏性のような性質が導入可能。 n を非確率的に操作するマスクの評価に適する。

多次元非定常時系列への新しいフィルタリング法¹

2019年1月

国友直人(明治大学)

1 はじめに

- 多くのマクロ経済データにおいては非定常性、季節性、定常的な循環、観測ノイズなどがしばしば観察される。例えばGDP統計の季節調整・設備投資系列の予測(内閣府)、複数のマクロ消費系列からの合成指標の構成(統計局)など、幾つかの具体的問題がある。
- 政府統計で広く利用されている季節調整法のX-12-ARIMAでは一次元Box-Jenkins法(Reg-ARIMAモデル)が古典的な移動平均法(Moving Average)とともに(十分な理解があるといえない中でも)広く利用されている。
- 統計数理研究所で開発された赤池・石黒のBAYSEA、北川のDECOMP(Kitagawa(2010))においてはガウス尤度関数に基づくカルマン・フィルタリングにより1次元時系列の成分分解が実用化されている。多次元化はかなり困難である。
- 多くのマクロ時系列では全く同一ではないが似たような季節性や(景気)循環変動などの共変動が観察されている。Engle-Granger(1987)では非定常系列においてVAR(vector AR)による共和分(co-integration)分析を提唱しているが、(現実的な)季節性や観測ノイズの存在は無視している。
- 四半期データ(20年)では80、月次データでは120程度の時系列データを想定、時系列としては小標本データである。
- 他方、Kunitomo, Sato and Kurisu (2018)では金融高頻度データ分析において分離最尤推定(SIML)法を提案している。(*Separating Information Maximum Likelihood Estimation for High Frequency Financial Data*, Springer) 本稿で説明するのはSIMLをマクロ(離散時間)時系列に応用する方法であり、Nishimura, Sato and Takahashi (2019)によるsmoothing-methodの一般化に対応している。
- 本稿では多次元非定常時系列の分布にあまり依存しないフィルタリング法を考察したので報告する。応用可能性は広範におよぶが、マクロ・トレンド抽出や季節調整の問題などが実例である。

¹2019-2-5. 佐藤整尚氏との共同研究 Kunitomo-Sato(2017), Kunitomo-Sato(2019)に基づく。高橋明彦氏からの指摘に感謝する。

2 マクロ経済データの例

マクロ経済データとして重要な系列である GDP 統計における四半期の実質 GDP と実質消費の二次元時系列、マクロ消費系列の月次系列 (家計調査、消動、第三次指数) を例として議論する。

3 非定常変数誤差モデル

簡単な離散時間の変数誤差モデル

$$(3.1) \quad \mathbf{y}_i = \mathbf{x}_i + \mathbf{v}_i \quad (i = 1, \dots, n),$$

を考察する。ここで \mathbf{x}_i ($i = 1, \dots, n$) は非定常 I(1) 過程であり

$$(3.2) \quad \Delta \mathbf{x}_i = (1 - \mathcal{L})\mathbf{x}_i = \mathbf{v}_i^{(x)}$$

ラグ作用素 $\mathcal{L}\mathbf{x}_i = \mathbf{x}_{i-1}$, $\Delta = 1 - \mathcal{L}$,

$$(3.3) \quad \mathbf{v}_i = \sum_{j=0}^{\infty} \mathbf{C}_j^{(v)} \mathbf{e}_{i-j}^{(v)},$$

$$(3.4) \quad \mathbf{v}_i^{(x)} = \sum_{j=0}^{\infty} \mathbf{C}_j^{(x)} \mathbf{e}_{i-j}^{(x)},$$

$\mathbf{e}_i^{(v)}$, $\mathbf{e}_i^{(x)}$ は i.i.d. 系列、 $\mathbf{E}(\mathbf{e}_i^{(v)}) = \mathbf{0}$, $\mathbf{E}(\mathbf{e}_i^{(x)}) = \mathbf{0}$, $\mathbf{E}(\mathbf{e}_i^{(v)} \mathbf{e}_i^{(v)'}) = \Sigma_e^{(v)}$ (正定). $\mathbf{E}(\mathbf{e}_i^{(x)} \mathbf{e}_i^{(x)'}) = \Sigma_e^{(x)}$ (非負定), 係数和の絶対値収束性を仮定する。

さらに $\mathbf{f}_{\Delta x}(\mu)$, $\mathbf{f}_v(\mu)$ をそれぞれ $\Delta \mathbf{x}_i, \mathbf{v}_i$ ($i = 1, \dots, n$) の $(p \times p)$ スペクトル密度行列

$$(3.5) \quad \mathbf{f}_{\Delta x}(\mu) = \frac{1}{\pi} \left(\sum_{j=0}^{\infty} \mathbf{C}_j^{(x)} e^{2\pi i \mu j} \right) \Sigma_e^{(x)} \left(\sum_{j=0}^{\infty} \mathbf{C}_j^{(x)'} e^{-2\pi i \mu j} \right) \quad \left(-\frac{1}{2} \leq \mu \leq \frac{1}{2} \right),$$

$$(3.6) \quad \mathbf{f}_v(\mu) = \frac{1}{\pi} \left(\sum_{j=0}^{\infty} \mathbf{C}_j^{(v)} e^{2\pi i \mu j} \right) \Sigma_e^{(v)} \left(\sum_{j=0}^{\infty} \mathbf{C}_j^{(v)'} e^{-2\pi i \mu j} \right) \quad \left(-\frac{1}{2} \leq \mu \leq \frac{1}{2} \right)$$

基準化 $\mathbf{C}_0^{(x)} = \mathbf{C}_0^{(s)} = \mathbf{C}_0^{(v)} = \mathbf{I}_p$, $i^2 = -1$ 。階差変換 $\Delta \mathbf{y}_i (= \mathbf{y}_i - \mathbf{y}_{i-1})$ の $p \times p$ スペクトル密度行列は

$$(3.7) \quad \mathbf{f}_{\Delta y}(\mu) = \mathbf{f}_{\Delta x}(\mu) + (1 - e^{2\pi i \mu}) \mathbf{f}_v(\mu) (1 - e^{-2\pi i \mu})$$

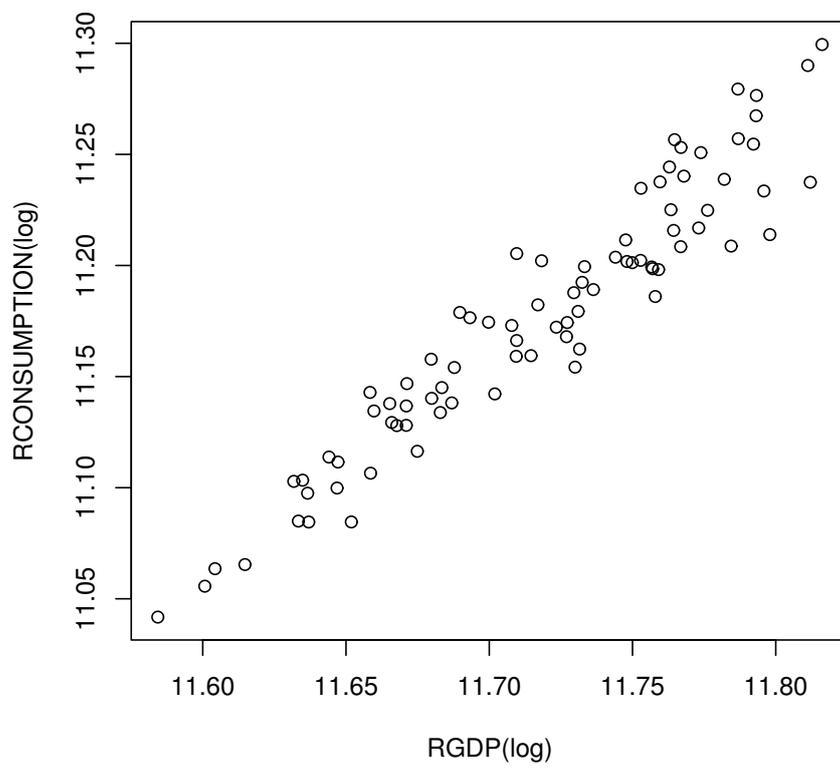


図 2.1 : 実質 GDP vs. 実質消費

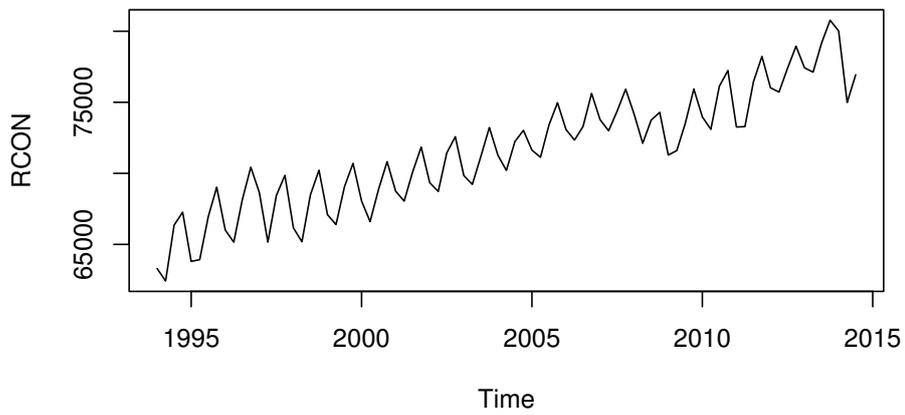
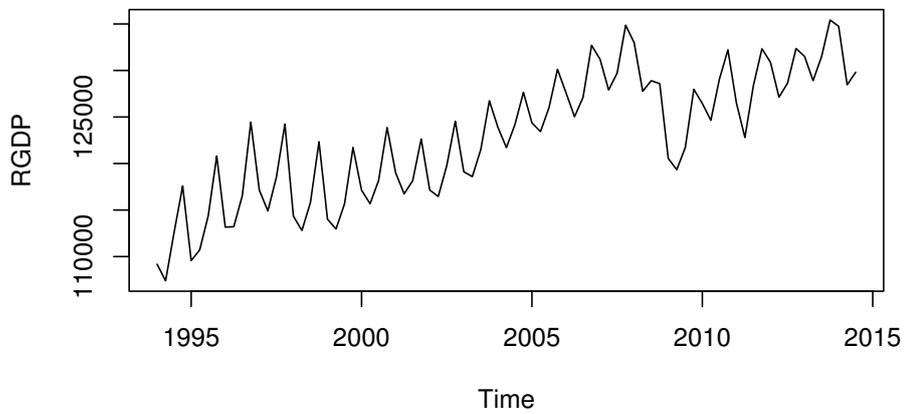


図 2.2 : 実質 GDP vs. 実質消費

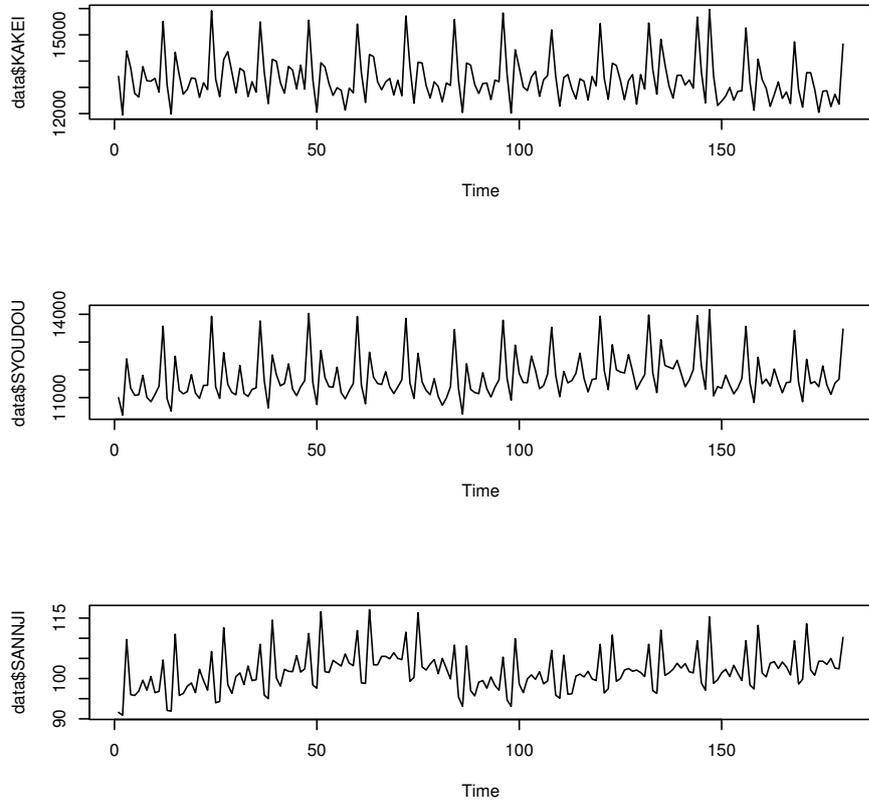


図 3.1 : マクロ消費系列

と表現される。各成分の長期分散共分散行列 ($g, h = 1, \dots, p$) は

$$(3.8) \quad \boldsymbol{\Omega}_x = \mathbf{f}_{\Delta x}(0) (= (\omega_{gh}^{(x)})),$$

$$(3.9) \quad \boldsymbol{\Omega}_v = \mathbf{f}_v(0) = (\omega_{gh}^{(v)})$$

で与えられる。

季節性を含む加法モデル

$$(3.10) \quad \mathbf{y}_i = \mathbf{x}_i + \mathbf{s}_i + \mathbf{v}_i \quad (i = 1, \dots, n),$$

において正整数 s ($s > 1$), $N, n = sN$ とした季節要素を \mathbf{s}_i ($i = 1, \dots, n$) 非定常確率過程 $\Delta \mathbf{s}_i = (1 - \mathcal{L})\mathbf{s}_i = \mathbf{v}_i^{(s)}$,

$$(3.11) \quad \mathbf{v}_i^{(s)} = \sum_{j=0}^{\infty} \mathbf{C}_{js}^{(s)} \mathbf{e}_{i-sj}^{(s)},$$

$\mathbf{e}_i^{(s)}$ は i.i.d. 系列、 $\mathbf{E}(\mathbf{e}_i^{(s)}) = \mathbf{0}$, $\mathbf{E}(\mathbf{e}_i^{(s)} \mathbf{e}_i^{(s)'}) = \boldsymbol{\Sigma}_e^{(s)}$ 、係数和の絶対値収束性を仮定する。

さらに $\mathbf{f}_{\Delta s}(\mu)$ を $\Delta \mathbf{s}_i$ の ($p \times p$) スペクトル密度行列とすると

$$(3.12) \quad \mathbf{f}_{\Delta s}(\mu) = \frac{1}{\pi} \left(\sum_{j=0}^{\infty} \mathbf{C}_{sj}^{(s)} e^{2\pi i \mu s j} \right) \boldsymbol{\Sigma}_e^{(s)} \left(\sum_{j=0}^{\infty} \mathbf{C}_{sj}^{(s)'} e^{-2\pi i \mu s j} \right) \quad \left(-\frac{1}{2} \leq \mu \leq \frac{1}{2} \right)$$

により与えられる。ここで基準化 $\mathbf{C}_0^{(s)} = \mathbf{I}_p$, $i^2 = -1$ 。階差変換 $\Delta \mathbf{y}_i (= \mathbf{y}_i - \mathbf{y}_{i-1})$ の $p \times p$ スペクトル密度行列は

$$(3.13) \quad \mathbf{f}_{\Delta y}(\mu) = \mathbf{f}_{\Delta x}(\mu) + \mathbf{f}_{\Delta s}(\mu) + (1 - e^{2\pi i \mu}) f_v(\mu) (1 - e^{-2\pi i \mu}) .$$

長期分散共分散行列 ($g, h = 1, \dots, p$) は

$$(3.14) \quad \boldsymbol{\Omega}_s = \mathbf{f}_{\Delta s}\left(\frac{1}{s}\right) (= (\omega_{gh}^{(s)}))$$

で与えられる。

4 \mathbf{K}_n 変換と \mathbf{Z}_n 過程

データ行列に対して \mathbf{K}_n -変換 (\mathbf{Y}_n より $\mathbf{Z}_n (= (\mathbf{z}_k'))$) は

$$(4.15) \quad \mathbf{Z}_n = \mathbf{K}_n (\mathbf{Y}_n - \bar{\mathbf{Y}}_0), \mathbf{K}_n = \mathbf{P}_n \mathbf{C}_n^{-1},$$

により定義。ただし

$$(4.16) \quad \mathbf{C}_n^{-1} = \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 \\ -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & 0 & \cdots \\ 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & -1 & 1 \end{pmatrix}_{n \times n},$$

$$(4.17) \quad \mathbf{P}_n = (p_{jk}^{(n)}), p_{jk}^{(n)} = \sqrt{\frac{2}{n + \frac{1}{2}}} \cos \left[\frac{2\pi}{2n+1} \left(k - \frac{1}{2} \right) \left(j - \frac{1}{2} \right) \right].$$

このときスペクトル分解により $\mathbf{C}_n^{-1} \mathbf{C}_n'^{-1} = \mathbf{P}_n \mathbf{D}_n \mathbf{P}_n' \mathbf{D}_n$ は対角行列、第 (k, k) 要素は $d_k = 2[1 - \cos(\pi(\frac{2k-1}{2n+1}))] = 4 \sin^2(\frac{\pi}{2}(\frac{2k-1}{2n+1}))$ ($k = 1, \dots, n$)。さらに \mathbf{K}_n -変換された系列 \mathbf{Z}_n に対するフィルタリング (あるいは Nishimura, et al. (2019) で提案された smoothing) 法を考える。 $m \times n$ 選択行列 $\mathbf{J}_m = (\mathbf{I}_m, \mathbf{O})$ として $n \times p$ 行列

$$(4.18) \quad \hat{\mathbf{X}}_n = \mathbf{C}_n \mathbf{P}_n' \mathbf{J}_m' \mathbf{J}_m \mathbf{P}_n \mathbf{C}_n^{-1} (\mathbf{Y}_n - \bar{\mathbf{Y}}_0)$$

ただし

$$(4.19) \quad \mathbf{Z}_n = \mathbf{P}_n \mathbf{C}_n^{-1} (\mathbf{Y}_n - \bar{\mathbf{Y}}_0)$$

$n \times n$ 変換

$$(4.20) \quad \mathbf{Q}_n = \mathbf{P}_n \mathbf{J}_m' \mathbf{J}_m \mathbf{P}_n$$

で与えられる。

次により一般の \mathbf{K}_n -変換された系列 \mathbf{Z}_n に対するフィルタリングを考える。例えば季節性は離散時間の季節周期 $s (> 1)$ と理解できる。 $m_2 \times [m_1 + m_2 + (n - m_1 - m_2)]$ 選択行列 $\mathbf{J}_{m_1, m_2, n} = (\mathbf{O}, \mathbf{I}_{m_2}, \mathbf{O})$, $n \times p$ 行列

$$(4.21) \quad \hat{\mathbf{W}}_n = \mathbf{C}_n \mathbf{P}_n' \mathbf{J}_{m_1, m_2, n}' \mathbf{J}_{m_1, m_2, n} \mathbf{P}_n \mathbf{C}_n^{-1} (\mathbf{Y}_n - \bar{\mathbf{Y}}_0)$$

$n \times n$ 行列

$$(4.22) \quad \mathbf{Q}_n = \mathbf{P}_n \mathbf{J}_{m_1, m_2, n}' \mathbf{J}_{m_1, m_2, n} \mathbf{P}_n.$$

を利用する。このとき (j, j') -要素 $\mathbf{Q}_n = \mathbf{P}_n \mathbf{J}_{m_1, m_2, n}' \mathbf{J}_{m_1, m_2, n} \mathbf{P}_n = (q_{j, j'})$ は

$$q_{j, j} = \frac{2m_2}{2n+1} + \frac{1}{2n+1} \left[\frac{\sin \frac{2(m_1+m_2)\pi}{2n+1} (2j-1) - \sin \frac{2(m_1)\pi}{2n+1} (2j-1)}{\sin \frac{\pi}{2n+1} (2j-1)} \right],$$

$$q_{i, j'} = \frac{1}{2n+1} \left[\frac{\sin \frac{2(m_1+1+m_2)\pi}{2n+1} (j+j'-1) - \sin \frac{2(m_1)\pi}{2n+1} (j+j'-1)}{\sin \frac{\pi}{2n+1} (j+j'-1)} + \frac{\sin \frac{2(m_1+m_2)\pi}{2n+1} (j-j') - \sin \frac{2(m_1)\pi}{2n+1} (j-j')}{\sin \frac{\pi}{2n+1} (j-j')} \right] \quad (j \neq j')$$

となる。

5 データの直交分解

記号 $\theta_{jk} = \frac{2\pi}{2n+1}(j - \frac{1}{2})(k - \frac{1}{2})$,

$$(5.23) \quad p_{jk}^{(n)} = \frac{1}{\sqrt{2n+1}}(e^{i\theta_{jk}} + e^{-i\theta_{jk}})$$

を利用して

$$(5.24) \quad \Delta_{\lambda \mathbf{z}}^{(n)}(\lambda_k^{(n)}) = \sum_{j=1}^n p_{jk}^{(n)} \mathbf{r}_j^{(n)}, \quad \mathbf{r}_j^{(n)} = \mathbf{y}_j - \mathbf{y}_{j-1},$$

と表現すると、変換系列 \mathbf{Z}_n はデータのある種の実フーリエ変換 $\Delta_{\lambda \mathbf{z}}^{(n)}(\lambda_k^{(n)})$ ($k = 1, \dots, n$) はデータの周波数 $\lambda_k^{(n)}$ ($= (k - 1/2)/(2n + 1)$), におけるフーリエ変換、データ直交増分過程である。

定理 5.1 : (離散時間) 確率過程 \mathbf{r}_j ($j = 1, \dots, n$) はエルゴード的定常過程、 $\mathbf{\Gamma}(h) = \mathcal{E}(\mathbf{r}_j \mathbf{r}_{j-h}')$ は

$$(5.25) \quad \sum_{h=0}^{\infty} \|\mathbf{\Gamma}(h)\| < \infty .$$

を満たす(有界性) ことを仮定する。

(i) $\Delta_{\lambda \mathbf{z}}^{(n)}(\lambda_k^{(n)}) = \sum_{j=1}^n p_{jk}^{(n)} \mathbf{r}_j^{(n)}$, $\mathbf{r}_j^{(n)}$ がエルゴード的定常過程で $\mathcal{E}[\mathbf{r}_j] = \mathbf{0}$, 対称化実スペクトル密度行列

$$(5.26) \quad \mathbf{f}_{SR}(\lambda) = \mathbf{\Gamma}(0) + \sum_{h=1}^{\infty} \cos(2\pi h\lambda)[\mathbf{\Gamma}(h) + \mathbf{\Gamma}(-h)],$$

は正定符号、有界性を仮定。 $\lambda_k^{(n)} \rightarrow s$, $\lambda_{k'}^{(n)} \rightarrow t$ ($0 < s < t < \frac{1}{2}$). $n \rightarrow \infty$ のとき

$$(5.27) \quad \begin{bmatrix} \Delta_{\lambda \mathbf{z}}^{(n)}(\lambda_k^{(n)}) \\ \Delta_{\lambda \mathbf{z}}^{(n)}(\lambda_{k'}^{(n)}) \end{bmatrix} \xrightarrow{w} N_{2p} \left[\mathbf{0}, \begin{bmatrix} \mathbf{f}_{SR}(s) & \mathbf{0} \\ \mathbf{0} & \mathbf{f}_{SR}(t) \end{bmatrix} \right].$$

(ii) 増分過程 $\mathbf{Z}_n(t) - \mathbf{Z}_n(s) = \frac{1}{\sqrt{n}} \sum_{k=[sn]}^{[tn]} \sum_{j=1}^n p_{jk}^{(n)} \mathbf{r}_j^{(n)}$ ($0 < s < t < 1$). $n \rightarrow \infty$ のとき

$$(5.28) \quad \mathbf{Z}_n(t) - \mathbf{Z}_n(s) \xrightarrow{w} N_p[\mathbf{0}, F_{SR}(t) - F_{SR}(s)],$$

ただし $F_{SR}(t)$ は $p \times p$ (対称化実) スペクトル分布行列

$$(5.29) \quad F_{SR}(t) = \int_0^t f_{SR}(\lambda) d\lambda .$$

で与えられる。

(実) スペクトル経験分布 (行列) は

$$(5.30) \quad \mathbf{F}_{SR,n}(t) = \frac{1}{n} \sum_{k \leq [2nt]} (\Delta_t \mathbf{z}_k^{(n)}(\lambda_k^{(n)})) (\Delta_t \mathbf{z}_k^{(n)}(\lambda_k^{(n)}))' .$$

となる。

KS(2017) における SIML 推定量は

$$(5.31) \quad \mathbf{G}_m = \frac{1}{m} \sum_{k=1}^m \Delta_0 \mathbf{z}_k^{(n)}(\lambda_k) \Delta_0 \mathbf{z}_k^{(n)}(\lambda_k)' .$$

より一般には

$$(5.32) \quad \mathbf{G}_m(t) = \frac{1}{m} \sum_{k=[2nt]-\frac{m}{2}+1}^{[2nt]+\frac{m}{2}} (\Delta_t \mathbf{z}_k^{(n)}(\lambda_k^{(n)})) (\Delta_t \mathbf{z}_k^{(n)}(\lambda_k^{(n)}))' .$$

定理 5.2: (離散時間) 確率過程 \mathbf{r}_j ($j = 1, \dots, n$) がエルゴード的定常過程、 $\mathbf{\Gamma}(h) = \mathcal{E}(\mathbf{r}_j \mathbf{r}_{j-h}')$ とする (有界性を仮定)。

(i) $m_n = [n^\alpha]$ ($0 < \alpha < 1$) とおくと、任意の $t \in (0, \frac{1}{2})$ に対し $n \rightarrow \infty$ のとき

$$(5.33) \quad \mathbf{G}_m(t) \xrightarrow{p} \mathbf{f}_{SR}(t)$$

(ii) $n \rightarrow \infty$ のとき

$$(5.34) \quad \mathbf{F}_{SR,m}(t) \xrightarrow{p} \mathbf{F}_{SR}(t) .$$

6 Filtering と Smoothing 法

伝統的時系列論ではスペクトル分布 F を持つ定常 (離散時間) ベクトル過程 \mathbf{r}_k^* に対して右連続な直交増分 (ベクトル, 複素数値) 過程 $\mathbf{z}^*(\lambda)$ ($-1/2 \leq \lambda \leq 1/2$) が存在し、

$$(6.35) \quad \mathbf{r}_k^* = \int_{(-1/2, 1/2]} e^{i2\pi k\nu} d\mathbf{z}^*(\nu) \quad (k = 1, \dots, n).$$

と表現されることが知られている。(Doob (1953), Hannan (1971), Brockwell and Davis (1990) など参照.)

本稿におけるような実ベクトル値をとる確率過程 (理論値) は

$$(6.36) \quad \mathbf{r}_k^{(u)} = \int_{(0, 1/2]} \cos(2\pi k\nu) h_u(\cos(2\pi k\nu)) d\mathbf{z}(\nu) \quad (k = 1, \dots, n)$$

($u = x, u = s$)、データからの直交過程による (離散時間) 平滑化 (smoothing) 値は

$$(6.37) \quad \mathbf{r}_k^{(n,u)} = \int_{(0,1/2]} \cos(2\pi k\nu) h_{u,n}(\cos(2\pi k\nu)) d\mathbf{z}_n(\nu) \quad (k = 1, \dots, n)$$

($u = x, u = s$) と表現される ($h(\cdot)$ はカーネル関数である)。

7 実例と議論

幾つかのマクロ系列を利用して本稿で考察しているフィルタリング法の妥当性が高いことが分かった。伝統的な時系列解析では定常過程にもとづくスペクトル密度の推定に高い関心を示していたが、その方法では非定常系列では上手くいかない理由が明らかになる。また例えば Box-Jenkin 法で推奨されている季節階差法には基本的な問題がある、ことなどがここでの分析から分かる。ここでは幾つかのデータ分析の例示にとどめる。

8 まとめと展望

- 多くのマクロ経済データにおいては非定常性、季節性、定常的な循環、観測ノイズなどが。例えば GDP 統計の季節調整・系列の合成 (内閣府)、複数のマクロ消費系列から総合指標の作成 (統計局) など、幾つかの問題がある。
- 季節調整法の X-12-ARIMA、DECOMP、VAR による共和分などはいずれも改善する必要がある。
- 本稿では多次元非定常時系列の分布にあまり依存しないマクロ SIML 法にもとづく新しいフィルタリング法を議論した。多次元非定常系列の実務的課題、例えば、トレンド抽出、季節調整、集計系列の合成、などへの応用可能性は広範と考えられる。
- 多期間の予測誤差を最小化するような最適なカーネル $h_{u,n}(\cdot)$ の選択を考察中、また離散時間確率過程 $\Delta_\lambda \mathbf{z}^{(n)}(\lambda_k^{(n)})$ は基準化すると右連続な直交増分 (ベクトル, 実数値) 過程 $\mathbf{z}(\lambda)$ ($0 \leq \lambda \leq 1/2$) に弱収束するが、filtering、smoothing の理論的誤差評価を考察中である。

References

- [1] Anderson, T.W. (1971), *The Statistical Analysis of Time Series*, John-Wiley.
- [2] Anderson, T.W. (1984), "Estimating Linear Statistical Relationships," *Annals of Statistics*, 12, 1-45.

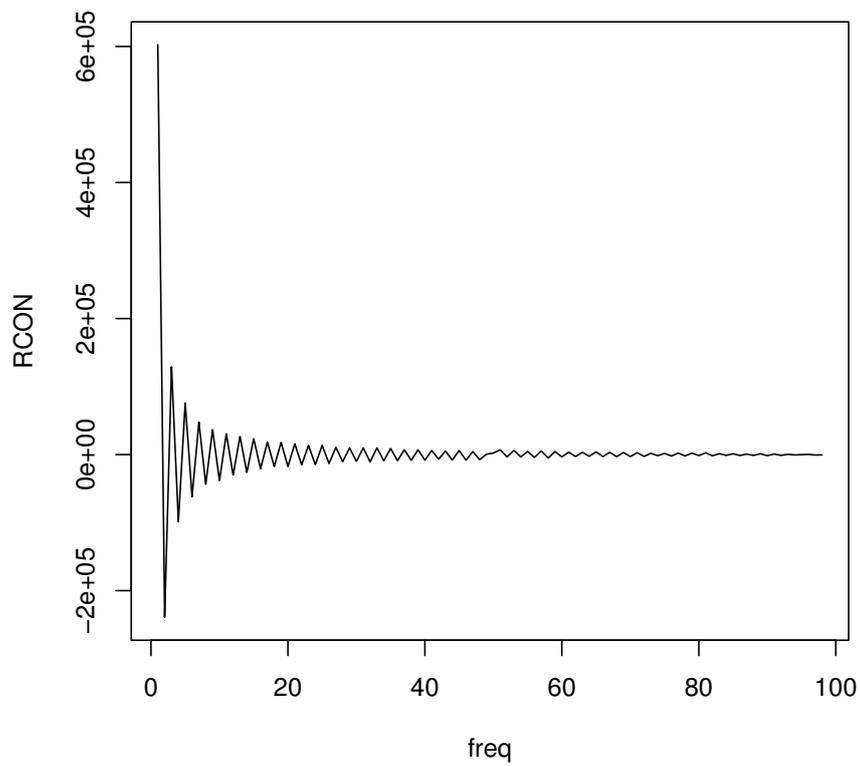


図 7.1 : 実質消費 (原系列) の z_n 過程

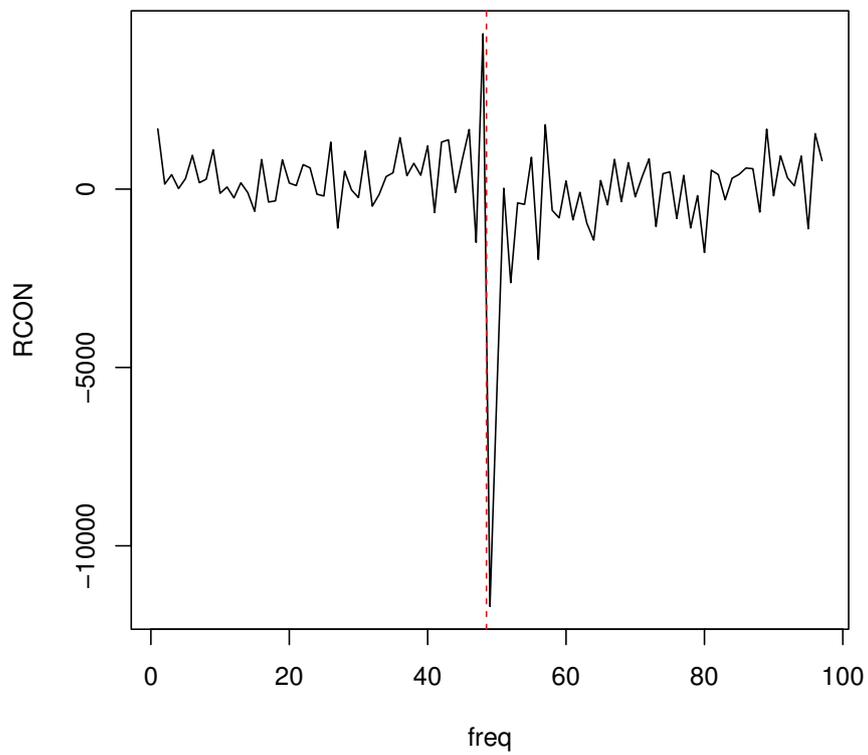


図 7.2 : 実質消費 (階差) の z_n 過程

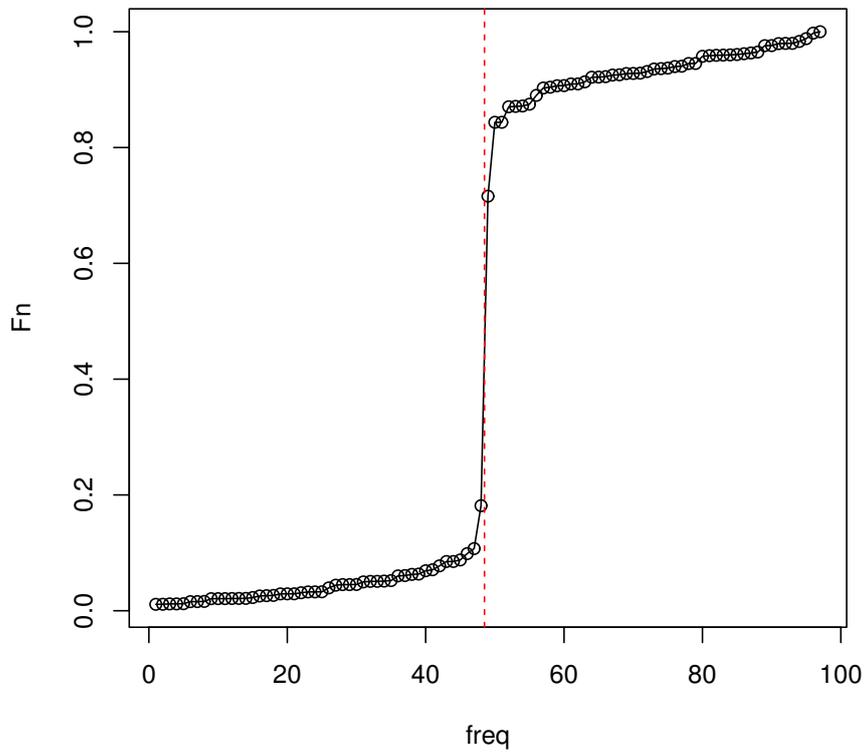


図 7.3 : 実質消費 (経験スペクトル分布)

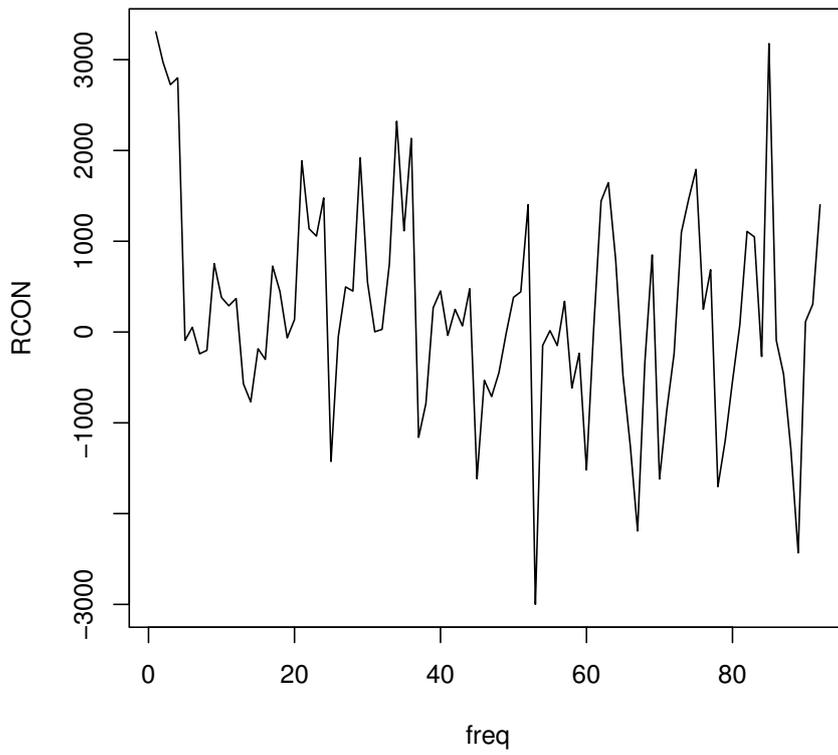


図 7.4 : 実質消費 (季節階差) の z_n 過程

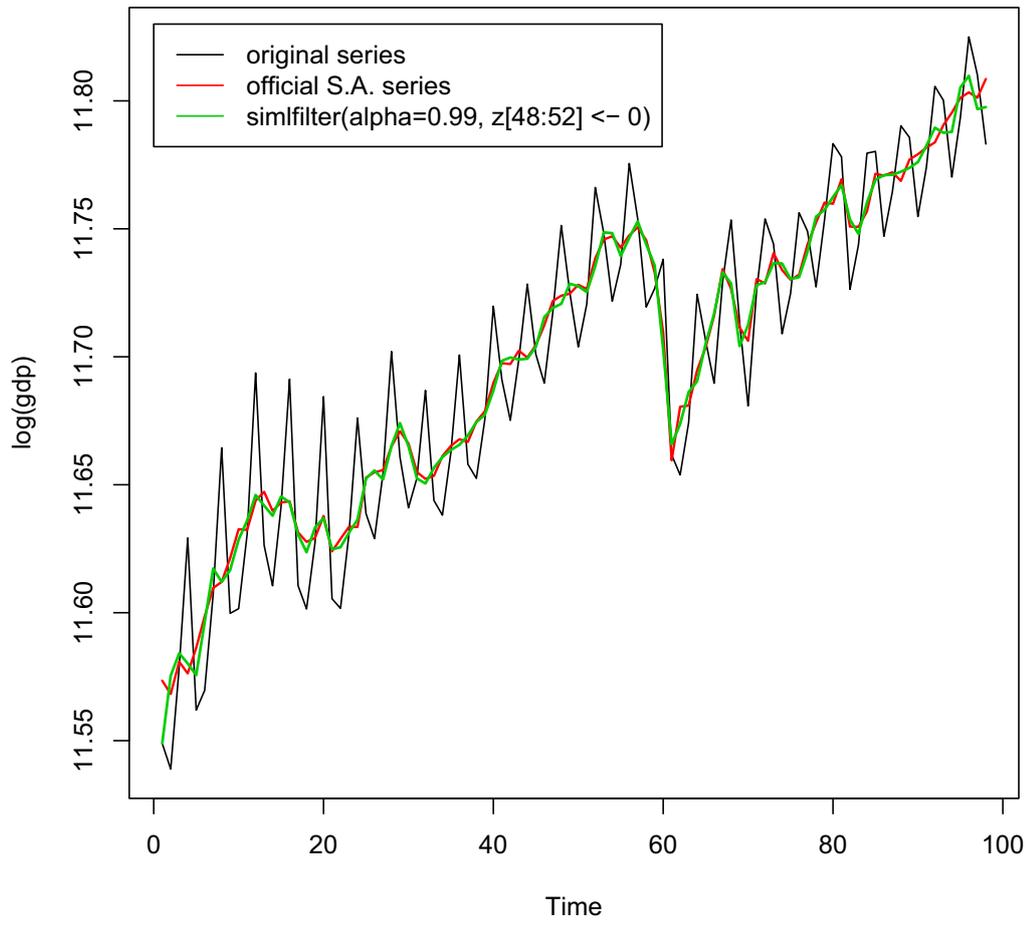


図 7.5 : 実質 GDP

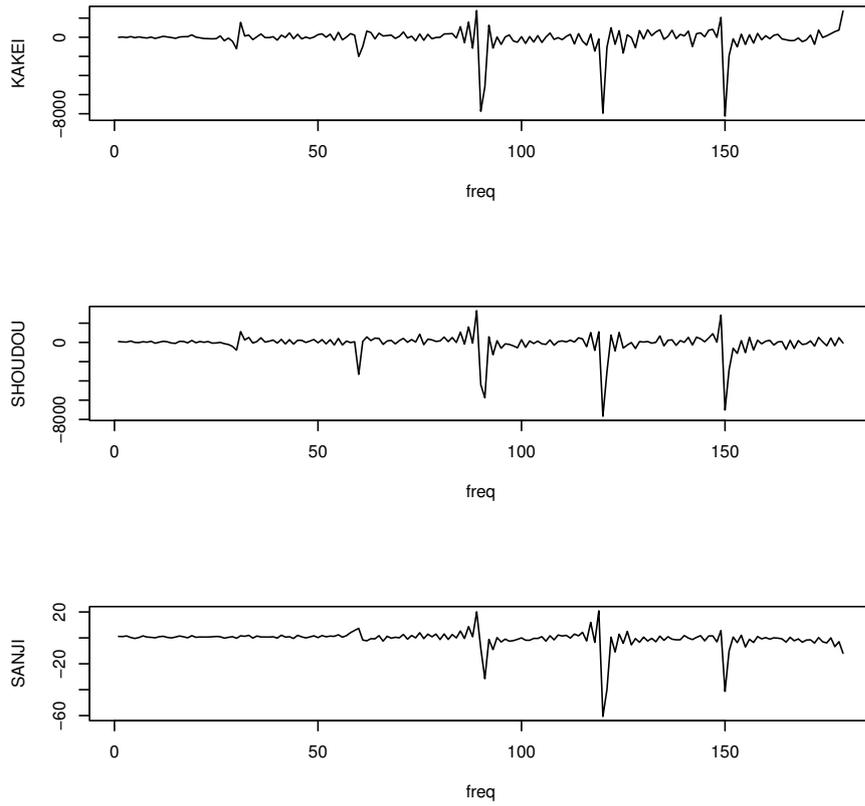


図 7.6 : 消費系列の z_n 過程

- [3] Engle, R. and C.W.J. Granger (1987), "Co-integration and Error Correction," *Econometrica*, Vol.55, 251-276.
- [4] Kitagawa, G. (2010), *Introduction to Time Series Analysis*, CRC Press.
- [5] Johansen, S. (1995), *Likelihood Based Inference in Cointegrated Vector Autoregressive Models*, Oxford UP.
- [6] Kunitomo, N. , Sato and D. Kurisu (2018), *Separating Information Maximum Likelihood Estimation for High Frequency Financial Data*, Springer.
- [7] Kunitomo, N. and S. Sato (2017), "Trend, Seasonality and Economic Time Series : the Non-stationary Errors-in-variables Models," SDS-4, MIMS, Meiji University, <http://www.mims.meiji.ac.jp/publications/2017-ds>.
- [8] Kunitomo, N., N. Awaya and D. Kurisu (2017), "Some Properties of Estimation Methods for Structural Relationships in Non-stationary Errors-in-Variables Models," SDS-3, MIMS, Meiji University.
- [9] Kunitomo, N. and S. Sato (2019), " A Robust-filtering Method for Noisy Non-Stationary Time Series," Unpublished Manuscript.
- [10] Müller, U. and M. Watson (2018), "Long-run Covariability," *Econometrica*, 86-3, 775-804.
- [11] Nishimura, K.G. S. Sato and A. Takahashi (2019), "Term Structure Models During the Global Financial Crisis: A Parsimonious Text Mining Approach," Asia-Pacific Financial Markets.

テキストデータからの情報抽出 を利用した時系列予測⁺

川崎能典

統計数理研究所

2019年1月31日

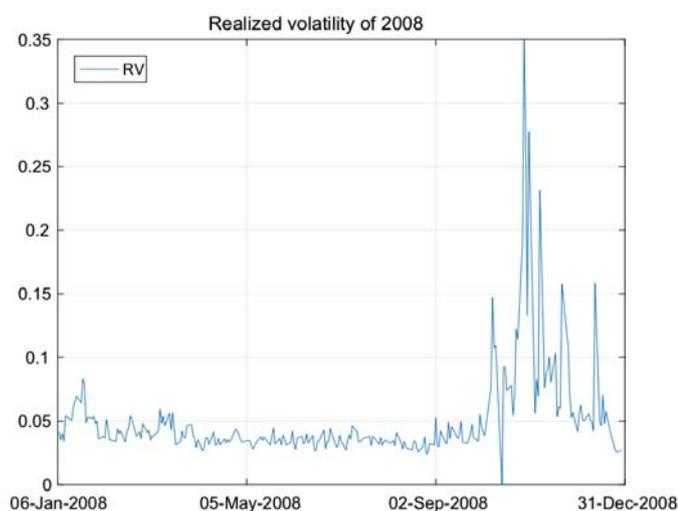
科研費基盤(A)「経済統計・政府統計の理論と応用」
研究集会@東京大学経済学部

⁺森本孝之氏 (関西学院大学理工学部)との共著

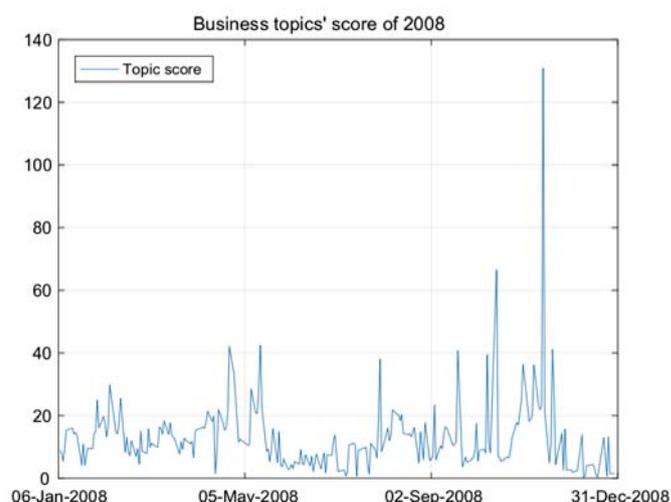
Motivation

- Counts of keywords sometimes helps
 - (Ex.) Google SVI (Search Volume Index)
- Have to find nice keywords.
- From news (text) data, we want to extract **topics** (defined by **distribution of words**) that may affect market sentiments
- Construct **topic score time series** SC_t
- Investigate if SC_t improves volatility forecasting

Illustration: topic score and realized volatility



Realized volatility estimated from high frequent data



Estimated topic score (one of 20 scores)

“Bag-of-Words” model

- We only focus on **word frequencies**, and neglect other information (order of words, dependency and so on).
- (Ex.) A document $D = \text{“It is fine today”}$ can be expressed $D = \{\text{“it”, “is”, “fine”, “today”}\}$.
- Usually we exclude so-called “stop words” such as “a”, “the”, “for”, etc.
- In this research, after morphological analysis, we choose **nouns only**, and remove numerals, suffixes, non-independent words, pronouns and symbols.

Latent Dirichlet Allocation Model

- A standard method for **topic modeling**
- Often abbreviated as LDA
- Distribution of words follows **multinomial distribution** (gives **likelihood**)
- **Dirichlet distribution** gives a **prior** distribution of words frequencies
- Word distribution ϕ_z characterizes a topic z , and each document d consists of many topics of which distribution θ_d .

Typical MCMC cycle for LDA

1. For $k \in \{1 \dots Z\}$:
 - Generate a word distribution for each topic, $\phi_z \sim \text{Dirichlet}(\phi|\beta)$,
2. For each document $d \in D$:
 - Generate a topic distribution for each document, $\theta_d \sim \text{Dirichlet}(\theta|\alpha)$,
 - for each word $w_i \in d$:
 - (a) Generate a topic from a multinomial distribution, $z_i \sim \text{Multinomial}(\theta_d)$,
 - (b) Generate a word from a multinomial distribution, $w_i \sim \text{Multinomial}(\phi_{z_i})$,

where α, β denote hyper parameters of a Dirichlet distribution.

This algorithm is for a single document. We do this day by day for Reuters news, and want to **ensure some continuity of topics** along time axis.

Multiscale Dynamic Topic Model

- Proposed by Iwata et al. (2010)
- Parameter $\phi_{t,z}$ of word distribution of topic z at time t has some time dependent structure.

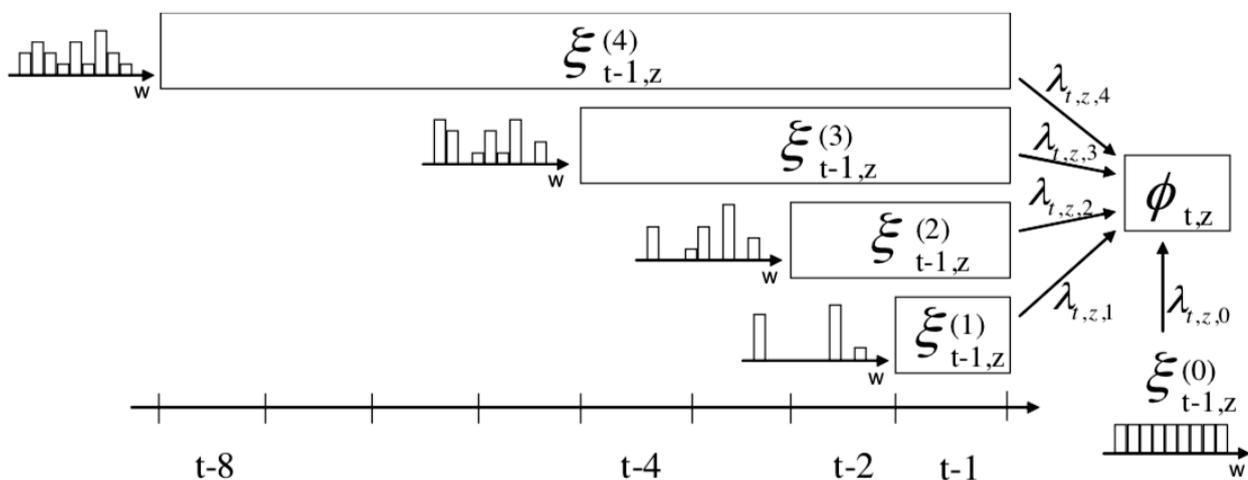
$$\phi_{t,z} \sim \text{Dirichlet} \left(\sum_{s=0}^S \lambda_{t,z,s} \hat{\omega}_{t-1,z}^{(s)} \right)$$

- $\hat{\omega}_{t-1,z}^{(s)}$: distribution of words over topic z with scale s at time $t - 1$
- $\lambda_{t,z,s}$: weight for scale s in topic z at time t

Dirichlet parameter in detail

- $\hat{\omega}_{t-1,z}^{(s)}$ indicated the word distribution (w.d.) from epoch $(t - 1) - 2^{s-1} + 1$
- If $S = 4$, s runs through 0,1,2,3,4.
- $s = 4 \rightarrow$ w.d. comes from $t - 8$ to $t - 1$
- $s = 3 \rightarrow$ w.d. comes from $t - 4$ to $t - 1$
- $s = 2 \rightarrow$ w.d. comes from $t - 2$ to $t - 1$
- $s = 1 \rightarrow$ word distribution comes at $t - 1$
- $s = 0$; assume uniform distribution for $\hat{\omega}_{t-1,z}^{(0)}$

Illustration of Multiscale Word Distribution



Word distributions are likely to be smoothed as the time scale becomes long.

Iwata, T. et al. (2000) Proceedings of 16th ACM SIGKDD, p.663-672.

MCMC cycle for MDTM

1. For each topic $k = 1, \dots, Z$:
 - (a) Draw word distribution of topic $\phi_{t,z} \sim \text{Dirichlet}(\sum_{s=0}^S \lambda_{t,z,s} \hat{\omega}_{t-1,z}^{(s)})$,
 - (b) Draw a hyper parameter of prior of topic distribution $\alpha_{t,z} \sim \text{Gamma}(\zeta \alpha_{t-1,z}, \zeta)$,
2. For each document $d = 1, \dots, D_t$:
 - (a) Draw topic proportions $\theta_{t,d} \sim \text{Dirichlet}(\alpha_t)$,
 - (b) For each word $n = 1, \dots, N_{t,d}$:
 - i. Draw topic $z_{t,d,n} \sim \text{Multinomial}(\theta_{t,d})$,
 - ii. Draw word $w_{t,d,n} \sim \text{Multinomial}(\phi_{t,z_{t,d,n}})$,

Weights $\{\lambda_{t,z,s}\}$ and hyperparameter $\alpha_{t-1,z}$ are estimated in an outer loop of this cycle, by stochastic EM algorithm and fixed point iteration method.

Construction of topic scores

- Topic scores are made up by estimated topic proportions $\theta_{t,j,i}$ (percentage of topic i included in j -th document at time t)

$$SC_t^i = \sum_{j=d}^{D_t} \theta_{t,j,i}$$

- SC_t^i : score for topic i at time t
- D_t : number of documents at time t
- $\theta_{t,j,i}$: i -th element of the topic distribution within j -th document at time t

Word distribution (June 2, 2008)

Topic 1		Topic 2	
Nikkei	0.109	Yen	0.208
Average	0.107	Present	0.069
Continued rise	0.043	Weekend	0.053
TSE	0.038	Temporary	0.040
Center	0.037	Higher quotation	0.038
Mutual fund	0.035	Feasible	0.036
Domestic	0.032	Session	0.034
Tokyo	0.032	Rebound	0.033
Opening	0.031	Late	0.030
Major	0.027	Holidays	0.024

- We consider 20 topics in all.
- Word distribution in Topic 1 and Topic 2
- Only top 10 words are shown

Data

- High frequent data of stock index (TOPIX)
- January 7th 2008 – December 28th 2012, $T = 1223$
- Generate 1 min return and calculate daily **realized volatility** (RV_t) and **realized quarticity** (RQ_t)
- News data taken from Reuter Japan's web site
- Language = Japanese
- 298,205 documents, 24,227 non-overlapping words excluding stop words

Forecasting models

- Heterogeneous Autoregressive (**HAR**) model
 - ✓ Baseline model, Corsi (2009)
- **HARQ** model, adding realized quarticity (RQ_{t-1}) in the coefficient of RV_{t-1}
 - ✓ Bollerslev, Patton and Quaedvleig (2016)
- HAR + topic score (**HAR-SC**)
- HARQ + topic score (**HARQ-SC**)
 - In our paper, we did **AR vs. AR-SC** and **ARQ vs. ARQ-SC** comparison which will be omitted here.

HAR vs. HAR-SC

- HAR-SC model is defined by

$$RV_t = \beta_0 + \beta_1 RV_{t-1} + \beta_2 RV_{t-1|t-5} + \beta_3 RV_{t-1|t-22} + \gamma SC_{t-1} + u_t$$

where $RV_{t-j|t-h} = \frac{1}{h+1-j} \sum_{i=j}^h RV_{t-i}$

- Omitting γSC_{t-1} reduces to HAR model

HARQ vs. HARQ-SC

- HARQ-SC model is defined by

$$RV_t = \beta_0 + (\beta_1 + \beta_{1Q} RQ_{t-1}^{1/2}) RV_{t-1} + \beta_2 RV_{t-1|t-5} + \beta_3 RV_{t-1|t-22} + \gamma SC_{t-1} + u_t$$

where $RV_{t-j|t-h} = \frac{1}{h+1-j} \sum_{i=j}^h RV_{t-i}$

- Omitting γSC_{t-1} reduces to HARQ model

Out-of-sample forecast losses

	HAR	HARQ	HAR-SC	HARQ-SC	
MSE (RW)	1.000	0.5562	0.9658	0.5369	$SC^{(11)}$
MSE (IW)	1.000	0.8408	0.9678	0.8175	$SC^{(11)}$
QLIKE (RW)	1.000	1.3781	0.9891	1.3439	$SC^{(3)}$
QLIKE (IW)	1.000	1.1529	0.9883	1.1292	$SC^{(18)}$

MSE: $L(RV_t, X_t) = (RV_t - X_t)^2$

QLIKE: $L(RV_t, X_t) = \frac{RV_t}{X_t} - \log\left(\frac{RV_t}{X_t}\right) - 1$

IW: increasing window in regression

RW: rolling regression with window size 400 days

Discussion

- SC_{t-1} is estimated based on information **up to $t - 1$** . So no cheat.
- We need some preliminary analysis to search **promising $SC^{(j)}$** .
- Forecasting performance depends on the choice of **error function** as well as the choice of **regression window**.

HAR-HARSC: another complication

- HAR-HARSC model is defined by

$$\begin{aligned}RV_t = & \beta_0 + \beta_1 RV_{t-1} + \beta_2 RV_{t-1|t-5} \\ & + \beta_3 RV_{t-1|t-22} + \gamma_1 SC_{t-1} \\ & + \gamma_2 SC_{t-1|t-5} + \gamma_3 SC_{t-1|t-22} + u_t\end{aligned}$$

where $SC_{t-j|t-h} = \frac{1}{h+1-j} \sum_{i=j}^h SC_{t-i}$

HARQ-HARQSC: yet another complication

- HARQ-HARQSC model is defined by

$$\begin{aligned}RV_t = & \beta_0 + (\beta_1 + \beta_{1Q} RQ_{t-1}^{1/2}) RV_{t-1} \\ & + \beta_2 RV_{t-1|t-5} + \beta_3 RV_{t-1|t-22} \\ & + \gamma_1 SC_{t-1} + \gamma_2 SC_{t-1|t-5} \\ & + \gamma_3 SC_{t-1|t-22} + u_t\end{aligned}$$

where $SC_{t-j|t-h} = \frac{1}{h+1-j} \sum_{i=j}^h SC_{t-i}$

Reference

- Morimoto, T. and Kawasaki, Y. (2017). Forecasting Financial Market Volatility Using a Dynamic Topic Model, *Asia-Pacific Financial Markets*, Vol. 24, pp. 149-167. DOI: 10.1007/s10690-017-9228-z
- And references therein



Statistical Inference in Regression Models with Possibly Integrated Processes (未定稿)

千木良 弘朗 山本 拓

東北大学 一橋大学

研究集会「経済統計・政府統計の理論と応用からの提言」

2019年1月31日 東京大学

- 0/26 -

1 はじめに

□モチベーション

フィリップス曲線の推定・検定:

$$inf_t = \alpha + \beta unem_t + \varepsilon_t, t = 1, \dots, T$$

データ:

$$\left\{ \begin{array}{l} inf_t \cdots \text{percentage change in CPI} \\ unem_t \cdots \text{civilian unemployment rate, \%} \\ \text{年次データで1948} \sim \text{1996年、} T = 49 \end{array} \right.$$

OLS推定の結果 (括弧内はt-値) :

$$\Rightarrow inf_t = 1.42 + 0.47 unem_t, \bar{R}^2 = 0.03$$

(0.85) (1.65)

- 1/26 -

⇒ 負のはずの $\hat{\beta}$ が正で、しかもあまり有意でない

何かおかしい?

ε_t の系列相関?

ε_t の AR(1) 係数は 0.57、t-値は 5.03 で強く有意

$unem_t$ の非定常性?

AR(1) 係数は 0.73、有意水準 5% の ADF 検定で H_0 採択

inf_t の非定常性?

AR(1) 係数は 0.67、有意水準 5% の ADF 検定で H_0 棄却

inf_t と $unem_t$ で和分の次数が違う? ADF 検定の低検出力のせい?

⇒ 非定常性の問題に、単位根・共和分検定せずに対処したい

- 2/26 -

□ フィリップス曲線の成立条件

$$inf_t = \alpha + \beta unem_t + \varepsilon_t$$

$\beta < 0$ の場合

		ε_t の次数	
		$I(1)$	$I(0)$
$unem_t$ の次数	$I(1)$	$inf_t \sim \begin{cases} I(1) : \text{見せかけの回帰} \\ I(0) : - \\ \text{不成立} \end{cases}$	$inf_t \sim \begin{cases} I(1) : \text{共和分} \\ I(0) : - \\ \text{成立} \end{cases}$
	$I(0)$	$inf_t \sim \begin{cases} I(1) : \text{和分の次数が異なる} \\ I(0) : - \\ \text{不成立} \end{cases}$	$inf_t \sim \begin{cases} I(1) : - \\ I(0) : \text{定常で関係あり} \\ \text{成立} \end{cases}$

- 3/26 -

$\beta = 0$ の場合

		ε_t の次数	
		$I(1)$	$I(0)$
$unem_t$ の次数	$I(1)$	$inf_t \sim \begin{cases} I(1) : \text{見せかけの回帰} \\ I(0) : - \\ \text{不成立} \end{cases}$	$inf_t \sim \begin{cases} I(1) : - \\ I(0) : \text{和分の次数が異なる} \\ \text{不成立} \end{cases}$
	$I(0)$	$inf_t \sim \begin{cases} I(1) : \text{和分の次数が異なる} \\ I(0) : - \\ \text{不成立} \end{cases}$	$inf_t \sim \begin{cases} I(1) : - \\ I(0) : \text{定常で関係無い} \\ \text{不成立} \end{cases}$

⇒ H_0 : 不成立 H_1 : 成立なる検定問題を上手く作り、 H_0 の棄却を以ってフィリップス曲線を立証

- 4/26 -

□ 階差を取ればいいのか?

$$inf_t = \alpha + \beta unem_t + \varepsilon_t, \varepsilon_t = \rho \varepsilon_{t-1} + u_t, u_t \sim I(0)$$

$(1 - \rho L)$ を

掛けて整理

$$\rightarrow inf_t = (1 - \rho)\alpha + \beta unem_t - \beta \rho unem_{t-1} + \rho inf_{t-1} + u_t$$

$\rho = 1$ と仮定

$$\rightarrow \Delta inf_t = \beta \Delta unem_t + u_t$$

⇒ 階差モデルは、 $\rho = 1$ なら正しいが、 $\rho < 1$ では mis-specification

⇒ もし $\rho = 1$ で β が非ゼロでも、 $\varepsilon_t \sim I(1)$ の意味で「見せかけの回帰」

- 5/26 -

$$inf_t = \alpha + \beta unem_t + \varepsilon_t, \varepsilon_t = \rho \varepsilon_{t-1} + u_t, u_t \sim I(0), \\ inf_t \sim I(1), unem_t \sim I(1), \rho < 1 \dots \text{共和分}$$

$$\rightarrow \Delta inf_t = \beta \Delta unem_t + u_t$$

⇒ 階差モデルは、共和分時には misspecification

- 6/26 -

$$inf_t = \alpha + \beta unem_t + \varepsilon_t, \varepsilon_t = \rho \varepsilon_{t-1} + u_t, u_t \sim I(0), \\ \begin{cases} inf_t \sim I(1), unem_t \sim I(0) \\ inf_t \sim I(0), unem_t \sim I(1) \end{cases}$$

$$\rightarrow \Delta inf_t = \beta \Delta unem_t + u_t$$

⇒ 階差を取ると、 inf_t と $unem_t$ で和分の次数が異なっているとしても「フィリップス曲線が成立」となりかねない

- 7/26 -

$\left\{ \begin{array}{l} inf_t \text{と} unem_t \text{の関係式} \cdots \text{フィリップス曲線} \\ \Delta inf_t \text{と} \Delta unem_t \text{の関係式} \cdots \text{expectations augmented} \\ (inf_t - inf_t^e \text{と} unem_t - \mu_0) \cdots \text{フィリップス曲線} \end{array} \right.$

⇒ 階差を取ると経済学的インプリケーションが異なる

- 8/26 -

□ 分析の目的

levelの単純回帰モデル（共和分回帰の最も簡単なケース）

$$y_t = \alpha + \beta x_t + \varepsilon_t, \varepsilon_t = \rho \varepsilon_{t-1} + u_t, u_t \sim i.i.d.(0, \sigma_u^2),$$

$$-1 < \rho \leq 1 \quad (1)$$

$$x_t = c + \xi_t, \xi_t = \phi \xi_{t-1} + v_t, v_t \sim i.i.d.(0, \sigma_v^2),$$

$$-1 < \phi \leq 1, u_t \perp v_s, \forall t, s$$

単位根検定・共和分検定・ β の有意性検定の3ステップ

⇒

- { 単位根・共和分検定が煩雑
- { 単位根・共和分検定の誤りにより有意性検定が歪む
- 単位根・共和分検定無しで1ステップで有意性検定

- 9/26 -

発表の構成

2. アイデア
3. 検定統計量の漸近分布
4. モンテカルロ実験
5. まとめと今後の課題

- 10/26 -

2 アイデア

(1) 式に $(1 - \rho L)$ を掛けて整理

$$y_t = (1 - \rho)\alpha + \beta x_t - \beta \rho x_{t-1} + \rho y_{t-1} + u_t, \quad u_t \sim i.i.d.(0, \sigma_u^2) \quad (2)$$

⇒ (2) 式で次の検定問題考える。

$$H_0 : \beta + (-\beta\rho) = 0 \quad H_1 : \beta + (-\beta\rho) \neq 0 \quad (3)$$

→ この H_0 は、次の表の ○ で正しく、× で正しくない

- 11/26 -

$$H_0 : \beta + (-\beta\rho) = 0 \quad H_1 : \beta + (-\beta\rho) \neq 0$$

$\beta \neq 0$ の場合

		ρ (ε_t の次数)	
		1 ($I(1)$)	-1 ~ 1 ($I(0)$)
ϕ (x_t の 次数)	1 ($I(1)$)	見せかけの回帰 ○	共和分 ×
	-1 ~ 1 ($I(0)$)	和分の次数が異なる ○	定常で関係あり ×

$\beta = 0$ の場合

		ρ (ε_t の次数)	
		1 ($I(1)$)	-1 ~ 1 ($I(0)$)
ϕ (x_t の 次数)	1 ($I(1)$)	見せかけの回帰 ○	和分の次数が異なる ○
	-1 ~ 1 ($I(0)$)	和分の次数が異なる ○	定常で関係無い ○

- 12/26 -

⇒ $\left\{ \begin{array}{l} \text{○は「}x\text{と}y\text{に経済学的に意味のある関係が無い場合」} \\ \text{×は「}x\text{と}y\text{に経済学的に意味のある関係がある場合」} \end{array} \right.$

- 13/26 -

→ (2) 式を OLS で推定し、t-検定統計量

$$t = \frac{\hat{\beta} + \widehat{-\beta\rho}}{\sqrt{\widehat{Var}(\hat{\beta}) + 2\widehat{Cov}(\hat{\beta}, \widehat{-\beta\rho}) + \widehat{Var}(\widehat{-\beta\rho})}} \quad (4)$$

で H_0 を検定

⇒ $\begin{cases} \times \text{に検出力を持つ} \\ \bigcirc \text{には検出力を持たない} \end{cases}$

⇒ 予備的な単位根・共和分検定無しで \times と \bigcirc を判別できる

⇒ 単位根・共和分検定は不要で1ステップで「経済学的に意味のある関係」を検出できる

- 14/26 -

3 検定統計量の漸近分布

(4) 式の t の分布は?

- $\rho = \phi = 1$ では non-standard な分布 (Hamilton (1994))
- ρ と ϕ の一方だけが 1 でも non-standard (モンテカルロ実験)

- 15/26 -

□ Toda and Yamamoto (1995) の LA (lag-augment) 法

(2) 式に、 x_{t-2} と y_{t-2} という本来は不要なラグ変数を追加

$$y_t = (1 - \rho)\alpha + \beta x_t - \beta\rho x_{t-1} + \rho y_{t-1} + 0 \times x_{t-2} + 0 \times y_{t-2} + u_t, u_t \sim i.i.d.(0, \sigma_u^2) \quad (5)$$

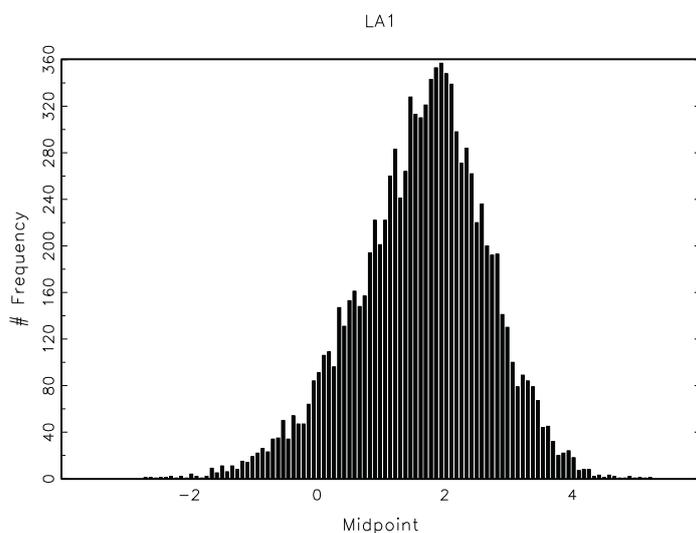
命題 (まだ証明はしていない)

(5) 式の OLS に基づく t-検定統計量 (4) は、(3) の H_0 の下で、 $\rho = 1$ and/or $\phi = 1$ でも、漸近的に標準正規分布に従う。

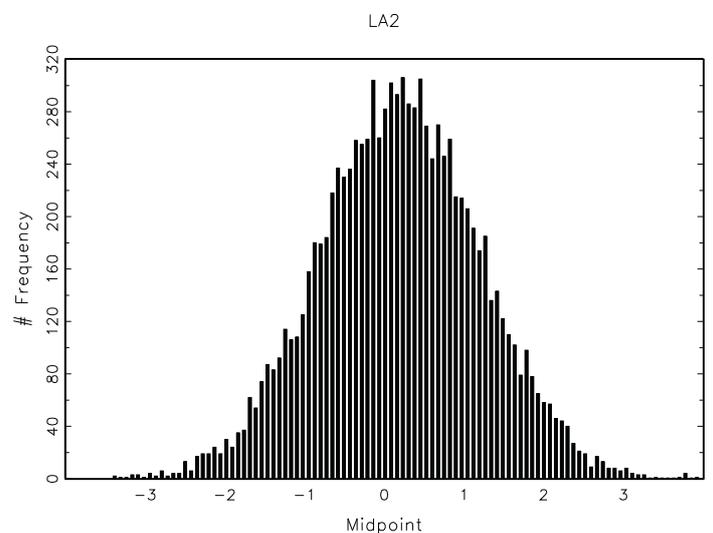
- 16/26 -

$\beta = 2$ 、 $\rho = 1$ 、 $\phi = 1$ での t の分布 ($T = 1000$ 、 $R = 10000$)

LA 無し: (2) 式



LA あり: (5) 式



- 17/26 -

4 モンテカルロ実験

□ 実験の設定

DGP:

$$y_t = 10 + \beta x_t + \varepsilon_t, \varepsilon_t = \rho \varepsilon_{t-1} + u_t, u_t \sim NID(0, 1), t = 1, \dots, T$$
$$x_t = 5 + \xi_t, \xi_t = \phi \xi_{t-1} + v_t, v_t \sim NID(0, 1), u_t \perp v_s, \forall t, s$$

$\beta = \{0, 2\}$ 、 $\rho = \{1, 0.9\}$ 、 $\phi = \{1, 0.9\}$ 、 $T = \{50, 1000\}$ 、
全ての検定のnominal sizeは5%、実験回数1万回

通常法: x にDF検定・ y にPP検定・ y on x の回帰でEG検定 or
コクラン-オーカット法

LA法: (5)式のOLSに基づき、(3)の H_0 をt-検定統計量(4)で
 $N(0, 1)$ で検定

- 18/26 -

⇒

通常法とLA法の「成功率」を比較する。「成功率」の定義
は脚注1~8

- 19/26 -

■ 実験結果

		$\beta = 2$ の場合					
		ρ (ε_t の次数)					
		1			0.9		
ϕ (x_t の 次数)	1		$T = 50$	$T = 1000$		$T = 50$	$T = 1000$
		通常法 ¹	0.86	0.87	通常法 ⁶	0.10	0.92
	LA法 ²	0.85	0.95	LA法 ⁷	0.25	0.83	
	(見せかけの回帰)			(共和分)			
0.9			$T = 50$	$T = 1000$		$T = 50$	$T = 1000$
		通常法 ³	0.08	0.49	通常法 ⁸	0.07	1.00
	LA法 ²	0.85	0.95	LA法 ⁷	0.23	0.83	
	(和分の次数が異なる)			(定常で関係あり)			

¹ x のDF検定で採択・ y のPP検定で採択・EG検定で採択で「成功」と判定。

² H_0 を採択で「成功」と判定。LA法の1-sizeのこと。

³ x のDF検定で棄却・ y のPP検定で採択で「成功」と判定。

⁴ x のDF検定で採択・ y のPP検定で棄却で「成功」と判定。

⁵ x のDF検定で棄却・ y のPP検定で棄却・コ克蘭-オーカット法で $\beta = 0$ を採択で「成功」と判定。

⁶ x のDF検定で採択・ y のPP検定で採択・EG検定で棄却で「成功」と判定。

⁷ H_0 を棄却で「成功」と判定。LA法のpowerのこと。

⁸ x のDF検定で棄却・ y のPP検定で棄却・コ克蘭-オーカット法で $\beta = 0$ を棄却で「成功」と判定。

- 20/26 -

		$\beta = 0$ の場合					
		ρ (ε_t の次数)					
		1			0.9		
ϕ (x_t の 次数)	1		$T = 50$	$T = 1000$		$T = 50$	$T = 1000$
		通常法 ¹	0.87	0.87	通常法 ⁴	0.14	0.95
	LA法 ²	0.91	0.95	LA法 ²	0.92	0.95	
	(見せかけの回帰)			(和分の次数が異なる)			
0.9			$T = 50$	$T = 1000$		$T = 50$	$T = 1000$
		通常法 ³	0.11	0.95	通常法 ⁵	0.02	0.95
	LA法 ²	0.92	0.95	LA法 ²	0.91	0.95	
	(和分の次数が異なる)			(定常で関係無し)			

⇒ LA法は、0.25では通常法より「成功率」が低いですが、他ではほぼ同等か上回る

- 21/26 -

□ モンテカルロ実験から解る「通常法」の大きな欠点

実験結果の のように、 $T = 1000$ でも「成功率」が49%と極めて低い

$$y_t = 10 + 2x_t + \varepsilon_t, \varepsilon_t = \varepsilon_{t-1} + u_t \rightarrow y_t \sim I(1)$$

→ PP検定で H_0 を51%reject (size distortionが46%)

$$x_t = 5 + \xi_t, \xi_t = 0.9\xi_{t-1} + v_t \rightarrow x_t \sim I(0)$$

→ DF検定で H_0 を100%reject (powerは100%)

⇒ なぜ y_t へのPP検定が極端に悪いのか?

- 22/26 -

(2) 式より、

$$y_t = (1 - 1)10 + 2x_t - 2 \times 1x_{t-1} + y_{t-1} + u_t$$
$$= 2\Delta x_t + y_{t-1} + u_t = 2\Delta x_t + y_{t-1} + u_t$$

⇒ PP検定における誤差項は $2\Delta x_t + u_t$ であり、 Δx_t には過剰階差によりMA単位根が発生

↓

⇒ $2\Delta x_t + u_t$ のMAの根は約**0.956** (u_t が混じるため完全なMA単位根にはならない)

⇒ PP検定はMA単位根が無いのが前提のため、極端にパフォーマンスが悪化

- 23/26 -

5 まとめと今後の課題

□ 本稿の貢献

元のモデル: $y_t = \alpha + \beta x_t + \text{誤差}$

↓2期分LA

$y_t = \beta_1 + \beta_2 x_t + \beta_3 x_{t-1} + \beta_4 y_{t-1} + \beta_5 x_{t-2} + \beta_6 y_{t-2} + \text{誤差}$

↓

$H_0 : \beta_2 + \beta_3 = 0$ をOLSでt-検定

⇒

誤差項系列相関・単位根・共和分・見せかけの回帰を気にせずに「元のモデルの β が有意か」解る

- 24/26 -

□ 課題

・LA法では、 H_0 を棄却した際に「 x と y が共和分」なのか「 x も y も定常で関係がある」のか解らない

・実験結果のように、LA法はpowerの上昇が鈍い

・共和分回帰の枠組みでは $u_t \sim MA(\infty)$ とできるが、LA法では $u_t \sim AR(p)$ としてADFタイプの拡張しかできない⁹

・いわゆる共和分回帰の枠組みでは u_t と v_s に相関を許せる(FM-OLS)が、LA法では相関を許すことはできない¹⁰

・(1)式の重回帰への拡張については、LA法は計量経済学の方法論的には問題無く行えるだろうが、 H_0 を棄却した際の経済学的解釈が難しくなる可能性がある

⁹ (5)式の誤差項 u_t に $MA(\infty)$ の系列相関が残ると、 y_{t-1} や y_{t-2} が説明変数にあるため $I(0)$ の世界では一致推定ができなくなるということ。

¹⁰ u_t と v_s の相関は(5)式での説明変数 x_t と誤差項 u_t ということになり、 $I(0)$ の世界では一致推定ができなくなるということ。

- 25/26 -

□ フィリップス曲線の推定・検定

$$inf_t = \alpha + \beta unem_t + \varepsilon_t, t = 1948, \dots, 1996$$

	α	β	\bar{R}^2
推定値	1.42	0.47	0.03
t-値	0.85	1.65	

(単位根検定では $unem_t \sim I(1)$ ・
 $inf_t \sim I(0)$)

$$inf_t = (1 - \rho)\alpha + \beta \times unem_t - \beta\rho \times unem_{t-1} + \rho \times inf_{t-1} + 0 \times unem_{t-2} + 0 \times inf_{t-2} + u_t$$

	$(1 - \rho)\alpha$	β	$-\beta\rho$	ρ	0	0	\bar{R}^2
推定値	1.67	-0.44	-0.22	0.84	0.51	0.00	0.58
t-値	1.35	-0.35	-0.57	6.44	-	-	

$H_0 : \beta + (-\beta\rho) = 0$ に対する t-検定統計量 $\dots -1.93$

Covariance Structure Analysis of Panel Regression Models

Kazuhiko Hayakawa
(joint with Takashi Yamagata(U. of York))

Preliminary and incomplete

January 31 2019

Outline

1. Introduction
2. Covariance Structure Analysis of Panel Regression Model
- (3. Extension)
4. Monte Carlo simulation
5. Conclusion

1. Introduction

Introduction

- Several panel data models have been proposed in the literature.
- One of the most recent models is the panel regression models with unobserved factor structure or interactive fixed effects.
- The literature can be divided into two categories:
 - Small T and large N
 - GMM** Ahn, Lee and Schmidt (2001, 2013), Robertson and Sarafidis (2015), Hayakawa (2012), etc.
 - ML** Hayakawa, Pesaran and Smith (2018), Bai (2013a,b), **this paper** etc.
 - Large T and large N
 - Pesaran (2006), Bai (2009), Moon and Weidner (2015), Bai (2013a,b), Chudik and Pesaran (2015) etc.
- This paper proposes the maximum likelihood(**ML**) and minimum distance(**MD**) estimators for **small** T and large N panel models with interactive fixed effects based on **covariance structure analysis (CSA)**.

Introduction

- In the literature of panel data analysis, covariance structure analysis has been used in estimation of income process (e.g. Lillard and Willis, 1978; Abowd and Card, 1989)
- However, recently, the covariance structure analysis has been applied to estimation of **panel regression models**.
- Previous studies are
 - Bollen and Brand (2010) :
 - demonstrate that FE/RE panel data model can be estimated by CSA.
 - small T and large N , no theoretical results.
 - Bai (2013a,b):
 - (mainly) panel AR(1) model with standard/interactive FE,
 - (mainly) large T and large N
 - Moral-Benito (2013), Moral-Benito, Allison and Williams (2019) :
 - Dynamic panel model with predetermined regressor and standard FE,
 - small T and large N

Introduction

- This paper proposes a unified approach to estimate for **small T** and large N linear panel regression models that allow
 - **static / dynamic** model
 - **endogenous / predetermined / strictly exogenous** regressors
 - **(known) standard / (unknown) interactive** fixed effects
- In this talk, we mainly consider **static/dynamic** panel model with **(unknown) interactive** fixed effects and **endogenous/strictly exogenous** regressors.
- The novelty of our CSA approach is that we do not need to use instrumental variables even in the presence of endogenous regressor since we use ML and MD.
- Thus, we are free from the weak/many instruments problem.

2. Covariance Structure Analysis of Panel Regression Model

Panel Regression Model

- Let us consider the following model

$$y_{it} = \mu_{y_t} + \beta' \mathbf{x}_{it} + \varepsilon_{it}, \quad (i = 1, \dots, N; t = 1, \dots, T) \quad (1)$$

$$\varepsilon_{it} = \mathbf{f}'_t \boldsymbol{\eta}_i + v_{it} \quad (2)$$

where β and \mathbf{x}_{it} are $K_x \times 1$, unobserved non-random factor \mathbf{f}_t and its random loading $\boldsymbol{\eta}_i$ are both $m \times 1$. $\mu_{y,t}$ denotes the time effects.

- If $\mathbf{f}_t = 1$ with $m = 1$, this model becomes the standard panel data model.
- Thus, this model can be seen as a generalization of standard panel data model.
- When y_{it} is wage, \mathbf{f}_t is business condition and $\boldsymbol{\eta}_i$ is unobserved ability of individual i , this model allows time-varying effect of ability to wage that is driven by business condition.

Panel Regression Model

- For the process of \mathbf{x}_{it} , we assume

$$\mathbf{x}_{it} = \boldsymbol{\mu}_{x_t} + \boldsymbol{\xi}_{x_t,i}, \quad (t = 1, 2, \dots, T) \quad (10)$$

where $E(\mathbf{x}_{it}) = \boldsymbol{\mu}_{x_t}$ and $\boldsymbol{\xi}_{x_t,i}$ is a continuous variable which is independent over i .

- We also let

$$\text{Cov}(\mathbf{x}_{it}, \boldsymbol{\eta}_i) = \boldsymbol{\Sigma}_{x_t \eta}, \quad t = 1, \dots, T \quad (11)$$

$$\text{Cov}(\mathbf{x}_{is}, \mathbf{x}_{it}) = E(\boldsymbol{\xi}_{x_s,i} \boldsymbol{\xi}'_{x_t,i}) = \boldsymbol{\Sigma}_{x_s x_t}, \quad s, t = 1, \dots, T \quad (12)$$

- Note that \mathbf{x}_{it} is correlated with unobserved individual effects $\boldsymbol{\eta}_i$ in an unrestricted way as in FE model.

Panel Regression Model

- For the correlation between \mathbf{x}_{is} and v_{it} , we consider two cases:
 - Strictly exogenous

$$\text{Cov}(\mathbf{x}_{is}, v_{it}) = \mathbf{0}, \quad s, t = 1, \dots, T \quad (13)$$

- Endogenous

$$\text{Cov}(\mathbf{x}_{is}, v_{it}) = \begin{cases} \mathbf{0} & s < t \\ \boldsymbol{\sigma}_{x_s v_t} & s \geq t \end{cases}, \quad s, t = 1, \dots, T \quad (14)$$

Panel Regression Model

- Then, we have

$$\mathbf{z}_i = \mathbf{B}^{-1}\boldsymbol{\mu} + \mathbf{B}^{-1}\mathbf{u}_i \quad (20)$$

- The covariance matrix of \mathbf{z}_i can be written as

$$\begin{aligned} \text{Var}(\mathbf{z}_i) &= \boldsymbol{\Sigma}_{zz}(\boldsymbol{\theta}_{all}) = \mathbf{B}^{-1}\boldsymbol{\Sigma}_{uu}\mathbf{B}^{-1'} = \begin{bmatrix} \boldsymbol{\Sigma}_{yy} & \boldsymbol{\Sigma}_{yx} \\ \boldsymbol{\Sigma}_{xy} & \boldsymbol{\Sigma}_{xx} \end{bmatrix} \\ &= \begin{bmatrix} \boldsymbol{\Sigma}_{\varepsilon\varepsilon} + \mathbf{B}_{12}\boldsymbol{\Sigma}_{xx}\mathbf{B}'_{12} - \boldsymbol{\Sigma}'_{x\varepsilon}\mathbf{B}'_{12} - \mathbf{B}_{12}\boldsymbol{\Sigma}_{x\varepsilon} & \boldsymbol{\Sigma}'_{x\varepsilon} - \mathbf{B}_{12}\boldsymbol{\Sigma}_{xx} \\ \boldsymbol{\Sigma}_{x\varepsilon} - \boldsymbol{\Sigma}_{xx}\mathbf{B}'_{12} & \boldsymbol{\Sigma}_{xx} \end{bmatrix} \end{aligned} \quad (21)$$

where

$$\boldsymbol{\Sigma}_{uu} = \text{Var}(\mathbf{u}_i) = \begin{bmatrix} \text{Var}(\varepsilon_i) & \text{Cov}(\mathbf{x}_i, \varepsilon_i)' \\ \text{Cov}(\mathbf{x}_i, \varepsilon_i) & \text{Var}(\mathbf{x}_i) \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Sigma}_{\varepsilon\varepsilon} & \boldsymbol{\Sigma}'_{x\varepsilon} \\ \boldsymbol{\Sigma}_{x\varepsilon} & \boldsymbol{\Sigma}_{xx} \end{bmatrix} \quad (22)$$

$$\begin{matrix} \boldsymbol{\Sigma}_{\varepsilon\varepsilon} \\ (T \times T) \end{matrix} = \boldsymbol{\Sigma}_{vv} + \mathbf{F}\boldsymbol{\Sigma}_{\eta\eta}\mathbf{F}' \quad (23)$$

$$\begin{matrix} \boldsymbol{\Sigma}_{vv} \\ (T \times T) \end{matrix} = \text{diag}[\sigma_{v_1}^2, \dots, \sigma_{v_T}^2], \quad \mathbf{F} = (\mathbf{f}_1, \dots, \mathbf{f}_T)' \quad (24)$$



Panel Regression Model

$$\begin{matrix} \boldsymbol{\Sigma}_{x\varepsilon} \\ (p_2 \times T) \end{matrix} = \text{Cov}(\mathbf{x}_i, \varepsilon_i) = \boldsymbol{\Sigma}_{xv} + \boldsymbol{\Sigma}_{x\eta}\mathbf{F}' \quad (25)$$

$$\begin{matrix} \boldsymbol{\Sigma}_{xv} \\ (p_2 \times T) \end{matrix} = \begin{bmatrix} \sigma_{x_1 v_1} & \mathbf{0} & \cdots & \mathbf{0} \\ \sigma_{x_2 v_1} & \sigma_{x_2 v_2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0} \\ \sigma_{x_T v_1} & \cdots & \sigma_{x_T v_{T-1}} & \sigma_{x_T v_T} \end{bmatrix}, \quad \begin{matrix} \boldsymbol{\Sigma}_{x\eta} \\ (p_2 \times m) \end{matrix} = \begin{bmatrix} \boldsymbol{\Sigma}_{x_1 \eta} \\ \vdots \\ \boldsymbol{\Sigma}_{x_T \eta} \end{bmatrix} \quad (26)$$

$$\begin{matrix} \boldsymbol{\Sigma}_{xx} \\ (p_2 \times p_2) \end{matrix} = \begin{bmatrix} \boldsymbol{\Sigma}_{x_1 x_1} & \cdots & \boldsymbol{\Sigma}_{x_1 x_T} \\ \vdots & \ddots & \vdots \\ \boldsymbol{\Sigma}_{x_T x_1} & \cdots & \boldsymbol{\Sigma}_{x_T x_T} \end{bmatrix} \quad (27)$$



Panel Regression Model

- By using the sample covariance matrix of \mathbf{z}_i , \mathbf{S}_N , we estimate the following parameters

$$\boldsymbol{\theta}_{all} = (\boldsymbol{\theta}'_1, \boldsymbol{\theta}'_2)' = (\boldsymbol{\theta}'_1, \boldsymbol{\theta}'_{\varepsilon\varepsilon}, \boldsymbol{\theta}'_{x\varepsilon}, \boldsymbol{\theta}'_{xx})' \quad (28)$$

where

$$\boldsymbol{\theta}_1 = (\boldsymbol{\beta}', \text{vec}(\mathbf{F}_1))'$$

$$\boldsymbol{\theta}_2 = (\boldsymbol{\theta}'_{\varepsilon\varepsilon}, \boldsymbol{\theta}'_{x\varepsilon}, \boldsymbol{\theta}'_{xx})'$$

with

$$\boldsymbol{\theta}_{\varepsilon\varepsilon} = (\sigma_{v_1}^2, \dots, \sigma_{v_T}^2, \text{vech}(\boldsymbol{\Sigma}_{\eta\eta}))'$$

$$\boldsymbol{\theta}_{x\varepsilon} = (\text{vec}(\boldsymbol{\Sigma}_{x_1\eta})', \dots, \text{vec}(\boldsymbol{\Sigma}_{x_T\eta})', \boldsymbol{\sigma}'_{x_1v_1}, \boldsymbol{\sigma}'_{x_2v_1}, \dots, \boldsymbol{\sigma}'_{x_Tv_T})'$$

$$\boldsymbol{\theta}_{xx} = \text{vech}(\boldsymbol{\Sigma}_{xx})$$

Panel Regression Model

- The order condition is given by $p(p+1)/2 \geq \dim(\boldsymbol{\theta}_{all})$.
- The minimum value of T required for order condition depends on m and the exogeneity property of \mathbf{x}_{it} .

Minimum number of T required for order condition
($K_x = 1, 2, 3$)

	$m = 1$	$m = 2$	$m = 3$
Endogenous	3	5	6
Strictly exogenous	2	3	4

- Throughout the paper, we assume that the order condition is satisfied.
- Next, we consider estimation of $\boldsymbol{\theta}_{all}$ based on ML and MD.

ML

- The likelihood function associated with \mathbf{z}_i can be written as

$$\log L_{ML}(\boldsymbol{\theta}_{all}) = -\frac{N}{2} \log |\boldsymbol{\Sigma}_{zz}(\boldsymbol{\theta}_{all})| - \frac{N}{2} \text{tr} \left[\mathbf{S}_N \boldsymbol{\Sigma}_{zz}(\boldsymbol{\theta}_{all})^{-1} \right] \quad (29)$$

where \mathbf{S}_N is sample covariance matrix of \mathbf{z}_i .

- The (Q)ML estimator is defined as

$$\hat{\boldsymbol{\theta}}_{all}^{ML} = \underset{\boldsymbol{\theta}_{all}}{\text{argmax}} \log L_{ML}(\boldsymbol{\theta}_{all})$$

- The dimension of $\boldsymbol{\theta}_{xx}$ can be very large ($K_x T(K_x T + 1)/2$).
- Solution for $\boldsymbol{\Sigma}_{xx}$ is given by

$$\boldsymbol{\Sigma}_{xx} = \frac{1}{N} (\mathbf{X} - \boldsymbol{\varepsilon} \boldsymbol{\Sigma}_{\varepsilon\varepsilon}^{-1} \boldsymbol{\Sigma}_{\varepsilon x})' (\mathbf{X} - \boldsymbol{\varepsilon} \boldsymbol{\Sigma}_{\varepsilon\varepsilon}^{-1} \boldsymbol{\Sigma}_{\varepsilon x}) + \boldsymbol{\Sigma}_{x\varepsilon} \boldsymbol{\Sigma}_{\varepsilon\varepsilon}^{-1} \boldsymbol{\Sigma}_{\varepsilon x} \quad (30)$$

$$\boldsymbol{\varepsilon}_{(N \times T)} = \begin{bmatrix} \varepsilon_{11} & \cdots & \varepsilon_{1T} \\ \vdots & & \vdots \\ \varepsilon_{N1} & \cdots & \varepsilon_{NT} \end{bmatrix}, \quad \mathbf{X}_{(N \times K_x T)} = \begin{bmatrix} \mathbf{x}'_{11} & \cdots & \mathbf{x}'_{1T} \\ \vdots & & \vdots \\ \mathbf{x}'_{N1} & \cdots & \mathbf{x}'_{NT} \end{bmatrix}$$

ML/MD estimators

- After concentrating out $\boldsymbol{\theta}_{xx}$, we have

$$\log L_{ML}^{con}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_{\varepsilon\varepsilon}, \boldsymbol{\theta}_{x\varepsilon}) = -\frac{N}{2} \log |\boldsymbol{\Sigma}_{\varepsilon\varepsilon}| - \frac{1}{2} \text{tr} [\boldsymbol{\Sigma}_{\varepsilon\varepsilon}^{-1} \boldsymbol{\varepsilon}' \boldsymbol{\varepsilon}] + \frac{N}{2} \log \left| \frac{1}{N} (\mathbf{X} - \boldsymbol{\varepsilon} \boldsymbol{\Sigma}_{\varepsilon\varepsilon}^{-1} \boldsymbol{\Sigma}_{\varepsilon x})' (\mathbf{X} - \boldsymbol{\varepsilon} \boldsymbol{\Sigma}_{\varepsilon\varepsilon}^{-1} \boldsymbol{\Sigma}_{\varepsilon x}) \right|.$$

MD

- Next, we consider minimum distance(MD) estimators.
- Let us consider the following moment condition:

$$E[\mathbf{s}_i - \boldsymbol{\sigma}_{zz}(\boldsymbol{\theta}_{all})] = \mathbf{0} \quad (31)$$

where $\mathbf{s}_i = \text{vech}[(\mathbf{z}_i - \bar{\mathbf{z}})(\mathbf{z}_i - \bar{\mathbf{z}})']$, $\bar{\mathbf{z}} = \frac{1}{N} \sum_{i=1}^N \mathbf{z}_i$ and $\boldsymbol{\sigma}_{zz}(\boldsymbol{\theta}_{all}) = \text{vech}(\boldsymbol{\Sigma}_{zz}(\boldsymbol{\theta}_{all}))$.

- The MD estimator is defined as

$$\hat{\boldsymbol{\theta}}_{all}^{MD} = \underset{\boldsymbol{\theta}_{all}}{\text{argmin}} [\mathbf{s}_N - \boldsymbol{\sigma}_{zz}(\boldsymbol{\theta}_{all})]' \mathbf{W}_N [\mathbf{s}_N - \boldsymbol{\sigma}_{zz}(\boldsymbol{\theta}_{all})] \quad (32)$$

where $\mathbf{s}_N = \text{vech}(\mathbf{S}_N)$, \mathbf{W}_N is a positive definite weighting matrix.

ML/MD estimators

- We consider three MD estimators depending on the choice of \mathbf{W}_N .
- Since the optimal weighting matrix under normality is given by

$$\mathbf{W}_0 = \mathbf{W}(\boldsymbol{\theta}_{all0}) = \frac{1}{2} \mathcal{D}'_p \left(\boldsymbol{\Sigma}(\boldsymbol{\theta}_{all0})^{-1} \otimes \boldsymbol{\Sigma}(\boldsymbol{\theta}_{all0})^{-1} \right) \mathcal{D}_p \quad (33)$$

we consider two weighting matrices which are optimal under normality:

$$\mathbf{W}_{N,1} = \frac{1}{2} \mathcal{D}'_p \left(\mathbf{S}_N^{-1} \otimes \mathbf{S}_N^{-1} \right) \mathcal{D}_p \quad (34)$$

$$\mathbf{W}_{N,2} = \frac{1}{2} \mathcal{D}'_p \left(\boldsymbol{\Sigma}(\tilde{\boldsymbol{\theta}}_{all})^{-1} \otimes \boldsymbol{\Sigma}(\tilde{\boldsymbol{\theta}}_{all})^{-1} \right) \mathcal{D}_p \quad (35)$$

where $\text{vec}(\mathbf{A}) = \mathcal{D}_p \text{vech}(\mathbf{A})$ and $\tilde{\boldsymbol{\theta}}_{all}$ is a preliminary estimator of $\boldsymbol{\theta}_{all}$.

- Since the weighting matrix that is optimal under both normality and non-normality is given by $\boldsymbol{\Omega} = \text{Var}(\mathbf{s}_i)$, we consider

$$\mathbf{W}_{N,opt} = \left(\frac{1}{N} \sum_{i=1}^N (\mathbf{s}_i - \bar{\mathbf{s}})(\mathbf{s}_i - \bar{\mathbf{s}})' \right)^{-1}, \quad \bar{\mathbf{s}} = \frac{1}{N} \sum_{i=1}^N \mathbf{s}_i \quad (36)$$

- Three MD estimators, “MD1”, “MD2” and “OMD” are defined as

$$\hat{\theta}_{all}^{MD1} = \underset{\theta_{all}}{\operatorname{argmin}} [\mathbf{s}_N - \sigma_{zz}(\theta_{all})]' \mathbf{W}_{N,1} [\mathbf{s}_N - \sigma_{zz}(\theta_{all})] \quad (37)$$

$$\hat{\theta}_{all}^{MD2} = \underset{\theta_{all}}{\operatorname{argmin}} [\mathbf{s}_N - \sigma_{zz}(\theta_{all})]' \mathbf{W}_{N,2} [\mathbf{s}_N - \sigma_{zz}(\theta_{all})] \quad (38)$$

$$\hat{\theta}_{all}^{OMD} = \underset{\theta_{all}}{\operatorname{argmin}} [\mathbf{s}_N - \sigma_{zz}(\theta_{all})]' \mathbf{W}_{N,opt} [\mathbf{s}_N - \sigma_{zz}(\theta_{all})] \quad (39)$$

- Note that $\hat{\theta}_{MD1}$ is a one-step estimator whereas $\hat{\theta}_{MD2}$ is a two-step estimator.
- In practice, we can set $\tilde{\theta} = \hat{\theta}_{MD1}$ to obtain $\mathbf{W}_{N,2}$.
- Next, we consider concentration out of MD estimators.

- To derive concentrated out MD estimators, we introduce a new vech operator that changes the order of a vector obtained by standard vech.

vech* operator

Let us define a $p \times p$ symmetric matrix Σ as follows:

$$\underset{(p \times p)}{\Sigma} = \begin{bmatrix} \Sigma_{11} & \Sigma'_{21} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \quad (40)$$

where Σ is $p \times p$, Σ_{11} is $p_1 \times p_1$, Σ_{21} is $p_2 \times p_1$, Σ_{22} is $p_2 \times p_2$. Define the vech* operator as follows:

$$\operatorname{vech}^*(\Sigma) = \begin{bmatrix} \operatorname{vech}(\Sigma_{11}) \\ \operatorname{vec}(\Sigma_{21}) \\ \operatorname{vech}(\Sigma_{22}) \end{bmatrix} \quad (41)$$

vech* operator

Note that $\text{vech}^*(\boldsymbol{\Sigma})$ and $\text{vech}(\boldsymbol{\Sigma})$ have a relationship as follows:

$$\text{vech}(\boldsymbol{\Sigma}) = \mathcal{R}_{p_1, p_2} \text{vech}^*(\boldsymbol{\Sigma}) \quad (42)$$

where

$$\mathcal{R}_{p_1, p_2} = \mathcal{D}_p^+ \begin{bmatrix} \mathcal{K}_{p_1, p} & \mathbf{0} \\ \mathbf{0} & \mathcal{K}_{p_2, p} \end{bmatrix} \begin{bmatrix} \mathcal{D}_{p_1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathcal{K}_{p_2, p_1} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{p_1 p_2} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathcal{D}_{p_2} \end{bmatrix} \quad (43)$$

and $\mathcal{K}_{m, n}$ is a commutation matrix such that $\text{vec}(\mathbf{B}') = \mathcal{K}_{m, n} \text{vec}(\mathbf{B})$ for an $m \times n$ matrix \mathbf{B} .

Note that \mathcal{R}_{p_1, p_2} is $p(p+1)/2$ dimensional full rank square matrix of zeros or ones.

ML/MD estimators

- Using this new vech^* operator, we have

$$\boldsymbol{\sigma}_{zz}(\boldsymbol{\theta}) = \text{vech}(\boldsymbol{\Sigma}_{zz}(\boldsymbol{\theta})) = \mathcal{R}_{T, p_2} \text{vech}^*(\boldsymbol{\Sigma}_{zz}(\boldsymbol{\theta})) \quad (44)$$

$$= \mathcal{R}_{T, p_2} \begin{bmatrix} \text{vech}(\boldsymbol{\Sigma}_{yy}) \\ \text{vec}(\boldsymbol{\Sigma}_{xy}) \\ \text{vech}(\boldsymbol{\Sigma}_{xx}) \end{bmatrix} = \mathcal{R}_{T, p_2} \mathbf{A}(\boldsymbol{\theta}_1) \boldsymbol{\theta}_2 \quad (45)$$

where $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ are defined in slide 17

$$\mathbf{A}(\boldsymbol{\theta}_1) = \begin{bmatrix} \mathbf{C}_{\varepsilon\varepsilon}(\mathbf{F}) & -2\mathcal{D}_T^+(\mathbf{I}_T \otimes \mathbf{B}_{12})\mathbf{C}_{x\varepsilon}(\mathbf{F}) & \mathcal{D}_T^+(\mathbf{B}_{12} \otimes \mathbf{B}_{12})\mathcal{D}_{p_2} \\ \mathbf{0} & \mathbf{C}_{x\varepsilon}(\mathbf{F}) & -(\mathbf{B}_{12} \otimes \mathbf{I}_{p_2})\mathcal{D}_{p_2} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} \end{bmatrix} \quad (46)$$

$$\mathbf{B}_{12} = -(\mathbf{I}_T \otimes \boldsymbol{\beta}') \quad (47)$$

$$\mathbf{C}_{\varepsilon\varepsilon}(\mathbf{F}) = [\mathbf{J}, \mathcal{D}_T^+(\mathbf{F} \otimes \mathbf{F})\mathcal{D}_m] \quad (48)$$

$$\mathbf{C}_{x\varepsilon}(\mathbf{F}) = [(\mathbf{F} \otimes \mathbf{I}_{p_2}), \mathcal{D}_{p_2}^\dagger] \quad (49)$$

where $\mathbf{J} = [\text{vech}(\mathbf{J}_{11}), \dots, \text{vech}(\mathbf{J}_{TT})]$ and \mathbf{J}_{st} is a $T \times T$ matrix whose (s, t) element is 1 and 0 otherwise.

- Using this alternative expression, we can solve for θ_2 as follows, which is a function of θ_1 :

$$\theta_2 = \mathbf{b}(\theta_1) = [\mathbf{A}(\theta_1)' \mathcal{R}'_{T,p_2} \mathbf{W}_N \mathcal{R}_{T,p_2} \mathbf{A}(\theta_1)]^{-1} \mathbf{A}(\theta_1)' \mathcal{R}'_{T,p_2} \mathbf{W}_N \mathbf{s}_{zz} \quad (50)$$

- Then, we have the concentrated objective function:

$$Q_{MD}^{con}(\theta_1) = [\mathbf{s}_N - \mathcal{R}_{T,p_2} \mathbf{A}(\theta_1) \mathbf{b}(\theta_1)]' \times \mathbf{W}_N [\mathbf{s}_N - \mathcal{R}_{T,p_2} \mathbf{A}(\theta_1) \mathbf{b}(\theta_1)] \quad (51)$$

where $\mathbf{s}_N = \text{vech}(\mathbf{S}_N)$.

- Compared with ML, $\theta_{\varepsilon\varepsilon}$ and $\theta_{x\varepsilon}$ are further concentrated out.
- Therefore, MD is computationally less demanding than ML.

Properties of ML/MD estimators

- Next, we investigate the theoretical properties of ML and MD estimators.
- If $\mathbf{G}(\theta_{all}) = \frac{\partial \sigma_{zz}(\theta_{all})}{\partial \theta_{all}'}$ is of full rank and other regularity conditions are satisfied, the ML and MD estimators are consistent and asymptotically normal:

$$\hat{\theta}_{all}^{est} \xrightarrow{p} \theta_{all0}, \quad (est = ML, MD1, MD2, OMD) \quad (52)$$

$$\sqrt{N} (\hat{\theta}_{all}^{est} - \theta_{all0}) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Sigma}_\theta), \quad (est = ML, MD1, MD2) \quad (53)$$

where

$$\boldsymbol{\Sigma}_\theta = \begin{cases} (\mathbf{G}'_0 \mathbf{W}_0 \mathbf{G}_0)^{-1} \mathbf{G}'_0 \mathbf{W}_0 \boldsymbol{\Omega} \mathbf{W}_0 \mathbf{G}_0 (\mathbf{G}'_0 \mathbf{W}_0 \mathbf{G}_0)^{-1} & \text{when } \mathbf{z}_i \text{ is non-normally distributed} \\ (\mathbf{G}'_0 \mathbf{W}_0 \mathbf{G}_0)^{-1} & \text{when } \mathbf{z}_i \text{ is normally distributed} \end{cases}$$

The asymptotic distribution of the OMD estimator is given by

$$\sqrt{N} (\hat{\theta}_{all}^{OMD} - \theta_{all0}) \xrightarrow{d} N\left(\mathbf{0}, (\mathbf{G}'_0 \mathbf{W}_0 \mathbf{G}_0)^{-1}\right). \quad (54)$$

- In practice, we have to determine the number of factors.
- The number of factors, m_0 , can be estimated by information criterion such as AIC, BIC and HQIC for ML:

$$AIC = -2 \log L(\theta) + 2q$$

$$BIC = -2 \log L(\theta) + \log(N)q$$

$$HQIC = -2 \log L(\theta) + 2.01 \log(\log(N))q$$

where q denotes the number of parameters.

- GMM version of AIC/BIC/HQIC proposed by Andrews and Lu (2001) can be used for MD estimators.

3. Extension

Dynamic panel data model with time-invariant regressor

- We extend the previous model to include dynamics and time-invariant regressors as follows:

$$y_{it} = \alpha y_{it-1} + \beta' \mathbf{x}_{it} + \gamma' \mathbf{w}_i + \mu_{y,t} + \mathbf{f}'_t \boldsymbol{\eta}_i + v_{it}, \quad (t = 1, \dots, T) \quad (56)$$

where $|\alpha| < 1$, γ and \mathbf{w}_i are $K_w \times 1$.

- For identification of γ , we assume that $\boldsymbol{\nu}_T$ and \mathbf{F} are linearly independent, i.e. \mathbf{f}_t is time-varying.
- For the initial condition y_{i0} , and the time-invariant regressor \mathbf{w}_i , we assume

$$y_{i0} = \mu_{y0} + \xi_{y0,i}, \quad (57)$$

$$\mathbf{w}_i = \boldsymbol{\mu}_w + \boldsymbol{\xi}_{w,i}, \quad (58)$$

where $E(y_{i0}) = \mu_{y0}$, $E(\mathbf{w}_i) = \boldsymbol{\mu}_w$ and $\xi_{y0,i}$ and $\boldsymbol{\xi}_{w,i}$ are independent over i .



Dynamic panel data model with factor error structure

- For the initial conditions y_{i0} , we assume that

$$\text{Cov}(y_{i0}, v_{it}) = 0 \quad (t = 1, \dots, T), \quad \text{Cov}(y_{i0}, \boldsymbol{\eta}'_i) = \boldsymbol{\sigma}'_{y0\eta} \quad (59)$$

- There are almost no restrictions in the initial conditions.
 - We do not need to impose a restriction such as mean-stationarity as in the system GMM estimator.
- For the time-invariant regressor \mathbf{w}_i , we assume that

$$\text{Cov}(\mathbf{w}_i, v_{it}) = \mathbf{0} \quad (t = 1, \dots, T), \quad \text{Cov}(\mathbf{w}_i, \boldsymbol{\eta}_i) = \boldsymbol{\Sigma}_{w\eta} \quad (60)$$

- Time-invariant regressor \mathbf{w}_i is allowed to be correlated with unobserved fixed effects $\boldsymbol{\eta}_i$ in an unrestricted way.
- The definitions of other variables are the same as before.



Dynamic panel data model with factor error structure

- The model can be written as

$$\mathbf{B}z_i = \boldsymbol{\mu} + \mathbf{u}_i$$

where

$$\mathbf{B} = \left[\begin{array}{cccc|cccccc} 1 & 0 & \cdots & 0 & -\alpha & -\boldsymbol{\beta}' & \mathbf{0}' & \cdots & \mathbf{0}' & -\boldsymbol{\gamma}' \\ -\alpha & 1 & & \vdots & 0 & \mathbf{0}' & -\boldsymbol{\beta}' & \ddots & \vdots & \vdots \\ 0 & \ddots & \ddots & \vdots & \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & \cdots & -\alpha & 1 & 0 & \mathbf{0}' & \cdots & \cdots & -\boldsymbol{\beta}' & -\boldsymbol{\gamma}' \\ \hline 0 & \cdots & \cdots & 0 & & & & & & \\ \vdots & & & \vdots & & & & & & \\ 0 & \cdots & \cdots & 0 & & & & & & \end{array} \right] \mathbf{I}_{TK_x+K_w+1}$$

$$\begin{aligned} \mathbf{z}_i &= (y_{i1}, \dots, y_{iT}, y_{i0}, \mathbf{x}'_{i1}, \dots, \mathbf{x}'_{iT}, \mathbf{w}'_i)' = (y_{i1}, \dots, y_{iT}, \mathbf{z}'_{2i})', \\ \mathbf{z}_{2i} &= (y_{i0}, \mathbf{x}'_{i1}, \dots, \mathbf{x}'_{iT}, \mathbf{w}'_i)' \\ \boldsymbol{\mu} &= (\mu_{y_1}, \dots, \mu_{y_T}, \mu_{y_0}, \boldsymbol{\mu}'_{x_1}, \dots, \boldsymbol{\mu}'_{x_T}, \boldsymbol{\mu}'_w)' \end{aligned}$$

Dynamic panel data model with factor error structure

-

$$\mathbf{u}_i = \begin{bmatrix} \varepsilon_i \\ \xi_{y_0,i} \\ \boldsymbol{\xi}_{x,i} \\ \xi_{w,i} \end{bmatrix} = \begin{bmatrix} \mathbf{F}\boldsymbol{\eta}_i + \mathbf{v}_i \\ \xi_{y_0,i} \\ \boldsymbol{\xi}_{x,i} \\ \xi_{w,i} \end{bmatrix}$$

$$\mathbf{F} = (\mathbf{f}_1, \dots, \mathbf{f}_T)', \quad \mathbf{v}_i = (v_{i1}, \dots, v_{iT})', \quad \boldsymbol{\xi}_{x,i} = (\xi'_{x_1,i}, \dots, \xi'_{x_T,i})'$$

- Since \mathbf{B} is invertible, we have

$$\mathbf{z}_i = \mathbf{B}^{-1}\boldsymbol{\mu} + \mathbf{B}^{-1}\mathbf{u}_i$$

- Therefore, noting that $E(\mathbf{z}_i) = \mathbf{B}^{-1}\boldsymbol{\mu}$, we have

$$\text{Var}(\mathbf{z}_i) = \boldsymbol{\Sigma}_{zz}(\boldsymbol{\theta}) = \mathbf{B}^{-1}\boldsymbol{\Sigma}_{uu}(\mathbf{B}^{-1})'$$

where $\boldsymbol{\Sigma}_{uu}$ is given by

-

$$\Sigma_{uu} = \begin{bmatrix} \Sigma_{\varepsilon\varepsilon} & \\ \Sigma_{z_2\varepsilon} & \Sigma_{z_2z_2} \end{bmatrix} = \begin{bmatrix} \Sigma_{vv} + \mathbf{F}\Sigma_{\eta\eta}\mathbf{F}' & \\ \Sigma_{z_2v} + \Sigma_{z_2\eta}\mathbf{F}' & \Sigma_{z_2z_2} \end{bmatrix} \quad (61)$$

with

$$\Sigma_{z_2\varepsilon} = E \begin{bmatrix} y_{i0}\varepsilon'_i \\ \xi_{x,i}\varepsilon'_i \\ \xi_{w,i}\varepsilon'_i \end{bmatrix} = \begin{bmatrix} \sigma'_{y_0\eta}\mathbf{F}' \\ \Sigma_{x\eta}\mathbf{F}' + \Sigma_{xv} \\ \Sigma_{w\eta}\mathbf{F}' \end{bmatrix} = \Sigma_{z_2\eta}\mathbf{F}' + \Sigma_{z_2v} \quad (62)$$

$$\Sigma_{z_2\eta} = \begin{bmatrix} \sigma'_{y_0\eta} \\ \Sigma_{x\eta} \\ \Sigma_{w\eta} \end{bmatrix}, \quad \Sigma_{z_2v} = \begin{bmatrix} \mathbf{0} \\ \Sigma_{xv} \\ \mathbf{0} \end{bmatrix} \quad (63)$$

- The definitions of Σ_{vv} , $\Sigma_{\eta\eta}$, $\Sigma_{x\eta}$, Σ_{xv} are the same as the static model.
- The estimation procedure is exactly the same as before.
- Note that $\theta_1 = (\alpha, \beta', \gamma', \text{vec}(\mathbf{F}_1)')$ and θ_2 are remaining parameters such as $\sigma_{y_0\eta}$, $\text{vec}(\Sigma_{x\eta})$ etc..

4. Monte Carlo simulation

Static panel data model

Setup

- For the sample size, we consider

$$T = \{5\}, \quad N = \{200, 500, 1000\}$$

- The number of replications is 1,000.
- Significance level is 5%.

Static panel data model

Setup

- We compare six estimators:
 - ML estimator (“ML”).
 - Three MD estimators (“MD1”, “MD2”, “OMD”).
 - “MD1” and “MD2” are asymptotically efficient only when \mathbf{z}_i is normal.
 - “OMD” is efficient regardless the distribution of \mathbf{z}_i .
 - One- and two-step GMM estimators proposed by Ahn, Lee and Schmidt (2013) (“GMM1”, “GMM2”).
 - When x_{it} is strictly exogenous, all periods of x_{it} , i.e. $\mathbf{z}_{it} = (x_{i1}, \dots, x_{iT})$ are used as instruments for each period.
 - When x_{it} is endogenous, all lagged x_{it} , i.e. $\mathbf{z}_{it} = (x_{i1}, \dots, x_{i,t-1})$ are used as instruments for each period.
 - When x_{it} is endogenous, $m = 2$ and $T = 5$, GMM estimators are not computed since the number of parameters(=5) is larger than that of moment conditions(=3).
- For ML and MD, two standard errors are computed
 - one is assuming normality (“Size”)
 - the other is robust to non-normality (“Size_{rob}”)

Static panel data model

Setup

- Note that $\Sigma_{zz,i}(\theta)$ is given by

$$\Sigma_{zz,i}(\theta) = \mathbf{B}^{-1} \Sigma_{uu,i} (\mathbf{B}^{-1})'$$

$$\Sigma_{uu,i} = \begin{bmatrix} \Sigma_{\varepsilon\varepsilon,i} & \Sigma'_{x\varepsilon,i} \\ \Sigma_{x\varepsilon,i} & \Sigma_{xx,i} \end{bmatrix} = \begin{bmatrix} \Sigma_{vv,i} + \mathbf{F}\Sigma_{\eta\eta}\mathbf{F}' & (\Sigma_{xv,i} + \Sigma_{x\eta}\mathbf{F}')' \\ \Sigma_{xv,i} + \Sigma_{x\eta}\mathbf{F}' & \Sigma_{xx,i} \end{bmatrix}$$

where

$$\Sigma_{vv,i} = \text{diag}(\sigma_{v,i1}^2, \dots, \sigma_{v,iT}^2), \quad \Sigma_{\varepsilon\varepsilon,i} = \Sigma_{vv,i} + \mathbf{F}\Sigma_{\eta\eta}\mathbf{F}'$$

$$\Sigma_{xv,i} = \Xi^v \check{\Sigma}_{vv,i}, \quad \check{\Sigma}_{vv,i} = \begin{bmatrix} \mathbf{0} \\ \Sigma_{vv,i} \end{bmatrix}, \quad \Sigma_{x\eta} = \tau_{\eta} \Xi^f \Sigma_{\eta\eta}$$

$$\Sigma_{x\varepsilon,i} = \Sigma_{xv,i} + \Sigma_{x\eta}\mathbf{F}' = \Xi^v \check{\Sigma}_{vv,i} + \tau_{\eta} \Xi^f \Sigma_{\eta\eta}\mathbf{F}'$$

$$\Sigma_{xx,i} = \tau_{\eta}^2 \Xi^f \Sigma_{\eta\eta} \Xi^{f'} + \sigma_{\omega_0,i}^2 \mathbf{b}_{\rho} \mathbf{b}'_{\rho} + \Xi^v \Sigma_{\check{v}\check{v},i} \Xi^{v'} + \sigma_r^2 (\mathbf{C}_{\rho} \mathbf{C}'_{\rho})$$

Static panel data model

Setup

-

$$\mathbf{b}_{\rho} = \rho \mathbf{C}_{\rho} \mathbf{e}_1 = \begin{bmatrix} \rho \\ \rho^2 \\ \vdots \\ \rho^T \end{bmatrix}, \quad \mathbf{C}_{\rho} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ \rho & 1 & \ddots & 0 \\ \vdots & \ddots & \ddots & 0 \\ \rho^{T-1} & \dots & \rho & 1 \end{bmatrix}$$

$$\Xi^f = \mathbf{b}_{\rho} \tilde{\mathbf{f}}_0' + \mathbf{C}_{\rho} \mathbf{F}, \quad \Xi^v = \tau_{v0} \mathbf{C}_{\rho} \mathbf{L}_0$$

Static panel data model

Findings for estimation of β

- All estimators have little bias.
- ML and MD perform very similarly, but computational cost of MD is smaller than ML.
- ML and MD are more efficient than GMM under both normality and non-normality when x is endogenous.
- When x is strictly exogenous, all estimators perform similarly when $N = 500, 1000$ and $m_0 = 1$, but not when $m_0 = 2$.
- In terms of RMSE, ML and MD perform (sometimes substantially) better than GMM.
- Inference based on ML and MD is more accurate than GMM.

Static panel data model

Results under normality ($T = 5$, $m_0 = \{1, 2\}$, Endogenous)

$z_i \sim \text{Normal}$, $T = 5$, $\beta = 1$, Endogenous

	$m_0 = 1$					$m_0 = 2$				
	ML	MD1	OMD	GMM1	GMM2	ML	MD1	OMD	GMM1	GMM2
$N = 200$										
Mean	1.002	0.997	0.998	1.004	1.004	1.004	1.005	0.969	—	—
StDev	0.046	0.041	0.043	0.079	0.074	0.074	0.075	0.199	—	—
RMSE	0.046	0.041	0.043	0.079	0.074	0.074	0.075	0.202	—	—
Size	7.8	5.1	—	—	—	8.7	7.6	—	—	—
Size _{rob}	8.2	5.6	8.0	2.7	5.1	9.7	9.0	8.7	—	—
$N = 500$										
Mean	1.001	0.999	0.999	1.002	1.000	1.002	1.002	0.993	—	—
StDev	0.027	0.026	0.027	0.046	0.041	0.043	0.043	0.104	—	—
RMSE	0.027	0.026	0.027	0.046	0.041	0.044	0.043	0.104	—	—
Size	6.4	5.3	—	—	—	7.1	6.4	—	—	—
Size _{rob}	6.3	5.3	6.4	3.5	3.9	7.4	6.8	7.0	—	—
$N = 1000$										
Mean	0.999	0.998	0.999	1.000	0.999	1.000	1.000	0.998	—	—
StDev	0.018	0.018	0.018	0.031	0.027	0.028	0.028	0.053	—	—
RMSE	0.018	0.018	0.018	0.031	0.027	0.028	0.028	0.053	—	—
Size	4.4	4.2	—	—	—	4.8	4.7	—	—	—
Size _{rob}	4.5	4.0	4.5	3.6	4.2	4.8	4.8	5.1	—	—

Static panel data model

Results under normality ($T = 5$, $m_0 = \{1, 2\}$, Strictly exogenous)

$z_i \sim \text{Normal}$, $T = 5$, $\beta = 1$, Strictly exogenous

	$m_0 = 1$					$m_0 = 2$				
	ML	MD1	OMD	GMM1	GMM2	ML	MD1	OMD	GMM1	GMM2
	$N = 200$					$N = 200$				
Mean	1.001	0.998	0.998	1.005	1.004	1.000	0.999	0.982	1.016	1.015
StDev	0.015	0.015	0.035	0.021	0.018	0.019	0.019	0.134	0.033	0.030
RMSE	0.015	0.015	0.035	0.021	0.019	0.019	0.019	0.136	0.037	0.034
Size	6.5	6.4	—	—	—	5.9	5.5	—	—	—
Size _{rob}	6.6	6.8	10.8	5.6	11.4	6.3	6.4	9.4	12.1	16.8
	$N = 500$					$N = 500$				
Mean	1.000	0.999	1.000	1.002	1.002	1.000	1.000	0.996	1.012	1.012
StDev	0.010	0.010	0.010	0.014	0.011	0.011	0.011	0.064	0.021	0.020
RMSE	0.010	0.010	0.010	0.014	0.011	0.011	0.011	0.064	0.025	0.023
Size	7.0	6.3	—	—	—	6.4	6.0	—	—	—
Size _{rob}	6.8	6.7	10.0	5.3	9.1	6.5	6.3	6.9	14.0	19.2
	$N = 1000$					$N = 1000$				
Mean	1.000	1.000	1.000	1.001	1.001	1.000	1.000	0.997	1.009	1.009
StDev	0.006	0.006	0.007	0.009	0.007	0.008	0.008	0.055	0.016	0.015
RMSE	0.006	0.006	0.007	0.009	0.007	0.008	0.008	0.055	0.019	0.017
Size	5.4	5.4	—	—	—	6.1	6.3	—	—	—
Size _{rob}	5.4	5.3	6.9	5.5	6.6	5.9	6.3	6.7	14.3	21.0

Navigation icons: back, forward, search, etc.

Static panel data model

Results under non-normality ($T = 5$, $m_0 = \{1, 2\}$, Endogenous)

$z_i \sim \chi^2$, $T = 5$, $\beta = 1$, Endogenous

	$m_0 = 1$					$m_0 = 2$				
	ML	MD1	OMD	GMM1	GMM2	ML	MD1	OMD	GMM1	GMM2
	$N = 200$					$N = 200$				
Mean	1.002	0.995	0.996	0.999	1.003	1.001	1.005	0.925	—	—
StDev	0.058	0.050	0.052	0.090	0.095	0.093	0.097	0.293	—	—
RMSE	0.058	0.050	0.052	0.090	0.095	0.092	0.097	0.302	—	—
Size	16.4	11.5	—	—	—	14.6	12.5	—	—	—
Size _{rob}	10.0	5.2	9.8	1.9	4.2	8.8	7.3	7.0	—	—
	$N = 500$					$N = 500$				
Mean	0.998	0.995	0.995	0.998	0.999	0.998	0.998	0.967	—	—
StDev	0.034	0.032	0.031	0.054	0.052	0.056	0.054	0.181	—	—
RMSE	0.034	0.033	0.031	0.054	0.052	0.056	0.054	0.184	—	—
Size	14.5	12.2	—	—	—	16.0	15.5	—	—	—
Size _{rob}	8.2	6.7	7.8	3.0	4.0	9.3	8.8	8.7	—	—
	$N = 1000$					$N = 1000$				
Mean	1.000	0.998	0.998	1.000	0.999	0.998	0.998	0.983	—	—
StDev	0.024	0.024	0.023	0.041	0.036	0.039	0.038	0.127	—	—
RMSE	0.024	0.024	0.023	0.041	0.036	0.039	0.038	0.128	—	—
Size	14.6	13.7	—	—	—	14.2	14.3	—	—	—
Size _{rob}	7.2	6.2	6.5	3.2	5.4	7.8	7.3	7.3	—	—

Navigation icons: back, forward, search, etc.

Static panel data model

Results under non-normality ($T = 5$, $m_0 = \{1, 2\}$, Strictly exogenous)

$z_i \sim \chi^2$, $T = 5$, $\beta = 1$, Strictly exogenous

	$m_0 = 1$					$m_0 = 2$				
	ML	MD1	OMD	GMM1	GMM2	ML	MD1	OMD	GMM1	GMM2
	$N = 200$					$N = 200$				
Mean	0.999	0.995	0.993	1.006	1.003	0.999	0.997	0.941	1.015	1.015
StDev	0.021	0.020	0.066	0.030	0.023	0.025	0.024	0.235	0.042	0.038
RMSE	0.021	0.021	0.067	0.030	0.023	0.025	0.025	0.242	0.044	0.041
Size	19.5	16.9	—	—	—	14.4	14.0	—	—	—
Size _{rob}	7.4	6.3	12.2	8.3	14.0	7.4	6.3	10.0	8.0	17.7
	$N = 500$					$N = 500$				
Mean	0.999	0.998	0.999	1.003	1.001	1.000	0.999	0.975	1.016	1.015
StDev	0.013	0.013	0.013	0.019	0.014	0.016	0.016	0.157	0.030	0.027
RMSE	0.013	0.013	0.013	0.019	0.015	0.016	0.016	0.159	0.034	0.030
Size	18.9	19.4	—	—	—	18.4	19.7	—	—	—
Size _{rob}	5.1	5.5	8.4	6.4	10.3	7.3	7.1	7.6	12.2	18.6
	$N = 1000$					$N = 1000$				
Mean	0.999	0.999	0.999	1.001	1.001	1.000	0.999	0.990	1.012	1.012
StDev	0.010	0.010	0.009	0.014	0.010	0.011	0.012	0.100	0.021	0.020
RMSE	0.010	0.010	0.009	0.014	0.010	0.011	0.012	0.101	0.025	0.024
Size	20.4	19.5	—	—	—	18.1	18.2	—	—	—
Size _{rob}	6.3	6.2	6.3	7.0	7.9	7.6	7.0	7.1	12.1	21.0

Navigation icons: back, forward, search, etc.

Static panel data model

Findings for estimating the number of factors

- The performance of model selection based on information criterion is investigated with $T = 7$, $N = \{200, 500, 1000\}$.
- Performance based on AIC, BIC, HQIC is investigated for the candidate $m = \{0, 1, 2\}$ where the true number of factors is $m_0 = \{0, 1, 2\}$ and known fixed effects ($f_t = 1$) are included.
- Findings are
 - ① AIC based on ML performs well under normality, but does not under non-normality.
 - ② BIC based on ML/MD2/OMD performs well when $m_0 = 1, 2$, but does not when $m_0 = 2$ and $N = 200$.
 - ③ HQIC based on ML/MD2 tends to perform best in many cases.

Navigation icons: back, forward, search, etc.

Dynamic panel data model with time-invariant regressor

Setup

- Consider the following DGP:

$$y_{it} = \mu_{y,t} + \alpha y_{i,t-1} + \beta x_{it} + \gamma w_i + \mathbf{f}'_t \boldsymbol{\eta}_i + v_{it}, \quad (t = 1, \dots, T)$$

$$x_{it} = \mu_{x,t} + \rho x_{i,t-1} + \tau_\eta \mathbf{f}'_t \boldsymbol{\eta}_i + \tau_{v0} v_{it} + r_{it}$$

$$x_{i0} = \tilde{\mu}_{x,0} + \tau_\eta \tilde{\mathbf{f}}'_0 \boldsymbol{\eta}_i + \varpi_{i0}$$

$$w_i = \mu_w + \kappa_\eta \frac{1}{m} \mathbf{v}'_m \boldsymbol{\eta}_i + s_i$$

- For parameter values, we set

$$(\alpha, \beta, \gamma) = (0.2, 0.8, 0.8), (0.8, 0.2, 0.2), \quad \rho = 0.5, \quad \tau_\eta = 0.2$$

$$\tau_{v0} = \begin{cases} 0 & \text{strictly exogenous} \\ 0.2 & \text{endogenous} \end{cases}, \quad \kappa_\eta = 0.2,$$

- Other variables are generated similarly to the static case with $\text{Var}(s_i) = \text{Var}(r_{it})$.
- For the sample size, we consider

$$T = 5, \quad N = \{200, 500, 1000\}$$

Dynamic panel data model with time-invariant regressor

Findings for estimation of α and β

- In terms of bias, ML, MD and OMD perform similarly well, and GMM performs poorly.
- ML is more dispersed than MD, and slightly size distorted when x_{it} is endogenous and $N = 200$.
- MD performs best in almost all cases in terms of RMSE and size.
- OMD performs worse than MD when $N = 200$, but similarly when $N = 500, 1000$.
- Even under non-normality, OMD does not perform best when $N = 200$.
- When $N = 1000$, OMD performs better than MD, but the difference is very small.
- GMM is (sometimes severely) biased and has large size distortions.

Dynamic panel data model with time-invariant regressor

Results under normality ($T = 5$, $m_0 = 1$, Endogenous)

$z_i \sim \text{Normal}$, $T = 5$, $(\alpha, \beta, \gamma) = (0.2, 0.8, 0.8)$, $m_0 = 1$, Endogenous

	$\alpha = 0.2$				$\beta = 0.8$				$\gamma = 0.8$			
	ML	MD	OMD	GMM	ML	MD	OMD	GMM	ML	MD	OMD	GMM
$N = 200$												
Mean	0.202	0.204	0.204	0.194	0.792	0.790	0.790	0.816	0.798	0.796	0.795	0.805
StDev	0.039	0.032	0.034	0.065	0.115	0.090	0.096	0.122	0.043	0.037	0.040	0.070
RMSE	0.039	0.032	0.035	0.065	0.115	0.090	0.097	0.123	0.043	0.037	0.040	0.070
Size	10.7	6.1	—	—	13.8	7.5	—	—	10.0	6.2	—	—
Size _{rob}	10.3	5.3	10.6	7.8	13.0	6.9	12.8	7.2	9.7	5.3	10.5	7.7
$N = 500$												
Mean	0.199	0.201	0.201	0.196	0.802	0.800	0.799	0.810	0.800	0.799	0.799	0.803
StDev	0.020	0.019	0.019	0.046	0.052	0.051	0.051	0.086	0.023	0.023	0.023	0.048
RMSE	0.020	0.019	0.019	0.046	0.052	0.051	0.051	0.086	0.023	0.023	0.023	0.048
Size	7.7	6.6	—	—	9.1	7.0	—	—	7.3	6.5	—	—
Size _{rob}	7.4	6.3	7.3	5.5	8.7	6.3	8.5	6.0	6.8	6.2	6.5	5.6
$N = 1000$												
Mean	0.200	0.200	0.200	0.197	0.801	0.800	0.800	0.806	0.800	0.800	0.800	0.802
StDev	0.013	0.013	0.013	0.035	0.036	0.035	0.035	0.065	0.016	0.016	0.016	0.036
RMSE	0.013	0.013	0.013	0.035	0.036	0.035	0.035	0.066	0.016	0.016	0.016	0.036
Size	5.8	6.2	—	—	5.4	4.9	—	—	5.9	5.5	—	—
Size _{rob}	5.5	5.6	6.0	6.1	5.2	4.7	5.6	6.0	5.7	5.8	5.3	5.7

Note: MD=MD1, GMM=GMM2



Dynamic panel data model with time-invariant regressor

Results under normality ($T = 5$, $m_0 = 1$, Endogenous)

$z_i \sim \text{Normal}$, $T = 5$, $(\alpha, \beta, \gamma) = (0.8, 0.2, 0.2)$, $m_0 = 1$, Endogenous

	$\alpha = 0.8$				$\beta = 0.2$				$\gamma = 0.2$			
	ML	MD	OMD	GMM	ML	MD	OMD	GMM	ML	MD	OMD	GMM
$N = 200$												
Mean	0.799	0.801	0.801	0.789	0.198	0.195	0.197	0.212	0.201	0.200	0.199	0.211
StDev	0.018	0.017	0.018	0.027	0.059	0.049	0.053	0.052	0.024	0.023	0.025	0.034
RMSE	0.018	0.017	0.018	0.029	0.059	0.049	0.053	0.053	0.024	0.023	0.025	0.035
Size	6.1	5.4	—	—	11.6	5.2	—	—	5.9	4.3	—	—
Size _{rob}	6.2	5.1	9.1	11.6	11.8	5.4	9.3	9.3	5.7	4.5	8.4	9.5
$N = 500$												
Mean	0.799	0.800	0.800	0.796	0.201	0.199	0.199	0.206	0.201	0.200	0.200	0.204
StDev	0.011	0.011	0.011	0.017	0.032	0.030	0.031	0.031	0.015	0.015	0.015	0.022
RMSE	0.011	0.010	0.011	0.018	0.032	0.030	0.031	0.032	0.015	0.015	0.015	0.022
Size	5.7	5.7	—	—	6.8	4.5	—	—	6.6	6.1	—	—
Size _{rob}	5.4	4.9	6.8	8.7	7.0	4.9	5.8	6.9	6.0	5.7	6.2	8.9
$N = 1000$												
Mean	0.800	0.800	0.800	0.797	0.201	0.200	0.200	0.203	0.200	0.200	0.200	0.202
StDev	0.007	0.007	0.007	0.012	0.023	0.022	0.022	0.023	0.010	0.010	0.010	0.014
RMSE	0.007	0.007	0.007	0.012	0.023	0.022	0.022	0.023	0.010	0.010	0.010	0.015
Size	6.6	6.6	—	—	5.9	5.0	—	—	5.1	5.1	—	—
Size _{rob}	6.3	5.8	6.2	7.1	6.0	5.3	6.3	6.7	5.1	5.0	5.9	5.1



5. Conclusion

Conclusion

- In this paper, we proposed a covariance structure analysis approach to estimation of panel data models with endogenous variables and factor error structure.
- We showed that rank deficiency problem arises when x_{it} is endogenous and proposed a method to address it.
- Monte Carlo simulation results showed that the ML and MD estimators perform better than GMM in most cases.

Future works

- Investigate the relationship between our ML/MD and the moment conditions proposed in the literature.
- In this paper, all variables are assumed to be continuous. However, in practice, it is important to extend to include discrete variables such as dummy variables.

Reference

- Abowd, J. M. and D. Card (1989) "On the Covariance Structure of Earnings and Hours Changes," *Econometrica*, 57, 411-445.
- Ahn, S. C., H.Y. Lee, and P. Schmidt (2001) "GMM estimation of Linear Panel Data Models with Time-Varying Individual Effects," *Journal of Econometrics*, 101, 219-255.
- Ahn, S. C., Y. H. Lee, and P. Schmidt (2013) "Panel Data Models with Multiple Time-Varying Individual Effects," *Journal of Econometrics*, 174, 1-14.
- Altonji, J. G. and L. M. Segal (1996) "Small-sample Bias in GMM Estimation of Covariance Structures," *Journal of Business & Economic Statistics*, 14, 353-366.
- Andrews, D. W. K. and B. Lu (2001) "Consistent Model and Moment Selection Procedures for GMM Estimation with Application to Dynamic Panel Data Models," *Journal of Econometrics*, 101, 123-164.
- Bai, J. (2009) "Panel Data Models with Interactive Fixed Effects," *Econometrica*, 77, 1229-1279.
- Bai, J. (2013a) "Fixed-Effects Dynamic Panel Models, a Factor Analytical Method," *Econometrica*, 81, 285-314.

Multivariate Stochastic Volatility with Realized Volatility and Pairwise Realized Correlation

Yasuhiro Omori

Faculty of Economics, University of Tokyo

with Yuta Yamauchi

January 31, 2019

University of Tokyo

1 / 51

Outline

1. Introduction
 - ▶ Univariate 'Realized' SV models
2. Multivariate SV Model
 - ▶ Basic MSV model
 - ▶ Dynamic Correlations
 - ▶ Realized Volatilities and Pairwise Realized Correlations
 - ▶ Leverage effects
3. MCMC algorithm
 - ▶ MCMC algorithm
 - ▶ Portfolio optimization
4. Empirical studies
 - ▶ Estimation results
 - ▶ Portfolio performances
5. Conclusion

2 / 51

1. Introduction

To model the multivariate asset returns, we consider

- ▶ Latent stochastic processes for time-varying variance and covariances.
→ **How to guarantee the positive definiteness for the covariance matrix?**
- ▶ ‘Asymmetry’ or ‘Cross leverage effects’ in stock market (the negative correlations between today’s return and tomorrow’s volatility).
→ **How to reduce too many parameters and latent variables?**

3 / 51

- ▶ Using observations: daily returns and ‘realized covariances’ (using high frequency data) to estimate many parameters efficiently for multivariate models.
→ **More accurate parameter estimates.** Posterior distributions are not sensitive to the specifications of priors.
- ▶ Forecasting time-varying covariances to optimize the portfolio of financial assets.
→ **Model comparison based on the portfolio performances**

4 / 51

Univariate SV model with leverage

y_t : observed log-return of stock prices.

$$\begin{aligned}
 y_t &= \epsilon_t \exp(h_t/2), \quad t = 1, \dots, n, \\
 h_{t+1} &= \mu + \phi(h_t - \mu) + \eta_t, \quad t = 1, \dots, n-1, \\
 (\epsilon_t, \eta_t)' &\sim N_2(\mathbf{0}, \Sigma), \\
 \Sigma &= \begin{pmatrix} 1 & \rho\tau \\ \rho\tau & \tau^2 \end{pmatrix}, \\
 |\phi| &< 1, \quad h_1 \sim N(\mu, \tau^2/(1 - \phi^2)).
 \end{aligned}$$

$\rho < 0 \rightarrow$ leverage effect

e.g. Omori, Chib, Shephard and Nakajima (2007) JOE, Omori and Watanabe (2008) CSDA

5 / 51

Realized Volatility

Assuming the log price $p(s)$ follows the diffusion process, $dp(s) = \mu(s)ds + \sigma(s)dW(s)$, the variance for day t is defined as the integral of $\sigma^2(s)$ over the interval $(t-1, t)$

$$IV_t = \int_{t-1}^t \sigma^2(s) ds,$$

If we have n intraday returns during a day t , $\{r_{t,i}\}_{i=0}^{n-1}$, the realized volatility is defined as their squared sum for the day t ,

$$RV_t = \sum_{i=0}^{n-1} r_{t,i}^2,$$

Then, $RV_t \rightarrow IV_t$ in probability.

6 / 51

Properties of Realized measures

- ▶ RV is the consistent estimator of Integrated variance in ideal markets
- ▶ In real markets, there are some problems in realized variances and covariances:
 - ▶ Non-trading hours (overnight)
 - may lead to the underestimation of the integrated variance.
 - ▶ Market microstructure noise
 - Bid-ask bounces, discreteness of price changes, differences in trade sizes ...
 - ▶ Nonsynchronous trading (for multiple returns)
 - may lead to the underestimation of the correlations (Epps effect).

Due to these noises, realized measures can be biased estimators.

Realized SV model

Using additional information ($x_t = \log RV_t$), we set RSV model

$$\begin{aligned}
 y_t &= \exp(h_t/2)\epsilon_t \\
 x_t &= \xi + h_t + u_t, \quad u_t \perp \epsilon_t, \eta_t \\
 h_{t+1} &= \mu + \phi(h_t - \mu) + \eta_t.
 \end{aligned}$$

- ▶ ξ and u_t correct the biases and noises of RV caused by the microstructure noise and non-trading hours automatically within models.
- ▶ Posterior standard deviations of parameters become smaller by using additional information. → **Efficient Estimation.**
 Takahashi, Omori, and Watanabe (2009). CSDA

- ▶ $x_t = \xi + \psi h_t + u_t$ may be used. But, in forecasting performances, the model $\psi \equiv 1$ performs better.
- ▶ Similar approach in GARCH model: Realized GARCH Hansen, Huang and Shek (2012) JAE

The correlations of multivariate returns are also important for

- ▶ portfolio optimization
- ▶ risk management

and thus, in the multivariate SV models, we will further consider

- ▶ time varying volatility, correlations and (cross) leverage effects.

2. Multivariate SV model

1. Basic MSV model.
2. Dynamic Correlations.
3. Realized volatilities and correlations.
4. Leverage effects.

$\mathbf{y}_t: p \times 1$ return vector at t .

\mathbf{h}_t : log volatility vector at t .

(1) Basic MSV model

$$\begin{aligned} \mathbf{y}_t &= \mathbf{m}_t + \mathbf{V}_t^{1/2} \boldsymbol{\varepsilon}_t, \quad t = 1, \dots, n, \\ \mathbf{h}_{t+1} &= \boldsymbol{\mu} + \boldsymbol{\phi} \odot (\mathbf{h}_t - \boldsymbol{\mu}) + \boldsymbol{\eta}_t, \quad t = 1, \dots, n-1, \\ \mathbf{m}_{t+1} &= \mathbf{m}_t + \mathbf{w}_t, \quad \mathbf{w}_t \sim \mathcal{N}_p(\mathbf{0}, \boldsymbol{\Sigma}_m), \boldsymbol{\Sigma}_m : \text{diagonal}, \end{aligned}$$

where

$$\begin{aligned} \mathbf{V}_t &= \text{diag}(\exp(h_{1t}), \dots, \exp(h_{pt})), \\ \mathbf{h}_t &= (h_{1t}, \dots, h_{pt})' \quad \boldsymbol{\phi} = (\phi_1, \dots, \phi_p)', \quad |\phi_i| < 1, \\ \begin{pmatrix} \boldsymbol{\varepsilon}_t \\ \boldsymbol{\eta}_t \end{pmatrix} &\sim \mathcal{N}_{2p}(\mathbf{0}, \boldsymbol{\Sigma}), \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{\varepsilon\varepsilon} & \boldsymbol{\Sigma}_{\varepsilon\eta} \\ \boldsymbol{\Sigma}_{\eta\varepsilon} & \boldsymbol{\Sigma}_{\eta\eta} \end{pmatrix}, \end{aligned}$$

with $\boldsymbol{\Sigma}_{\varepsilon\eta} \neq \mathbf{0}$, and **correlations are assumed to be constant**.

Ishihara and Omori (2012). CSDA without \mathbf{m}_t .

(2) Dynamic Correlations

We consider dynamic correlations

$$\boldsymbol{\varepsilon}_t \sim \mathcal{N}_p(\mathbf{0}, \mathbf{R}_t),$$

where we assume

$$\begin{aligned} \rho_{ijt} &= \frac{\exp(g_{ijt}) - 1}{\exp(g_{ijt}) + 1}, \\ g_{ij,t+1} &= g_{ijt} + \zeta_{ijt}, \quad \zeta_{ijt} \sim \mathcal{N}(0, \sigma_\zeta^2), \end{aligned}$$

- ▶ Many parameters and latent variables
- ▶ Correlation matrices should be positive definite.
- ▶ Alternative MSV models: Cholesky SV (Shirota et al. (2017)), Matrix exponential SV (Ishihara, Omori and Asai (2017)), Dynamic equicorrelation SV (Kurose and Omori (2016,2017)).
- ▶ DCC-GARCH (Engle (2002)).

(3) Realized Volatilities and Correlations

Let RV_{jt} and $RCorr_{ijt}$ denote the realized volatility of the j -th asset return and the realized correlation between the i -th and j th asset returns at time t . Define

$$x_{it} = \log RV_{jt}, \quad w_{ijt} = \log\left\{\frac{1 + RCorr_{ijt}}{1 - RCorr_{ijt}}\right\},$$

and consider additional measurement equations:

$$\begin{aligned} x_{it} &= \xi_{it} + h_{it} + u_{it}, & u_{it} &\sim \mathcal{N}(0, \sigma_{uj}^2), \\ w_{ijt} &= \delta_{ij} + g_{ijt} + v_{ijt}, & v_{ijt} &\sim \mathcal{N}(0, \sigma_{v,ij}^2), \end{aligned}$$

Coefficients for h_{it} and g_{ijt} are set to ones for simplicity.

Positive definite correlation matrix

We sample ρ_{ijt} (g_{ijt}) to guarantee the positive definiteness of the correlation matrix \mathbf{R}_t given other correlations and parameters. Let $\boldsymbol{\rho}_{jt} = (\rho_{j1,t}, \dots, \rho_{j,j-1,t}, \rho_{j,j+1,t}, \dots, \rho_{jpt})'$ and $\mathbf{R}_{-j,t}$ denote the matrix which excludes j -th row and j -th column of \mathbf{R}_t . Noting that

$$|\mathbf{R}_t| = |\mathbf{R}_{-j,t}| \times |1 - \boldsymbol{\rho}'_{jt} \mathbf{R}_{-j,t}^{-1} \boldsymbol{\rho}_{jt}|$$

we sample ρ_{ijt} under the constraint $\boldsymbol{\rho}'_{jt} \mathbf{R}_{-j,t}^{-1} \boldsymbol{\rho}_{jt} < 1$. Initial values for ρ_{ijt} 's are zero to satisfy the positive definiteness.

Proposition 1

The condition for $\rho_{ij,t}$ to guarantee that \mathbf{R}_t is positive definite is $\rho_{ij,t} \in (L_{ijt}, U_{ijt})$ where bounds L_{ijt} and U_{ijt} are given by

$$\frac{-\mathbf{b}'_j \boldsymbol{\rho}_{i,-j,t} \pm \sqrt{(\mathbf{b}'_j \boldsymbol{\rho}_{i,-j,t})^2 - a_j(\boldsymbol{\rho}'_{i,-j,t} \mathbf{C}_j \boldsymbol{\rho}_{i,-j,t} - 1)}}{a_j},$$

and $\boldsymbol{\rho}_{i,-j,t}$ is the vector excluding the j -th element of $\boldsymbol{\rho}_{it}$, a_j is the (j, j) -th element of \mathbf{R}_{it}^{-1} , \mathbf{b}_j is the vector excluding a_j from the j -th column of \mathbf{R}_{it}^{-1} , and \mathbf{C}_j is the matrix excluding the j -th row and j -th column from \mathbf{R}_{it}^{-1} .

15 / 51

Remark:

1. Dynamic volatilities and correlations. → General dynamics.
2. Additional information through realized volatilities and pairwise realized correlations (when some of realized variances and correlations are missing, we can use only available realized measures). → Stable and accurate estimates.
3. Proposed model is independent of the ordering of assets.
4. Covariance matrices are always positive definite.
5. Leverages between \mathbf{y}_t and \mathbf{h}_{t+1} ?

16 / 51

(4) Extension to the model with leverage effects

Consider

$$\begin{pmatrix} \mathbf{y}_t \\ \mathbf{h}_{t+1} \end{pmatrix} \sim \mathcal{N} \left(\begin{pmatrix} \mathbf{m}_t \\ \boldsymbol{\mu} + \boldsymbol{\Phi}(\mathbf{h}_t - \boldsymbol{\mu}) \end{pmatrix}, \begin{pmatrix} \mathbf{V}_t^{1/2} \mathbf{R}_t \mathbf{V}_t^{1/2} & \mathbf{V}_t^{1/2} \mathbf{R}_t^{1/2} \boldsymbol{\Lambda}' \\ \boldsymbol{\Lambda} \mathbf{R}_t^{1/2} \mathbf{V}_t^{1/2} & \boldsymbol{\Psi} + \boldsymbol{\Lambda} \boldsymbol{\Lambda}' \end{pmatrix} \right).$$

so that

$$\begin{aligned} E(\mathbf{h}_{t+1} | \mathbf{y}_t) &= \boldsymbol{\mu} + \boldsymbol{\Phi}(\mathbf{h}_t - \boldsymbol{\mu}) + \boldsymbol{\Lambda} \mathbf{R}_t^{-1/2} \mathbf{V}_t^{-1/2} (\mathbf{y}_t - \mathbf{m}_t), \\ \text{Var}(\mathbf{h}_{t+1} | \mathbf{y}_t) &= \boldsymbol{\Psi}. \end{aligned}$$

17 / 51

Remark:

1. Alternative specification:

$$\begin{pmatrix} \mathbf{y}_t \\ \mathbf{h}_{t+1} \end{pmatrix} \sim \mathcal{N} \left(\begin{pmatrix} \mathbf{m}_t \\ \boldsymbol{\mu} + \boldsymbol{\Phi}(\mathbf{h}_t - \boldsymbol{\mu}) \end{pmatrix}, \begin{pmatrix} \mathbf{V}_t^{1/2} \mathbf{R}_t \mathbf{V}_t^{1/2} & \mathbf{V}_t^{1/2} \boldsymbol{\Lambda}' \\ \boldsymbol{\Lambda} \mathbf{V}_t^{1/2} & \boldsymbol{\Psi} + \boldsymbol{\Lambda} \mathbf{R}_t^{-1} \boldsymbol{\Lambda}' \end{pmatrix} \right),$$

so that

$$\begin{aligned} E(\mathbf{h}_{t+1} | \mathbf{y}_t) &= \boldsymbol{\mu} + \boldsymbol{\Phi}(\mathbf{h}_t - \boldsymbol{\mu}) + \boldsymbol{\Lambda} \mathbf{R}_t^{-1} \mathbf{V}_t^{-1/2} (\mathbf{y}_t - \mathbf{m}_t), \\ &= \boldsymbol{\mu} + \boldsymbol{\Phi}(\mathbf{h}_t - \boldsymbol{\mu}) + \boldsymbol{\Lambda} \mathbf{R}_t^{-1/2} \mathbf{R}_t^{-1/2} \mathbf{V}_t^{-1/2} (\mathbf{y}_t - \mathbf{m}_t), \\ \text{Var}(\mathbf{h}_{t+1} | \mathbf{y}_t) &= \boldsymbol{\Psi}. \end{aligned}$$

However, we do not use this specification since it is difficult to interpret $\boldsymbol{\Lambda}$ and to justify the time-varying unconditional covariance matrix for \mathbf{h}_{t+1} 's where \mathbf{h}_t may not be stationary and its initial distribution is unknown.

18 / 51

Remark:

2. Need to select the decomposition $\mathbf{R}_t = \mathbf{R}_t^{1/2} \mathbf{R}_t^{1/2'}$.
 (1) Spectral decomposition (independent of the ordering of asset returns) (2) Cholesky decomposition
 → For example, we may select the decomposition with best portfolio performance for a practical purpose.
3. Parsimonious parameterization.

$$\mathbf{\Lambda} = \{\lambda_1, \dots, \lambda_q, \mathbf{0}, \dots, \mathbf{0}\},$$

where we set $q = 1$ in our empirical studies. This would imply the leverage effect between the market factor and individual latent volatilities.

3. MCMC algorithm

1. Initialize.
2. Generate $\{\mathbf{h}_t\}_{t=1}^n | \cdot$.
3. Generate $\{\mathbf{g}_t\}_{t=1}^n | \cdot$.
4. Generate $\{\mathbf{m}_t\}_{t=1}^n | \cdot$.
5. Generate $\boldsymbol{\mu} | \cdot, \boldsymbol{\xi} | \cdot, \boldsymbol{\delta} | \cdot, \mathbf{\Lambda} | \cdot \cdot$.
6. Generate $\phi | \cdot$.
7. Generate $\boldsymbol{\Psi} | \cdot$.
8. Generate $\{\sigma_{u,i}^2\}_{i=1}^p | \cdot, \{\sigma_{v,ij}^2\}_{i,j=1}^p | \cdot, \{\sigma_{\zeta,ij}^2\}_{i,j=1}^p | \cdot, \{\sigma_{mi}^2\}_{i=1}^p | \cdot$.
9. Go to 2.

- ▶ Generation of $\{\mathbf{h}_t\}_{t=1}^n$. → Single move sampler using MH algorithm. Sample \mathbf{h}_t given \mathbf{h}_s ($s \neq t$). The multi-move sampler may not be necessary since the sampling efficiency is high enough by using the additional measurement equations based on high frequency data.
- ▶ Generation of $\{\mathbf{g}_t\}_{t=1}^n$. Single move sampler using Gibbs sampler subject to the constraint for the positive definiteness of \mathbf{R}_t .
- ▶ Generation of $\{\mathbf{m}_t\}_{t=1}^n$. Consider the linear and Gaussian state space model using the auxiliary variables $\tilde{\mathbf{y}}_t$

$$\begin{aligned}\tilde{\mathbf{y}}_t &= \mathbf{m}_t + \tilde{\boldsymbol{\epsilon}}_t, & \tilde{\boldsymbol{\epsilon}}_t &\sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Gamma}_t), \\ \mathbf{m}_{t+1} &= \mathbf{m}_t + \boldsymbol{\nu}_t, & \boldsymbol{\nu}_t &\sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_m),\end{aligned}$$

Generate \mathbf{m} using a simulation smoother.

- ▶ Generation of $\boldsymbol{\mu}, \boldsymbol{\xi}, \boldsymbol{\delta}, \boldsymbol{\Lambda}$. Assuming normal priors, the conditional posterior distributions are conditionally independent normal distributions.
- ▶ Generation of $\boldsymbol{\phi}$. Generate a candidate for $\boldsymbol{\phi}$ from a normal distribution truncated on the region $|\phi_{ij}| < 1$ and conduct MH algorithm.
- ▶ Generation of $\boldsymbol{\Psi}$. Assuming inverse Wishart priors, we conduct MH algorithm using the the inverse Wishart distribution as the proposal.
- ▶ Generation of $\{\sigma_{u,i}^2\}_{i=1}^P, \{\sigma_{v,ij}^2\}_{i,j=1}^P, \{\sigma_{\zeta,ij}^2\}_{i,j=1}^P, \{\sigma_{mi}^2\}_{i=1}^P$. Assuming inverse gamma priors, their conditional posterior distributions are conditionally independent inverse gamma distributions.

Proposition 2

Suppose that the prior distribution of Λ given Ψ is $N_{p,p}(\mathbf{M}_0, \Psi \otimes \Gamma_0)$. Then the conditional posterior distribution of Λ given other parameters and latent variables is $N_{p,p}(\mathbf{M}_1, \Psi \otimes \Gamma_1)$ where

$$\mathbf{M}_1 = (\mathbf{A} + \Gamma_0^{-1})^{-1} (\mathbf{B} + \Gamma_0^{-1} \mathbf{M}_0), \quad \Gamma_1 = (\mathbf{A} + \Gamma_0^{-1})^{-1},$$

$$\mathbf{A} = \sum_{t=1}^{T-1} \mathbf{z}_t \mathbf{z}_t', \quad \mathbf{B} = \sum_{t=1}^{T-1} \mathbf{z}_t \eta_t',$$

and $\mathbf{z}_t = \mathbf{R}_t^{-1/2} \mathbf{V}_t^{-1/2} (\mathbf{y}_t - \mathbf{m}_t)$ and $\eta_t = \mathbf{h}_{t+1} - \mu - \Phi(\mathbf{h}_t - \mu)$.

23 / 51

Proposition 3

Let $\Lambda = [\lambda_1, \dots, \lambda_q, \mathbf{0}, \dots, \mathbf{0}]$ and $\lambda = (\lambda_1', \dots, \lambda_q')$. If the prior distribution of λ is assumed to be normal, $\lambda \sim N(\mathbf{m}_0, \Gamma_0)$, then the conditional posterior distribution of λ is $\lambda | \cdot \sim N(\mathbf{m}_1, \Gamma_1)$ where where

$$\mathbf{m}_1 = \Gamma_1 \{ \Gamma_0^{-1} \mathbf{m}_0 + (\mathbf{I}_q \otimes \Psi^{-1} \mathbf{B}') \text{vec}(\{\mathbf{e}_1, \dots, \mathbf{e}_q\}) \},$$

$$\Gamma_1 = (\Gamma_0^{-1} + \mathbf{A}_{1:q,1:q} \otimes \Psi^{-1})^{-1},$$

\mathbf{A}, \mathbf{B} are defined in (1), $\mathbf{A}_{1:q,1:q}^{-1}$ denotes the first q rows and the q columns of \mathbf{A} , $\text{vec}(\mathbf{X}) \equiv (\mathbf{x}_1', \dots, \mathbf{x}_m')$ denotes a vectorization of the matrix $\mathbf{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_m\}$, and \otimes denotes Kronecker product.

24 / 51

Predictive mean and covariance for portfolio optimization

Let $\mathbf{m}_{t+1|t} \equiv E[\mathbf{y}_{t+1}|\mathcal{F}_t]$ and $\boldsymbol{\Sigma}_{t+1|t} \equiv \text{Var}[\mathbf{y}_{t+1}|\mathcal{F}_t]$ given the current information set \mathcal{F}_t . Then, add the following several steps to each MCMC iteration:

1. Generate

$$\mathbf{h}_{n+1}^{(i)}, \mathbf{g}_{n+1}^{(i)}, \mathbf{m}_{n+1}^{(i)} | \{\mathbf{z}_t\}_{t=1}^n, \boldsymbol{\theta}^{(i)}, \{\mathbf{h}_t^{(i)}\}_{t=1}^n, \{\mathbf{g}_t^{(i)}\}_{t=1}^n, \{\mathbf{m}_t^{(i)}\}_{t=1}^n.$$

2. Store $\mathbf{m}_{n+1|n}^{(i)}$, $\mathbf{V}_{n+1|n}^{(i)}$, $\mathbf{R}_{n+1|n}^{(i)}$ and $\boldsymbol{\Sigma}_m^{(i)}$.

3. Compute

$$\hat{\mathbf{m}}_{n+1|n} = \frac{1}{N} \sum_{i=1}^N \mathbf{m}_{n+1|n}^{(i)},$$

$$\hat{\boldsymbol{\Sigma}}_{n+1|n} = \frac{1}{N} \sum_{i=1}^N \boldsymbol{\Sigma}_{n+1|n}^{(i)} = \frac{1}{N} \sum_{i=1}^N \left(\mathbf{V}_{n+1|n}^{1/2(i)} \mathbf{R}_{n+1|n}^{(i)} \mathbf{V}_{n+1|n}^{1/2(i)} + \boldsymbol{\Sigma}_m^{(i)} \right).$$

25 / 51

Portfolio strategies

We consider the minimum-variance strategy for the portfolio optimization.

- ▶ $\mu_{p,t+1}$, $\sigma_{p,t+1}^2$: the conditional mean and variance of the portfolio return, $r_{p,t+1}$.
- ▶ r_f : the risk free asset return (the federal funds rate).
- ▶ $\boldsymbol{\omega}_t$: the vector of portfolio weights for stock returns.

$$\mu_{p,t+1} = \boldsymbol{\omega}_t' \mathbf{m}_{t+1|t} + (1 - \boldsymbol{\omega}_t' \mathbf{1}) r_f, \quad \sigma_{p,t+1}^2 = \boldsymbol{\omega}_t' \boldsymbol{\Sigma}_{t+1|t} \boldsymbol{\omega}_t.$$

26 / 51

Portfolio strategies

Minimum-variance strategy

$$\min_{\omega_t} \sigma_{p,t+1}^2 \quad \text{s.t.} \quad \mu_{p,t+1} = \mu_p^*,$$

where μ_p^* is the target expected return. The solution is

$$\hat{\omega}_t = \Sigma_{t+1|t}^{-1} (\mathbf{m}_{t+1|t} - r_f \mathbf{1}) \frac{\mu_p^* - r_f}{\kappa_t},$$

where $\kappa_t = (\mathbf{m}_{t+1|t} - r_f \mathbf{1})' \Sigma_{t+1|t}^{-1} (\mathbf{m}_{t+1|t} - r_f \mathbf{1})$.

4. Empirical Studies – Data

- ▶ 9 US stocks. JP Morgan (JPM), International Business Machine (IBM), Microsoft (MSFT), Exxon Mobil (XOM), Alcoa (AA), American Express (AXP), Du Pont (DD), General Electric (GE), and Coca Cola (KO) in NYSE.
- ▶ Daily returns (**close-to-close**) and realized covariances (**open-to-close**) for $p = 9$ stocks from Oxford Man Institute Realized Library (e.g. Noureldin, Shephard and Sheppard (2012) JAE).
- ▶ The dataset also includes Bank of America (BA). However, since it has the extremely high volatility period after the financial crisis, it is excluded from our empirical studies.

4. Empirical Studies – Data

- ▶ The realized covariance is calculated via 5 minutes intraday returns with subsampling.
- ▶ The number of obs: is $n = 2242$ (Feb 1, 2001– Dec 31, 2009) and we estimate the proposed model with parsimonious leverage specification $q = 1$).
- ▶ Realized measures may have biases due to the microstructure noise, non-trading hours, nonsynchronous trading and so forth.
- ▶ Flat priors or typical priors are used in empirical studies.

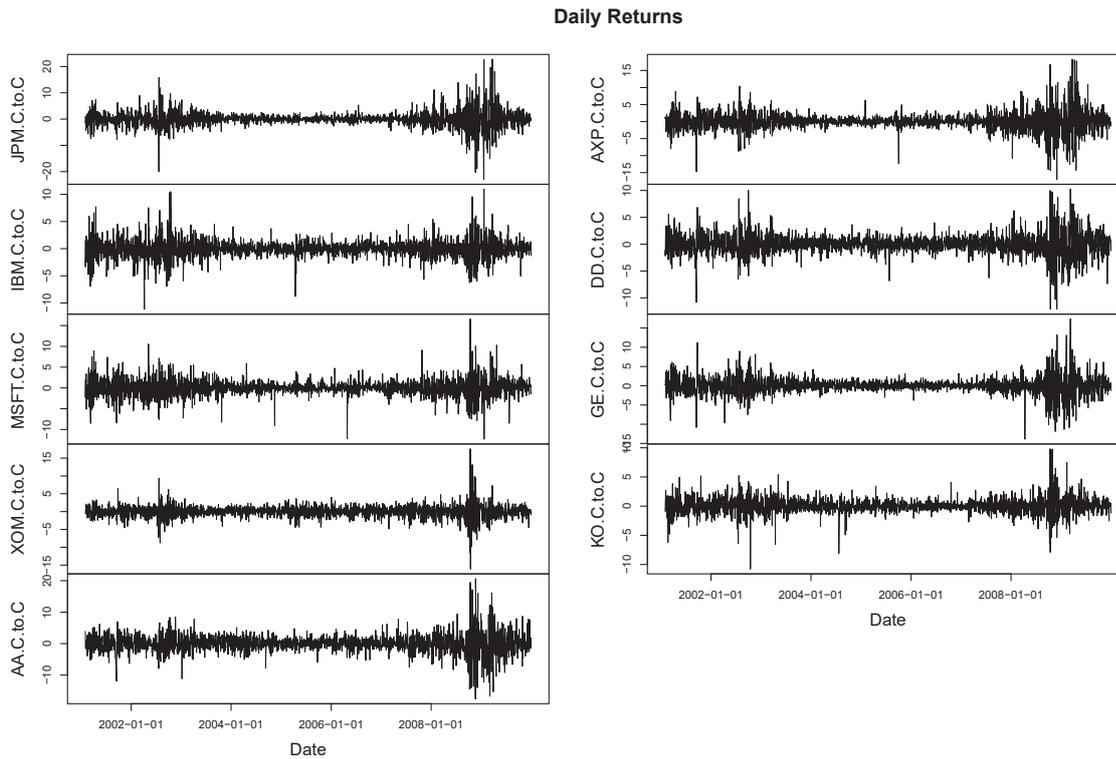
29 / 51

4. Empirical Studies – Data

- ▶ For the portfolio optimization, we conduct the rolling window estimation using past 1742 observations for 500 forecasting periods (Jan 9, 2008– Dec31, 2009). The results are obtained for the basic MSV model and the proposed model with and without leverage.

30 / 51

Application to U.S. stock returns



31 / 51

Application to U.S. stock returns

Parameter	Mean	95% interval	IF
ξ_1	-0.520	[-0.582,-0.47]	101.7
ξ_2	-0.554	[-0.61,-0.495]	100.5
ξ_3	-0.549	[-0.594,-0.501]	89.6
ξ_4	-0.442	[-0.487,-0.394]	88.1
ξ_5	-0.537	[-0.582,-0.494]	77.6
ξ_6	-0.586	[-0.651,-0.533]	108.8
ξ_7	-0.428	[-0.48,-0.376]	101.8
ξ_8	-0.535	[-0.589,-0.474]	100.1
ξ_9	-0.322	[-0.376,-0.263]	93.2

*Negative biases in realized variances (due to non-trading hours).

32 / 51

Application to U.S. stock returns

Parameter	Mean	95% interval	IF
μ_1	1.221	[1.02,1.43]	7.99
μ_2	0.644	[0.495,0.794]	16.13
μ_3	0.914	[0.759,1.07]	8.50
μ_4	0.684	[0.541,0.827]	11.09
μ_5	1.582	[1.43,1.73]	7.73
μ_6	1.136	[0.918,1.36]	9.05
μ_7	0.873	[0.726,1.02]	12.46
μ_8	0.861	[0.67,1.05]	10.38
μ_9	0.203	[0.0553,0.352]	15.08

33 / 51

Application to U.S. stock returns

Parameter	Mean	95% interval	IF
ϕ_1	0.914	[0.904,0.924]	19.2
ϕ_2	0.888	[0.874,0.901]	22.0
ϕ_3	0.900	[0.887,0.913]	19.1
ϕ_4	0.890	[0.876,0.904]	20.9
ϕ_5	0.907	[0.895,0.92]	19.7
ϕ_6	0.926	[0.916,0.935]	24.9
ϕ_7	0.899	[0.886,0.911]	21.5
ϕ_8	0.908	[0.897,0.92]	21.3
ϕ_9	0.903	[0.889,0.916]	21.4

*High persistence in log volatilities.

34 / 51

Application to U.S. stock returns

Parameter	Mean	95% interval	IF
$\sigma_{u,1}$	0.285	[0.27,0.3]	39.8
$\sigma_{u,2}$	0.286	[0.271,0.303]	57.4
$\sigma_{u,3}$	0.291	[0.278,0.304]	13.9
$\sigma_{u,4}$	0.274	[0.261,0.287]	20.3
$\sigma_{u,5}$	0.318	[0.304,0.332]	16.5
$\sigma_{u,6}$	0.307	[0.293,0.321]	20.3
$\sigma_{u,7}$	0.291	[0.278,0.304]	19.6
$\sigma_{u,8}$	0.303	[0.289,0.317]	17.9
$\sigma_{u,9}$	0.294	[0.281,0.308]	17.6

35 / 51

Application to U.S. stock returns

Parameter	Mean	95% interval	IF
$\sigma_{m,1}$	0.0873	[0.0673,0.108]	118
$\sigma_{m,2}$	0.0702	[0.0533,0.089]	126
$\sigma_{m,3}$	0.0756	[0.0532,0.106]	129
$\sigma_{m,4}$	0.0871	[0.0669,0.106]	117
$\sigma_{m,5}$	0.0976	[0.0734,0.135]	129
$\sigma_{m,6}$	0.0882	[0.0709,0.118]	119
$\sigma_{m,7}$	0.0874	[0.0627,0.111]	123
$\sigma_{m,8}$	0.0776	[0.0602,0.108]	125
$\sigma_{m,9}$	0.0666	[0.0481,0.0879]	124

*Small variances for the mean process

36 / 51

Application to U.S. stock returns (spectral decomp)

Parameter	Mean	95% interval	IF
λ_1	-0.0626	[-0.0852,-0.0403]	6.82
λ_2	-0.0541	[-0.0757,-0.033]	6.64
λ_3	-0.0430	[-0.0638,-0.0216]	7.30
λ_4	-0.0518	[-0.0722,-0.0311]	5.78
λ_5	-0.0424	[-0.0625,-0.0219]	7.24
λ_6	-0.0518	[-0.0735,-0.0303]	6.32
λ_7	-0.0536	[-0.0736,-0.0331]	9.07
λ_8	-0.0538	[-0.0767,-0.0308]	7.92
λ_9	-0.0436	[-0.0637,-0.0235]	9.72

*Leverage effects exist

Application to U.S. stock returns (Cholesky decomp)

Parameter	Mean	95% interval	IF
λ_1	-0.0552	[-0.0769,-0.0337]	7.64
λ_2	-0.0375	[-0.0576,-0.0173]	4.03
λ_3	-0.0280	[-0.0483,-0.00741]	7.33
λ_4	-0.0351	[-0.0548,-0.0151]	5.64
λ_5	-0.0329	[-0.0525,-0.0133]	7.10
λ_6	-0.0383	[-0.0596,-0.0172]	6.58
λ_7	-0.0385	[-0.058,-0.0191]	8.84
λ_8	-0.0467	[-0.0685,-0.0243]	9.38
λ_9	-0.0365	[-0.0562,-0.0171]	7.92

*Leverage effects exist

Application to U.S. stock returns (spectral decomp)

Parameter	Mean	95% interval	IF
ρ_1^*	-0.199	[-0.306,-0.113]	8
ρ_2^*	-0.187	[-0.302,-0.100]	8
ρ_3^*	-0.153	[-0.261,-0.067]	7
ρ_4^*	-0.199	[-0.327,-0.101]	7
ρ_5^*	-0.179	[-0.322,-0.076]	8
ρ_6^*	-0.177	[-0.288,-0.089]	8
ρ_7^*	-0.224	[-0.374,-0.114]	10
ρ_8^*	-0.171	[-0.278,-0.087]	8
ρ_9^*	-0.187	[-0.334,-0.083]	11

where $\rho_i^* = \text{Corr}(z_{1t}, h_{i,t+1})$, $\mathbf{z}_t = \mathbf{R}_t^{-1/2} \mathbf{V}_t^{-1/2} (\mathbf{y}_t - \mathbf{m}_t)$ with $q = 1$. **Leverage effects exist.**

Application to U.S. stock returns (Cholesky decomp)

Parameter	Mean	95% interval	IF
ρ_1^*	-0.170	[-0.266,-0.092]	7
ρ_2^*	-0.118	[-0.199,-0.049]	4
ρ_3^*	-0.091	[-0.174,-0.021]	8
ρ_4^*	-0.118	[-0.206,-0.045]	7
ρ_5^*	-0.123	[-0.239,-0.044]	7
ρ_6^*	-0.119	[-0.205,-0.048]	7
ρ_7^*	-0.141	[-0.244,-0.061]	9
ρ_8^*	-0.144	[-0.237,-0.067]	10
ρ_9^*	-0.148	[-0.271,-0.058]	8

where $\rho_i^* = \text{Corr}(z_{1t}, h_{i,t+1})$, $\mathbf{z}_t = \mathbf{R}_t^{-1/2} \mathbf{V}_t^{-1/2} (\mathbf{y}_t - \mathbf{m}_t)$ with $q = 1$. **Leverage effects exist, but with small absolute values.**

Application to U.S. stock returns (spectral decomp)

Par.	Mean	95% interval	IF	Par.	Mean	95% interval	IF
ρ_{11}^*	-0.195	[-0.305,-0.107]	6	ρ_{21}^*	-0.009	[-0.103,0.071]	35
ρ_{12}^*	-0.183	[-0.305,-0.091]	10	ρ_{22}^*	0.016	[-0.084,0.096]	41
ρ_{13}^*	-0.147	[-0.254,-0.063]	8	ρ_{23}^*	-0.001	[-0.141,0.106]	64
ρ_{14}^*	-0.199	[-0.334,-0.097]	6	ρ_{24}^*	0.037	[-0.054,0.116]	41
ρ_{15}^*	-0.187	[-0.343,-0.080]	11	ρ_{25}^*	0.025	[-0.096,0.114]	51
ρ_{16}^*	-0.166	[-0.276,-0.079]	7	ρ_{26}^*	-0.013	[-0.109,0.070]	31
ρ_{17}^*	-0.225	[-0.400,-0.111]	8	ρ_{27}^*	-0.012	[-0.125,0.081]	45
ρ_{18}^*	-0.173	[-0.284,-0.084]	8	ρ_{28}^*	0.021	[-0.079,0.099]	39
ρ_{19}^*	-0.183	[-0.337,-0.075]	13	ρ_{29}^*	0.054	[-0.066,0.143]	55

where $\rho_{1i}^* = \text{Corr}(z_{1t}, h_{i,t+1})$ and $\rho_{2i}^* = \text{Corr}(z_{2t}, h_{i,t+1})$ with $q = 2$. **Leverage effects exist only for the first component.**

Application to U.S. stock returns (Cholesky decomp)

Par.	Mean	95% interval	IF	Par.	Mean	95% interval	IF
ρ_{11}^*	-0.168	[-0.271,-0.087]	7	ρ_{21}^*	-0.029	[-0.089, 0.024]	5
ρ_{12}^*	-0.132	[-0.230,-0.056]	6	ρ_{22}^*	-0.113	[-0.200,-0.043]	7
ρ_{13}^*	-0.100	[-0.189,-0.029]	7	ρ_{23}^*	-0.073	[-0.147,-0.009]	6
ρ_{14}^*	-0.120	[-0.212,-0.046]	4	ρ_{24}^*	-0.019	[-0.080, 0.036]	7
ρ_{15}^*	-0.128	[-0.240,-0.045]	9	ρ_{25}^*	-0.044	[-0.117, 0.018]	8
ρ_{16}^*	-0.125	[-0.215,-0.051]	8	ρ_{26}^*	-0.028	[-0.093, 0.029]	5
ρ_{17}^*	-0.149	[-0.262,-0.064]	7	ρ_{27}^*	-0.081	[-0.160,-0.016]	8
ρ_{18}^*	-0.145	[-0.239,-0.066]	6	ρ_{28}^*	-0.011	[-0.071, 0.042]	5
ρ_{19}^*	-0.151	[-0.269,-0.058]	10	ρ_{29}^*	-0.003	[-0.069, 0.052]	7

where $\rho_{1i}^* = \text{Corr}(z_{1t}, h_{i,t+1})$ and $\rho_{2i}^* = \text{Corr}(z_{2t}, h_{i,t+1})$ with $q = 2$. **Leverage effects exist for two components.**

Application to U.S. stock returns

Table: Posterior means (standard deviation)

δ	$i = 1$	$i = 2$	$i = 3$	$i = 4$	$i = 5$	$i = 6$	$i = 7$	$i = 8$
$j = 2$	-0.313 (0.0285)							
$j = 3$	-0.265 (0.0305)	-0.333 (0.0529)						
$j = 4$	-0.301 (0.0482)	-0.18 (0.0349)	-0.149 (0.065)					
$j = 5$	-0.41 (0.0651)	-0.227 (0.0273)	-0.246 (0.038)	-0.287 (0.0425)				
$j = 6$	-0.629 (0.04)	-0.265 (0.032)	-0.295 (0.0476)	-0.353 (0.0449)	-0.391 (0.0359)			
$j = 7$	-0.472 (0.0473)	-0.276 (0.0293)	-0.219 (0.0299)	-0.312 (0.0252)	-0.532 (0.0351)	-0.477 (0.032)		
$j = 8$	-0.526 (0.0534)	-0.37 (0.0405)	-0.319 (0.0647)	-0.34 (0.0367)	-0.44 (0.0288)	-0.563 (0.0478)	-0.478 (0.0331)	
$j = 9$	-0.194 (0.0428)	-0.0807 (0.033)	-0.0983 (0.0258)	-0.249 (0.0501)	-0.15 (0.0293)	-0.246 (0.0301)	-0.158 (0.026)	-0.172 (0.0573)

*Negative biases in realized correlations (due to non-synchronous trading). – Epps effect

Application to U.S. stock returns

Table: Posterior means (standard deviation)

σ_v	$i = 1$	$i = 2$	$i = 3$	$i = 4$	$i = 5$	$i = 6$	$i = 7$	$i = 8$
$j = 2$	0.329 (0.00554)							
$j = 3$	0.325 (0.0053)	0.325 (0.00531)						
$j = 4$	0.329 (0.00538)	0.329 (0.0057)	0.31 (0.00524)					
$j = 5$	0.329 (0.00544)	0.322 (0.00527)	0.314 (0.00509)	0.328 (0.00546)				
$j = 6$	0.35 (0.00586)	0.331 (0.00555)	0.315 (0.00532)	0.319 (0.00565)	0.332 (0.00554)			
$j = 7$	0.338 (0.0056)	0.338 (0.00575)	0.319 (0.00547)	0.334 (0.00562)	0.344 (0.00571)	0.338 (0.00579)		
$j = 8$	0.334 (0.00565)	0.32 (0.00553)	0.304 (0.00523)	0.322 (0.00553)	0.317 (0.00518)	0.332 (0.00576)	0.326 (0.00548)	
$j = 9$	0.314 (0.00531)	0.329 (0.00544)	0.307 (0.00516)	0.317 (0.00562)	0.32 (0.00518)	0.321 (0.00532)	0.337 (0.0058)	0.328 (0.00555)

Application to U.S. stock returns

Table: Posterior means (standard deviation)

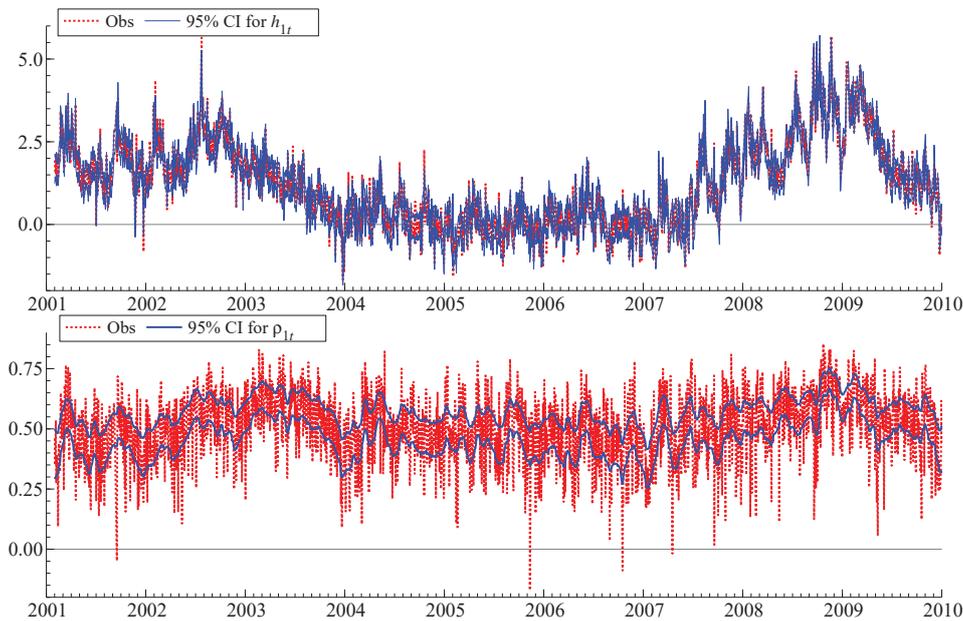
σ_ζ	$i = 1$	$i = 2$	$i = 3$	$i = 4$	$i = 5$	$i = 6$	$i = 7$	$i = 8$
$j = 2$	0.047 (0.00497)							
$j = 3$	0.0387 (0.00398)	0.0431 (0.00426)						
$j = 4$	0.0606 (0.00448)	0.0615 (0.00566)	0.0595 (0.0046)					
$j = 5$	0.0468 (0.00414)	0.0437 (0.00376)	0.042 (0.00397)	0.0465 (0.00446)				
$j = 6$	0.0578 (0.0048)	0.0455 (0.00506)	0.0442 (0.00478)	0.0686 (0.0055)	0.0471 (0.00446)			
$j = 7$	0.0411 (0.00421)	0.0436 (0.00483)	0.041 (0.00537)	0.0611 (0.00486)	0.0531 (0.00509)	0.0492 (0.00486)		
$j = 8$	0.0484 (0.00485)	0.0521 (0.00494)	0.0549 (0.00527)	0.0707 (0.00512)	0.0485 (0.00353)	0.0535 (0.0055)	0.049 (0.00477)	
$j = 9$	0.0485 (0.0049)	0.0422 (0.00465)	0.0402 (0.00404)	0.0641 (0.0055)	0.0432 (0.00418)	0.0452 (0.00456)	0.0441 (0.00512)	0.0476 (0.0043)

Application to U.S. stock returns

Table: Posterior means (standard deviation)

Ψ	$i = 1$	$i = 2$	$i = 3$	$i = 4$	$i = 5$	$i = 6$	$i = 7$	$i = 8$	$i = 9$
$j = 1$	0.165 (0.0105)								
$j = 2$	0.129 (0.00851)	0.142 (0.00944)							
$j = 3$	0.119 (0.00793)	0.114 (0.00743)	0.127 (0.00866)						
$j = 4$	0.111 (0.00774)	0.105 (0.00708)	0.0966 (0.00671)	0.124 (0.00839)					
$j = 5$	0.104 (0.00729)	0.0882 (0.00648)	0.082 (0.00614)	0.0891 (0.00624)	0.103 (0.00773)				
$j = 6$	0.137 (0.00882)	0.112 (0.00788)	0.105 (0.00751)	0.102 (0.00736)	0.0935 (0.00697)	0.142 (0.0101)			
$j = 7$	0.116 (0.00797)	0.102 (0.00719)	0.0967 (0.00674)	0.101 (0.00681)	0.0891 (0.00631)	0.108 (0.00737)	0.115 (0.00805)		
$j = 8$	0.141 (0.00917)	0.124 (0.00831)	0.116 (0.00778)	0.109 (0.00758)	0.1 (0.00712)	0.128 (0.00852)	0.113 (0.00775)	0.157 (0.0105)	
$j = 9$	0.101 (0.00715)	0.0954 (0.00657)	0.0866 (0.00612)	0.0861 (0.00606)	0.069 (0.00545)	0.091 (0.00648)	0.0866 (0.00616)	0.0992 (0.00695)	0.103 (0.00727)

95 % credible intervals (blue, dashed) of h_{1t} and ρ_{1t}



Top: $x_{1t} - \xi_1$ (red, dotted). Bottom:
 $\{\exp(w_{21,t} - \delta_{21}) - 1\} / \{\exp(w_{21,t} - \delta_{21}) + 1\}$ (red, dotted).

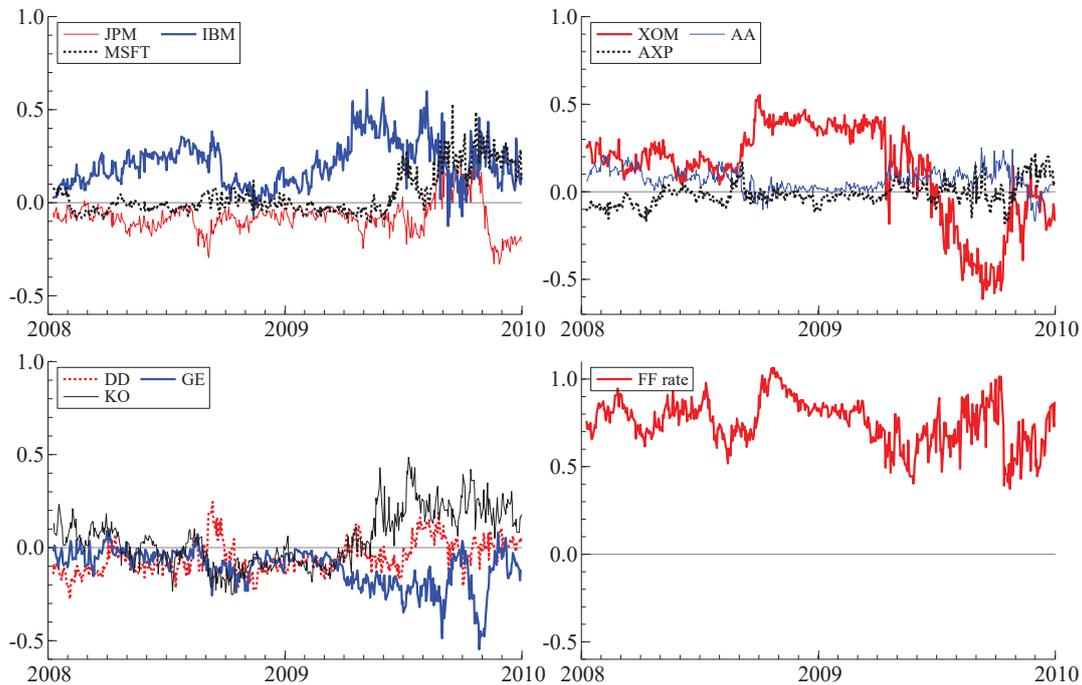
Portfolio performances for competing models

Cumulative values of realized objective functions.

Minimum-Variance

	$\mu_p^* = 0.004$	$\mu_p^* = 0.01$	$\mu_p^* = 0.1$
MSV	1.172	6.536	1184
CRSV	0.748	4.448	730
MRSV	0.526	2.943	510
MRSV-L1-C	0.272	1.601	262
MRSV-L2-C	0.264	1.552	255
MRSV-L1-S	0.249	1.430	232
MRSV-L1-S (constant mean)	0.568	3.032	543
DCC-GARCH	2.662	11.962	2537
Equal weight	1425	1425	1425

Portfolio weights in MRSV-L1-S ($\mu_p^* = 0.01$)



49 / 51

Conclusion

We propose

- ▶ Multivariate SV model with leverage.
- ▶ Dynamic correlations.
- ▶ Efficient parameter estimation method using realized covariances.
- ▶ The model is independent of order of asset returns.

In empirical studies,

- ▶ The model seems to capture the dynamics of log conditional variances and correlations.
- ▶ High persistences in h_{it} (log conditional volatilities) are found.
- ▶ Better portfolio performances than other volatility models.

50 / 51

- ▶ Negative biases are found for realized volatilities and correlations (negative x_i 's and δ_{ij} 's)
- ▶ Variances of measurement errors u_{it} 's and $v_{ij,t}$'s are larger than those of latent h_{it} 's and $g_{ij,t}$'s.
- ▶ The mean processes of the returns are almost constant.