

多様化するデータをどう使いこなすか ～米国での動向からの知見～

平田英明

東京財団政策研究所主席研究員／法政大学経営学部教授

h-hirata@hosei.ac.jp



東京財団政策研究所
THE TOKYO FOUNDATION FOR POLICY RESEARCH

2024年2月13日
14:00 ～ 15:30
東京財団政策研究所

目次

1. オルタナティブ・データ(AD)事始
2. コロナ期の利用状況(短期分析)
3. ポストコロナ期の課題(短期分析)
4. 学術研究におけるADの利用動向(中長期分析)
5. ビジネス分野におけるAD

オルタナティブ・データの定義

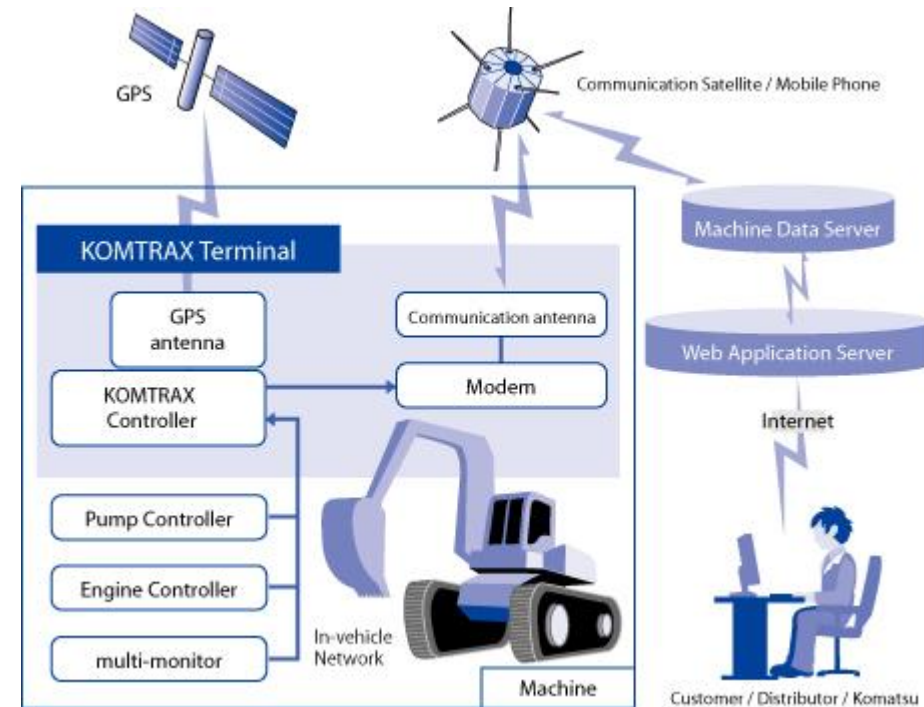
- 「POSデータ、クレジットカードデータ、位置情報、衛星画像など、これまで利活用されてこなかった代替的なデータ」
 - 辻中 仁士氏(株式会社ナウキャスト)による定義
- 財務情報や経済統計のような伝統的データとの比較
 - ビジネスの副産物
 - 外向けに出す場合、プライバシー等への配慮要
 - 元のデータはUnstructuredな場合が多い
 - そもそも企業はin-houseでどこまで使えているか

ADの潜在的ニーズ

- 伝統的データでは把握できない**新規業態等の情報把握**
 - 従来の経済データで経済を追跡するだけでは、経済が現在どこにあるかを理解するのに十分ではない
- **速報性と高い更新頻度**
 - 伝統的データはリリースの**タイムラグ & 更新**が発生
 - 集計の問題と「割り切り」の問題のトレードオフ
- 伝統的データの整備が遅れている国の情報が使いやすく

日本の先駆例

- Komtrax
(コマツ産機株式会社)
 - GPSを搭載したコマツの建設機械(世界で数十万台)の機械稼働管理システム
 - どの機械がどの場所にあって、エンジンが動いているか止まっているか、燃料がどれだけ残っているか、昨日何時間仕事をしたか、すべてがコマツのオフィスで分かる仕組み



出所:コマツカスタマーサポート ウェブサイト
https://kcsj.komatsu/service_support/komtrax

日本の先駆例(続)

- Komtrax
(コマツ産機株式会社)
 - 一台当たり月間平均稼働時間

出所:コマツ産機株式会社ウェブサイト
<https://www.komatsu.jp/ja/ir/library/demand-orders>

時間		[A]22/1-22/12	[B]23/1-23/12	[A] vs.[B]
日本	1月	58.8	51.5	-12.5%
	2月	57.6	52.6	-8.6%
	3月	52.5	47.6	-9.4%
	4月	45.1	41.7	-7.7%
	5月	41.6	41.5	-0.2%
	6月	48.2	45.5	-5.6%
	7月	48.1	47.3	-1.8%
	8月	45.3	44.0	-3.0%
	9月	47.3	46.6	-1.5%
	10月	48.2	46.9	-2.7%
	11月	47.0	45.6	-3.0%
	12月	51.7	47.7	-7.7%
北米	1月	65.2	62.0	-5.0%
	2月	62.3	60.7	-2.6%
	3月	72.9	71.0	-2.6%
	4月	67.1	62.9	-6.2%
	5月	69.6	71.9	3.4%
	6月	77.3	72.2	-6.5%
	7月	69.5	68.1	-2.1%
	8月	78.4	77.0	-1.8%
	9月	73.3	69.1	-5.8%
	10月	73.6	72.2	-1.8%
	11月	66.5	65.8	-1.0%
	12月	57.8	55.6	-3.8%

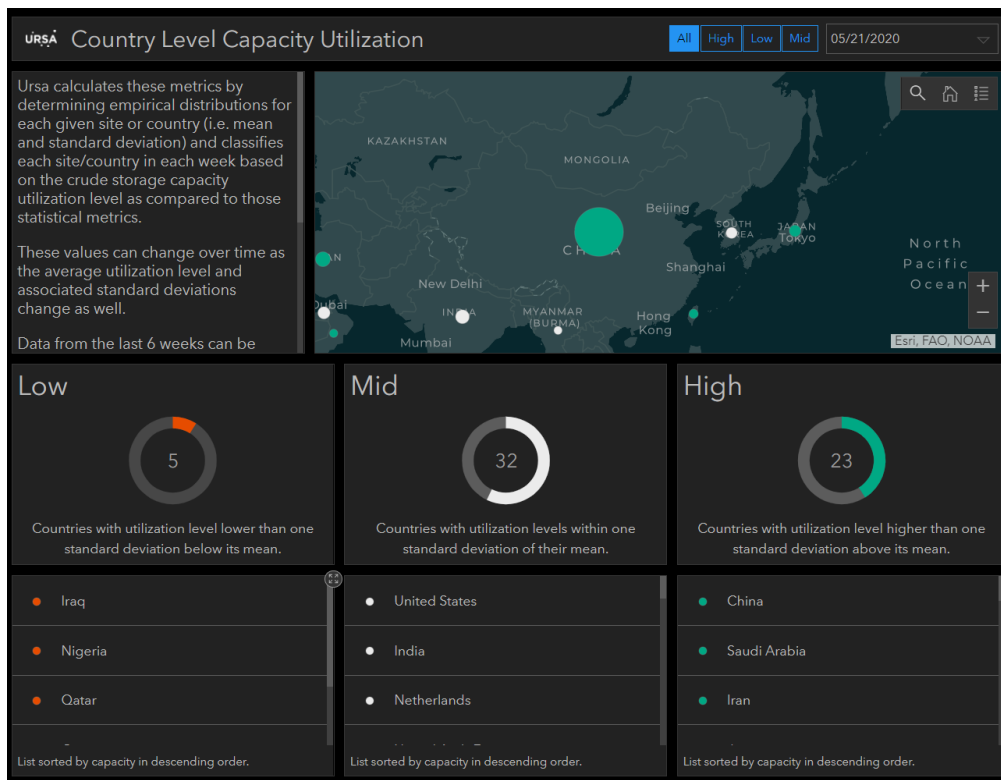
目次

1. オルタナティブ・データ(AD)事始
2. コロナ期の利用状況(短期分析)
3. ポストコロナ期の課題(短期分析)
4. 学術研究におけるADの利用動向(中長期分析)
5. ビジネス分野におけるAD

コロナ期とポストコロナ期の利用動向

- コロナ期
 - 2020年4月 IMF⇒数個の図表での利用
 - 国際通貨基金(IMF)の*World Economic Outlook*
 - 2020年6月 WB⇒全図表の1/4での利用
 - 世界銀行(WB)の*Global Economic Prospects*

利用例：人工衛星を使った貯蔵量推定



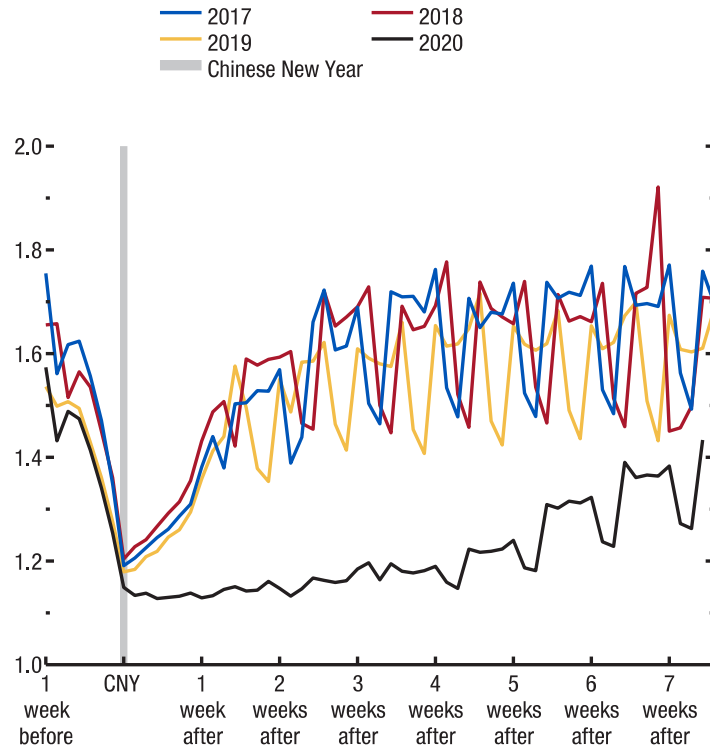
Sources: Argus; Bloomberg L.P.; Thomson Reuters Datastream; Ursa Space Systems; and IMF staff calculations.

出所：Ursa SpaceのUrsa Space Platformのメイン画面(2020年5月)。
<https://www.ursaspace.com/>

出所：IMF, *World Economic Outlook* (2020年4月)

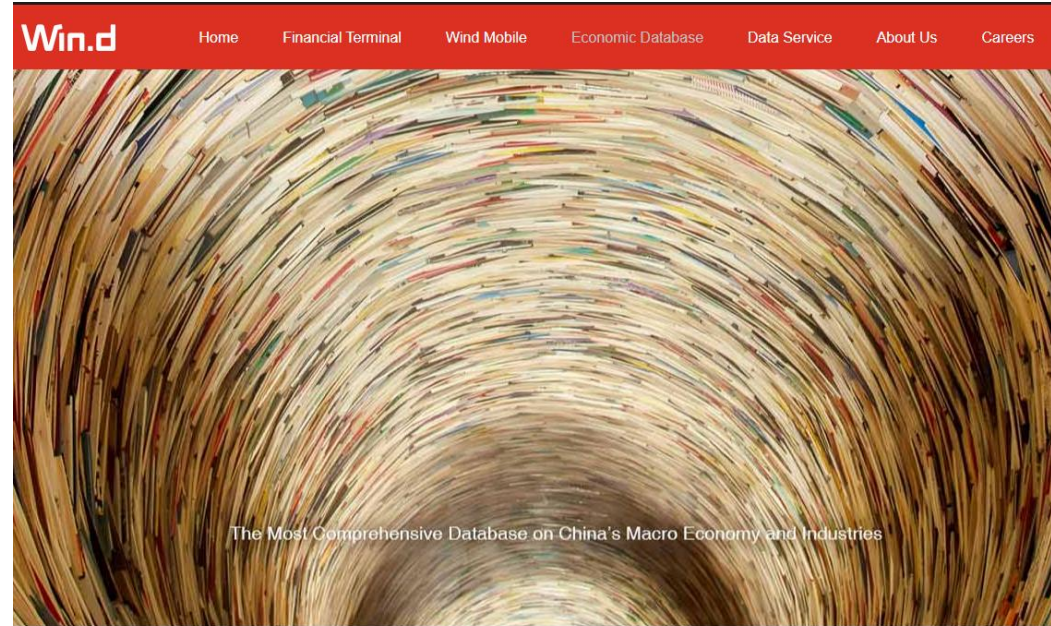
利用例：中国の交通量推定

Figure 1.SF.2. China Transport Congestion Index
(100 cities, average)



Sources: Wind; and IMF staff calculations.
Note: CNY = Chinese New Year.

出所：IMF, *World Economic Outlook* (2020年4月)



Wind Economic Database

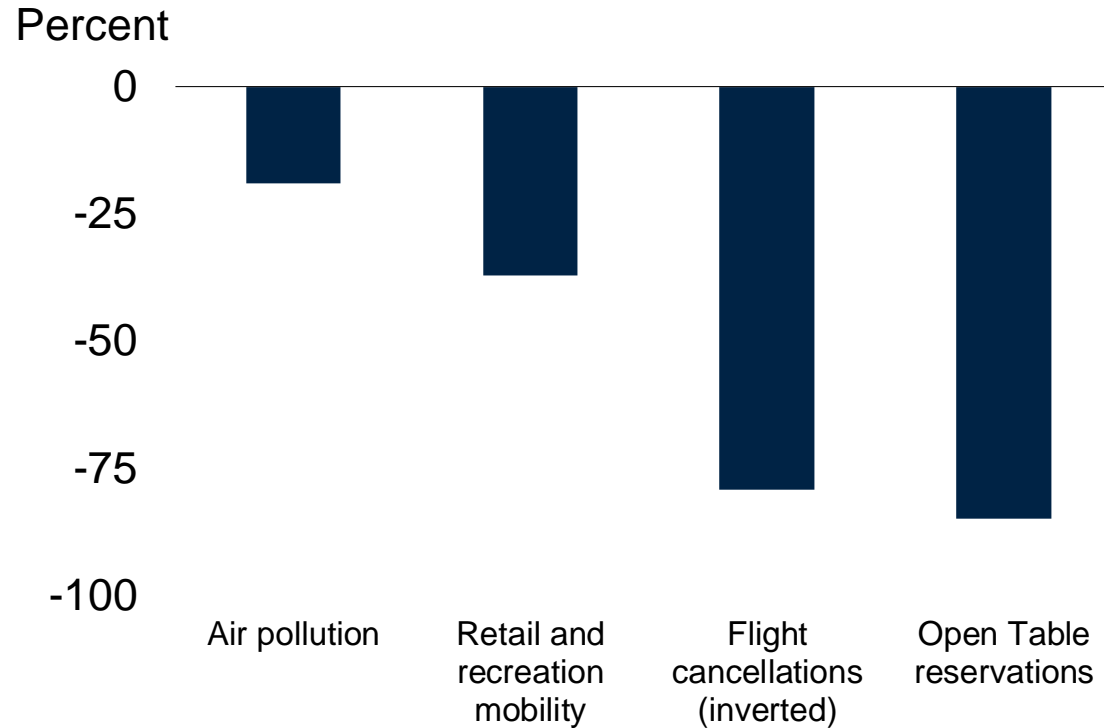
The Wind Economic Database pairs over 1.3 million macroeconomic and industry time series with powerful graphics and data analysis tools to give financial professionals the most comprehensive insights into China's economy.

No one else aggregates as many official, quasi-official, and private data sources in one place as we do.

出所：Windウェブサイト
<https://www.wind.com.cn/>

利用例: WB① 稼働率の代替指標

Figure 1.1.C. Change in global activity indicators in 2020



Source: Air Quality Open Data Platform; Airportia; Google; OpenTable; World Bank.

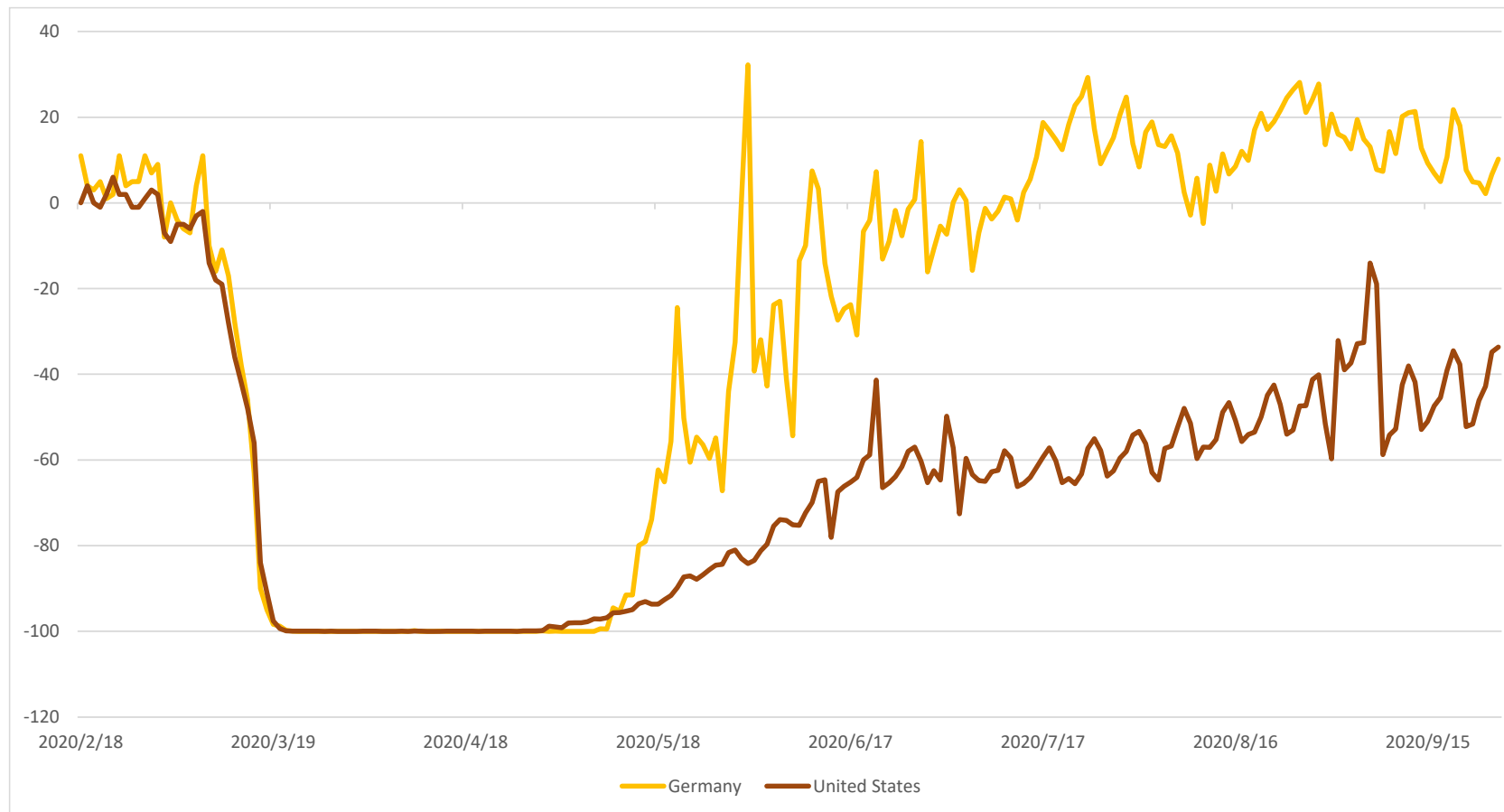
Note: Air pollution is the change in NO2 emissions over January 1 to May 28 in 2019 and 2020. Retail and recreation mobility is the percent change for May 21, 2020 from baseline, which is the median value for the corresponding day of the week during the 5-week period January 3-February 6, 2020, based on data from Google. Flight cancellations shows the cancellations relative to total planned flights based on comparing currently operating flights in 2020 with flights that were operating 52 weeks ago in 2019 as of May 27, 2020. Open Table reservations shows the change in seated diners at restaurants on the OpenTable network on May 27 in 2019 and 2020. For more information on flight cancellations data, go to <https://www.airportia.com/coronavirus/>.

[Return to Read Me](#)

出所: World Bank, *Global Economic Prospects* (2020年6月)

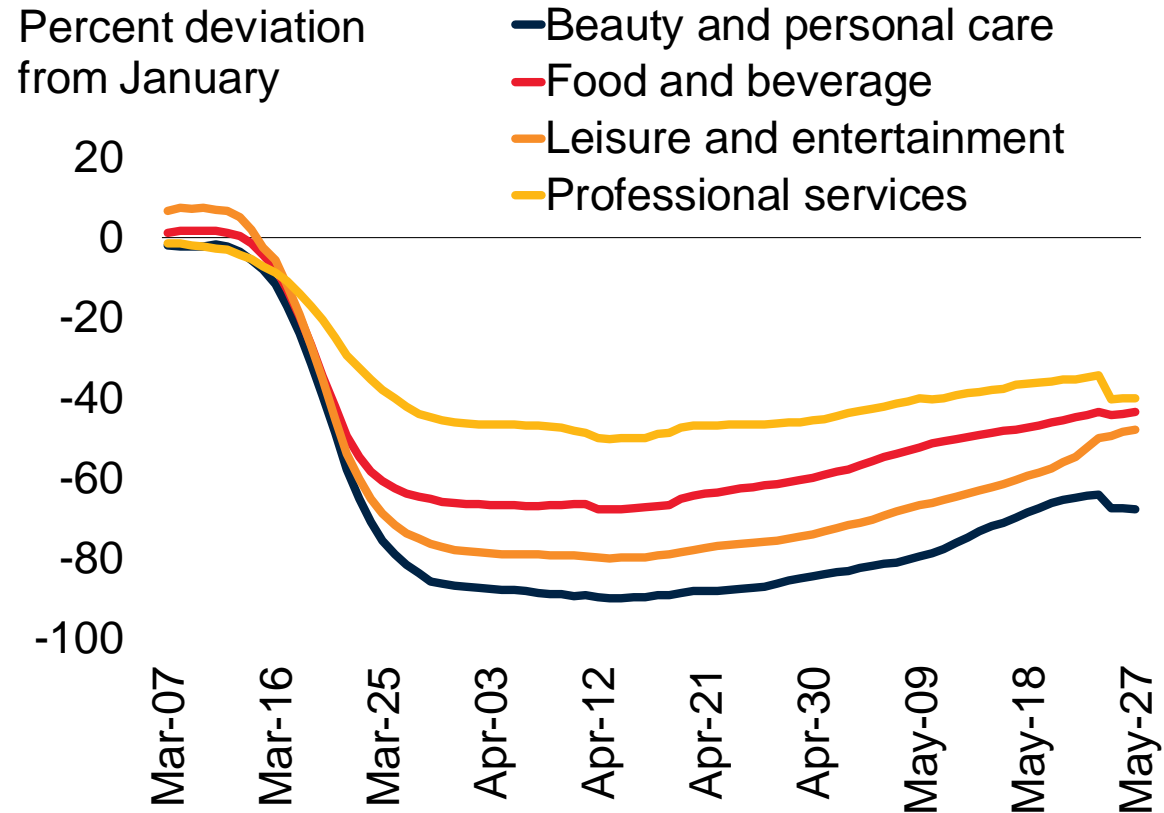
利用例：WB①' レストラン稼働率比較

OpenTableのデータ(前年比、YoY_Seated_Diner_Data)



利用例: WB② 中小企業の労働時間

Figure 1.4.A. Reduction in hours worked



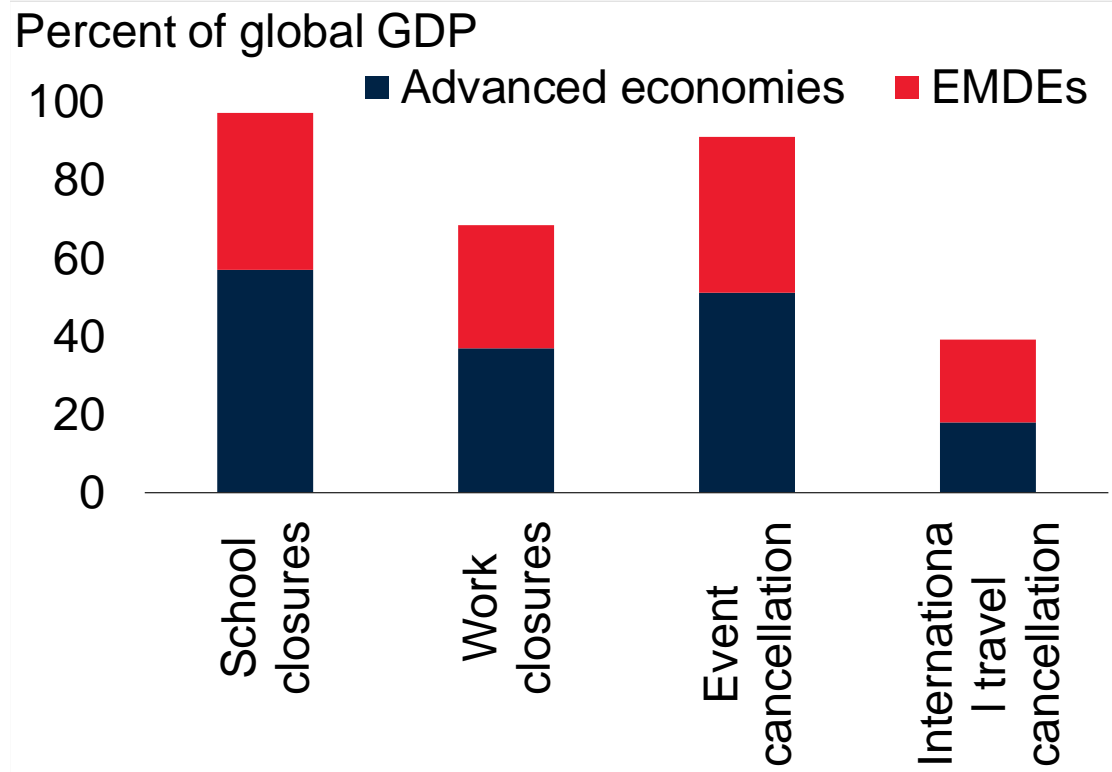
Source: Homebase; World Bank.

Note: Figure shows 7-day moving average. Sample covers 60,000 small businesses and 1 million hourly employees in the U.S. The data compare the hours worked for the observed day against the median hours worked for the same day of the week during the period January 4, 2020 to January 31, 2020 in order to compare the level of activity to pre-COVID-19 levels. Last observation is May 27, 2020. For more information on the data, go to <https://joinhomebase.com/data/covid-19/>.

出所: World Bank,
Global Economic Prospects
(2020年4月)

利用例: WB③ 一斉臨時休業

Figure 3.1.D. Share of global GDP affected by mitigation measures



Source: University of Oxford COVID-19 Government Response Tracker; World Bank.

Note: EMDEs = emerging market and developing economies. Figure shows share of GDP accounted for by economies with restrictions. Restrictions are counted if required (i.e., not only recommended) and, for school and work closures, if applied across all levels and sectors, respectively. Travel restrictions are counted if they entail a ban on arrivals from all regions or a total border closure. Data is for April 1, 2020. Sample includes 125 EMDEs and 34 AEs.

[Back to Read me](#)

出所: World Bank, *Global Economic Prospects* (2020年4月)

コロナ期とポストコロナ期の利用動向

- コロナ期
 - 2020年4月 IMF⇒数個の図表での利用
 - 国際通貨基金(IMF)の*World Economic Outlook*
 - 2020年6月 WB⇒全図表の1/4での利用
 - 世界銀行(WB)の*Global Economic Prospects*
- ポストコロナ期(2023年～)
 - IMF、WBともにほぼ利用せず
 - ただし、広く定例利用される指標も
 - 米国オフィス稼働率(警備会社Kastle Systemsの出退勤データを用いたもの。2020年3月に比べて直近は半分程)
 - ニューヨーク鉄道利用動向(MTA; Metropolitan Transportation Authority)
 - 国際サプライチェーン圧力指数(ニューヨーク連銀、インフレ指標の一つとして)
 - 日銀展望レポートでは、2データを定例利用
 - JCB/ナウキャスト「JCB消費NOW」⇒カード消費
 - Agoop⇒人流

目次

1. オルタナティブ・データ(AD)事始
2. コロナ期の利用状況(短期分析)
3. **ポストコロナ期の課題(短期分析)**
4. 学術研究におけるADの利用動向(中長期分析)
5. ビジネス分野におけるAD

AD利用が頭打ちとなっている背景

- 伝統的データで事足りるから
 - 短期的に急激な変化が生じやすいような状況では、それを捕捉していく上でADはパワフルなツール
 - コロナに関係なく、ADのコモディティ化も進んでいる
- データの水準を客観的に評価することが難しいから
 - 長期データが乏しいことが多い
- 統計に求められる属性を満たさないことがあるから
 - 公的統計の主要な品質要素：ニーズ適合性、正確性、適時性、解釈可能性・明確性
 - データの突然の非公表化

伝統的データ vs. AD

- 伝統的データとの作成コンセプトの違い

- 公共財か否か

- 非競合性(ある人が消費したからといって他の人の消費機会が減るわけではない)
- 排除不可能性(費用負担をしない人の排除が困難)

- データ調査のルールの厳格さ vs. 柔軟さ

- ①生データ取得
- ②集計作業
- ③①と②のグランドデザイン

- 不正の発生する可能性

- 意図せざる「不正」(もしかしたら≈ サンプルバイアス、かも)

目次

1. オルタナティブ・データ(AD)事始
2. コロナ期の利用状況(短期分析)
3. ポストコロナ期の課題(短期分析)
4. 学術研究におけるADの利用動向(中長期分析)
5. ビジネス分野におけるAD

データを巡る3つの視点

① データの粒度 (granularity)

– 粒度の高いデータにリーチできることが分析の強みに

- ADと粒度は好相性

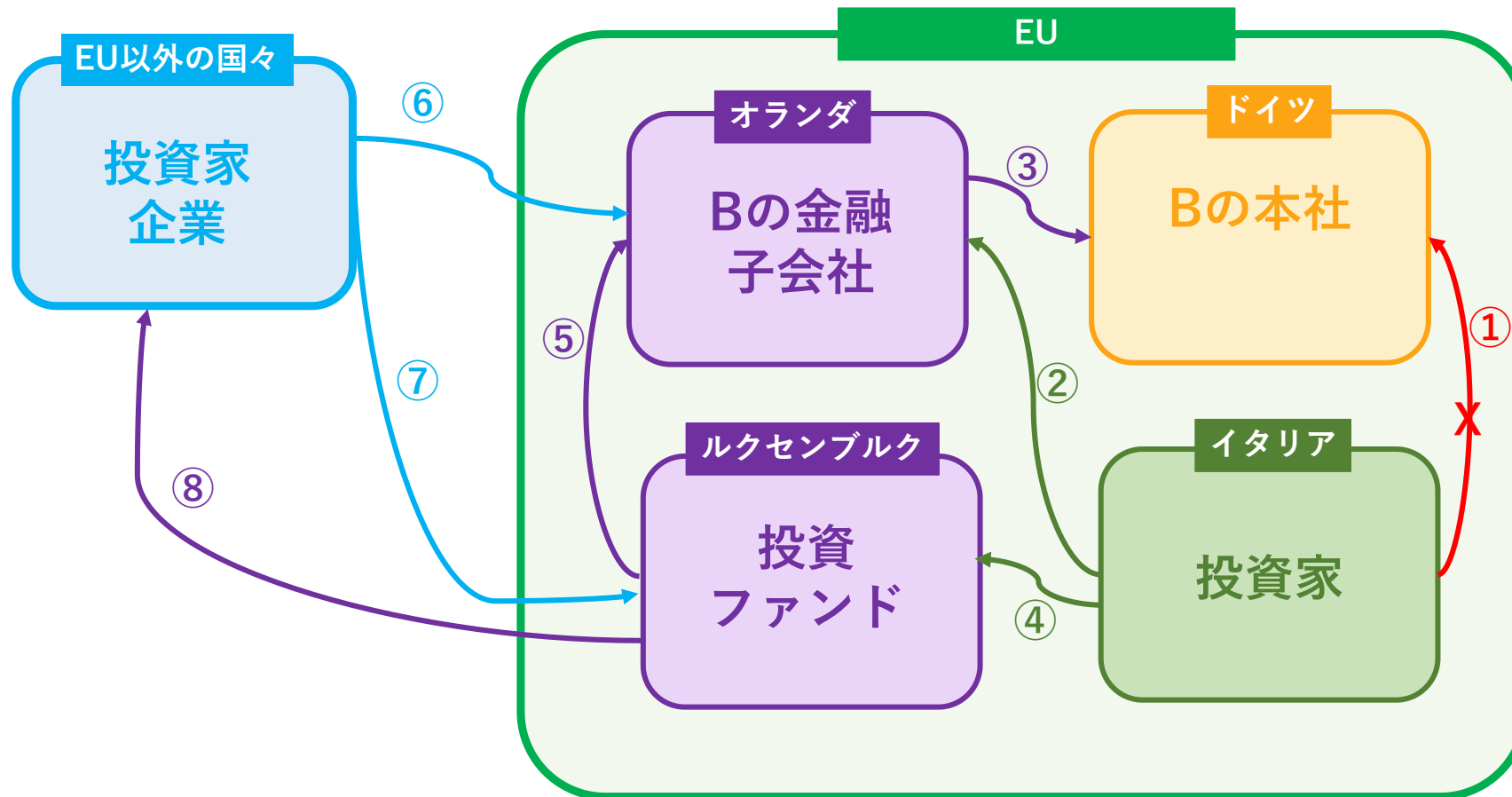
– 研究例a: フィラデルフィア連銀による**消費者向けフィンテック貸出企業の個別貸出データを検証した分析**

- パンデミック期においてLending ClubのスコアがFICOスコアを上回るパフォーマンスを示していたことを明らかに
- 両スコア間の相関関係は低く、Lending Clubが独自のアルゴリズムやデータ利用によって、よりデフォルトリスクの低い先への貸し付けを成功させていた可能性

データを巡る3つの視点

① データの粒度 (granularity) 続

- 研究例b: 欧州中央銀行とアカデミアの研究者による**クロスボーダーでの欧州内の資金フロー**に関する分析
 - (投資ファンドの仲介拠点であると同時に、外国企業による株式や社債発行の中心地としての役割を担っている)ルクセンブルク、アイルランド、オランダ等を経由した取引を調整した上で、ユーロ圏の金融統合パターンの姿をあぶり出す
 - **実際のユーロ圏のグロスの対外ポジションは、株式、債券いずれも公式統計の半分程度**(つまり、金融のグローバル化はそこまで進んでいない)



出所: Beck et al. (2023)を元に発表者作成

データを巡る3つの視点

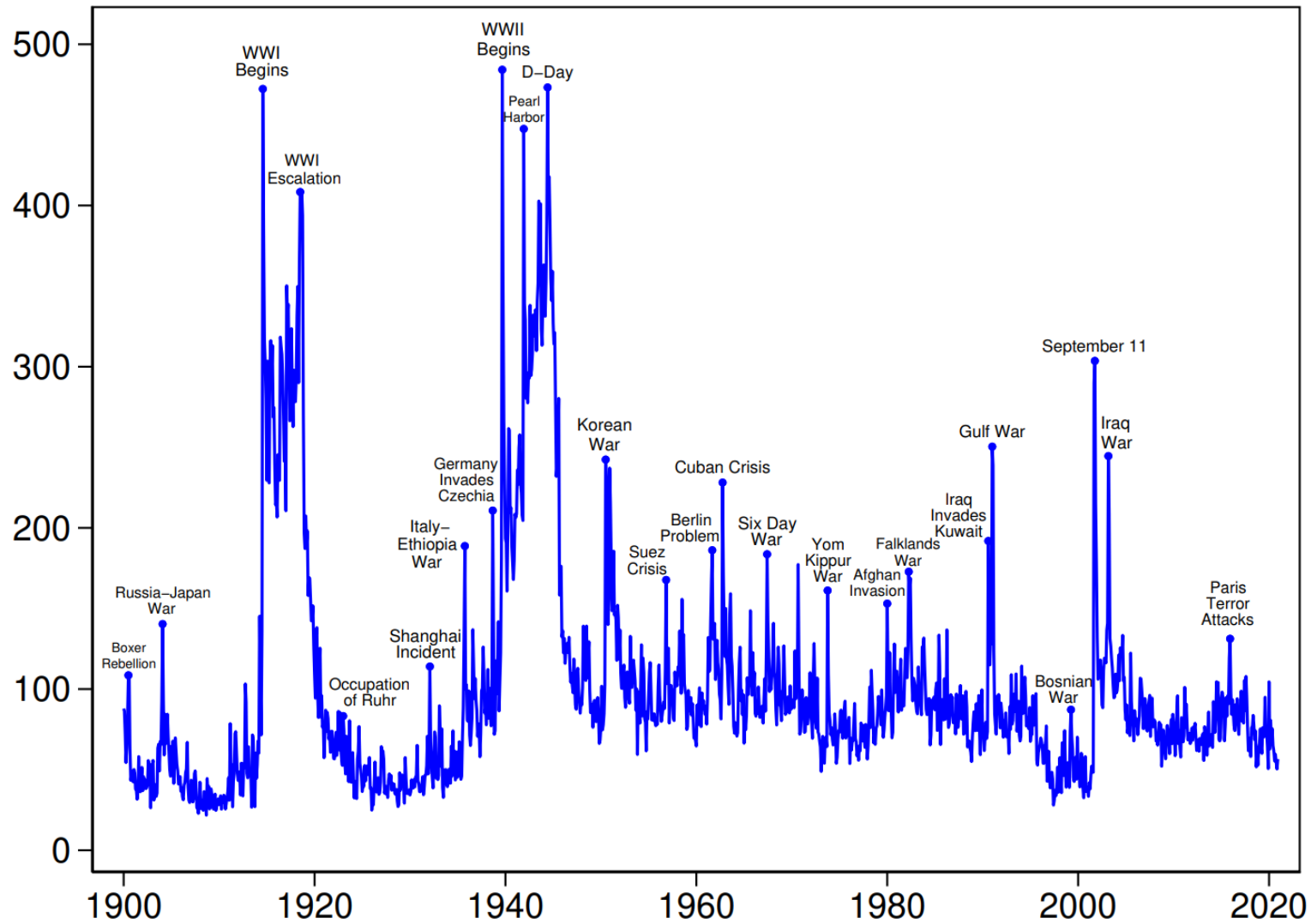
② 超長期データ

– 歴史とデータの関係性を垣間見る⇒不確実性分析

- 新聞データのテキスト分析～不確実性を表すキーワードを拾って指標作成

– 研究例c: Caldara et al. (2023)

- 地政学的リスクインデックスとインフレの関係を考察
- 米国以外: 地政学的リスク↑ ⇒ 経済活動↓
- 米国: 地政学的リスク↑ ⇒ GDP ↑ (∵ 軍事費の増加)
- いずれの地域においても、地政学的リスク↑ ⇒ 公的債務↑ & 国際貿易↓
& マネー↑ & インフレ↑



出所: 連邦準備制度理事会 国際金融部門 シニア・アソシエイト・ディレクター Matteo Iacoviello 氏 ウェブサイト
www.matteoiacoviello.com/gpr.htm

データを巡る3つの視点

② 超長期データ 続

– 研究例d: Baker et al., (2022)

- 日次での19カ国についての長期での株価のジャンプ(±2.5%の変化と仮定)に影響を与える「市場の受け止め(新聞のテキスト分析から抽出)」は何かを探る
- 主に**金融政策と政府支出に関連する政策ニュースは、すべての国で下方へのジャンプよりも上方へのジャンプを引き起こす割合が高い**
- 金融政策のニュースによって引き起こされたジャンプの後の方が、他のジャンプの後よりも、市場の変動が小さい傾向

データを巡る3つの視点

③ データがマクロ経済活動の決め手に

- eコマースの一般化に伴い、データを制する者がビジネスを制するという傾向が強まっている
- マクロ経済分析の中にも、データを生産の一要素として明示的に織り込む理論研究
 - “The Data Economy”セッション @ AEA (American Economic Association)
 - データに強い企業の生産性に影響を与えることを通じて、マクロ経済に与えるインパクトに注目 (Groh and Pfäuti, 2023)
 - 正確なデータを活用できる企業は、それをビジネスに活かし、より確実性の高いビジネス展開ができる
⇒質の高いデータを活用できる企業は生産性の向上が見込まれ、さらに生産性の変動も小さく
 - より多く生産する企業ほど、より多くのデータを蓄積していくメカニズム (Farboodi and Veldkamp, 2022) ⇒
例: スマートデバイスを通じたデータ等の分析を蓄積かつアップデートしていった企業は、よりデータの質を高めることができる

目次

1. オルタナティブ・データ(AD)事始
2. コロナ期の利用状況(短期分析)
3. ポストコロナ期の課題(短期分析)
4. 学術研究におけるADの利用動向(中長期分析)
5. ビジネス分野におけるAD

競争優位と専有データ

- 専有データ (proprietary data) の重要性
 - Davenport and Redman (DR, 2020)の整理をベースに
 - 専有データは、自社のみに固有⇒持続可能な競争優位の構築に利用可
 - 他者による複製が容易ではないため
 - その獲得、開発、活用の方法に関する戦略を持つ必要
 - 外部のデータと組み合わせるとメリット↑
 - 外部データの例：業界情報やSNSからのスクレイピング情報
 - AIを使えばいいというわけでもない
 - まずはunstructured⇒structuredに
 - 人の目を通じた普段のチェックの必要性
 - AIの導入には十も一も

DR (2020)の戦略の策定との関係

- 専有データはどの事業目的に役立つのか
 - 関係者内での情報の共有(データの中央化)、当局への提出情報、恣意的な情報集約の排除...
 - 決め手となる変数: 透明性、プライバシー配慮、情報の頻度、データの信頼性、コスト
- どの種類のデータが、自社にとって有益か
 - Trial and Error
 - スタンダード化による組織内での情報共有
- 有益な内部データにどんな方法で価値を付加し、精選し、保護するのか
 - 専門性の必要、必要に応じた人材の確保

参考文献

- Baker, Scott R., Nicholas Bloom, Steven J. Davis, and Marco Sammon. 2022. “What Triggers Stock Market Jumps?” mimeo.
- Beck, Roland, Antonio Coppola, Angus Lewis, Matteo Maggiori, Martin Schmitz, and Jesse Schreger. 2023. “The Geography of Capital Allocation in the Euro Area.” mimeo.
- Caldara, Dario, Sarah Conlisk, Matteo Iacoviello, and Maddie Penn. 2023. “Do Geopolitical Risks Raise or Lower Inflation?” mimeo.
- Davenport, Thomas H., and Thomas C. Redman. 2020. “Your Organization Needs a Proprietary Data Strategy.” *Harvard Business Review*.
- Farboodi, Maryam and Laura Veldkamp. 2022. “Data and Markets.” *Annual Review of Economics*, Vol. 15, pp. 23-40.
- Groh, Carl-Christian, and Pfäuti, Oliver Pfäuti. 2023. “Data, Business Cycles, and the Cyclical Effectiveness of Monetary Policy.” mimeo.
- Jagtiani, Julapa, Catharine Lemieux, Brandon Goldstein. 2023. “Did Fintech Loans Default More During the COVID-19 Pandemic? Were Fintech Firms “Cream Skimming” the Best Borrowers?” mimeo.
- 辻中仁士(2020)「COVID-19でにわかに注目を集めるオルタナティブデータ ～オルタナティブデータで捉える経済(1)」『経済セミナー』、2020年9月号、pp.52-57
- 平田英明(2020)「オルタナティブ・データは万能か？」『東京財団政策研究所 政策データウォッチ』(34)
- 平田英明(2023)「景気動向分析におけるオルタナティブ・データの現在地」東京財団政策研究所『Review』 R-2022-133