

# 経済・社会を読み解く ビッグデータ分析

鹿野繁樹

大阪府立大学 大学院経済学研究科 経済学専攻

現代システム科学域 マネジメント学類

2020年11月17日

# ○本日のコンテンツ

1. 経済・ビジネスとデータ分析
2. 社会経済変数の予測
3. 予測から因果関係へ
4. 結語

# ○講演者の紹介

- 鹿野繁樹 (Kano Shigeki)、大阪府立大学 准教授
  - 所属: 大学院 経済学研究科 / 同 マネジメント学類 経済データサイエンス課程
  - 博士(社会経済)、筑波大学 大学院 社会工学研究科
  - 専攻: ミクロ計量経済学、パネルデータ分析
  - e-mail: < [kano@eco.osakafu-u.ac.jp](mailto:kano@eco.osakafu-u.ac.jp) >

- 近年の著書・研究論文

- 鹿野繁樹(2015)「新しい計量経済学: データで因果関係に迫る」、日本評論社
- Kishi, T., & Kano, S. (2018). Labour market transitions in Australia and Japan: A panel data analysis. *Australian Journal of Labour Economics*, 20(3), 175.
- Kano, S. (2019). Indirect Inference for Linear Panel Data Models with Predetermined Variables. Presented at the 2019 Autumn Meeting of Japanese Economic Association

# 1. 経済・ビジネスとデータ分析

# ○データ分析の有用性

- 学術・実務を問わず、あらゆる分野に広まる**データ分析**。
  - 経済、経営、マーケティング、会計、ファイナンス。
  - 教育、心理、社会学。
  - 医療、公衆衛生、疫学。
  - 犯罪捜査、司法。
  - 歴史、音楽、美術、文学、スポーツ。

- **計量経済学**: 経済学(と関連分野)のデータを扱う統計学。
  - データを観測。⇒ 経済主体(消費者や企業)の行動・戦略、市場構造を知る。
  - データ = 経済主体が残した「足跡」。
  - 「足跡」をたどって、経済主体やマーケットの原理を突き止める。

・ データ分析をおこなう主な目的は？

1. 現状・実態の把握。
2. 予測。
3. 因果関係の評価。



- 1. データによる現状・実態の把握
  - 「今、どういう状況なのか」⇒意思決定に使う。データ分析の最も基本的な役割。
  - 例：患者の健康診断データ⇒健康状態。
  - 例：企業の財務データ⇒経営状態。
  - 例：気象データ⇒地球温暖化の進行。

- 経済学では？ : 計量経済学 = econometrics の和訳。
  - Metrics = 計測・測定。
  - ∴ 計量経済学のはじまり = 経済変数の計測。
  - 経済変数 = 国内総生産 (GDP) や失業率、物価水準。
  - 政府・経済学者の重要な役割。経済統計。

- 2. データによる予測
  - まだ起こらない・全体が見えない物事について、特定の条件のもとで何が起こるか、見通しを立てる。
  - 蓄積されたデータから、背後にあるパターン・法則性を発見。⇒ 予測に使う。
  - **機械学習**の主な目的は、予測。

- 経済学では：経済やファイナンス、マーケティングに関わる予測
  - 例：過去の来客データから、時間帯ごとの来客数を予測。
  - 例：膨大な信用データ（年齢、性別や過去の延滞）から、ローン申請者の破産確率を求める。
  - 例：株価収益率の時系列データから、ボラティリティ（分散）が拡大する時期を予測する。

- 3. データ分析による因果関係の評価
  - 近年、さまざまな場面で重視される**統計的エビデンス**。
  - 始まりは医療分野⇒医療以外の分野へ。
  - 新しい医薬品や治療法の効果を評価。

- 経済学では？ : 政府の政策や企業の経営・マーケティング戦略、社会的なイベントの効果を評価。
  - 例 : 公的職業訓練で、参加者の生産性(賃金)がいくら上がる？
  - 例 : 洗剤の割引販売で、洗剤の売り上げがどれだけ上がる？
  - 例 : オリンピック開催で、開催都市の地価や雇用がどれだけ影響を受ける？
  - 例 : 小学校の少人数教育は、児童の学力を向上させる？

# ○経済学とデータ分析の現状

- 近年の経済学：経済理論＋データ分析⇒社会経済・ビジネスの問題に対する「解」。
  - 経済学者の学術的関心⇌実務家・政策担当者の利益。
  - ∴問題解決型、実践的な「テクノロジー」の側面が強くなる。
  - 先進国（日本を除く）の経済学部では、計量経済学が必修科目。

- 世界的なTech企業で、優れた経済学者(≠経済評論家)が顧問や重役に。(2020年現在。)
  - ハル・ヴァリアン(Google／カリフォルニア大学バークレー校)
  - スーザン・アティ(Microsoft／スタンフォード大学)
  - パトリック・バジヤリ(Amazon／ワシントン大学)
  - ...いずれも経済学・計量経済学で画期的な研究業績。



- 上記企業は、統計学や計算機科学だけでなく、経済学の学位を持つ人材(学士～博士)を広く採用。

- 例: amazonの求人

[https://www.amazon.jobs/jp/job\\_categories/economics](https://www.amazon.jobs/jp/job_categories/economics)

<http://www.amazoneconomistjobs.com/>

# エコノミクス

36求人

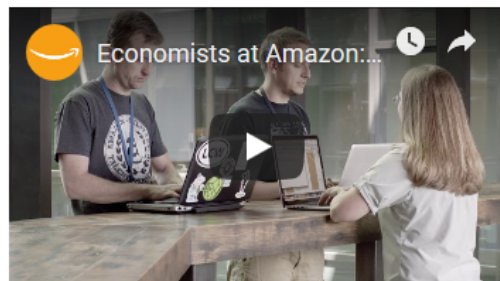
## データとテクノロジーを基に、戦略的な決断を導く

賢明な決断力が、自宅のガレージから始まったAmazonを、テクノロジー、リテール、コンテンツ産業のリーダーへと押し上げました。会社が成長するとともに、洗練された方法やシステムを使って、ビジネスが抱える問題に回答を出すことが求められています。エコノミストたちは、ビジネス、エンジニアリング、ファイナンス、そしてサイエンスのリーダーたちと協同し、膨大な量のデータから、世界中にいるお客様のために解決策を導き出します。

Amazonのエコノミストは、科学的な手法で、アマゾンの組織全体が直面するキービジネスでの問題のために、複雑で大規模なモデルを開発します。キービジネスとは、リテール、クラウドコンピューティング、サーチ、Kindle、ビデオストリーミングやオペレーションです。バックグラウンドが、応用ミクロ経済理論、応用マクロ経済理論、実証産業組織論、理論経済学、またはその他の専門分野に関わらず、膨大なデータを目的達成のため、学んだことを活用できます。ビジネス組織の中、セントラル・エコノミクス・チーム内に、エコノミストのポジションが存在します。Amazon内のどの部署で働いても、チーフエコノミストのPat Bajariや、博士号を持った才能あるエコノミストチームと共に、テクノロジーやEコマースにおける難しい問題に立ち向かいます。

Amazonでは、ビジネスへ多大なインパクトをもたらす問題に対して、学んだスキルを活用する機会が提供されます。そして、分析のためのクオリティの高いデータを生み出す苦勞はありません。

## Economists at Amazon



## Economists at Amazon

When you operate at the scale of Amazon, even

## Find jobs in エコノミクス

### フィルタ内容

36件中1 - 10件の仕事を表示

並べ替え：最も関連 ▼

仕事の種類 <span>▲</span>
フルタイム (35)
パートタイム (1)
仕事のカテゴリー <span>▲</span>
エコノミクス (36)
都市 <span>▲</span>
Seattle, Washington, USA (29) <input type="checkbox"/>
Palo Alto, California, USA (2) <input type="checkbox"/>
Tokyo, JPN (2) <input type="checkbox"/>
London, England, GBR (1) <input type="checkbox"/>
New York, New York, USA (1) <input type="checkbox"/>
Vancouver, British Columbia, CAN (1) <input type="checkbox"/>
事業カテゴリー <span>▲</span>
Human Resources (15) <input type="checkbox"/>
Amazon Web Services (3) <input type="checkbox"/>
ファイナンスおよびアカウンティング (3) <input type="checkbox"/>

### Economist

送信November 7, 2018

(更新5 days前)

US, WA, Seattle | 職種ID:744718

Economists at Amazon will be expected to work directly with our Chief Economists and senior management on key business problems faced in retail, international retail, cloud computing, third party merchants,[もっと読む](#)

### Economist

送信November 2, 2018

(更新10 days前)

US, WA, Seattle | 職種ID:741873

Amazon.com strives to be Earth's most customer-centric company where people can find and discover anything they want to buy online. We hire the world's brightest minds, offering them a fast paced, technologically[もっと読む](#)

### Economist

送信November 2, 2018

(更新10 days前)

US, WA, Seattle | 職種ID:741872

Amazon.com strives to be Earth's most customer-centric company where people can find and discover anything they want to buy online. We hire the world's brightest minds, offering them a fast paced, technologically[もっと読む](#)

### Senior Economist, Amazon Flex

送信October 31, 2018

(更新12 days前)

US, WA, Seattle | 職種ID:739527

Amazon aims to exceed the expectations of our customers by ensuring that their orders, no matter how large or small, are delivered as quickly, accurately, and cost effectively as possible. To meet this[もっと読む](#)

# ○経済・ビジネスにおけるビッグデータ利用

## ・ビッグデータの利用可能性

- ・IT技術の進展⇒データの収集・保存・計算能力が爆発的に成長。
- ・コンピュータ・サイエンスや統計学、工学など、分野を超えた研究。
- ・経済学：2010年頃より、ビッグデータの利用に関する本格的な研究が始まる。

- ビッグデータの特徴

- データのサイズ(サンプル数、変数の次元)が大きい⇒通常のソフトウェアによる管理や計算が困難。
- リアルタイムで観測: ネット検索やSNS、オンライン取引など、インターネット通信の副産物。
- 文字・数字だけでなく画像や映像も含む。
- 機械学習による予測。“Forecast”から“Nowcast”へ。

- 経済学におけるビッグデータの利用: 2010年頃より、本格的な研究が始まる。∴経済学の先端領域のひとつ。
  - 計量経済学: 従来は古典的統計学が中心。⇒機械学習・AI技術の応用へ。
  - おすすめの文献(英語): Varian(2014)、Athey(2017)、Athey and Luca(2019)、Taddy(2019)。
  - おすすめの文献(日本語): 照井(2018)、スティーヴンズ=ダヴィドウィッツ(2018)、タディ(2020)。

- 本日の目的: 経済学・ビジネス分野におけるビッグデータの利用例を紹介。
  - ビッグデータによる、社会経済変数の予測
  - ビッグデータによる、因果関係の評価

## 2. 社会経済変数を予測する



# ○社会経済をビッグデータで予測する

- 多くの重要な経済変数は、個人をベースに計測される。
  - 所得、資産、消費支出、就業状態、etc。
  - 伝統的に社会調査や、政府の持つ管理データ(行政データ)により計測。
  - 例:総務省統計局の労働力調査⇒「2017年9月の失業率は2.5パーセント」。

- 社会調査データ・政府管理データの欠点。
  - 調査実施に高額なコスト。
  - 観測の頻度が少ない。年次、四半期、月次。
  - 調査者の信頼性：大規模な調査を委託。⇒不正な手続きによる調査の可能性。
  - 回答者の信頼性：非協力、虚偽報告、適当に答える、など。

- **メディアデータ**に着目：テレビ視聴、ネット・携帯電話通信、IoT、etc。
  - スマートフォンの通信記録やネットの履歴⇒個人の移動、社会関係、消費パターンを色濃く反映。
  - 調査データ・管理データと異なり、リアルタイム・高頻度で観察される。
  - 副産物としてのデータ。∴追加的なコストは少ない。
  - 経済変数の、「早くて安い」予測が可能では？

- 機械学習(回帰分析)による予測
  - $Y$  = 予測したい変数(所得)。⇒観測が高コスト・低頻度。
  - $X = \{X_1, X_2, \dots, X_K\}$  = 予測の材料(通話記録や位置情報、購入履歴、支払い履歴など)。⇒観測が低コスト・高頻度。
  - $X$ と $Y$ のデータから、予測の方程式・回帰式
$$\hat{Y} = a + b_1X_1 + b_2X_2 + \dots + b_KX_K$$
の係数 $\{a, b_1, b_2, \dots, b_K\}$ および必要な変数を**学習**(推定)。
  - いったん予測式ができれば、 $X$ から $Y$ を推定できる。∴無理に $Y$ を観測しなくてよい。

- いくつかの実例

- Choi and Varian (2012): Googleの検索データ(Google Trends)⇒さまざまな国の、景気や経済活動に関する時系列を予測。
- Blumenstock, Cadamuro, and On (2015): 携帯電話の通信データ⇒ルワンダの地域別資産を予測。(のちに詳しく紹介。)
- Engstrom, Hersh, and Newhouse (2017): 衛星の高解像度画像⇒スリランカの地域別貧困率を予測。

# ○ケーススタディ: 通話記録で資産の予測

- Blumenstock, Cadamuro, and On (2015)より。
  - $X$  = 携帯電話会社の保有する、膨大な通話記録: だれが、どれだけ、いつ、だれに、どこで。
  - 小規模な電話調査で、個人の資産を尋ねる。⇒資産 $Y$ と通話記録 $X$ を紐づけ。
  - 個人の資産を予測する回帰式を作る。⇒資産 $Y$ を予測。

- 表1: データの概要。
  - 政府の実施した (Demographic and Health Survey、DHS): 資産状況を回答者に直接尋ねる、伝統的な社会調査のアプローチ。
  - DHSと比べ、携帯会社の調査は小規模。⇒ただし通話記録は膨大。

# 表1: データの要約

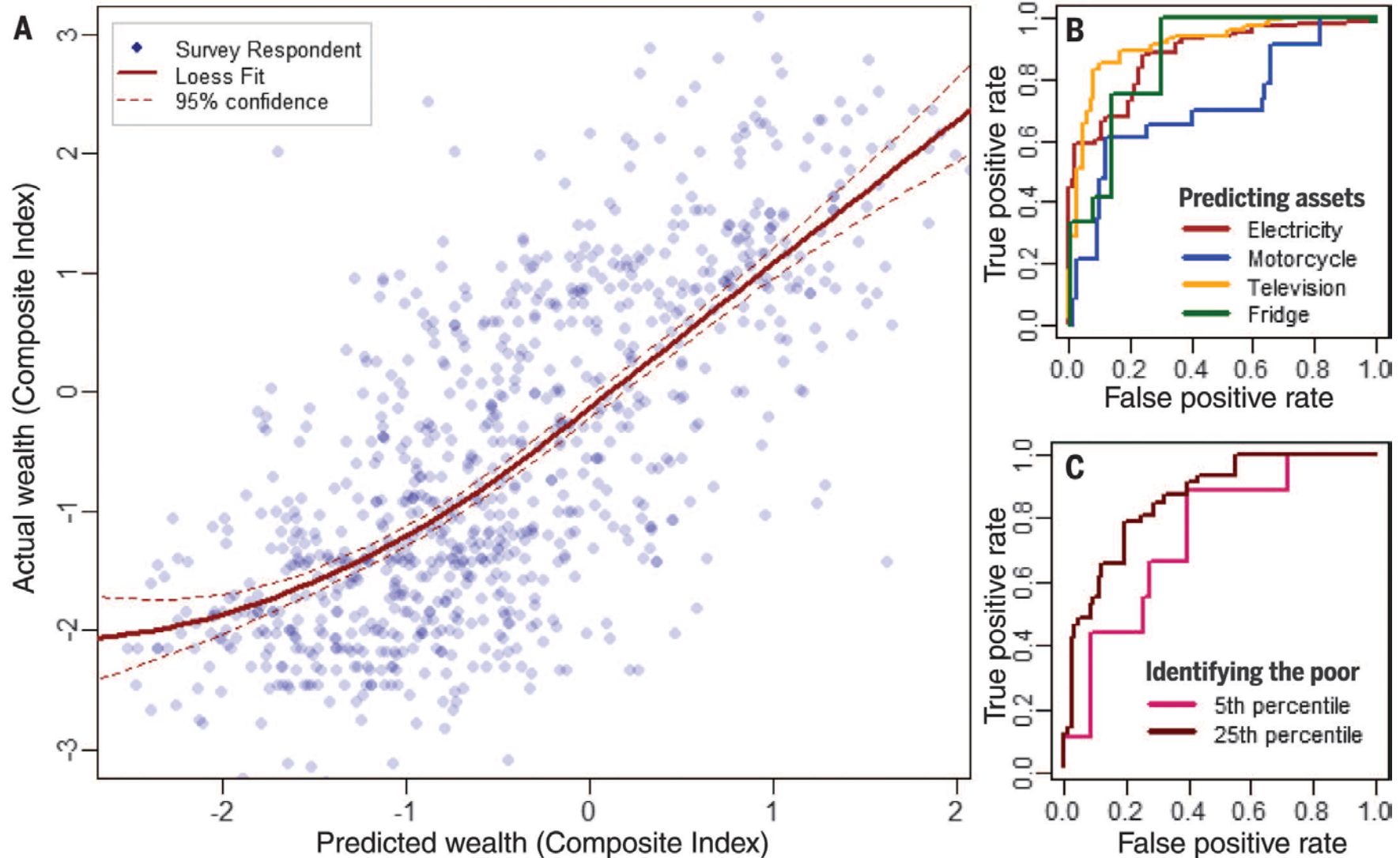
**Table 1. Summary statistics for primary data sets.** Phone survey data were collected by the authors in Kigali, in collaboration with the Kigali Institute of Science and Technology. Call detail records were collected by the primary mobile phone operator in Rwanda at the time of the phone survey. Demographic and Health Survey (DHS) data were collected by the Rwandan National Institute of Statistics. N/A, not applicable.

Summary statistic	Phone survey	Call detail records	DHS (2007)	DHS (2010)
Number of unique individuals	856	1.5 million	7377	12,792
Data collection period	July 2009	May 2008–May 2009	Dec. 2007–Apr. 2008	Sept. 2010–Mar. 2011
Number of questions in survey	75	N/A	1615	3396
Primary geographic units	30 districts	30 districts	30 districts	30 districts
Secondary geographic units	300 cell towers	300 cell towers	247 clusters	492 clusters

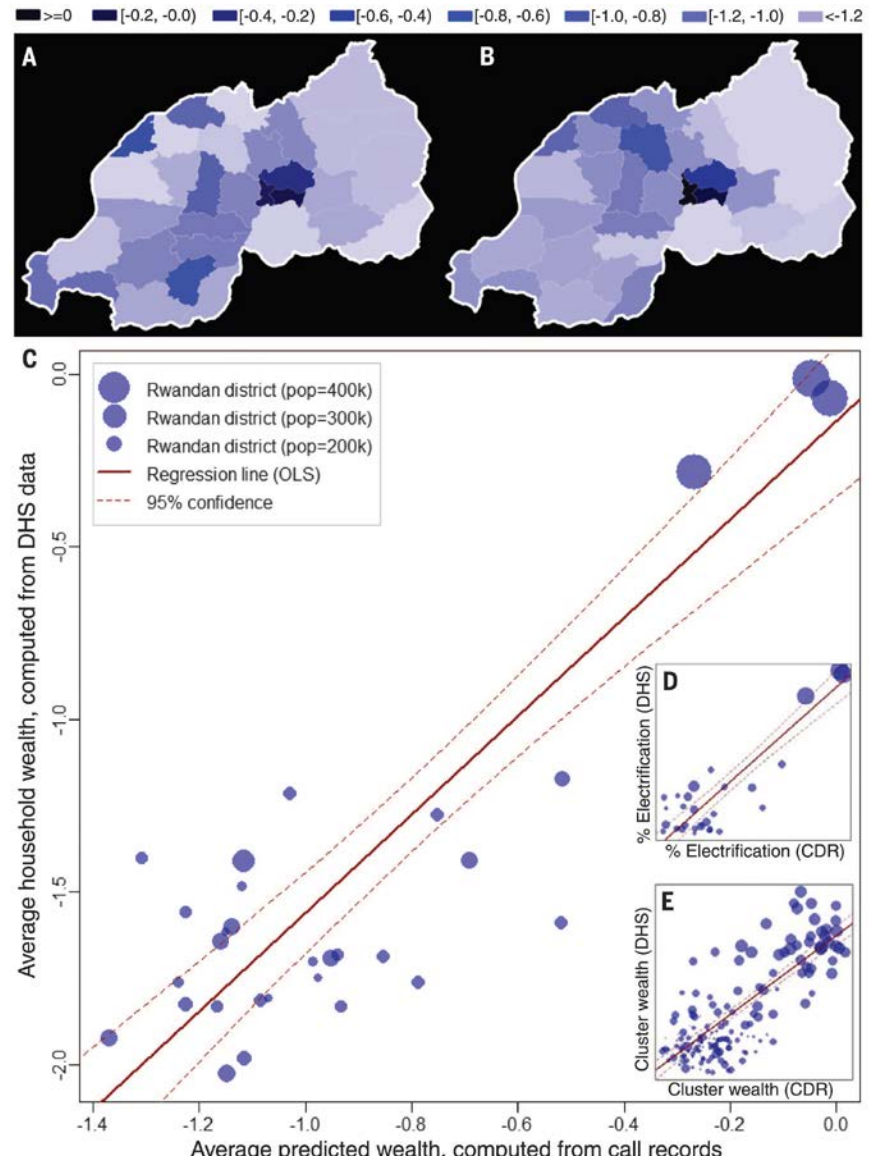


- 図1: 予測された資産(横軸)と実際の資産(縦軸)の関係。⇒予測値で、実測値をよく補足している。
  - この研究の特徴: 通話記録の予測値と、伝統的な調査による実測値を比較し、「答え合わせ」。
  - 図2: 通話記録で予測された資産(パネルA)と、調査データに基づく資産(パネルB)の、地域的な分布。⇒非常に良く似た結果。

# 図1：資産の予測値（横軸）と実測値（縦軸）



# 図2: 予測値 (A) と調査 (B) の比較



# 3. 予測から因果関係へ

# ○プログラム評価：因果関係の推論

- 機械学習の目的：予測と分類
  - 基本的に、観測されたデータ(学習データ)のパターンを利用。
  - パターン=変数間・観測間の相関関係。
  - 相関関係は、因果関係にあらず。∴機械学習だけでは、因果関係の分析はできない。
  - **統計的因果推論**：データから、変数間の因果関係を検証。

- **プログラム評価**: 政府や私企業が行った介入 (intervention) や、社会的なイベントが、人々の行動やアウトカムに与える影響を評価。
  - 例(再掲): 公的職業訓練で、参加者の生産性(賃金)がいくら上がる?
  - 例(再掲): 洗剤の割引販売で、洗剤の売り上げがどれだけ上がる?
  - **処置** = 介入・イベントの発生。(医療分野の研究に由来。)
  - 重要なパラメータ = 平均処置効果。

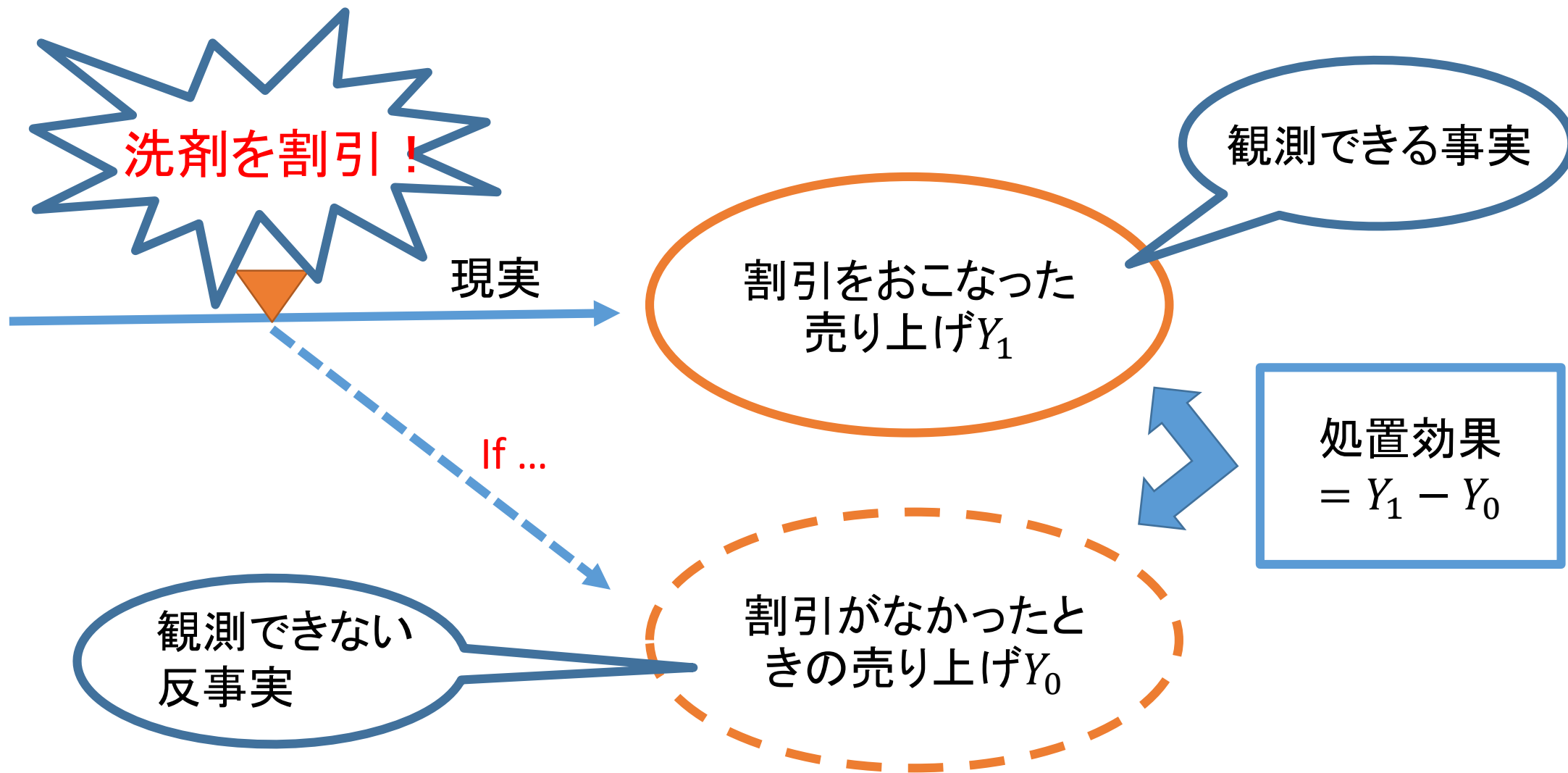
- 「洗剤の割引が売り上げに与える効果」を例に、プログラム評価の枠組みを考える。
  - $Y_1$  = 処置(割引)があった場合の売り上げ。
  - $Y_0$  = もしも処置(割引)がなかった場合の売り上げ。
  - $D$  = 処置を受けた・受けないを示すダミー(受けたなら  $D = 1$ 、そうでなければ  $D = 0$ )。
  - 現実には割引があったので、 $D = 1$ 。

- 割引が売り上げに与える**平均処置効果** (average treatment effect、**ATE**)は、

$$\begin{aligned} \text{ATE} &= \text{処置ありの売り上げ平均} - \text{仮に処置がない売り上げ平均} \\ &= E(Y_1|D = 1) - E(Y_0|D = 1) \end{aligned}$$

- $E(Y_1|D = 1)$ は現実の売り上げ平均。⇒データから推定可能。
- 一方 $E(Y_0|D = 1)$ は、直接推定できない**反事実** (counter-factual)。
- ∴単に売り上げのデータだけでは、ATEを計測できない！





- プログラム評価の課題: いかにして反事実  $E(Y_0|D=1)$  を埋めるか。
  - **無作為化実験**: あらかじめ処置あり ( $D=1$ )・処置なし ( $D=0$ ) を、ランダムに被験者に割り当て。⇒  $D=0$  グループの平均  $E(Y_0|D=0)$  を、反事実  $E(Y_0|D=1)$  とする。
  - 両グループは、処置を除けば均一・比較可能。

- 注意：実験データではない場合、 $D = 1$ グループと $D = 0$ グループの性質が、均一でない可能性。
  - 例：先月の売上げが落ちている店舗で、率先して割引。⇒割引の前に、両グループにすでに違いが発生。
  - 処置効果か、元々の性質の違いによる差か、区別できない！
  - ∴ 平均処置効果の推定＝「よい比較対象（反事実）」を用意すること。
  - 参考文献：伊藤（2017）、スミス（2019）、安井（2020）。

# ○ケース・スタディ: 移民と失業率

- 1979年から1981年ごろ、フロリダ州マイアミにキューバからの難民が押し寄せる。
  - 難民(=安価な労働力)は、受け入れ地域の住民の職を奪い、失業率が上がる可能性。
  - 確かにこの時期、マイアミの失業率は1.3%上昇(8.3%⇒9.6%)。
  - ポイント:「難民が来なくとも」マイアミの失業率は上昇していた?  
⇒難民受け入れ後の失業率と受け入れ前の失業率の差自体は、あまり意味がない。

- Card (1990) : マイアミと類似の社会経済属性・規模で、かつ1979年から1981年にかけて何も大きな出来事の無かった都市に着目。
  - 比較都市 : アトランタ、ヒューストン、ロサンゼルス、タンパ。
  - これら比較都市の失業率を、「仮に難民が押し寄せなかったマイアミの失業率」(=反事実)として使う。

- 表2: 比較都市(≡難民が来なかったマイアミ)では、失業率が2.3%上昇。
  - ∴難民受け入れが失業率に与える平均処置効果は $1.3 - 2.3 = -1\%$ 。
  - 難民により、むしろマイアミの失業率が低下。
  - **差分の差分法**(Difference in differences、DID)。政策・イベントの効果を評価するための、基本的な手法。

## 表2: マイアミと比較都市の失業率変化

	A: 1979	B: 1981		B-A	C-D
マイアミ	8.3	9.6	C: マイアミ分差	1.3	
比較都市	10.3	12.6	D: 比較都市差分	2.3	-1

# ○ビッグデータとDID

- Blake, Nosko, and Tadelis (2015) : 古典的な手法を、ビッグデータに適用。
  - eBay(ネット・オークション会社)で、paid search ad(検索広告)が販売収益に与える影響の評価。
  - アメリカの一部地域で試験的に、検索広告を停止。⇒停止地域と通常運営地域の、停止前後の収益を比較。
  - 通常運営地域=停止地域の反事実。
  - 結論: 広告は、効果がない。...ライトユーザーは広告に反応するが、ヘビーユーザーは反応しない。

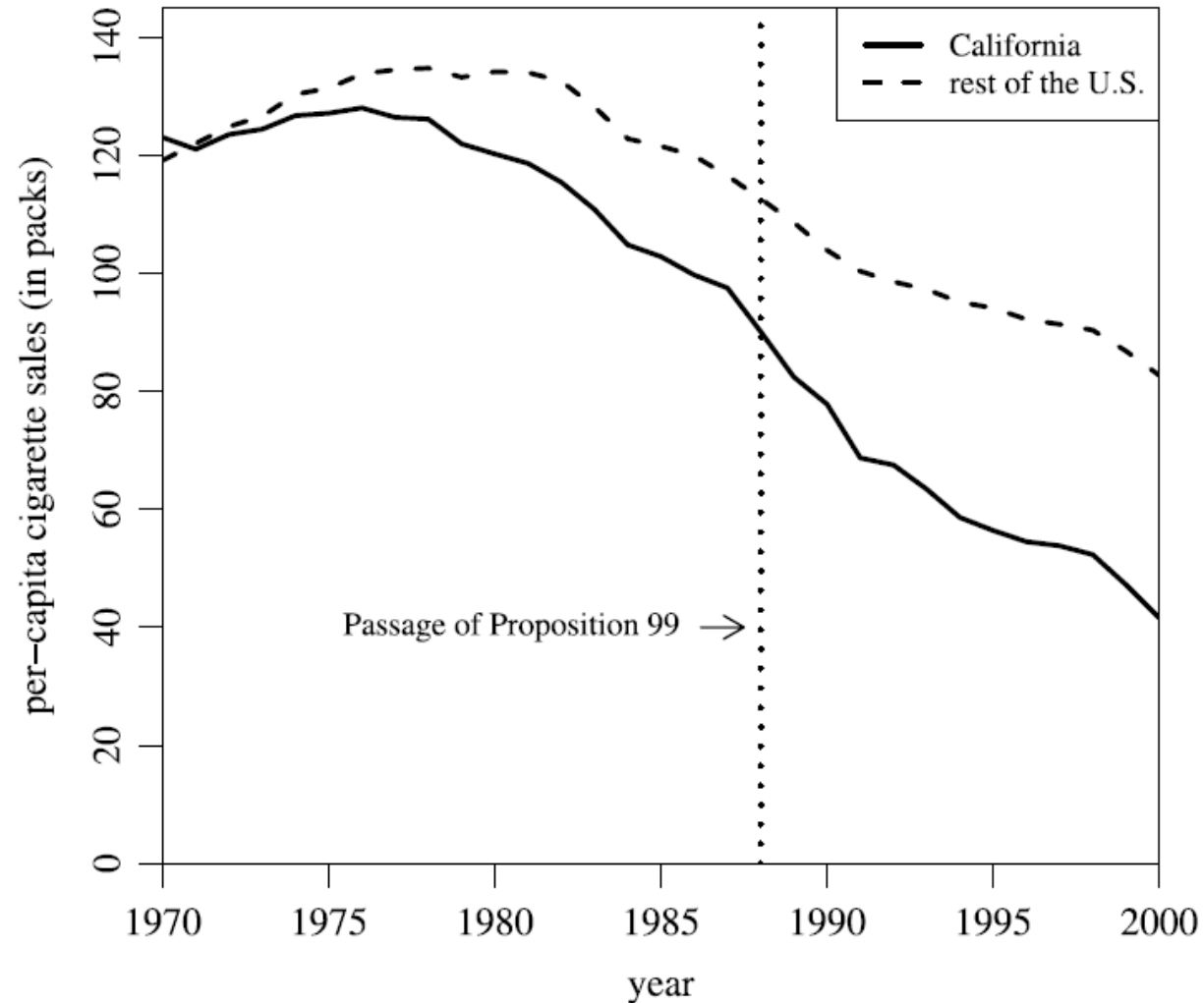


# ○人造コントロール法

- **人造コントロール法** (synthesis control method)。
  - Abadie, Diamond, and Hainmueller (2010) : 機械学習に基づく、新しい分析手法を開発。
  - 処置(政策・イベント)が起こった地域の反事実を、処置のない地域から推定。

- 1988年にカリフォルニアで施行された「Proposition 99」、大規模なタバコ規制。
  - 「Proposition 99」が、カリフォルニアのタバコ売上げに与えた影響を評価したい。
  - 図3:カリフォルニアとアメリカ全土のタバコ売上げ推移。
  - 規制の前と後の比較は無意味。⇒規制がなくとも、タバコの売上げは減少していた可能性。

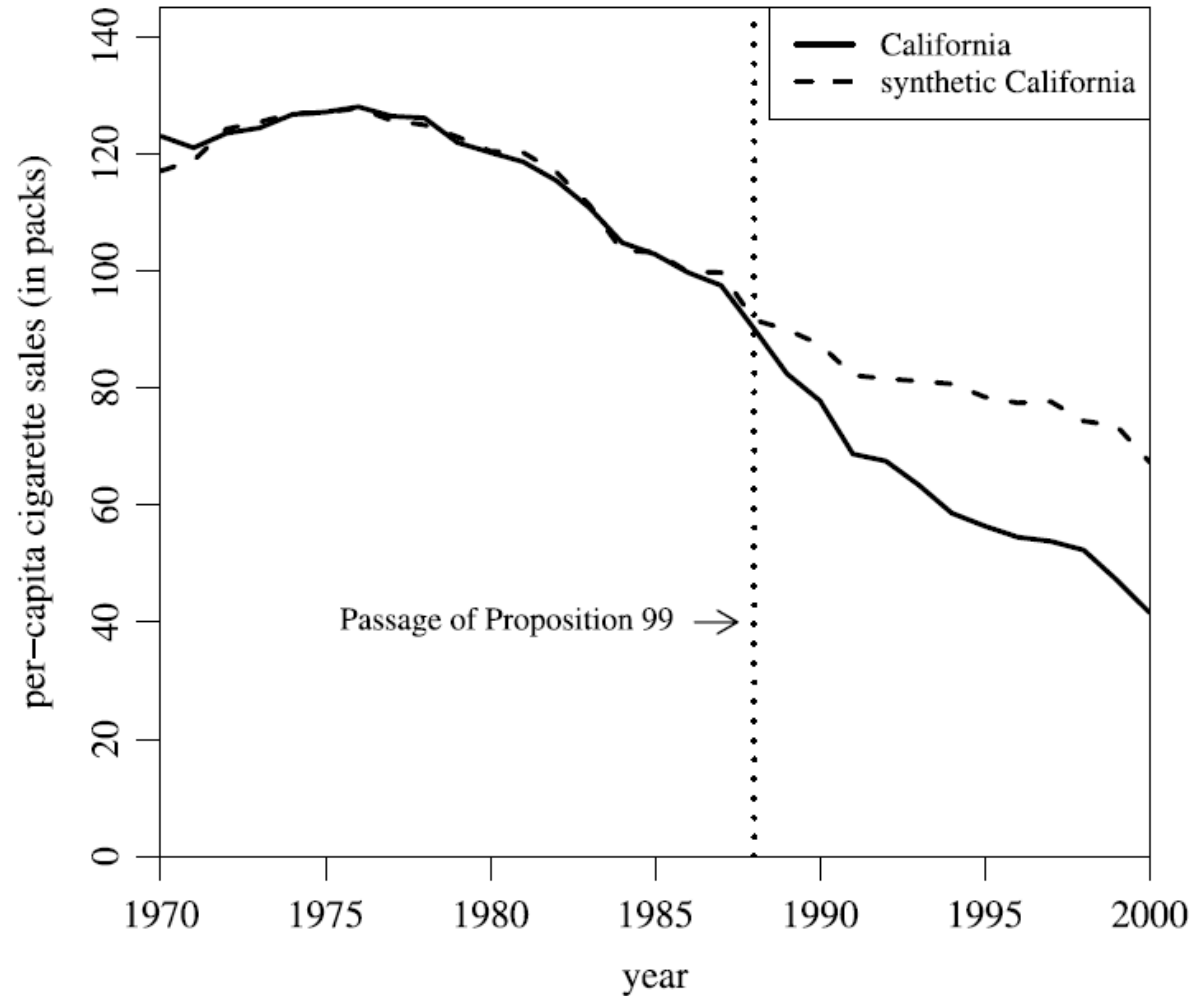
# 図3: カリフォルニアのタバコ売上げ推移



- 図4: ADHの分析結果

- カリフォルニア以外の州のタバコ売上げデータから、「仮に規制がなかった場合のカリフォルニア」(synthetic California)を生成。
- 「Proposition 99」法案通過前: 現実のカリフォルニアとsynthetic Californiaは同じレベルで変動。
- 法案通過後: 現実のカリフォルニアとsynthetic Californiaが乖離。
- 結論: 「Proposition 99」は、カリフォルニアのタバコ売上げを低下させた。

# 図4: 現実vs.人造カリフォルニア



# 4. 結語

# ○今後の展望

- 経済学のビッグデータ研究
  - 当面の目的＝機械学習による予測を超えた、因果関係の推論。
  - 因果関係の推論は、経済学が得意とする問題。⇒需要曲線と供給曲線、大規模マクロ連立方程式モデルの識別、etc。
  - 経済理論と、現実の政策・ビジネスに関する見識の重要性。

- ビッグデータで人間への理解が深まる
  - 機械学習・人工知能による判断と人間(プロフェッショナル)の判断を比較。⇒人間特有の偏り・偏見・非合理性が明らかに。
  - Kleinberg, et al.(2018)。



## ■ 参考文献 (英語)

1. Abadie, A., Diamond, A., & Hainmueller, J. (2010). Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program. *Journal of the American statistical Association*, 105(490), 493-505.
2. Athey, S. (2017). Beyond prediction: Using big data for policy problems. *Science*, 355(6324), 483-485.
3. Athey, S., & Luca, M. (2019). Economists (and economics) in tech companies. *Journal of Economic Perspectives*, 33(1), 209-30.
4. Blake, T., Nosko, C., & Tadelis, S. (2015). Consumer heterogeneity and paid search effectiveness: A large-scale field experiment. *Econometrica*, 83(1), 155-174.

5. Blumenstock, J., Cadamuro, G., & On, R. (2015). Predicting poverty and wealth from mobile phone metadata. *Science*, 350(6264), 1073-1076.
6. Card, D. (1990). The impact of the Mariel boatlift on the Miami labor market. *ILR Review*, 43(2), 245-257.
7. Choi, H., & Varian, H. (2012). Predicting the present with Google Trends. *Economic Record*, 88, 2-9.
8. Engstrom, R., Hersh, J., & Newhouse, D. (2017). Poverty from space: Using high-resolution satellite imagery for estimating economic well-being.

9. Kleinberg, J., Lakkaraju, H., Leskovec, J., Ludwig, J., & Mullainathan, S. (2018). Human decisions and machine predictions. *The quarterly journal of economics*, 133(1), 237-293.
10. Taddy, M. (2019). *Business Data Science: Combining Machine Learning and Economics to Optimize, Automate, and Accelerate Business Decisions*, McGraw-Hill Education
11. Varian, H. R. (2014). Big data: New tricks for econometrics. *Journal of Economic Perspectives*, 28(2), 3-28.

## ■参考文献(日本語)

1. 伊藤公一朗(2017)『データ分析の力 因果関係に迫る思考法』、光文社
2. ゲアリー・スミス(2019)『データは騙る: 改竄・捏造・不正を見抜く統計学』、早川書房。
3. セス・スティーヴンズ=ダヴィドウィッツ(2018)『誰もが嘘をついている ビッグデータ分析が暴く人間のヤバい本性』、光文社。
4. 照井伸彦(2018)『ビッグデータ統計解析入門 経済学部/経営学部で学ばない統計学』、日本評論社。

5. マット・タディ(2020)『ビジネスデータサイエンスの教科書』、すばる舎。
6. 安井 翔太(2020)『効果検証入門 正しい比較のための因果推論/計量経済学の基礎』、技術評論社。