

# リサーチテーマとしての ビッグデータと機械学習



一橋大学大学院 経営管理研究科  
准教授 宮川大介

Dec. 5<sup>th</sup>, 2020

日本ファイナンス学会秋季研究大会  
「オルタナティブデータによるファイナンス研究の新潮流」

# 報告の目的

## □ 学術研究の観点から

「ビッグデータ」

定義・利用範囲

と

「機械学習手法」

用途の確認

を用いた

「経済学・ファイナンス研究の展望」

実証研究の  
方向性×4

(必要な範囲で)  
+ 実務との関係

を議論する (実務的な応用例は次セッション)

# 構成

1. 現状の整理

2. 四つの方向性

3. 私的展望

4. 事例



「大勢」に影響なし？

# 1. 現状の整理

## □ 大規模・高次元・高頻度のビッグデータ利用が拡大

研究の方向性

- 経済主体レベルの大規模パネルデータ
- 関係を記述したネットワークデータ
- 高頻度の価格データ
- テキスト・画像データ
- 特定のアクティビティ（のみ）を計測したexoticな（高頻度）データ

移動、人出、活動量、登記、決済、送金、入出金、返済・・・

## □ 計測（2a）と検知・予測（2b）の取り組み

- 要約・描写
- 教師有り学習による検知・予測
- 教師無し学習によるクラスタリング・異常検知

## □ 因果関係（2c）を意識した比較的新しい取り組み

- 教師有り学習の「亜種」による統計的因果推論（例：DML、HTE）
- 強化学習による方策関数の推定

## □ これらの技術が持つ「含意」の検討（2d）

## 2a. 計測

### □ リアルタイム計測の困難さ

- 経済主体の各種アクティビティ（例：売上高、退出、消費、移動）  
⇒ 人流、通信、決済、入出金、衛星画像
- 取引関係の中身、取引関係の生成・消滅、ショック波及  
⇒ 送金、物流

### □ 概念的な計測の困難さ

- センチメント、トーン  
⇒ テキスト

研究上のポイント  
見えなかったものを  
可視化する意義

### □ 異なる粒度で計測されたデータの取り扱い

- POSと年商
- 仕訳と決算
- ノードとネットワーク

⇒ 計測ニーズとデータ利用可能性 + 計測技術の進歩

## 2b. 検知・予測

- 教師有り学習の「主戦場」
  - 企業成長、企業退出
  - Bad event (例：不正会計、詐欺、犯罪、暴動)
  - 需要予測（「レコメンデーション」も含む）
  - 市場価格、相対取引価格

- 教師無し学習にとっても「主戦場」
  - パターン認識
  - クラスタリング
  - 異常検知

- 追加的な 이슈ー
  - Omitted payoff bias
  - 解釈可能性
  - 安定性、モデルの陳腐化



研究上のポイント  
ファクトの提示  
+  
経済価値の実現

⇒ 検知・予測ニーズとデータ利用可能性 + 成熟した技術

## 2c. 因果推論

### □ 新しい方向性

- Double/debiased ML  
⇒ 交絡要因の排除、CIA条件
- Heterogeneous Treatment Effects  
⇒ RCTとの組み合わせ  
⇒  $\tau(X)$ を推定する工夫  
⇒ 応用例：政策処置、個別プライシング、「要約」

例：COVID-19 & Firm exit  
(規範的な評価)

### □ 経済理論との接合

- 方策関数の推定
- 最適化問題の解  
⇒ ゲームAIの例

研究上のポイント  
理論的検討とのinteraction  
必然性のある利用

⇒ 社会科学分野での自然な応用 (の可能性)

## 2 d. 技術の含意

### □ データ×技術の影響（マクロ）

- ロボティクスの文脈
- 資本と労働の代替・補完
- 人的資本の意義
  
- ビジネスダイナミズム

研究上のポイント  
既存研究との接続  
新しい理論的視点  
頑健なエビデンス

### □ データ×技術の影響（ミクロ）

- タスク分配問題
- 高精度の予測が経済主体の行動・インセンティブに与える影響
  
- Fundamental attenuation

⇒ （単に）一つの新しい技術としての位置づけ

## 3. 私的展望

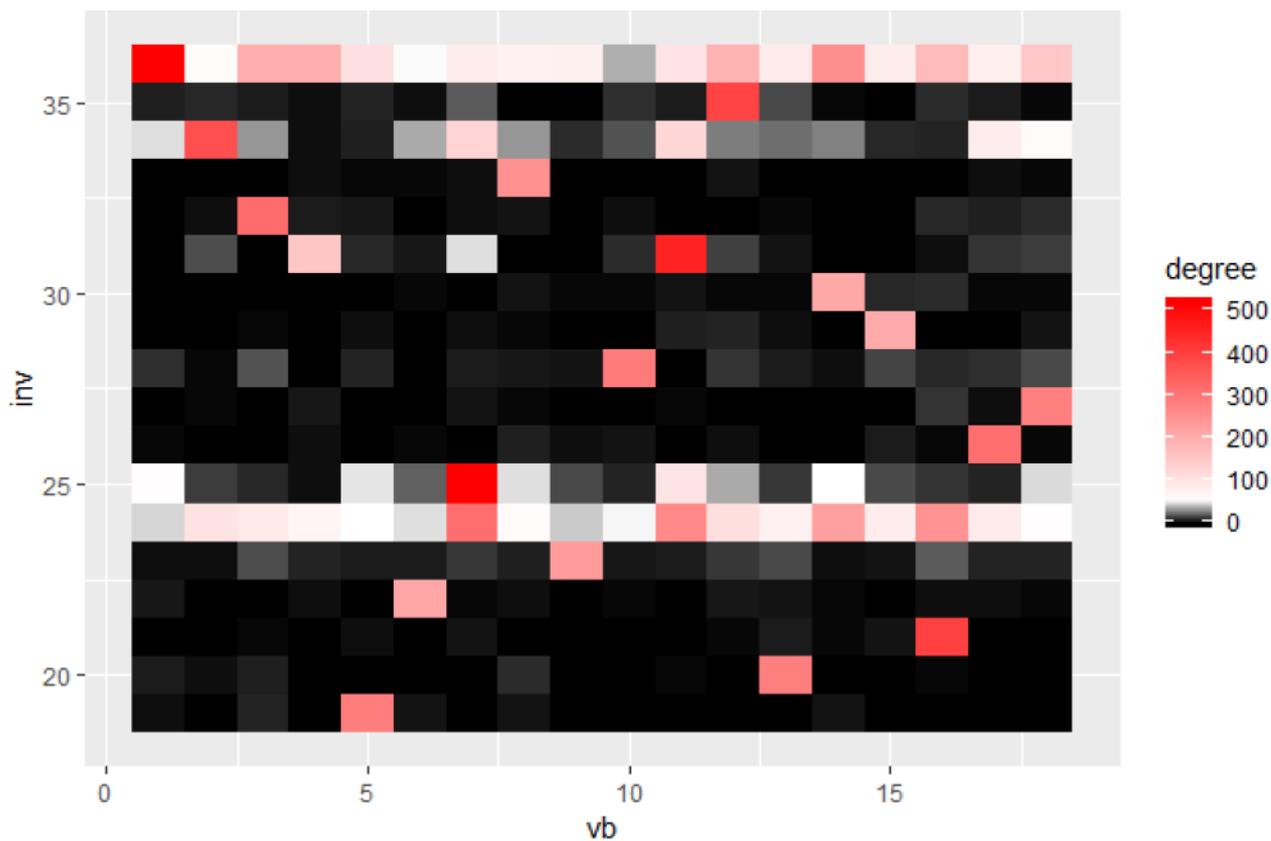
- 計測、検知・予測について
  - 既に標準的なツールとして広く用いられている
  - ビッグデータの価値を生かすために有効
  - 「Exoticデータとの掛け合わせ ⇒ 学術的な貢献」 ではない
  - 実務的要請と（狭義の）学術的意義の温度感
  - 分野における差異（investment vs. CF vs. Econ）
  
- 因果推論について
  - データと手法の利用に必然性が求められる
  
- 実務との関係？
  - データを含む研究資源
  - 代替不能な実験フィールド

標準的かつ必要な  
スキルセット  
(テーマの重要性！)

⇒ **より柔軟な学術的検討のための一つの追加的要素**

# 4. 事例

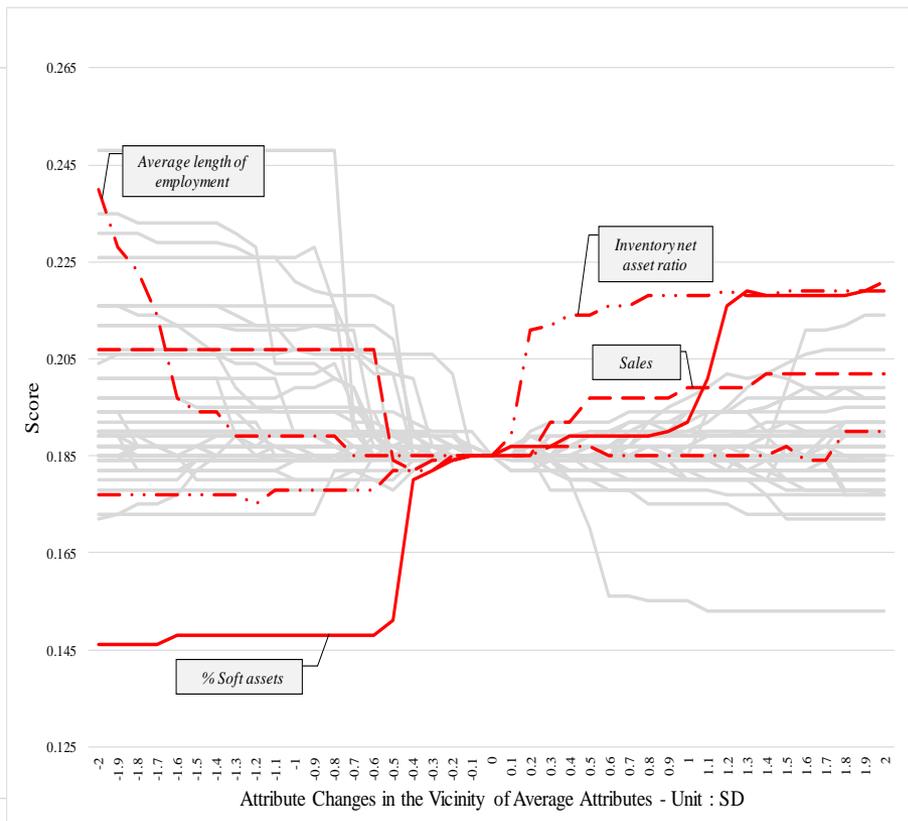
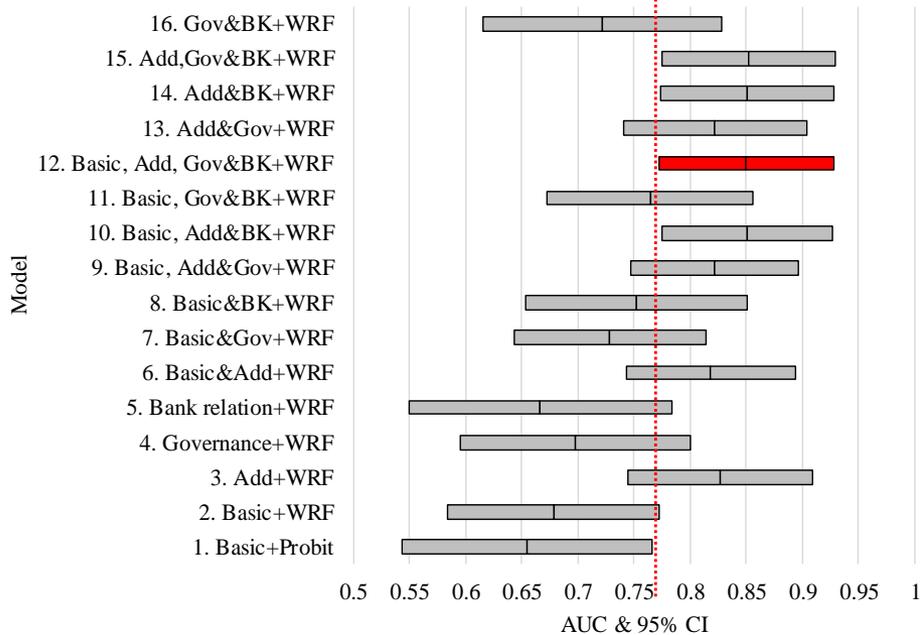
- 描写 : Koujaku & Miyakawa '19



# 4. 事例

□ 予測 : Kondo, Miyakawa, Shiraki, Suga, and Usuki '20

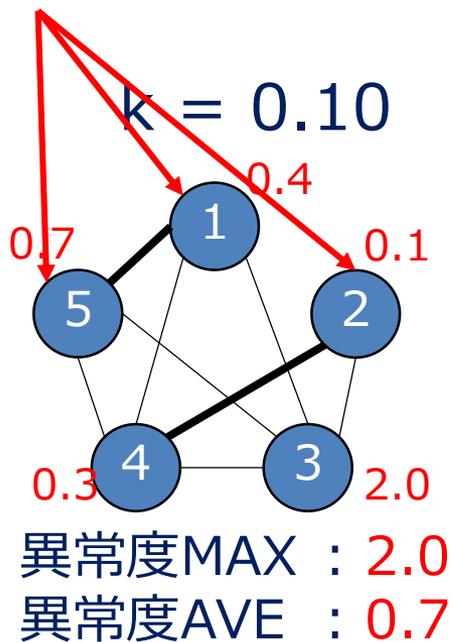
Prediction & Accounting Fraud Flag 2



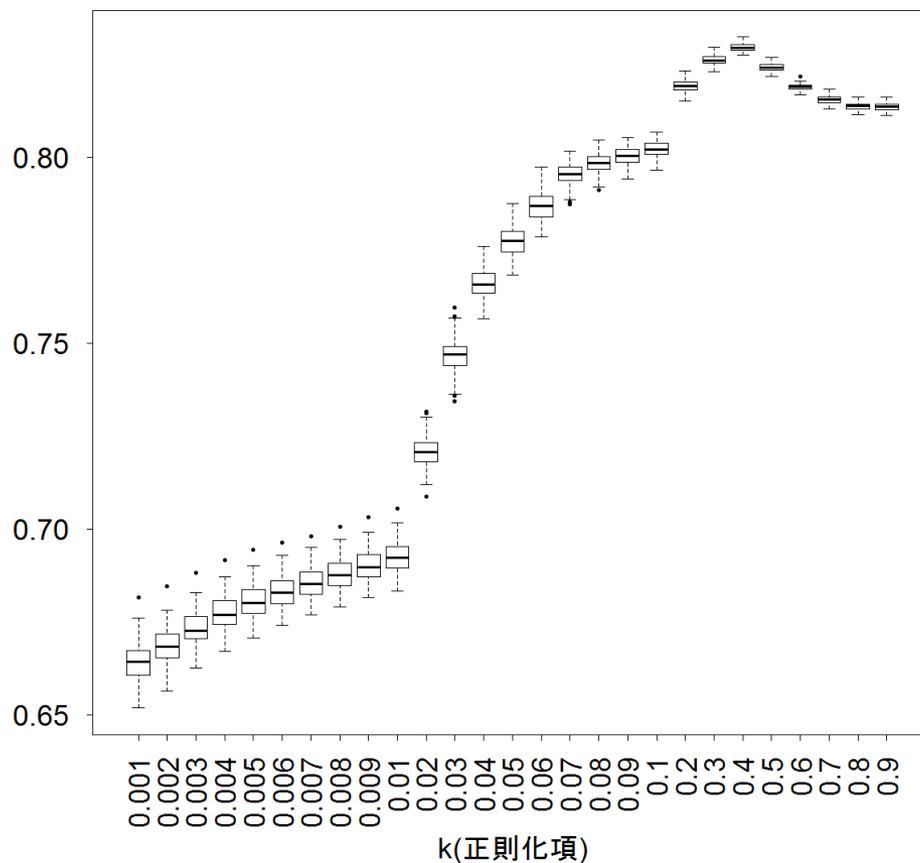
# 4. 事例

□ 検知：特許第667865号

勘定科目毎に  
異常計算

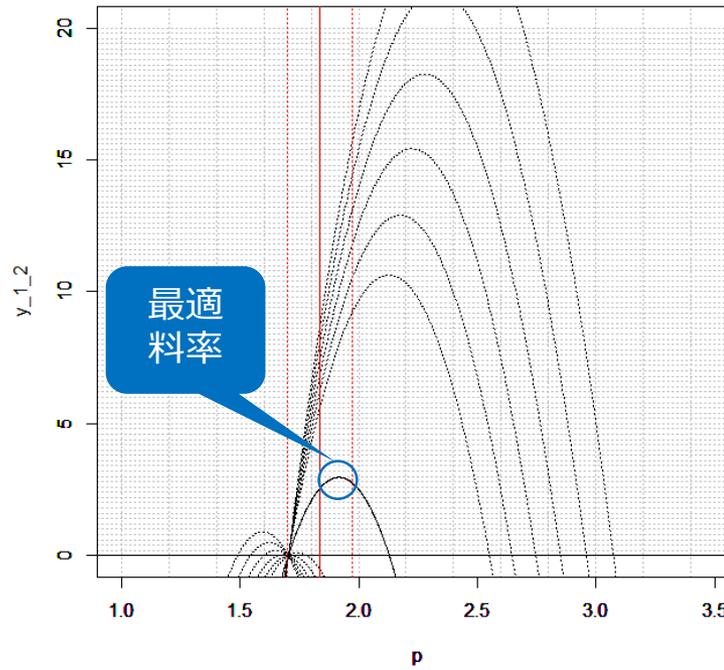
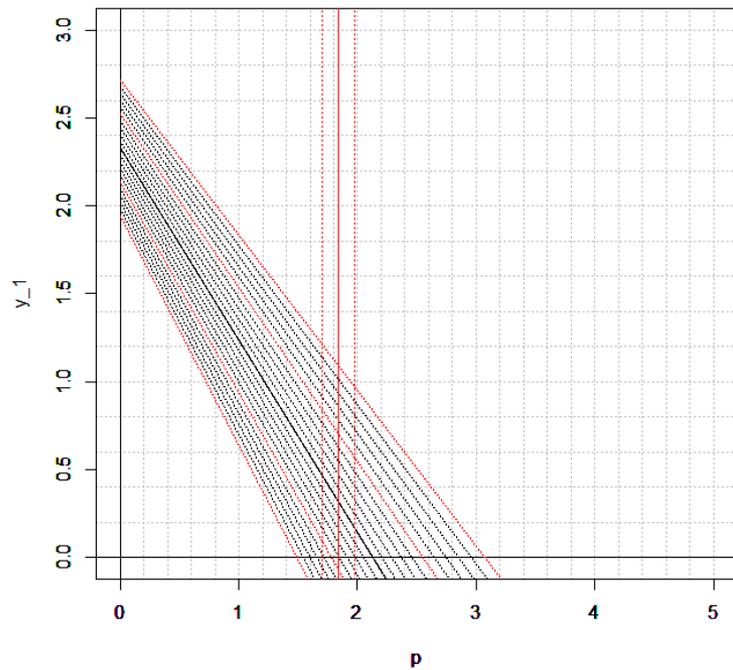


Sales\_Ave



# 4. 事例

## □ 因果推論：特許第4764178号



# 3. 事例

- 含意 : Miyakawa & Shintani '20 ⇒ 他分野への応用 (Algorithmic triage)

	<i>Prediction for default</i>			<i>Prediction for voluntary closure</i>		
	M = default H = not default (1)	M = not default H = default (2)	(2)/(1)	M = closure H = not closure (1)	M = not closure H = closure (2)	(2)/(1)
~20 %tile	49,117	23,068	0.47	25,206	19,453	0.77
20~40 %tile	36,094	54,446	1.51	28,326	23,667	0.84
40~60 %tile	37,362	46,368	1.24	28,370	28,134	0.99
60~80 %tile	33,409	39,218	1.17	20,249	30,962	1.53
80 %tile~	11,652	30,608	2.63	8,026	34,406	4.29

<Contact Information>

宮川大介 Daisuke Miyakawa:

一橋大学大学院経営管理研究科 (HUB) 金融戦略プログラム (FS)  
准教授

〒101-8439 東京都千代田区一ツ橋2-1-2 学術総合センター内

東京大学エコノミックコンサルティング株式会社 (UTEcon)  
チーフエコノミスト

〒113-0033 東京都文京区本郷7-3-1 国際学術総合研究棟 922号室

E-mail: [dmiyakawa@hub.hit-u.ac.jp](mailto:dmiyakawa@hub.hit-u.ac.jp)

Web: <https://sites.google.com/site/daisukemiyakawaphd/>