

# 制度的制約下におけるデータベース構造化、モデリング、モデル評価

## Database structuring, modeling, and model evaluation under institutional constraints

データ科学研究系 山下智志 (Satoshi Yamashita)

### 1. はじめに

私個人の今世紀に入って以降の研究活動の特徴は、社会ニーズを吸上げ、データの取得、理論構築、モデリング、モデル評価、社会実装まで一貫通貫的に行い、各ステップの整合性を重視することにあった。特に、秘匿性データや統合データなどのデータベースの構築を長期間にわたって実行してきた。また、社会実装を想定しているため、法律、条約、規制、会計ルールなどのコンプライアンス面を考慮しながらモデリングを行うところに特徴がある。以下、研究活動実績を紹介したい。

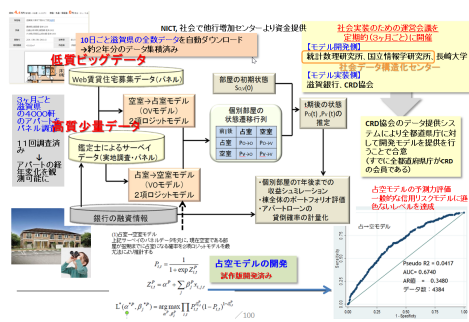
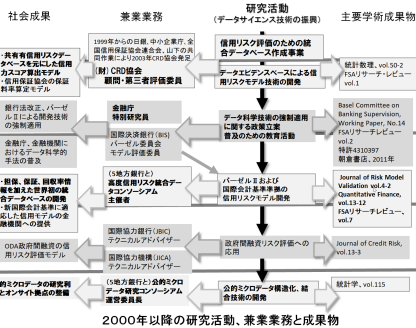
### 2. 統計モデル、データベース、政策の整合性と効率性

公開された研究業績は、いわゆる数理ファイナンスの分野や企業データ分析の方法論に関する著作が多い [1]。これは数理ファイナンス系方法論の研究が学術論文執筆と親和性が高く、書きやすいもしくは審査に通りやすいということが原因である。実態としては方法論構築に費やしたエフォートは、データベースの作成、高度化やニーズの引き上げや政策化に費やしたエフォートよりも少ないのが実感である。

特に、金融庁特別研究員を 10 年近く併任していたため、全世界の銀行の共通ルールであるバーゼル II (新 BIS 規制) [2] や銀行法改正に関する統計学的根拠の整理を長年続けていた。また、一般的な信用リスクモデルを改良したカントリーリスク計量化モデルは、国際協力機構 (JICA) や国際協力銀行 (JBIC) における政府間金融 (いわゆる ODA) の意思決定に利用されている。

データベース構築に関わる活動としては、日本最大の信用リスクデータベースである CRD 協会を日銀、中小企業庁、全国信用保証協会連合会との連携により 2003 年に設立し、現在も顧問を行っている。そこでは日本の企業の過半数である 160 万社の財務データを蓄積しているだけでなく、そのまた地方銀行とのデータ共有コンソーシアムを立ち上げ、世界で初めて秘匿性の高い与信データの統合を統計数理研究所内で実現した。これにより銀行間で与信構造に差異があることが確認された。現在は総務省との共同で、公的マイクロデータの統合化に取り組んでいる。以下、現在運営管理中のデータベースである。

- 公的マイクロデータ (オンサイト拠点): 政府統計のマイクロデータ (法人統計データ、労働統計データ) を利用するため、統数研内に政府データの窓口である (財) 統計センターと直接つながるオンサイト拠点
- 国際マイクロデータ: アジア 6 カ国の政府マイクロデータ (主として労働統計)



- 高度信用リスク統合データベース： 地方銀行5行（滋賀、群馬、八十二、伊予、北陸）の貸出先法人の全数データの統合データベース。与信情報と回収情報があるのは世界唯一。お互いに秘匿情報のため、分析方法に工夫が必要。
- CRD協会法人データベース：1996年より収集を開始した中小企業財務データベース（法人としては2003年から）。毎年約160万社、与信データベースとして日本最大である。
- 民間の購入データ（帝国データバンク、日経 Needs Company）2017年から開始した政府企業統計マイクロデータと民間信用データとの結合実験として帝国データバンクの企業情報（約3万件）がある。その他、民間の企業情報データとして日経 Needsと契約を続けている。

### 3. 賃貸住宅データを用いた、質の異なるデータベースの接続実験

様々なデータベースが増えるとともに、複数のデータベースを用いた統計モデルの構築方法が問題となる。特に質の異なるデータを利用するための方法論が必要とされている。具体的な研究としては、Webデータを定時観測（楽天不動産、SUUMOから10日ごとにデータをダウンロード）することによって大規模データベースを作成し、賃貸住宅の入居化要因を分析する。一方、Webデータの精度を補完するため、不動産鑑定士による現地パネル調査（賃貸住宅定点観測データ：約4400戸）を行った。それぞれのデータを統合することにより、より正確な賃貸住宅の収益予測を行う。2つの質の異なるデータベースを接続することにより、これまでリスク計量化モデルが考案されてこなかったアパートローン与信リスクについて、より正確な賃貸住宅の収益予測を行うことを目指している。現在、モデル精度は個人ローンのデフォルト予測と同等であり、実用化への具体的な準備が進んでいる。

### 参考文献

[1] Satoshi Yamashita, Toshinao Yoshiba, "A Collateralized Loan's Loss Under a Quadratic Gaussian Default Intensity Process", Quantitative Finance, vol.13-12, p.1935-1946, 2013年6月  
 [2] Satoshi Yamashita, 他 "Studies on the Validation of Internal Rating Systems", Basel Committee on Banking Supervision, Working Paper, No.14, 2005年5月