

統計数理 第67巻第1号

(通巻129号)

PROCEEDINGS OF THE INSTITUTE OF STATISTICAL MATHEMATICS

目次

特集「農林業の生態系サービスの経済・統計分析」

「特集 農林業の生態系サービスの経済・統計分析」について

田中 勝也・吉本 敦 1

日本における森林生態系サービスの経済評価 [原著論文]

佐藤 真行・栗山 浩一・藤井 秀道・馬奈木 俊介 3

農林業の生態系サービスと自然資本のグローバル分析 [研究ノート]

池田 真也・山口 臨太郎・馬奈木 俊介 21

外来種駆除の生物多様性保全効果：保全優先地域と脅威動態の関係 [原著論文]

楠本 聞太郎・南木 大祐・久保田 康裕 39

インターネット調査におけるモニター情報の利用による非回答バイアスの補正
—国内草資源を利用した放牧飼養牛肉に対する消費者評価への適用— [研究ノート]

楠戸 建・後藤 貴文・高橋 義文・矢部 光保 51

ミツバチの送粉サービスと景観構造との関係解析
—宮崎県綾町における日向夏生産の事例— [研究ノート]

光田 靖・湯村 昂広・平田 令子・伊藤 哲 63

関数データに基づく統計的モデリング [総合報告]

松井 秀俊 73

農地と森林の生態系サービスの経済評価手法 [研究詳解]

柘植 隆宏 97

個人企業向けクレジットスコアリングモデルにおける業歴の有効性 [原著論文]

尾木 研三・内海 裕一・枇々木 規雄 121

2019年6月

大学共同利用機関法人 情報・システム研究機構 統計数理研究所

〒190-8562 東京都立川市緑町 10-3 電話 050-5533-8500(代)

本号の内容はすべて <https://www.ism.ac.jp/editsec/toukei/> からダウンロードできます

ISSN 0912-6112

統
計
数
理

PROCEEDINGS OF THE INSTITUTE OF STATISTICAL MATHEMATICS

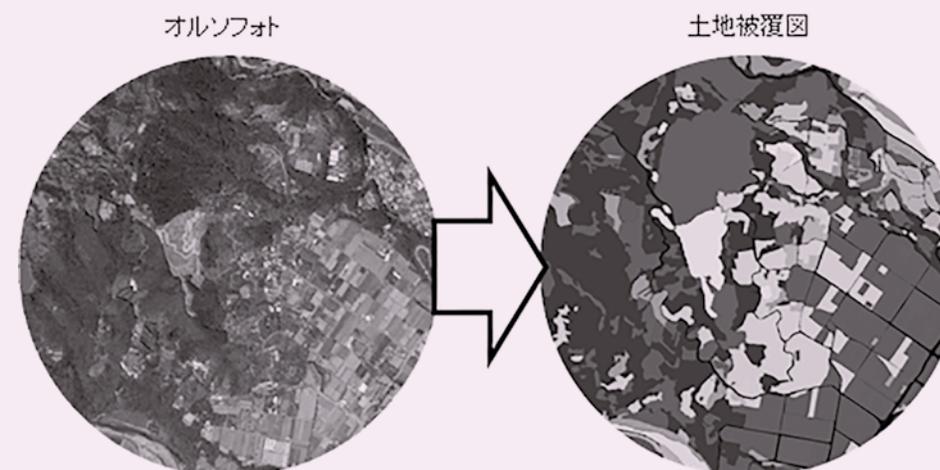
第
67
巻
第
1
号

2019

統計数理

Vol. 67, No. 1

PROCEEDINGS OF THE INSTITUTE OF STATISTICAL MATHEMATICS



土地被覆の凡例

果樹園	草地	人口緑地	河川
人工林	竹林	建物	河川敷
天然林	新植地	道路	
農地	伐採跡地		

統計数理研究所

統計数理

(年2回発行)

編集委員長 間野 修平
編集委員 伊藤 陽一
志村 隆彰
庄 建倉
中野 慎也
朴 堯星
特集担当編集委員 吉本 敦

編集室

池田 広樹 長嶋 昭子 脇地 直子

「統計数理」は、統計数理研究所における研究成果を掲載する統計数理研究所「彙報」として1953年に歴史を始め、1985年に誌名を変更し今の形となりました。現在は、統計数理研究所の研究活動に限らず、広く統計科学に関する投稿論文を掲載し、統計科学の深化と発展、そして統計科学を通じた社会への貢献を目指しています。

投稿を受け付けるのは、次の6種です。

- a. 原著論文
- b. 総合報告
- c. 研究ノート
- d. 研究詳解
- e. 統計ソフトウェア
- f. 研究資料

投稿された原稿は、編集委員会が選定・依頼した査読者の審査を経て、掲載の可否を決定します。投稿規程、執筆要項は、本誌最終頁をご参照ください。

また、上記以外にも統計科学に関して編集委員会が重要と認める内容について、編集委員会が原稿作成を依頼することがあります。

その他、「統計数理」に関するお問い合わせは、各編集委員にお願いします。

All communications relating to this publication should be addressed to associate editors of the Proceedings.

大学共同利用機関法人 情報・システム研究機構
統計数理研究所

〒190-8562 東京都立川市緑町10-3 電話050-5533-8500(代)

<https://www.ism.ac.jp/>

© The Institute of Statistical Mathematics 2019

印刷：笹氣出版印刷株式会社

PROCEEDINGS OF THE INSTITUTE OF STATISTICAL MATHEMATICS

Vol. 67, No. 1

Contents

Special Topic : Economic and Statistical Analysis on Ecosystem Services Induced from Agricultural and Forestry Activities

On the Special Topic “Economic and Statistical Analysis on Ecosystem Services
Induced from Agricultural and Forestry Activities”

Katsuya TANAKA and Atsushi YOSHIMOTO	1
Economic Valuation of Forest Ecosystem Service in Japan Masayuki SATO, Koichi KURIYAMA, Hidemichi FUJII and Shunsuke MANAGI	3
Global Analysis of Ecosystem Services from Agroforestry and Natural Capital Shinya IKEDA, Rintaro YAMAGUCHI and Shunsuke MANAGI	21
Effectiveness of Invasive Species Eradication Efforts on Biodiversity Conservation : Spatial Congruence between Conservation Priority Areas and Threat Buntarou KUSUMOTO, Daisuke NANKI and Yasuhiro KUBOTA	39
Adjusting for Non-response Bias Using Registration Data from an Internet Research Company —Application to Consumers' Evaluation of Japanese Grass-fed Beef— Takeru KUSUDO, Takafumi GOTOH, Yoshifumi TAKAHASHI and Mitsuyasu YABE	51
The Relationship between Pollination Service by the Native Honeybee (<i>Apis cerana</i>) and Landscape Structure —A Case Study of Hyuganatsu (<i>Citrus tamurana</i>) in Aya Town, Miyazaki Prefecture— Yasushi MITSUDA, Takahiro YUMURA, Ryoko HIRATA and Satoshi ITO	63
Statistical Modeling via Functional Data Analysis Hidetoshi MATSUI	73
Economic Valuation Methods for Ecosystem Services Provided by Agricultural Land and Forest Takahiro TSUGE	97

Paper

Effect of Firm Age in Credit Scoring Model for Loans to the Self-employed Kenzo OGI, Yuichi UTSUMI and Norio HIBIKI	121
--	-----

表紙の図は本誌 66 ページを参照

June, 2019

Research Organization of Information and Systems
The Institute of Statistical Mathematics
10-3 Midori-cho, Tachikawa, Tokyo 190-8562, JAPAN

「特集 農林業の生態系サービスの経済・統計分析」 について

田中 勝也¹・吉本 敦² (オーガナイザー)

農林業の舞台となる農地・森林は、単に農産物・材木などを生産する他にも、社会にさまざまな便益を提供している。日本学術会議(2001)によれば、それらは生物多様性保全機能、地球環境保全機能、土砂災害防止機能(土壌保全機能)、水源涵養機能、快適環境形成機能、保健・レクリエーション機能、文化機能など多岐に及んでいる。これらに農産物・林産物などを生産する機能(物質生産機能)をあわせて、生態系サービス(ecosystem services)と呼ばれている。この用語は専門家・実務家の間では広く用いられているが、一般に親しみやすいように、Díaz et al. (2018)は“nature’s contributions to people”とも呼んでいる。本特集では、前者の「生態系サービス」を用いることとする。

農地・森林は多くの機能を有しているにも関わらず、価格を伴い市場で取引されているのは物質生産機能などごく一部に限られており、その他の機能は市場メカニズムのもとでは十分に考慮されてない。このことが農地・森林に拘る様々な問題の引き金となり、例えば里山の劣化や農村部の荒廃、豪雨による災害リスクの増加など農山村の社会問題につながっている。これらの問題を改善するためには、農林業における生態系サービスの価値を適切に評価し、その結果を社会システムに内生化する必要がある。

こうした背景を踏まえ、本特集「農林業の生態系サービスの経済・統計分析」では、生態系サービスの評価・保全に関する統計手法の活用に関する計7編の論文から構成されている。その内訳は原著論文が2編、研究ノートが3編、総合報告および研究詳解が各1編である。以下に簡単に紹介する。

佐藤 他論文では、森林がもつ生態系サービスを評価する手法の理論的枠組を整理した上で、表明選好法(アンケートによる調査・分析)における代表的手法である仮想評価法(CVM)およびコンジョイント分析により、森林がもつ生態系サービスの価値を推定している。また池田・山口・馬奈木論文では、国連報告書が提供するグローバルな農林業データを活用して、地域ごとの生態系サービスの傾向を評価・分析している。

楠本・南木・久保田論文では、外来種の駆除と生物多様性の保全の問題について、沖縄北部におけるマングース捕獲事業を事例として分布確率の時空間動態を階層ベイズモデルにより推定している。その結果を基に、外来種の分布確率と保全優先地域を比較検討し、保全アクションの戦略的適用を検討している。

楠戸 他論文では、アンケートなどを用いる表明選好法における欠測データ(非回答者の情報)について考察している。牧畜における消費者調査を事例とした逆重み付き推定量による分析から、欠測データが従来の推定手法に与える影響は限定的であることを示している。

光田 他論文では、農業においてしばしば重要視されているミツバチによる送粉サービスについて、周囲の景観構造との関係性を解析している。ベイズ推定による解析の結果から、天然

¹ 滋賀大学 環境総合研究センター：〒522-8522 滋賀県彦根市馬場 1-1-1

² 統計数理研究所：〒190-8562 東京都立川市緑町 10-3

林面積が多くなるほど、また農地と草地の合計面積が多くなるほど、送粉サービスの水準が向上することを示している。

生態系サービスの分析手法に関するレビュー・詳解として、松井論文は、農林業における関数データ解析に基づく分析手法を紹介している。関連する手法の中でも回帰分析、時系列解析、空間データ解析を取り上げ、各手法の適用例を示しこれらの手法により得られる結果について説明している。また、柘植論文では、農林業の生態系サービスにおける価値評価に適用可能な評価手法として、代替法、ヘドニック価格法、トラベルコスト法、CVM、コンジョイント分析を取り上げ、その経済理論と推定方法について解説している。これらの手法は、本特集でも佐藤 他論文や楠戸 他論文で用いられており、技術的な詳細を理解する上でも有用である。

最後に、本特集により農林業の生態系サービスに対する評価手法の理解が進むとともに、関連分野の定量的研究が一層盛んになることを願って止まない。また、本特集の査読者・編集者の方々のご尽力に、この場を借りて厚くお礼申し上げたい。

参 考 文 献

日本学術会議 (2001). 地球環境・人間生活にかかわる農業及び森林の多面的な機能の評価について (答申), 日本学術会議, <http://www.scj.go.jp/ja/info/kohyo/pdf/shimon-18-1.pdf> (アクセス日: 2019年1月15日).

Díaz, S. et al. (2018). Assessing nature's contributions to people, *Science*, **359**(6373), 270-272.

日本における森林生態系サービスの経済評価

佐藤 真行¹・栗山 浩一²・藤井 秀道³・馬奈木 俊介⁴

(受付 2018 年 6 月 25 日; 改訂 12 月 25 日; 採択 12 月 26 日)

要 旨

本研究では、日本における森林資源の生態系サービスを定量的に評価することを目的とし、経済価値評価を実施した。その際に、これまで発展してきた環境の経済評価論の観点から本目的に合致する評価手法の検証を行い、生態系サービスの評価における効用理論的枠組を整理したうえで、経済学的理論背景と照らし合わせて概観した。その上で、生態系サービスの経済価値評価を行う上で重要となる非利用価値を含めた幅広い価値を推定可能とする表明選好法のうち仮想評価法(CVM)を利用し、支払カード型 CVM によって森林の原単位価値評価を行った。CVM は支払意思額(WTP)に基づいて評価を行うものであることに鑑み、ランダム効用モデルを基礎にもち同様に WTP に基づくコンジョイント分析を応用することで属性別のウェイトを推定した。こうした分析から、非利用価値を考慮した森林生態系サービスの経済価値を原単位および属性単位で推定した。こうした評価は、生態系サービスの可視化のひとつの方法である生態系勘定への応用に資するものである。

キーワード：生態系サービス、経済価値評価、仮想評価法、森林資源、原単位価値、属性別価値。

1. はじめに

1.1 環境の経済的価値

生態系ストックや生態系サービスは、典型的な「市場価格を持たない価値物」(植田 1996)である。こうした自然資本の持つ非市場的価値を評価するために、環境の経済評価手法が開発されてきた。環境の経済評価手法は、非市場価値を含めて環境・資源・生態系のもつ価値を貨幣単位という一元化された尺度で計測しようとするものである。こうした取り組みは、環境や生態系サービスの可視化につながり、現在進められている生態系サービス勘定などの開発においても重要な役割を担っている。

しかしながら環境の経済評価手法は万能ではなく、方法論的に克服すべき問題点も多く提起されている。そこで本研究では、環境の経済評価手法の理論的背景と、その目的を明確にしたうえで、生態系サービス評価に応用するための条件や実践を検討する。

非市場価値をもつ生態系サービスは、市場においては不当に低く評価されているため開発における意思決定において常に破壊の圧力にさらされる。生態系サービスをふくめて、環境評価

¹ 神戸大学 大学院人間発達環境学研究所：〒 657-8501 兵庫県神戸市灘区鶴甲 3-11

² 京都大学 大学院農学研究科：〒 606-8502 京都府京都市左京区北白川追分町

³ 九州大学 大学院経済学研究院：〒 812-8581 福岡県福岡市東区箱崎 6 丁目 19 番 1 号

⁴ 九州大学 大学院工学研究院：〒 819-0395 福岡県福岡市西区元岡 744 West2 号館

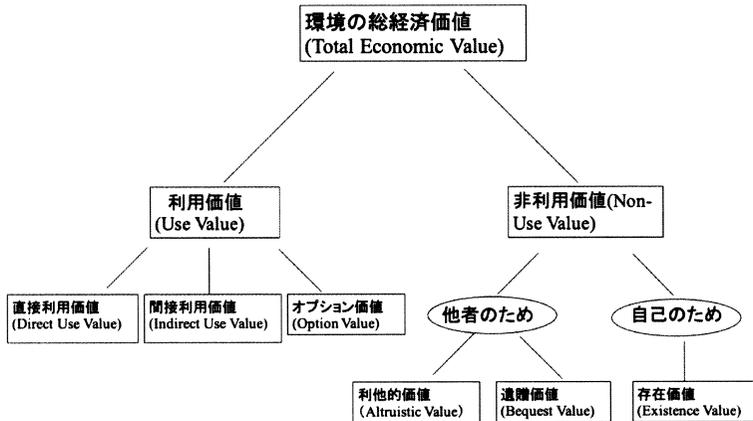


図 1. 環境経済価値の分類 (Bateman et al., 2002 から作成).

手法が評価対象とする価値は, Turner et al. (1994) により一般的に次のように定められている.

総経済価値 (TEV; Total economic value) = 利用価値 (Use value) + 非利用価値 (Non-use value)

利用価値には, 実際に消費して得られる価値 (Direct use value; 直接利用価値) のほかに, 物的な消費を伴わない, いわば環境機能の利用を意味する間接利用価値 (Indirect use value) や将来の利用のために保持する価値 (Option value; オプション価値) がある. 一方で, 非利用価値とは, その環境を全く利用しないにもかかわらず, それに価値があると感じるときに認められるものである. その中で, それが消失することを「損失」と感じる場合に認められるのが存在価値と呼ばれるものである. また, 自分は利用しないが, 他人のために在ったほうがよいと感じる場合, その他者が同世代の場合の利他的価値 (Altruistic value) や将来世代の場合の遺贈価値 (Bequest value) がある. これらは, 図 1 のようにまとめられる.

このように, 環境は多面的な諸価値を有しているが, それらのすべてが市場で取引される性質のものとは限らない. このために, 市場価格 (= 利用のためのコスト) は, 社会的価値ないしは福祉を反映したものである計算価格 (Accounting Price, Shadow Price ともいう) から下方に乖離し, その結果, 市場が機能せず, 環境の過剰利用が発生するのである.

生態系サービスの価値評価において市場価値ベースで評価付けを行っていくと, 深刻な過小評価につながる. そこで森林などの環境の価値を正しく評価する必要が生じるのだが, これは簡単な問題ではない. たとえば, 世界銀行の World Development Indicators (WDI) などでは, 森林資源の価値評価において木材価格と伐採費用に基づく市場レントを用いている. これは図 1 における直接利用価値に相当するものであるが, その他の価値を反映しているものとはいえない. すなわち, これは森林資源の価値の一部の評価にとどまっているものであり, 適切な評価を行うためには非利用価値を明示的に考慮した価値評価手法を適用する必要がある.

こうした問題に対して, 環境の経済評価論は効用理論に基づく消費者余剰による測定を提案した. こうして測られた価値は余剰価値 (surplus value) と呼ばれる. 通常は, Hicks (1943) の定式化に従った補償余剰によって (1.1) 式のように定義される¹⁾.

$$(1.1) \quad V(p, Q^0, Y) = V(p, Q^1, Y - CS) = U^0$$

(1.1) 式は, 環境質変化 $Q^0 \rightarrow Q^1$ が「改善」である場合, CS は環境を改善させるための支払意

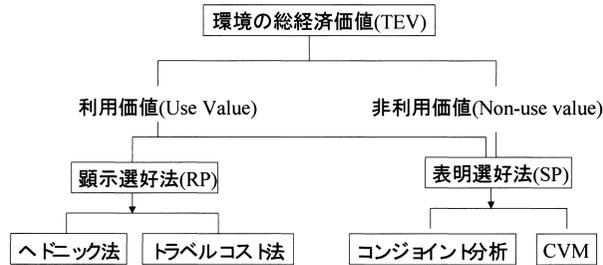


図 2. 評価のアプローチと評価する価値(出典 Bateman et al., 2002 から作成).

思額(WTP; Willingness to Pay)であることを意味する。また、環境質変化 $Q^0 \rightarrow Q^1$ が「悪化」の場合は、CS が受入許容額(WTA; Willingness to Accept)となる。

1.2 環境の経済評価手法

経済理論の枠組みにもとづいて生態系サービスを評価するためには、利用価値だけでなく非利用価値を合わせて WTP を測定することが求められる。そのために適用可能な環境評価手法として、表明選好法が開発されてきた。経済学が提供してきた評価手法としては図 2 が示す通り、環境の経済評価手法には表明選好法だけでなく顕示選好法もある。しかしながら顕示選好法は、レクリエーション価値など生態系サービスの重要な要素を評価することができる一方で、存在価値などの評価対象を測れない性質が指摘されている。国内外で顕示選好法による生態系サービス源の評価が蓄積しているが、その評価値を用いる際には注意を要する。

それに対して表明選好法は、生態系サービスの幅広い価値を測定することが可能である。表明選好法には仮想評価法(CVM; Contingent Valuation Method)やコンジョイント分析がある。これらは環境評価手法として、とりわけ発展の目覚ましい手法である。CVM の歴史は比較的古く、Ciriacy-Wantrup (1947) がその概念を最初に提案したと言われている。また、CVM の手続きを体系的にまとめた Mitchell and Carson (1989) によれば、CVM を用いた最初の実証研究は Davis (1963) であると言われている。

CVM は生態系の価値を経済的に評価する要請(1989年のエクソン=バルディーズ号事件およびオハイオ裁判)を受けて発展してきた経緯がある²⁾。エクソン=バルディーズ号事件は、アラスカ沖で座礁したエクソン社のバルディーズ号から大量の原油が流出し、深刻な海水汚濁や沿岸レクリエーション地の破壊などをもたらした事件であり、その損害賠償額を算定するにあたり、非利用価値を含めるか否かで大論争を引き起こしたものであった。オハイオ裁判は、スーパーファンド法における損害評価の手続きに関する内務省のルールを巡って、オハイオ州政府および環境保護団体と産業界側とが争った裁判であり、評価対象を非利用価値まで範囲を広げるとともに CVM 適用の妥当性を判決として下したものである。その後、1993年にアメリカ商務省国家海洋大気管理局(NOAA; National Oceanic and Atmospheric Administration)によってまとめられた NOAA パネル報告書で CVM の有効性を認める結論が出され、非利用価値の認知がさらに進んだ。1989年からのこれら一連の事件を経て、非利用価値を推定する手法として表明選好法が大きく発展するに至った背景がある。これをうけて、その後の生態系サービス評価を目指したミレニアム生態系評価(MEA, 2005)や TEEB 報告書(TEEB, 2010)などの大型研究プロジェクトでも、かならず注目される手法となっており、本研究プロジェクトにおいても重要な手法と位置づけられる。

CVMの最大の欠点は、様々な評価バイアスが入り込む余地があることにある³⁾。従って、CVM研究の主要な課題は、いかにしてバイアスを最小に抑えるか、という点であった。そのために、調査票の作成や回答方式の工夫などに研究の力点が置かれている。調査計画の策定や調査票作成において問題が含まれている場合、その評価値の利用可能性は大きく損なわれてしまうことになる。こうした観点から、既存の評価研究を概観することが本研究の課題でもある。

CVMは、調査票における回答方法によって解析モデルが異なり、大きく分けて、自由回答型、付け値ゲーム型、支払カード型、二肢選択型がある。どのタイプを用いるかによって、支払意思額の定式化やバイアスの入りやすさ、その結果として推定値が異なることがあるため、その選択は重要である。例えば Welsh and Poe (1998)では、自由回答型、支払カード型、二肢選択型のそれぞれでの評価値を推定し比較した。その結果、それぞれ、\$54、\$37、\$98という結果を示し、およそ165%程度の乖離率を示した。Ready et al. (2001)も同様に、支払カード型で低い評価値を示した。こうしたことから、しばしば大きくなりがちな非利用価値の推定に対して「控えめな推定」を行いつつ、国民経済計算体系への接続を意識したマクロ規模での生態系サービス評価を行うためには、支払カード型は有力な手法の一つである。CVMの詳細な解説文献はかなりの蓄積が進んでいるため、どのような手順を踏んだCVMサーベイ調査において利用性が高いかについての判断材料となる(Mitchell and Carson, 1989; Bjornstad and Kahn, 1996; Bateman and Willis, 1999; Bateman et al., 2002; Champ et al., 2003; Haab and McConnell, 2002; Nocera et al., 2003; Herriges and Kling, 1999; Alberini and Kahn, 2006; Wilis and Garrod, 2012; 栗山, 1997, 1998; 鷺田, 1999; 鷺田 他, 1999; 柘植 他, 2011)。

また、マクロ的な原単位価値だけでなく、質的な要素をより詳細に価値付ける際にはコンジョイント分析の利用が考えられる。コンジョイント分析の利点は下記のような点があり、生態系サービス評価を進めるうえで質的な要因を考慮する際には有用である。

- (1) 広範な対象について評価ができる。
- (2) 非利用価値を評価できる。
- (3) 多属性を有する環境について、属性ごとの評価ができる。

(1)、(2)は顕示選好法と比較して、(3)はCVMと比較して、コンジョイント分析の特徴が明らかとなるであろう。環境は本質的に多属性を有するものが多く、環境にインパクトを与える政策プロジェクトについて、属性ごとに評価されるべきケースが多い。そのときに(3)多属性評価が可能なコンジョイント分析が注目されることになる。また、コンジョイント分析のなかでは、選択型実験と呼ばれる手法が優れていることが近年の研究で明らかとされてきており(柘植 他, 2011)、生態系サービスの経済価値評価手法として生態系サービス源の多様な属性をより詳細に把握・評価する際には有力な手法である。

2. 生態系サービスの評価

生態系サービスを評価するにあたって、環境評価手法がどの価値を測定対象にしているか、そしてどのように評価しているかによって、適用可能かどうか判断される。ここでは、評価手法、評価主体単位、評価の時間的単位、評価の空間的単位の観点を提起する。その後、本稿では生態系サービス源としての森林資源を測定するこれまでの評価事例を概観したうえで、日本における森林生態系サービスの経済評価に取り組む。

2.1 評価手法

前節で述べたとおり、いずれの手法を採用するかによって評価対象や評価結果は異なる。生

態系サービスの評価を行う際には、こうした手法の差異を考慮して進める必要がある。これまで利用されてきた手法とその特徴は次のようにまとめられる。

- 置換法(特定の生態系サービスに着目して、その機能を人工物で代替するために掛かる費用から推定)
- 家計生産法(生態系サービス保全のための家計支出から推定)
- ヘドニック価格法(生態系サービスが地価や不動産価格に与える影響から推定)
- トラベルコスト法(生態系サービスが訪問行動に与える影響から推定)
- 仮想評価法(アンケート調査により生態系サービスに対する WTP から推定)
- コンジョイント分析(アンケート調査により生態系サービスに対する WTP を、生態系の属性ごとに推定)
- 市場価格・レント(市場取引される生態系サービスを、その交換価格から推定)

2.2 評価の主体

生態系サービス評価については、評価主体の単位についても注意が必要である。これまでの評価事例では次のようなものがある。

- (1) 国家・自治体単位で評価して公共部門の支出として生態系サービスを評価。国民経済計算、県民経済計算、産業連関などのマクロデータから推計するため、評価主体は国あるいは県全体となる。
- (2) 世帯・家計に着目したもの。家計生産法や世帯単位での WTP を推計した場合に、評価主体は世帯となる。
- (3) 個人に着目したもの。WTP を個人一人あたりで算出したもの。

こうした評価主体の設定方法は、生態系サービスを評価する際に非常に重要である。国全体(あるいは県全体)に当てはめて考える時、個人別に推計された WTP に家計数を乗じる場合(2)と、人口を乗じる場合(3)では、かなり大きな差異が発生する。たとえば表明選好法などで支払意思額を測定する場合に、分析枠組みにおいて評価の主体があいまいな場合、そうした影響が生態系サービスの評価結果に重要な影響を与える。こうしたことから、生態系サービスの評価に応用可能な手法を精査する場合には、評価主体の取り扱いには慎重な配慮を要する。

2.3 評価の時間的単位

評価の時間的単位として、大きく分けて次の2つがある。

- (ア) 一回の評価値として評価する。例えば CVM の WTP の調査で一回だけの支払として分析されており、その後の支払は求められないようなシナリオでの評価。
- (イ) 毎年の負担額として評価する。例えば CVM の WTP の調査で税あるいは基金等による毎年の支払としての評価。あるいは国民経済計算、県民経済計算、産業連関などによる毎年の生態系サービスフローとしての評価。

こうした違いも、生態系サービスの評価を行う際には配慮を要する。生態系サービスを評価する目的に合わせて、ワンショットの評価値(ア)として評価すべきか、国民経済計算のように毎年の価値(イ)として評価すべきかを選択する必要がある。経済理論における合理性の仮定によれば、将来割引などの考慮によって(ア)と(イ)は変換可能ではあるが、合理性の仮定はかなり強い仮定であり、評価手法における時間単位の選択によって評価値は左右されることを踏まえておいたほうがよい。

2.4 評価の空間的単位

評価値が評価対象の空間的広さを考慮しているか、また便益の波及範囲をどの程度想定しているかについても考慮する必要がある。前者はスコープ無反応性という環境の経済評価手法の問題に関わる。スコープ無反応性は、Kahneman et al. (1999)によって環境の経済評価手法への批判として指摘された性質であり、2,000羽の渡り鳥保護と2,000,000羽の渡り鳥保護に対するWTPに差異がなかったという実験結果とともに示されたものである。こうした問題は他の生態系の価値評価にも応用可能であり、森林評価の場合も保全対象の空間的単位を無視して評価してしまう場合、国全体あるいは県レベルで森林面積等を乗じて評価することが不可能となる。したがって、生態系サービスの評価に応用可能な評価手法として、空間的単位の取扱いが適切かどうかについて慎重な配慮が求められる。

2.5 評価値の妥当性

生態系サービスの評価手法への適用可能性を検討する際に、生態系サービス評価値の妥当性を検証することは重要なステップである。評価結果の妥当性について、特に表明選好法における検証点としてMitchell and Carson (1989)やBateman et al. (2002)では、趣意妥当性(Content Validity)、基準関連妥当性(Criterion Validity)、構成概念妥当性(Construct Validity)の観点から検討することを推奨している。構成概念妥当性は、収束妥当性(Convergent Validity)と理論的妥当性(Theoretical Validity)に分けられる。

趣意妥当性とは、もともとの調査の意図を妥当とするかということであり、これは最終的には調査者の主観的判断に基づくことになるが、専門家やフォーカスグループとの間で議論すべき項目であるといえる。基準関連妥当性は、実際の市場データと比較して、評価値が妥当な値であるかを検討するものであり、同一対象を扱う顕示選好法の結果と比較することでチェックされる。実際には多くの研究で表明選好法と顕示選好法とで評価値が乖離することが確認されている(Carson et al., 1996; Carlsson and Martinsson, 2001)。しかし、そもそもの問題として表明選好法は、現実市場データがないために開発された手法であり、これをチェックすることはしばしば不可能である。竹内(1999)が述べるとおり、もともと市場価格のないものへ市場価格をつける試みであるため、それが正しい値かどうかは決定できないのであり、全般的な評価研究の蓄積に依存するのである。

収束妥当性は、他の手法による評価の結果と比較することで確認され、環境評価研究の場合には、CVMとコンジョイント分析という二つの異なる表明選好法の比較対照により検討できる。つまり、同じ対象を測るならば同じ額で評価されねばならない(Stevens et al., 2000)ということである。Roe et al. (1996)では、分析対象や、分析の意図および目的に応じてCVMとコンジョイント分析を使い分けることが最も望ましいとしている。従って、国単位・県単位での生態系サービスの評価を目指すためにはシンプルなCVMで、そして質的要因を詳細に評価するためにはコンジョイント分析の利用が考えられる。

最後に、理論的妥当性については、社会経済的属性の係数符号のように経済理論による予測と一致しているかをチェックすることにより確認されるものであり、一般に学術研究として報告されている評価研究は理論的な検証が踏まえられているものがほとんどである。

妥当性ととも検討されるべきは、評価値の信頼性(Reliability)である。多くの場合、信頼性とは再現可能性(Replicability)の度合いを意味する(Bateman et al., 2002)。藤本(1998)では、しばらくの期間を置いて再度同様の調査を行うという再現調査により、双方の評価値が高く相関すれば信頼性が検証されるとしている。適切にデザインされた調査は、同一の評価値が再現される。エクソン=バルディーズ号事件に関するCVM調査(Carson et al., 1994)では、2年後に同一の調査をしたときに同一の結果を得ているため、信頼が高いことが実証されている

(Carson et al., 1997). しかし、この信頼性を調査するに当たり、かなりの費用がかかるという現実的問題が存在する。

妥当性や信頼性を低下させる最大の原因は、調査におけるバイアスであり、表明選好法を現実に利用可能なものとするためには、適切な調査票作成が必要条件である。本研究では、従来の環境評価研究で使用されたサーベイ調査票をできる限り収集、精査し、また著者らが過去に実際に実施した森林生態系評価研究からの知見と照らし合わせて、調査票設計を行い、プレテストを行ったうえで次節で述べる調査データを収集した。

3. 森林資源の経済価値評価

3.1 調査方法とデータ

本研究では、CVMによる森林資源の原単位経済価値評価を目的として2015年11月16日から12月4日までインターネット調査を実施し、全国の192,704人から有効回答を得た。このサーベイ対象者は、株式会社日経リサーチが保有するモニターパネルであり、全国の都道府県の人口比率および年齢比率に合うよう条件のもとでモニターパネルから無作為に抽出・依頼し、回答を集めた。その結果、サンプルは各都道府県の人口構成を反映したものになっているが、モニターパネルからの抽出であることによるバイアスは免れない点に注意が必要である。

また、3.2節で議論する多属性評価に用いられるコンジョイント分析のデータは2016年12月2日から12月9日にかけて収集され、CVMと同様に各都道府県から人口比率と年齢比率を考慮したパネルからの無作為抽出によって6,843人から回答を得た。ただし、一人あたり8回の繰り返し部分プロフィール選択質問を行ったために、解析に使われたデータは54,744となる。

サーベイ調査票では、第2節における生態系サービスへの環境評価手法への応用についての議論にもとづいて、支払カード型CVMを採用し、評価主体を世帯として明確化し、毎年の支払として時間的範囲を定め、居住地域における1haの森林面積の増加に対する評価として空間的範囲を定めた。この時間的範囲は、国民経済計算体系への接続を意識したものであり、空間的範囲は都道府県単位で勘定表を作成しやすいようにしたものである。1年間を時間単位にした価値評価の、調査票における質問文は次のようなものである。最初に、評価対象の性質の再認識を促すために、森林資源のもつ環境機能や生態系機能に関する重要性についての質問を行う(表1)。

表1. 調査票における森林の機能の重要性に関する設問。

<p>森林には減災機能（浸水防止・土砂災害防止）がありますが、レクリエーションの場の提供や緊急時には建築資材になるなど、他にも様々な機能を持ちます。</p> <p>あなたは、あなたのお住まいの地域（市町村）の森林について、以下のどの機能が重要だと思いますか。重要と思う順に、1位～5位をお答えください。</p> <ul style="list-style-type: none"> ・木材供給 ・減災機能（土砂災害防止など国土保全） ・水質浄化 ・生態系・生物多様性保全 ・レクリエーション・遊び場 ・文化的・宗教的価値（神社仏閣など） ・景観 ・CO₂吸収 ・遺伝子資源（薬品の材料など） ・その他（自由回答）

表 4. 森林の多面的機能(出所: 学術会議, 2001).

1. 生物多様性保全	遺伝子保全、生物種保全、生態系保全
2. 地球環境保全	地球温暖化の緩和(二酸化炭素吸収、化石燃料代替エネルギー)、地球の気候の安定
3. 土砂災害防止/土壌保全	表面侵食防止、表層崩壊防止、その他土砂災害防止、雪崩防止、防風、防雪
4. 水源涵養	洪水緩和、水資源貯留、水量調節、水質浄化
5. 快適環境形成	気候緩和、大気浄化、快適生活環境形成(騒音防止、アメニティー)
6. 保険・レクリエーション	療養、保養(休養、散策、森林浴)、行楽、スポーツ
7. 文化	景観・風致、学習・教育(生産・労働体験の場、自然認識・自然との触れ合いの場)、芸術、宗教・祭礼、伝統文化、地域の多様性維持

表 5. 想定した森林生態系サービス.

1	水源涵養機能
2	土砂災害防止機能
3	レクリエーション機能
4	地球温暖化防止機能
5	生態系保全機能
6	木材生産機能

ている(栗山, 1999). しかし近年, 調査票提示の技術進歩とともに, より多くの属性数を回答者への負荷を抑制しながら分析する方法が開発された. 本研究で採用した部分プロファイル分析がその1つである.

部分プロファイル分析とは, 多くの評価対象属性のうち, 認知的負荷が大きくなりすぎない程度の属性数を抽出し, その属性間のトレードオフを分析していくものである. こうした演算を繰り返して, 全体としての多属性の間でのトレードオフを網羅的に推定することが可能となる. 本研究では, 森林の生態系サービスとして表5にまとめられる主要な6つの森林属性に, 貨幣属性としての年間支払額を加えた合計7つの属性を評価対象として想定し, 部分プロファイル分析により, 各属性の相対的なウェイトを分析した. それぞれの属性について, 評価に必要な最小限の認識を共有するために, 表6のような説明を加えている.

なお, これらの説明のあとに, 現時点での認識の度合いや, 重要性についての質問を行い, コンジョイント分析の質問自体の理解を促すとともに知識や認識のデータを入手した. これらの属性から部分プロファイルを作成し, 回答データから各属性のウェイトを推定する. 部分プロファイル分析の質問例を図3に示す.

こうした選択質問に対する回答データをロジットモデルで分析する. コンジョイント分析は効用理論と整合的なかたちで定式化することができるため, 支払い意思額の推定に利用される. ここでは効用関数(3.1)式を仮定する. これは観察不可能な攪乱項を含むため, ランダム効用関数と呼ばれる.

$$(3.1) \quad U_{in} = V_{in} + \varepsilon_{in}$$

表 6. 森林生態系サービスの属性に関する説明内容.

<p><水源かん養機能> 森林に雨が降ると、土の中に雨水が染み込み、染み込んだ雨水はゆっくりと森林から流出していきます。このため、雨が降った時には洪水が緩和され、雨が降っていないときでも森林から川に水が流出していきます。さらに、土の中を水が通る際にろ過され、適度にミネラルを含むことから、おいしい水ができていきます。森林のこのような働きは水源かん養機能と呼ばれています。</p>
<p><土砂災害防止機能> 森林の土壌は、樹木や下草、落ち葉に覆われており、これらが雨のしずくの衝撃を吸収し土砂の流出を防いでいます。また、樹木の根は地中深く伸び、岩の亀裂にまで入り込み、土壌と基岩層との境界を根がしっかりと固定するので山崩れが起こりにくくなります。森林のこのような働きは土砂災害防止機能と呼ばれています。</p>
<p><生態系保全機能> 森林は、多様な植物、動物、昆虫、菌類、微生物などが生息・生育しており、一つの森林生態系ができています。生物が多数いることも重要ですが、いろいろな種類の生物がいることも重要であり、いろいろな種類の生物がいることで森林生態系が安定し、人間の生活や経済活動にも良い影響もたらされます。森林のこのような働きは生態系保全機能と呼ばれています。</p>
<p><温暖化防止機能> 森林は光合成により二酸化炭素を吸収することにより、地球温暖化防止に貢献する働きを果たしています。森林のこのような働きは地球温暖化防止機能と呼ばれています。</p>
<p><木材生産機能> 森林からは住宅資材や家具などの原材料となる木材を生産することができます。木材を生産するために森林を伐採しても、植林を行い、適切に管理を行えば、再び木材を生産することができます。森林のこのような働きは木材生産機能と呼ばれています。</p>

	整備案1	整備案2	整備案3	
水源かん養機能	現状の75% (25%減少)	現状を維持する	現状の150% (50%増加)	この中からは選ばない
土砂災害防止機能	現状の75% (25%減少)	現状の125% (25%増加)	現状を維持する	
地球温暖化防止機能 (二酸化炭素吸収)	現状の75% (25%減少)	現状の150% (50%増加)	現状を維持する	
生態系保全機能	森林内の生物の種数が 現状より75%となる (25%減少)	森林内の生物の種数が 現状より150%となる (50%増加)	森林内の生物の種数が 現状より125%となる (25%増加)	
1年あたりの負担金	2,000円	20,000円	5,000円	
	↓	↓	↓	↓
	○	○	○	○

図 3. 部分プロフィール質問例.

ここで、 U_{in} は個人 n が選択肢 i を選んだときの効用を表し、 V_{in} は提示されたプロフィールに関する効用の確定項とする。 ε_{in} は誤差項であり、条件付きロジットモデルで解析する場合は第一種極値分布 (EVI) を想定する。

選択型コンジョイント分析は、幾つかの選択肢集合の中から、最もよいと思った選択肢を1つだけ選択することを要求するものである。回答者に提示された選択肢の集合を C_n とすると、個人 n が C_n の中から選択肢 i を選ぶ確率 P_{in} は、(3.2)式のように表現できる。

$$(3.2) \quad P_{in} = \text{prob}(U_{in} > U_{jn}, \text{ for all } j \in C_n)$$

(3.1)式を代入して変形すれば、次のようになる。

$$(3.3) \quad \begin{aligned} P_{in} &= \text{prob}(V_{in} + \varepsilon_{in} > V_{jn} + \varepsilon_{jn}, \text{ for all } j \in C_n) \\ &= \text{prob}(V_{in} - V_{jn} > \varepsilon_{jn} - \varepsilon_{in}) \end{aligned}$$

McFadden (1974)により条件付きロジットモデルは次のように定式化されることが示されている。

$$(3.4) \quad P_{in} = \frac{\exp(V_{in})}{\sum_{j \in C_n} \exp(V_{jn})}$$

この式を利用して、ランダムに与える部分プロフィールから、効用関数のパラメータ、すなわち各属性のウェイトが求められる。

4. 評価結果

4.1 CVM による森林生態系の原単位評価

支払いカード型の CVM 質問の回答データを利用して、森林生態系サービスの価値評価を実施した。推計結果は図 4 のように得られた。

評価結果として、全国で平均約 2,447 円(標準偏差 171 円)であり、最高値は 2,813 円、最低値は 1,967 円であった。森林 1ha あたりの評価を家計単位で行った先行研究は非常に少なく、この値の比較は困難であるが、森林・水源環境税がおよそ 500~1,000 円で実施したことを踏まえると、本研究の評価結果は比較的高い評価値であると考えられる。家計あたりで見た場合、鹿児島県屋久島を対象とした栗山 他 (2000) の 1,500 円や宮城県蕪栗沼を対象とした馬奈木・地球環境戦略研究機関 (2011) の 917 円と比べても本研究の評価値は高い値を示している。しかし森林 1ha に対する家計あたりの WTP という評価基準が異なるため、単純な比較は難しい。本研究では、表明された WTP の水準自体を考察するのではなく WTP がどのような要因

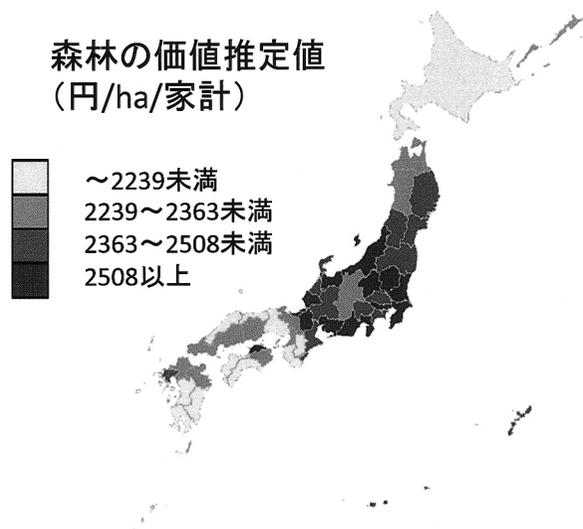


図 4. 森林 1ha に対する支払意思額の分布(世帯 1 年あたり)。

によって規定されているかを明らかにすることで、森林生態系の価値推定がなぜ地域によって異なるかを分析する。

図 3 が示す通り、日本全体で評価額は散らばっており、全体的に東日本のほうが比較的高い評価値を示している。Fujii et al. (2017) は、西日本は東日本に比べて私有林の割合が高く、森林面積あたりの利益率が高いこと、また東日本は商業利用以外の森林の公益的機能が相対的に重視されていることを示唆している。一般に商業価値は所有者のみに帰属する価値であるのに対して、公益的価値は一種の外部経済であり、住民に遍く享受される価値であるため、一般住民に対するサーベイ結果として東日本の方が総じて高い価値となったと考えられる。

しかしながら、所得などの社会経済属性や、森林の種別なども評価値に影響を与えていると考えるのが妥当であろう。そこで、なぜこのように支払意思額が分布したのかを回帰分析で検証する。まず社会属性が影響していることが考えられる。先行研究を踏まえてもっとも典型的な変数は所得である。所得効果とも呼ばれ、所得が高いほど支払意思額は高い値を示すことがある。従って、所得変数は森林に対する WTP の規定要因として外すことはできない。本調査データは、評価主体の単位として世帯を設定しているため、世帯所得を回帰分析の説明変数として設定する。このデータは平成 24 年度の内閣府の県民所得データと、総務省統計局の世帯数データをもちいて構築した。また、地域の特徴を表すデータとして、性別と年齢を導入した。性別については地域人口における女性率を表すものとなる。

次に、天然林／人工林という区別に対応するデータとして、面積に関する多重共線性を考慮して割合をとり、天然林率を採用した。また、広葉樹／針葉樹という区別に対応するものとして、同様に広葉樹林率を採用した。これらによって、天然林率の上昇や、広葉樹林率の上昇がどの程度 WTP の評価値に影響するかを示すものとなる。

さらに、森林の質的要因を表す要素として、樹齢を考慮し、年齢別の森林面積を考慮した加重平均として算出した。一般に、日本の森林の問題として間伐の遅れや林業の衰退に伴う樹齢の高齢化が挙げられている。樹齢の高齢化は山林の荒廃にも関連する要素であるのと同時に、二酸化炭素の吸収の遅れや生物多様性の棲息地としての機能にも関わる要因であるため、回帰分析に含めた⁴⁾。

以上の定式化により、最小二乗法による WTP の回帰式は次のように推定された⁵⁾。

$$\begin{aligned} \text{WTP}_{\text{Forest}} = & 2878.91^{***} + 710.98^{***} \times \text{広葉樹林率} - 348.78^{***} \times \text{天然林率} \\ & - 199.21^{***} \times \text{森林率} - 9.27^{***} \times \text{樹齢} + 0.14^{***} \times \text{世帯所得} \\ & - 355.61^{***} \times \text{性別} - 19.89^{***} \times \text{年齢} \end{aligned}$$

推定結果はいずれの係数も 1% 水準で有意に推定された。広葉樹林率および天然林率は、それぞれ回答者の居住している都道府県の森林面積に対する広葉樹林および天然林の割合を表している⁶⁾。推定結果から、広葉樹林率は WTP に正の影響を及ぼしている事がわかる。一般に広葉樹林は四季の変化に反応して景観を変えレクリエーションなどの非利用価値を提供したり、生物の生息地としての機能を果たしていると考えられ、そのことが正の影響をもたらしたと解釈される。一方で天然林率は WTP に負の影響をもたらしていた。この点はさらなる考察や議論が求められるが、天然林は生物の生息地としての機能がある一方で、人工林ほど森林の植栽が整っておらずアクセスが容易でないことに伴う利用の難しさが反映したのかもしれないが、その検証については今後の課題として残った。

森林率は、回答者の居住している都道府県面積に対する森林の割合を表している。この影響が負であるのは、森林が希少な地域ほど 1ha の森林に対する WTP が高いということから、事前の予想と整合的な結果である。樹齢は、回答者の居住している都道府県面積に存在する森林の加重平均樹齢であり、推定に用いられたデータの範囲では、樹齢の高まりに伴う森林の質の

表 7. 各属性係数の推定結果と限界支払意思額。

属性	推定値	標準誤差	t 値	P 値	限界支払意思額 (円)
1 水源涵養機能	0.0131	0.0002	52.862	0.000	143.6
2 土砂災害防止機能	0.0121	0.0003	46.185	0.000	132.2
3 レクリエーション機能	0.0038	0.0003	14.455	0.000	41.0
4 地球温暖化防止機能	0.0113	0.0003	43.900	0.000	123.9
5 生態系保全機能	0.0084	0.0003	33.367	0.000	92.0
6 木材生産機能	0.0051	0.0011	19.566	0.000	56.0
7 1年あたりの負担金 (世帯あたり)	-0.0914	0.0013	-83.964	0.000	
LogL					-75406.575
Sample size					54,744

変化が WTP に負の影響を及ぼしていることを示唆している。

世帯所得は回答者の支払能力に関係する。世帯所得は WTP に正の影響を及ぼしており、これは所得効果として先行研究でも観察される影響である。性別は女性であれば 1 を取るダミー変数である。性別や年齢も WTP に影響していることがわかる。このように、CVM によって評価された WTP を回帰分析することによって、自治体が森林整備を行う際に、許容されやすい森林特性について情報が入手できるが、生態系サービスごとの評価を行うために次節で多属性評価を行う。

4.2 コンジョイント分析による森林生態系の属性別評価

ここまで CVM によって森林生態系の原単位評価を行ったが、森林生態系のどのようなサービスが評価されているのかを分析するためには、森林の機能別に評価を行う必要がある。多属性評価手法であるコンジョイント分析を適用することによって、森林の価値を構成要素別に分解して評価することができる(坂上・栗山, 2009)。ここで推定される各属性に対する限界 WTP は、各属性が追加的に 1 単位改善することに対する価値額であると理解される。各属性の評価値を比べれば、現時点に存在する森林資源のもつ各属性の相対的重要性が示される。これにより、それぞれの都道府県で推定された森林の評価値に、どの属性が強く反映されているのかが推察される。この相対的ウェイトを用いて、生態系サービスごとの評価値を推計した。

表 7 の結果は、木材生産やレクリエーションといった市場において評価される価値よりも、水源涵養機能、土砂災害防止機能、地球温暖化防止機能、生態系保全機能といった非市場的な価値が高く評価されていることがわかる。また、必ずしも林業に従事していない一般の回答者が評価主体であるにもかかわらず木材の供給機能に対しても正の価値が示されたことは、国産材に対するニーズが背景にあると推察される⁷⁾。レクリエーション機能に対しては評価対象属性のなかでもっとも低い評価値となったが、これは今回の評価対象が必ずしもツーリズムを志向していない一般的な森林であるためと考えられる。

5. まとめ

本稿では、生態系サービスの経済学的評価手法について、環境の経済評価論の観点から精査し、生態系サービスの価値づけにおける効用理論的枠組を整理したうえで、理論的背景をもつ経済価値評価手法について概観し、生態系サービスの評価においては、商業的な利用価値だけ

でなく、非利用価値を評価することが重要であることを論じた。そして、生態系サービスの評価として適用する手法選択の際に、(1)環境評価の手法、(2)評価の主体、(3)評価の時間的単位、(4)評価の空間的単位に関する検討を行った上で社会調査を実施し、非利用価値の推定可能な表明選好法のうち支払カード型 CVM を実施し、森林の原単位価値評価を行った。

また、CVM が表明選好法であり WTP の評価を行うものであることに鑑み、同様のランダム効用モデルを基礎にもつコンジョイント分析を応用することで属性別のウェイトを推定した。森林については既存研究で指摘されている生態系機能の多面性から、属性数が多く単純なコンジョイント分析の適用は困難であるため、部分プロファイル分析を適用した。その結果、水源涵養機能、土砂災害防止機能、生態系保全機能、温暖化防止機能といった、森林生態系が提供する機能には市場ベースでは必ずしも評価されない社会的価値を有することが示された。これらの価値には公共性があり、市場における交換や消費によって便益が発生する利用価値とは異なり、多数の受益者に享受される性質がある。したがって便益受益者が多いほどこうした価値は高く評価されるのである。このことはまた、森林を市場価値ベースで評価することは過小評価につながる危険を示唆する。林業やツーリズムなどの森林の産業利用の可能性を議論する際には市場価値が重要であり、そうした評価に特化した顕示選好法も利用可能であるが、こうした公共性を考慮した評価にはやはり本研究で用いた表明選好法が有用である。ただし、表明選好法はバイアスやヒューリスティクスに影響されやすいため、慎重な調査票デザインを必要とする。本稿では、これまでの表明選好法に関する研究を踏まえて、全国レベルの森林生態系を対象としながら、比較的回答負荷の低い支払いカード型の CVM によって WTP を推定することを通じて森林生態系の原単位評価を行い、さらにその WTP の規定要因を回帰分析により特定した。本結果のように回帰式が特定されたことは、支払いカード型 CVM による WTP 推定値に一定の合理性があり、森林特性や社会属性によって規定されうることを示唆するが、二肢選択型 CVM など異なる手法を採用した場合との比較などを行うことは評価結果の妥当性を確認するうえで有用であり、今後の課題でもある。また、本研究では樹種や樹齢など森林に関して入手可能なデータを用いて、それら変数についての係数を有意に推定できたが、今後はリモートセンシングなどを用いてさらに詳細なデータを変数として導入し、変数選択を含めて回帰モデルの探索を行い、具体的な森林管理の指針に対して情報をより提供できる回帰式へと発展させることも期待される。また、本稿においては生態系サービスごとの評価を行う際に、回答負荷が低くバイアスが少なくと考えられている部分プロファイル選択型のコンジョイント分析を採用したが、他の手法との比較を含めて手法的な発展させることが今後求められる。さらに、回答負荷を減らしバイアスを回避する手法的改善だけでなく、樹種や樹齢といった森林特性に対する評価者の知識や理解の影響も検討する必要がある。

こうした環境の経済評価は、外部性や非利用価値などの市場メカニズムでは対応できない要因に対して、環境政策として対応するためにそれらの価値を推定する際に求められる。本研究は自然資本が持つ価値について、市場ベースの価値からの乖離を測定し、環境政策の必要性や妥当性を議論する上で有用な情報を提供するものである。例えば森林生態系サービスの価値評価により、森林資源の保全策の必要性や妥当性が示唆され、政策手段の議論へと発展する。今後は、森林整備について、どのような森林をどの空間に整備するかという個別具体的な政策立案に資するような、森林便益の空間的分布などを詳細に分析することが求められる。こうした研究は、水源税や森林環境税など、いわゆる生態系サービスに対する支払い (PES; Payment for Ecosystem Services) を制度に組み入れる際の経済学的基礎を与えることにもつながるため、政策的ニーズや重要性は大きいのである。

注.

- 1) 権利の所在などによって、等価余剰で定義する場合もある。鷲田 (1999) に詳細な解説がある。NOAA パネルなどで推奨されている生態系価値の評価枠組みは、WTP に基づくものである。これは環境の「控えめな」評価に対応するものである (NOAA パネル, Arrow et al., 1993)。
- 2) CVM の発展の経緯については、栗山 (1998) に詳しい。
- 3) 様々なバイアスについては、Mitchell and Carson (1989) や栗山 (1998) にまとめられている。
- 4) もちろん、非常に高い樹齢を持つ樹木は特殊な価値を帯びてくる。しかし本研究で対象としているのは平均樹齢であるため、マクロ的に平均して非常に高い樹齢となることはなく、およそ 20~50 年ほどの範囲に限られた議論となる。
- 5) 回帰式に導入した変数がすべて有意に推定されたことから、変数選択は行わなかった。
- 6) 各都道府県の森林データについては、林野庁 (2012) を用いた。
- 7) この価値は、林業従事者にとっての木材価値とは異なるため、市場における木材価格とみなすことはできない。

参 考 文 献

- Alberini, A. and Kahn, J. R. (2006). *Handbook on Contingent Valuation*, Edward Elgar, Cheltenham.
- Arrow, K., Solow, R., Portney, P. R., Leamer, E. E., Radner, R. and Schuman, H. (1993). Report of the NOAA panel on contingent valuation, *Federal Register*, **58**(10), 4601-4614.
- Bateman, I. J. and Willis, K. G. (1999). *Valuing Environmental Preferences*, Oxford University Press, Oxford.
- Bateman, I. J., Carson, R. T., Day, B., Hanemann, M., Hanley, N., Hett, T., Jones-Lee, M., Loomes, B., Mourato, S., Özdemiroğlu, E., Pearce, D. W., Sugden, R. and Swanson, J. (2002). *Economic Valuation with Stated Preference Techniques*, Edward Elgar, Cheltenham.
- Bjornstad, D. J. and Kahn, J. R. (1996). *The Contingent Valuation of Environmental Resources*, Edward Elgar, Cheltenham.
- Carlsson, F. and Martinsson, P. (2001). Do hypothetical and actual marginal willingness to pay differ in choice experiments?, *Journal of Environmental Economics and Management*, **41**, 179-192.
- Carson, R. T., Hanemann, W. M., Kopp, R. J., Krosnick, J. A., Mitchell, R. C., Presser, S., Ruud, P. A. and Smith, V. K. (1994). Prospective interim lost use value due to DDT and PCB contamination in the Southern California Bight, *Report to the National Oceanic and Atmospheric Administration*, Natural Resource Damage Assessment Inc.
- Carson, R. T., Flores, N., Martin, K. and Wright, J. (1996). Contingent valuation and revealed preference methodologies, *Land Economics*, **72**, 80-99.
- Carson, R. T., Hanemann, W. M., Kopp, R. J., Krosnick, J. A., Mitchell, R. C., Presser, S., Ruud, P. A. and Smith, V. K. (1997). Temporal reliability of estimates from contingent valuation, *Land Economics*, **73**(2), 151-161.
- Champ, P. A., Boyle, K. J. and Brown, T. C. (2003). *A Primer on Nonmarket Valuation*, Kluwer Academic Publishers, Boston.
- Ciriacy-Wantrup, S. V. (1947). Capital returns from soil-conservation practices, *Journal of Farm Economics*, **29**, 1181-1996.
- Davis, R. K. (1963). Recreational planning as an economic problem, *Natural Resources Journal*, **3**,

- 239-249.
- Fujii, H., Sato, M. and Managi, S. (2017). Decomposition analysis of forest ecosystem services values, *Sustainability*, **9**(5), p.687.
- 藤本高志 (1998). 『農がはぐくむ環境の経済評価』, 農林統計協会, 東京.
- 学術会議 (2001). 地球環境・人間生活にかかわる農業及び森林の多面的な機能の評価について, 学術会議答申.
- Haab, T. C. and McConnell, K. E. (2002). *Valuing Environmental and Natural Resources*, Edward Elgar, Cheltenham.
- Herriges, J. A. and Kling, C. L. (1999). *Valuing Recreation and the Environment*, Edward Elgar, Cheltenham.
- Hicks, J. R. (1943). The four consumer's surplus, *Review of Economic Studies*, **11**, 31-41.
- Kahneman, D., Ritov, I. and Schkade, D. (1999). Economic preferences or attitude expressions? An analysis of dollar responses to public issues, *Journal of Risk and Uncertainty*, **19**, 203-235.
- 栗山浩一 (1997). 『公共事業と環境の価値: CVM ガイドブック』, 築地書館, 東京.
- 栗山浩一 (1998). 『環境の価値と評価手法』, 北海道大学出版会, 札幌.
- 栗山浩一 (1999). 環境評価の現状と課題—CVM, コンジョイント分析を中心に—, 『環境評価ワークショップ』(鷺田豊明, 栗山浩一, 竹内憲司 編), 第 2 章, 築地書館, 東京.
- 栗山浩一, 北島能房, 大島康行 (2000). 『世界遺産の経済学』, 勁草書房, 東京.
- 馬奈木俊介, 地球環境戦略研究機関 (2011). 『生物多様性の経済学』, 昭和堂, 京都.
- McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior, *Frontiers in Econometrics* (ed. P. Zarembka), Academic Press, New York.
- MEA (2005). *Millennium Ecosystem Assessment*, Island Press, Washington, D.C.
- Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information, *The Psychological Review*, **63**(2), 81-97.
- Mitchell, R. C. and Carson, R. (1989). *Using Surveys to Value Public Goods: The Contingent Valuation Method*, Resources for the Future, Washington, D. C. (環境経済評価研究会 訳. 『CVM による環境質の経済評価: 非市場財の価値計測』, 山海堂, 東京.)
- Nocera, S., Telsler, H. and Bonato, D. (2003). *The Contingent Valuation Method in Health Care: An Economic Evaluation of Alzheimer's Disease*, Kluwer Academic Publishers, Boston.
- Ready, R., Navrud, S. and Dubourg, W. (2001). How do respondents with uncertain willingness to pay answer contingent valuation questions?, *Land Economics*, **77**, 315-326.
- 林野庁 (2012). 森林資源の現況(平成 24 年 3 月 31 日現在), <http://www.rinya.maff.go.jp/j/keikaku/genkyou/index1.html>.
- Roe, B., Boyle, K. J. and Teisl, M. F. (1996). Using conjoint analysis to derive estimates of compensating variation, *Journal of Environmental Economics and Management*, **31**, 145-159.
- 坂上雅治, 栗山浩一 (2009). 『エコシステムサービスの環境価値』, 晃洋書房, 京都.
- Stevens, T. H., Belkner, R., Dennis, D., Kittredge, D. and Willis, C. (2000). Comparison of contingent valuation and conjoint analysis in ecosystem management, *Ecological Economics*, **32**, 63-74.
- 竹内憲司 (1999). 『環境評価の政策利用』, 勁草書房, 東京.
- TEEB (2010). *The Economics of Ecosystems and Biodiversity Ecological and Economic Foundations* (ed. Pushpam Kumar), Routledge, London.
- 柘植隆宏, 三谷羊平, 栗山浩一 (2011). 『環境評価の最新テクニック』, 勁草書房, 東京.
- Turner, K., Pearce, D. and Bateman, I. (1994). *Environmental Economics*, Harvester Wheatsheaf. (大沼あゆみ 訳 (2001). 『環境経済学入門』, 東洋経済新報社, 東京.)
- 植田和弘 (1996). 『環境経済学』, 岩波書店, 東京.
- 鷺田豊明 (1999). 『環境評価入門』, 勁草書房, 東京.
- 鷺田豊明, 栗山浩一, 竹内憲司 (1999). 『環境評価ワークショップ』, 築地書館, 東京.

- Welsh, M. P. and Poe, G. L. (1998). Elicitation effects in contingent valuation: Comparisons to a multiple bounded discrete choice approach, *Journal of Environmental Economics and Management*, **36**, 170-185.
- Wilis, K. G. and Garrod, G. (2012). *Valuing Environment and Natural Resources*, Edward Elgar, Cheltenham.

Economic Valuation of Forest Ecosystem Service in Japan

Masayuki Sato¹, Koichi Kuriyama², Hidemichi Fujii³ and Shunsuke Managi⁴

¹Graduate School of Human Development and Environment, Kobe University

²Graduate School of Agriculture, Kyoto University

³Graduate School of Economics, Kyushu University

⁴Graduate School of Engineering, Kyushu University

This study estimates both unit and attribute values of forest ecosystem services in Japan. In this estimation, we examined the applicable valuation techniques from the viewpoint of environmental economics, in particular regarding the economic valuation method of ecosystem services. Our results revealed that direct use, indirect use, and non-use values are important in evaluating ecosystem services. In addition, it is important to incorporate indirect use and non-use values into the ecosystem evaluation framework. In order to reflect such values, we used the payment card Contingent Valuation Method (CVM) to estimate the unit value of forest in Japan. In addition, to estimate the attribute value of each forest ecosystem service, we conducted choice experiments that share the same theoretical foundation of a random utility model with CVM. Our findings visualized the ecosystem service value of forests in Japan.

農林業の生態系サービスと自然資本の グローバル分析

池田 真也¹・山口 臨太郎²・馬奈木 俊介^{3,4}

(受付 2018 年 6 月 30 日; 改訂 9 月 20 日; 採択 10 月 9 日)

要 旨

生態系サービスは人間の福祉にとって必須であるため、その価値評価は環境政策立案のためにも重要である。本稿では最新の国連報告書(Inclusive Wealth Report 2018)で公開された農林業の生態系サービスをもたらす再生可能な自然資本(森林と農地)のデータを用い、その資本ストックの変化からグローバルな傾向を明らかにした。主要な結果の一つとして、自然資本は総じて減少傾向にあるものの、南アメリカの農業国では農地の価値が増加する一方で、ヨーロッパを中心とした先進国では農地の価値が減少しているなどの地域差が明らかになった。今後の課題として、生態系の負のサービスや、都市部における生態系サービスなどに評価対象を拡張することが挙げられる。

キーワード：農地資本，森林資本，国富，資本評価。

1. はじめに

自然資本は森林、生態系、鉱物、大気といった物質、ないし情報のストックとして構成され、ある時点、ある場所に存在するものである(Costanza et al., 1997)。自然資本の中でも、農林業に関連する森林と農地は、それぞれ世界の地表の 30.8% と 37.3% を占めることから、人間の福祉にとっての価値を容易に窺い知ることができる¹⁾。自然資本としての生態系から生じるサービスフローが生態系サービスであり、中でもサンゴ礁、熱帯雨林、沿岸湿地がもたらす生態系サービスの価値は他の土地資源がもたらすサービス価値に比べても高く、農地などへの土地利用変化で生態系サービスの世界全体の価値は 1997 年から 2011 年の間に年間 4~20 兆ドル減少したとする研究もある(Costanza et al., 2014; de Groot et al., 2012)²⁾。なかでも農林業に直接関連する熱帯雨林の減少は、かねてより注目を集めてきた問題であり、途上国で多く見られる熱帯地域の森林資源は、商品作物(コーヒー、パーム油、大豆)生産用の農地へ転換されることも多い(DeFries et al., 2010; Gibbs et al., 2010; Peters et al., 1989)。また、湿地は多様な生物の生息地の提供、水源涵養や二酸化炭素隔離といった機能を持つが、20 世紀中に北米、ヨーロッパ、オーストラリア、ニュージーランドなどで主に農地へと転換された結果、その面積の 50% 以上が失われた(MEA, 2005)。そして土地利用変化で得られる農作物生産の利益以上に、

¹ 茨城大学 農学部：〒300-0393 茨城県稲敷郡阿見町中央 3-21-1

² 国立研究開発法人国立環境研究所 社会環境システム研究センター：〒305-8506 茨城県つくば市小野川 16-2

³ 九州大学大学院 工学研究院：〒819-0395 福岡県福岡市西区元岡 744

⁴ 九州大学 都市研究センター：〒819-0395 福岡県福岡市西区元岡 744

生物多様性やその他生態系サービスの損失を招いたのである (Hassan et al., 2005). このように農林業が内包する生態系サービスの源泉となる自然資本は減少傾向にあり, その対策は喫緊の課題として広く認識されている.

生態系サービスと再生可能な自然資本は, フローとストックの対応関係にある. 具体的には, 生態系サービスの年間のフロー価値の将来までの総和が, 自然資本の価値となる. その際, 重要な論点がいくつかある.

第一に, 生態系サービスと自然資本は表裏の関係にあるものの, 分析のバウンダリーが異なる. 自然資本は, 再生不能資源 (石油・天然ガス, 石炭, 鉱物資源) と再生可能資源 (森林, 農地, 漁業資源) とに分かれるが, 生態系サービスが対象とするのは, 再生可能資源のみである. 再生不能資源も, 動植物が何億年もかけて化石化した資源としてサービスをもたらすが, 生態系サービスの範疇には通常入らない.

逆に, 生態系サービスであればすべて自然資本になるわけでもない. たとえば水は, 貴重な生態系サービスの一つとされるが (Onofri et al., 2017), 水ストックそのものを自然資本と見なすことは困難である. これは, 水というサービスを一つの資本に帰着させるのが難しいだけでなく, 水の限界的な価値が状態に大きく依存し価値づけが難しいためである (Perry, 2012). ただし, 空間的なスコープを限定すれば, 価値づけが可能であることもある (Fenichel et al., 2016).

第二に, 自然資本の単位当たり価値は, 対象となる自然資本が追加的にあと 1 単位 (たとえば森林 1 ヘクタール) 増えたときに社会的福祉がどれだけ変わるかで表す (Arrow et al., 2012). これを自然資本のシャドウ価格と呼ぶ. 自然資本のシャドウ価格は, 生態系サービスの年間価値を, 妥当な割引率を使って資本化した値になる. t 時点における生態系サービス $D(t)$ の限界価値を $MB(\cdot)$ と表すと, 自然資本の t 時点におけるシャドウ価格 $p_N(t)$ は,

$$p_N(t) = \sum_{\tau=t}^{\infty} \frac{MB(D(\tau))}{(1+r(t+1))(1+r(t+2)) \cdots (1+r(\tau))}$$

となる (Arrow et al., 2003). 上式の分母において, 再生可能な自然資本の場合, 割引率 $r(\tau)$ は, 純粋時間選好率だけでなく, 自然資本の相対的な成長率も反映したものになる (Fenichel and Abbott, 2014). この成長率には, 人間による農地や森林の使い方が成長に与える影響も含む. 自然資本の希少性が大きく相対的な成長率がプラスであれば, 割引率は小さくなり, シャドウ価格は大きくなる. 逆に, 自然資本が比較的豊かにあり, 相対的な成長率がマイナスで絶対値が大きいほど, 割引率は大きくなり, シャドウ価格は小さくなる.

また上式の分子において, 生態系サービスの限界価値は, 今年追加的に得られるサービス価値のことである. そのため, 自然資本ストック全体の価値は, 資本ストックが希少になるにつれて上がる. 単純化した例で言うと, ある地域の森林の乱伐が進んだ結果, 最後に残された一本の木の価値は, 豊富な森林の中の一本の木よりも極めて大きいだろう. 経済学の用語を使うと, 生態系サービスの今年の価値には限界価格を使うが, 自然資本の全体のストックには平均価格を使うことになる. ただし, 実際に自然資本のストックを限界価値で評価することは難しいため, 実際の会計では, 限界価格にすべての数量を掛けたものとなっている. このため, 自然資本ストックの総価値には意味がなく, 限界価格に数量の増減を掛けたもの (自然資本の価値の変化) だけに意味があることに注意が必要である.

第三に, 広範な「生態系」の定義の難しさ, 利用可能なデータの制約 (特に途上国のマイクロデータ) により, 自然資本, およびその生態系サービスの評価には依然として多くの課題がある³⁾. もともと, 「生態系」の定義は広範で, 市場で取引されるものが少なく, 投入と産出が不明瞭なことも多いために, 経済学的観点からの評価は難しかった (Boyd and Banzhaf, 2007).

たとえば、生態系からの生産物(食料や水)、生態系の機能とプロセス(栄養循環や生息地供給)、生態系からの便益(洪水管理など)がまとめて議論されることが多い。MEA (2005)も、供給サービス、調整サービス、文化サービス、基盤サービスという分類を提示している。経済学の観点からは、人の福祉に直接的に影響する生態系サービス、いわば最終消費材・サービスのみを対象とすることが必要である。そうすることで生態系の機能とプロセスが対象から外れるとの批判もあるが、機能やプロセスは中間財と位置付けることにより、これまでの評価過程で生じた重複計上の問題を回避できる(Kroeger and Casey, 2007)。また、人の福祉に直接影響する生態系サービスだけを対象とするにしても、一般に市場で取引が行われていないサービスの価値づけは、生産関数アプローチや、その他の環境評価(顕示選好、表明選好、便益移転)を用いざるを得ない(Barbier, 2012)。これらの課題は、生態系サービスを生み出す自然資本のシャドウ価格にも共通する。

このように生態系サービスと自然資本の評価方法には議論もあるが、特に価値の変化は、人の福祉が向上しているかどうかに影響することから、何らかの評価を行うことが重要である(Costanza et al., 1997)。また、生態系サービスを生み出す自然資本を貨幣評価することは、他の資本とのトレードオフに関する情報を通じて、俯瞰的な環境・経済政策の立案の基礎となり得る。たとえば、市場取引が行われない生態系サービスと自然資本の貨幣価値を推計し、国民経済計算体系(SNA: System of National Accounts)に取り入れる動きは2010年の生態系サービスと生物多様性の経済学(TEEB)勧告後に加速している(林・佐藤, 2016)。国連統計局の環境・経済会計体系(SEEA: System of Environmental-Economic Accounting)はその一例である(Obst et al., 2016; Hamilton, 2016)。

さらに、厚生経済学の理論に基づいて(Dasgupta, 2009)、生態系サービスを生み出す自然資本だけでなく、人的資本や人工資本も含めた富の評価枠組みを用いた、グローバルレベルのデータベース作成も進められている。たとえば、世界銀行の調整純貯蓄(ANS: Adjusted Net Saving)と富会計の一連の報告書、そして国連環境計画(UNEP)を中心とした新国富指標(IWI: Inclusive Wealth Index)が挙げられる⁴⁾。この二つのプロジェクトは、評価対象国や計測期間、各資本や富の計測方法に共通点もあるが、相違点もある(Arrow et al., 2012; Engelbrecht, 2016)⁵⁾。

以下では、これらのうちIWIを対象に富と自然資本の分析を行う。最新の自然資本ストックの現状をこれらの指標から理解するためには、富の変化を把握することは重要であり、それゆえに計測期間の長いデータベースを用いることが必要である。その上で二つのプロジェクトにおける計測期間を比較し、世界銀行の富会計(1995-2014年)よりも5年長いIWI(1990-2014年)を分析対象とするのが適当だろう⁶⁾。以下ではIWIにおけるグローバルな富の変化と再生可能な自然資本の取り扱いを概観したのち、農林業における生態系サービスおよび自然資本のグローバルな特徴を明らかにする。その上で、今後の生態系の価値評価に重要と思われる課題を検討する。

2. Inclusive Wealth Report 2018 における富と自然資本

自然資本のみならず、人工資本(工場や道路や住宅など)や人的資本(教育と健康)と合わせた社会全体の富を算出し、その変化に基づいて持続可能性を判断する最新のInclusive Wealth Report 2018 から、過去四半世紀の自然資本の変化について報告する。

この報告書では、140か国を対象に、1990年から2014年までの人工資本、人的資本(教育、健康)、自然資本が算出されている。自然資本は、再生不能な自然資本(石油、ガス、石炭、鉱物資源)と再生可能な自然資本(森林、農地、漁業資源)とに分かれる。後者のうち、森林には、木材の市場価値や、非市場価値として扱われる森林の生態系サービスの価値が、農地には耕地

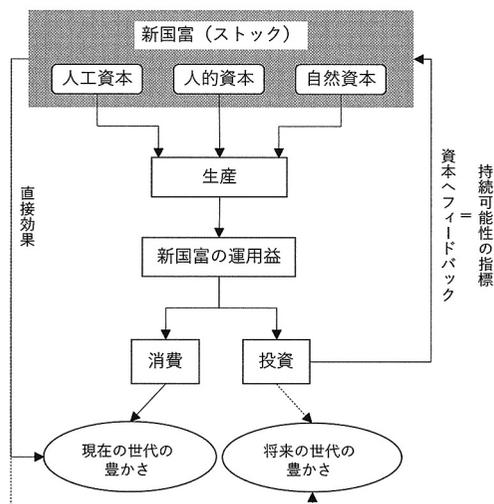


図 1. 新国富における福祉(豊かさ)の基本的概念. 出所) 山口 他 (2016) 図 1 より筆者作成.

や牧草地の市場価値がそれぞれ含まれている。図 1 に示したように、これらの資本ストックは生産活動に用いられ、その運用益が消費と投資に充てられ、それぞれ現在および将来世代の福祉(豊かさ)に繋がっていく。

持続可能な発展の定義は識者により異なるが、Inclusive Wealth Report 2018 では「福祉の増進(正の変化)」として定義しており、本稿も基本的には同じ定義に従う⁷⁾。ここで福祉とは、現在世代だけでなく将来世代を含んだ「社会的福祉」とする。福祉を直接測ることは難しいので、福祉に影響する三つの資本の価値を測るとというのが基本的なアイデアである。その際、暗黙の裡に自然資本、人工資本、人的資本間の代替性を前提としている。しかし実際、自然資本には他の資本で置き換えることができないものもあるため、自然資本そのものの増減、および自然資本の各要素の増減も同時に注視すべきである。このような捉え方を「強い持続可能性」基準と呼ぶ。

まず、再生不能資源と再生可能資源の合計である自然資本が増加していた国は、対象 140 か国中 31 か国のみであった。また図 2 の第 2 象限に示すように、過去四半世紀に 1 人当たり富は増えたが 1 人当たり自然資本は減少したという国が大半であった。非常におおざっぱな言い方をすると、自然資本を消耗することで他の資本を増やしてきた国が大半であるということになる。国レベルでのマクロデータに基づいた会計であるため、ボトムアップでの自然資本の分析は行われていないが、評価の難しい生態系サービスを生み出す再生可能な自然資本だけでなく、再生不能な自然資本も併せて、富の全体像を把握しておくことは重要である。

再生可能な自然資本の価値の算出は、毎年の生態系サービスの限界価値の割引現在価値を用いる。上式に示した通り、再生可能資源の場合、本来であれば相対的な成長率を勘案すべきだが、国全体のマクロ的な価値の計算で個別に計算することはほぼ不可能であるため、 r は一律 5% とされている。

既に述べたように、自然資本の価値の増減が問題であるため、自然資本ストックの価値そのものに経済学的な意味はない。そのためあくまで参考値ではあるが、世界全体の自然資本の内訳としては、再生可能な自然資本がほぼ半分を占めている(図 3)。特に森林の価値が全体の 3

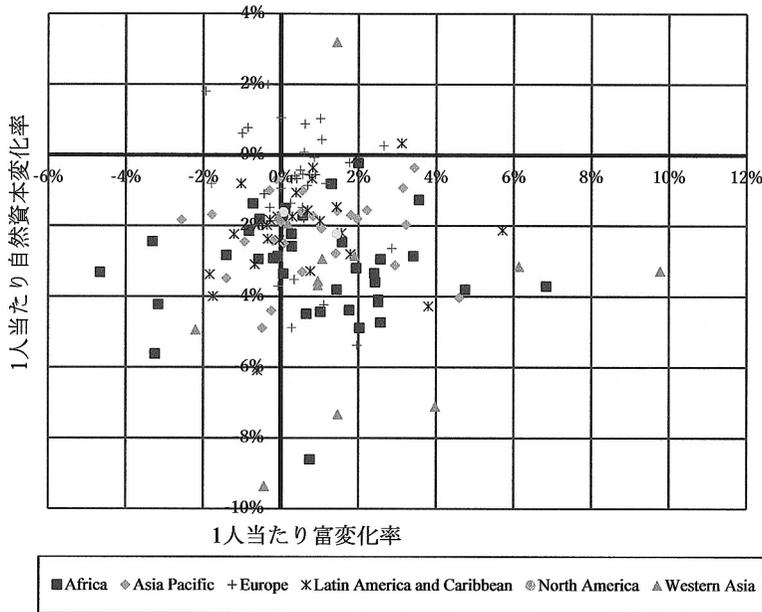


図 2. 1人当たり富変化率と1人当たり自然資本変化率.

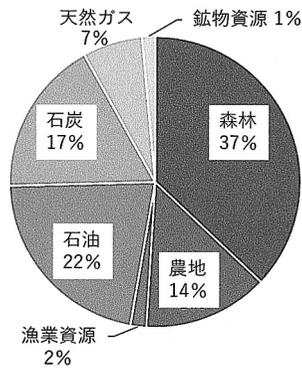


図 3. 世界の自然資本ストック価値の内訳.

分の1を占めることは、様々な生態系サービスが反映された結果ともいえる。これに対して農地がもたらすサービスとしては、農作物の生産という供給サービスのみ計上されている。

3. 農林業における生態系サービスと自然資本

3.1 森林資本と農地資本の時系列変化

Inclusive Wealth Report 2018の付録より、4つの所得水準で国グループを作成し、各々の一人当たり集計値を用いた時系列変化を図4に示した。なお、農地資本データは耕地、牧草地の二つの土地資本から成り立ち、また森林資本データは木材生産由来の資本(森林資本(木材)と

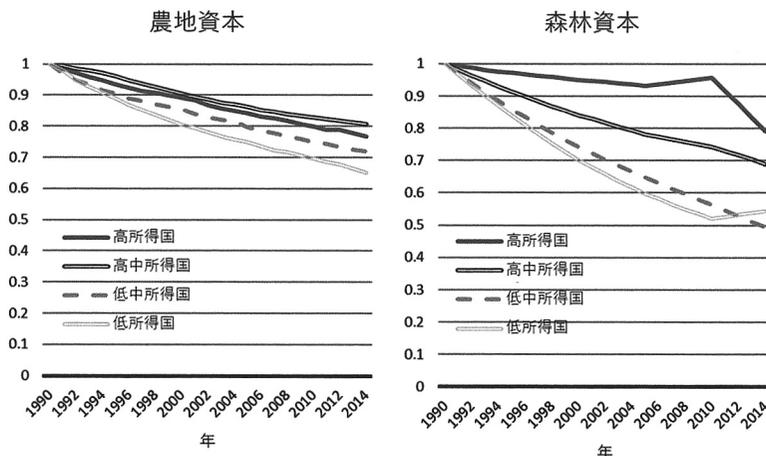


図 4. 所得グループ別の自然資本の変動。注) 農地資本には耕地, 牧草地を含む。また, 森林資本には木材, 非木材を含む。

記載)と, 木材供給以外の生態系サービスに由来する資本(森林資本(非木材)と記載)から成り立つ。ただし, Managi and Kumar (2018)で対象とした 140 か国の中に欠損データがあるため, 高所得国 25 か国, 高中所得国 24 か国, 低中所得国 20 か国, 低所得国 10 か国の計 79 か国を図示している。

まず, 図 4 左に示した農地資本の変化を概観すると, 全ての所得グループで農地資本が低下傾向にあることが指摘できる。また, 低所得国グループが最も大きく農地資本を減らしており, 24 年間で 35% 程度減少している。他方で, 高中所得国グループの農地資本の低下が緩やかであり, 24 年間で 19% の低下であった。興味深いのは高所得国グループの農地資本は同期間に 23% と, 高中所得国グループより大きく低下していた点である。その要因には, 高所得国グループに属し, かつ農地資本の総額が高いオーストラリアとアメリカで 24 年間にそれぞれ 13%, 5% 減少している点を指摘したい。人口増加も踏まえれば, 2 国が農地資本の減少に与えた影響は小さくないだろう。そして, シャドウ価格は一定と仮定しているため, これらの国の農地自体が減少しているのである。後で触れる熱帯雨林地域とは逆の現象が起こっていると言える⁸⁾。

次に森林資本の変化を示したのが図 4 右であるが, 農地資本と同様, 全ての所得グループで低下傾向を示している。ただし, 低所得国グループの森林資本は 2010 年以降改善傾向にあり, 2014 年時点では低中所得国グループよりも森林資本の減少幅が小さくなっていた点は注目に値する。図 5 から, これは主に森林資本(木材)の増加が要因と推察される。他方で, 生態系サービスがもたらす価値を計測している森林資本(非木材)の価値においては, 低所得国グループで改善は見られなかった。1990 年から一貫して減少傾向にあり, その価値は 40% 以上失われていた。むしろ, 森林資本(非木材)の価値は高所得国で 2010 年ごろから減少しておらず, FAO (2016)で明らかにされたような世界的な森林保護政策の一定の成果と整合的である。なお, 高所得国グループの木材価値が 2010 年以降急激に減少している。その要因として, アメリカ, ニュージーランドなどの木材の大生産地での資本価値の減少傾向を指摘できるが, 最も大きく減少していたのはカナダであった。しかし, その背景を現状では確認できなかった。

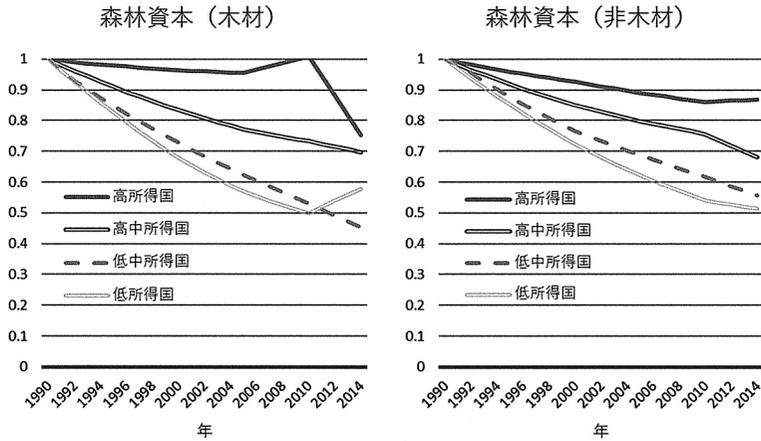


図 5. 森林資本の変動.

3.2 森林の生態系サービス

森林資本の非市場価値を計測する際に使われるシャドウ価格は、温帯林より熱帯雨林のほうが高いため、森林に占める熱帯雨林の割合が高いほど高く計算される⁹⁾。本稿の趣旨は農林業に係る生態系サービスの価値であることから、世界 140 か国の中で、1990 年時点で森林に占める熱帯雨林の比率が 50% を超える 13 か国 (以下、熱帯雨林グループ) と、それ以外 (以下、非熱帯雨林グループ) とにグループ分けして森林資本の分析を行った¹⁰⁾。グループ分けの基準年を 1990 年時点としたのは、一般的に熱帯雨林が減少傾向にあることを受け、失われた森林資本の大きさに注目するためである。

図 6 に示したように、熱帯雨林グループでは耕地資本の価値がほぼ一定に保たれており、他の資本と比べて減少していない点特徴的である。また、森林資本 (非木材) は、農地資本 (耕地、牧草地) や森林資本 (木材) よりも減少している。熱帯雨林グループに関してよく指摘される森林から耕地への土地利用変化と整合的な結果である。そして、仮にそのような土地利用変化が起きていたのだとしたら、自然資本全体を維持する強い持続可能性の立場からは、森林資本の非市場価値の低下分を補うほどの耕地資本の増加が不可欠である。しかし、残念ながら結果的に自然資本は全体として減少しており、強い持続可能性は損なわれている¹¹⁾。また、森林資本 (非木材) と耕地資本に何らかの代替性がないかを分析するために、各資本の成長率に関する散布図を図 7 に示した。何らかの代替性があれば、図 7 の第 4 象限にプロットが集中するはずであるが、そのような傾向は、熱帯雨林グループか否かにかかわらず観察されなかった。

1 点注意しなくてはならないのは、ここでは牧草地の価値をあえて除外して分析を行った点である。その理由としては、UNU-IHDP and UNEP (2014) と同様に Managi and Kumar (2018) では牧草地の単位面積当たりのシャドウ価格が耕地のそれと同一という仮定が置かれており、詳細な価値評価になっているとは言えないからである。なお、非熱帯雨林グループで 2010 年以降に森林資本 (木材) の価値が減少しているのは前述の 3 か国における森林面積の減少によるものである。

3.3 農業資本の変化と持続可能性

ところで、熱帯雨林グループで農地資本の減少幅が小さかったことは、国によってその傾向が異なり、異質性が存在する可能性を示唆している。そこで、いくつかの切り口から対象国を

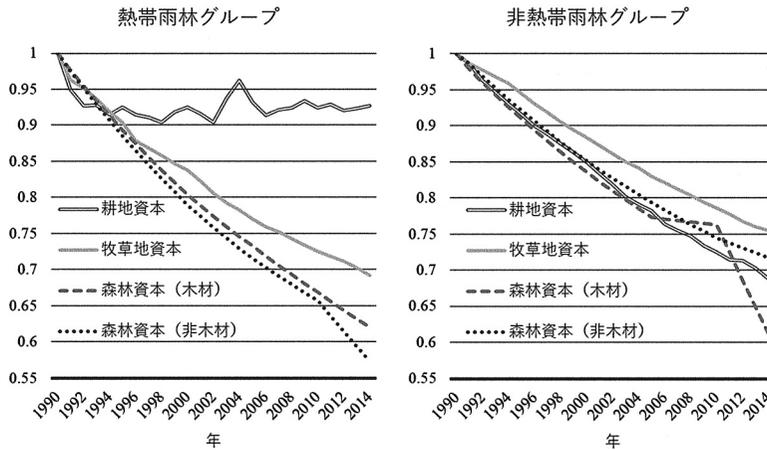


図 6. 熱帯雨林グループ別の自然資本の変動.

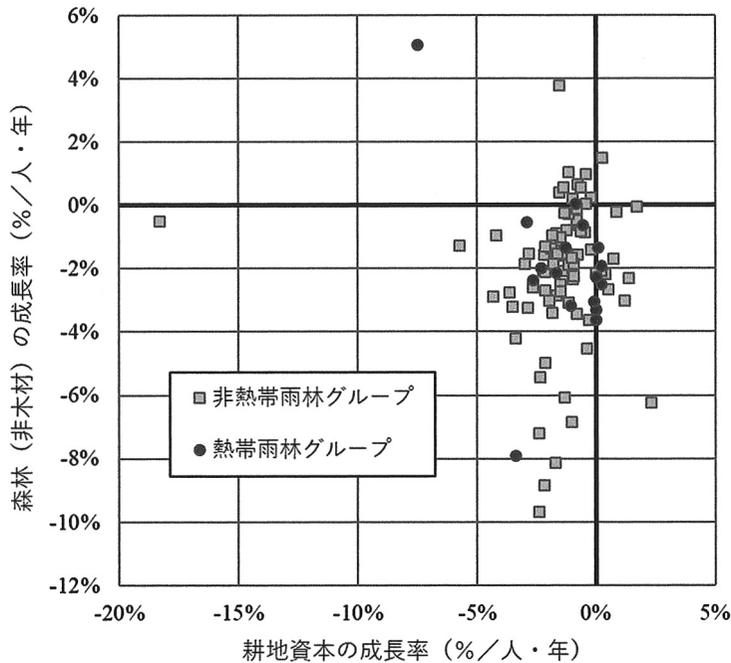


図 7. 耕地資本と森林資本(非木材)の成長率(1990~2014年).

選択し、農地資本のより詳細な分析を行う。

まず、データが揃っている 83 か国に関して、その農業生産額と耕地資本の成長率との関係を示したのが図 8 である¹²⁾。農業生産額が増加し、なおかつ耕地資本も増加していたのは 13 か国であり、農業生産額は減少しているものの、耕地資本が増加しているのがガンビアとオランダの 2 か国だった¹³⁾。これらを合わせた 15 か国は耕地資本を増加させているという点で

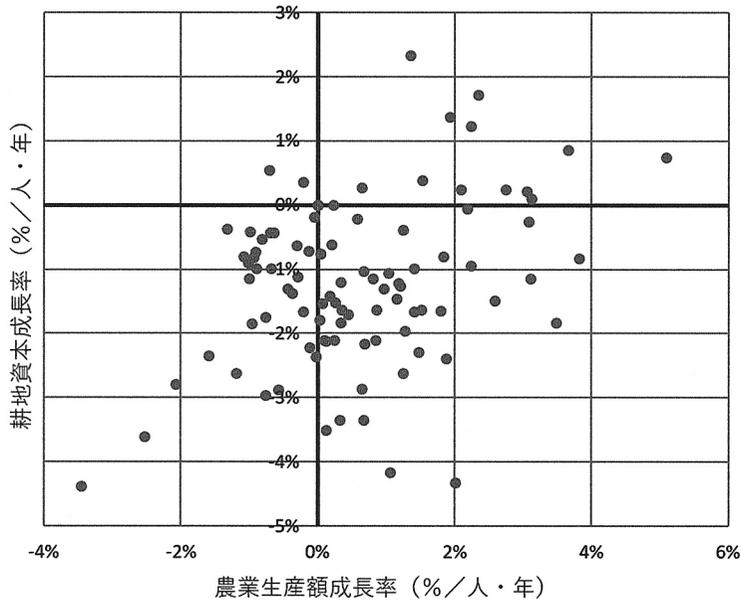


図 8. 農業生産額と耕地資本の成長率(1990～2014年)．注)モルディブ，ニュージーランド，シンガポールを除外して記載．

—強い持続可能性の意味で—持続可能な農業生産ができていると言える．15か国の特徴としては、低中所得国(7か国)、低所得国(5か国)と相対的に所得が低い国が多い点にある．また地域的には南アメリカが多い(6か国)．これらの地域は国際的な農業生産国であり農地の拡大が続いていると推察されるが、一方でルーマニアとオランダを除く13か国で森林資本は減少していた．この点は熱帯雨林グループと同様に、森林から農地への土地利用変化が生じている可能性を示唆している¹⁴⁾．

他方で、分析対象国のうち約3分の1に相当する32か国において農業生産額が低下し、かつ耕地資本も減少していた¹⁵⁾．32か国のうち23か国が高所得国であり、また約半数の15か国がヨーロッパ地域の国であった．たしかにヨーロッパ諸国では放棄された農地に関する問題が長年議論されており、主原因に小規模農家による粗放農業が減少している点(Henle et al., 2008)と整合的である．またこれら15か国の半数にあたる7か国では森林資本が増加していることから、生物多様性の保護を背景に農地よりも森林面積の増加が進んでいる国が多いことが推察される．しかし、これらの土地利用転換の背後には、社会、経済、環境的なプロセスが複雑に交差しており(Renwick et al., 2013)、その要因を詳細に明らかにすることは本研究の射程を大きく超えている．その要因分析に向けて、農地と森林の土地利用に関する詳細な地理データとそれに対応するシャドウ価格の精緻な実証研究の積み重ねが重要である．

上記の分析においては、1990年から2014年までの年平均変化率をみたが、世界有数の農業生産国に焦点を当て、その耕地資本の変動を観察することも重要だろう．図9に17か国の耕地資本の変動を示した．17か国は2014年の農業生産額上位国から選択した．アルゼンチン、ブラジル、インドネシアの3か国が1990年比で増加を示しているが、特にアルゼンチンは2002年以降に耕地資本を増加させており、その変動幅が大きい．そして、ドイツ、フランスに関しては耕地資本が減少しているものの、他の農業国に比べてその減少は緩やかだった．

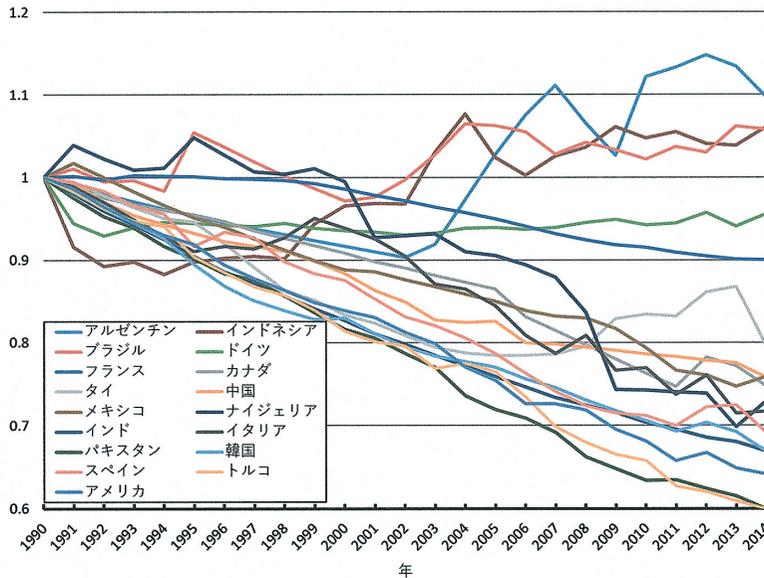


図 9. 主要農業国の耕地資本の変動.

4. 今後の課題

4.1 生態系のディサービス—メタンガスの例

MEA (2005) の分類に従えば、生態系は供給サービスだけでなく、生命の基礎となる基盤サービス、気候・洪水制御や水源涵養などの調整サービス、レクリエーションなどの文化サービスを通じて、人間の福祉の向上に貢献している（本稿では人間中心主義を前提とする）。一方で、害虫や野生動物の鳥獣害など、生態系は人間にとってディサービス（負のサービス）ももたらすが、価値化が困難であることもあり、今のところ自然資本の価値としては計上されていない。

負のサービスとしてグローバルに最も計上しやすく、かつ影響も大きいのが、耕地や放牧地が放出する温室効果ガスであろう。メタンや窒素酸化物は、二酸化炭素よりも温室効果係数が大きいものの、炭素の社会的費用(SCC)には定義上含まれておらず、富会計でも考慮されていない。

そこで 1990～2015 年のメタンや窒素酸化物の二酸化炭素相当排出量(ktCO_2e)を、炭素 1t 当たり 38 ドル(USD38/tC)の SCC を用いて金銭換算したのが図 10 である¹⁶⁾。過去四半世紀に、農地がもたらす温室効果ガスの費用は、年間 6,000 億ドルから 7,000 億ドルに増加している。ヨーロッパとオセアニアで減少している一方、アジア、アメリカ、アフリカの 3 大陸では増加している。

仮に図 10 の 2015 年の水準が永遠に続くとする、5%の割引率で農地の温室効果ガス排出による被害フロー額を単純に資本化した値は、世界全体で 1.4 兆ドルとなる。世界全体の農地の自然資本としての価値は、2010 年に 12.5 兆ドルであるとされている (UNU-IHDP and UNEP, 2014)。そのため温室効果ガスというディサービスも考慮した農地の価値は、大まかに 1 割程度少なく見積もるべきであろう。

上記の計算は、“将来の”生態系サービスのフローを農地という自然資本の価値に反映させたものである。これに対して、“過去の”サービスのフローにより過去の富の動きを調整すること

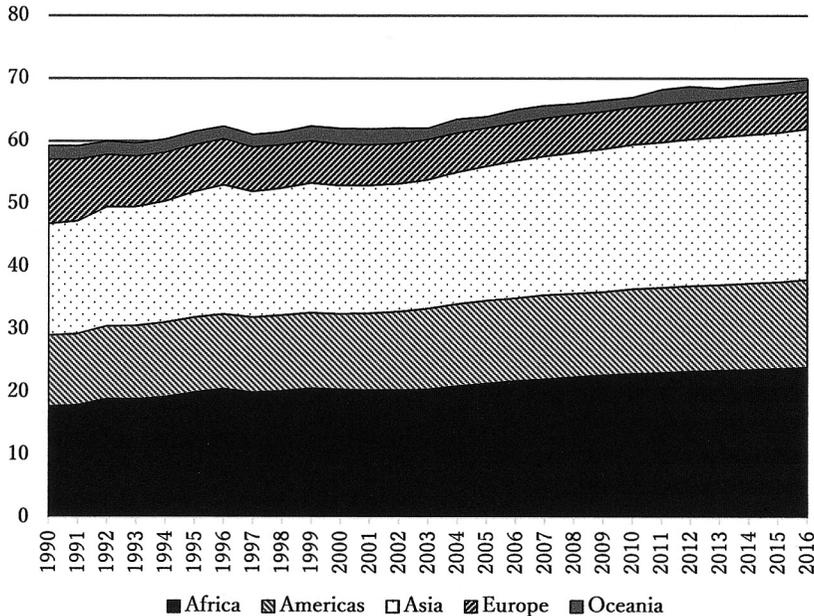


図 10. 農地が排出するメタン・窒素酸化物による被害の推移(単位：十億ドル)。

もできる。人工・人的・自然資本には入らないが、福祉に影響する要因として、炭素の被害・石油価格のキャピタルゲイン・全要素生産性の三つが計上され、調整済みのIWIとして報告されている。そこで、炭素の被害と同様に、過去に世界の農地から排出された温室効果ガスによる調整を行うとすると、世界の富の変化率を0.01%のオーダーで減じることになる。炭素排出によるIWIの調整の10分の1ほどの被害を通じて、各国の福祉の持続可能性に影響を与えてきたことがわかる。

また、農地のデイスサービスそのものではないが、森林など他の土地利用が農地に転換される活動はアジアやアフリカで活発に見られる。前節でも議論したように、たとえばインドネシアのスマトラ島やカリマンタン島などでは、森林からパーム油プランテーションへの大規模な転換が進んでいるが、森林伐採の際に泥炭土から大量の炭素が放出されている(Carlson et al., 2013)。

こうした森林伐採を通じた土地利用変化の二酸化炭素排出は事後的に計上されているため、会計そのものの課題ではない。追加的に増えた農地のシャドウ価格が、減った森林のシャドウ価格を上回っているかどうかが問題である。とは言え、自然資本全体、もしくは再生可能資源全体だけを見ていると、自然資本と人工資本だけでなく、自然資本“内”での代替も進んでいることは見落とされがちである。

4.2 都市の生態系サービスと自然資本

現状、Inclusive Wealth Report 2018における自然資本としての森林や農地の計上は、国連食糧農業機関(FAO)の土地利用カテゴリーに基づいたものである。このカテゴリーでは、土地利用によって分類が行われるため、たとえば都市部における森林などは計上されていないと考えられる。これに対して近年注目されているのが、都市内部の生態系サービスである。人口の多い都市部に隣接する自然資本がもたらす生態系サービスは、アクセスがしやすいことから、ア

表 1. 米国の都市部の森林がもたらす生態系サービスの年間価値(単位: 億ドル).

生態系サービス	価値	備考
大気汚染物質の除去	54	オゾン, 二酸化窒素, 二酸化硫黄, PM2.5
炭素の固定	48	炭素の社会的費用は\$129.8/tC
建物エネルギー消費の削減	54	日陰や冷気の提供
汚染物質の排出の回避	27	エネルギー消費の削減

出所) Nowak and Greenfield (2018) より作成.

メンティやレクリエーションの価値を中心に, 社会にとっての限界的な便益が大きいと考えられる. こうした都市の自然資本も富の一部として認識し, 保全や整備の政策に役立てることが重要だろう. これは, 人口減少と高齢化が進む先進国の都市部での政策のみならず, 都市への人口流出とスラム化の問題を抱えるアジアやアフリカの都市計画にとっても大きな意味を持つ.

たとえば全米 50 州の都市部の森林を調べた Nowak and Greenfield (2018) は, 表 1 のような 4 つの生態系サービスを計上し, その合計が年間 183 億ドルに相当するとした. このサービスが仮に年率 5% で無限の将来までもたらされるとすると, 自然資本としての価値は単純計算で 3,600 億ドルのオーダーとなる. UNU-IHDP and UNEP (2014) では, 米国全体の森林の価値を 3.6 兆ドルとしており, その 1 割に相当することになる. ただし, 国内のボトムアップデータを用いて構築した指標は, 計上の対象が重複している可能性もあり, 国際比較を行う際は定義などを慎重に確認する必要がある.

5. 結語

農地や森林といった再生可能な自然資本がもたらす生態系サービスは, 人間の福祉に大きな影響を与える. 自然資本の減少は世界的な傾向となっているが, 本稿で見たように, 地域によってその内情は異なる. 自然資本が人工資本等に代替される過程だけでなく, 森林から農地への変化をはじめ, 自然資本の内部でどのように代替が行われているかについてのさらなる分析が必要である. 最終節において, 生態系のデイスサービスと都市の生態系サービスに触れたが, これらは生態系サービスと自然資本ストックのスコープとして, 今後拡張されるべき方向性の例にすぎない. シャドウ価格の精緻化, ボトムアップとトップダウンの整合性など, データと会計に関する課題もクリアしていくことが期待される.

注.

- 1) 2015 年時点のデータを World Bank Open Data (<https://data.worldbank.org/>, 2018 年 6 月 10 日アクセス) より得た.
- 2) そのうち, 農地をはじめ森林, 草原, 灌木地への土地利用変化による土地劣化によって生態系サービスは年 6.3 兆ドル減じている (Sutton et al., 2016).
- 3) そのような複雑な問題に対処するために多くの環境評価手法が存在している (仮想評価法, 選択実験法, ヘドニック法, トラベルコスト法など).

- 4) 調整純貯蓄と富会計に関しては, Hamilton and Clemens (1999), Hamilton et al. (2005), World Bank (2011)を先鋒とした一連の世界銀行の報告書があり, 直近のものに Lange et al. (2018)がある. 一方で新国富指標は UNU-IHDP and UNEP (2012, 2014)で公開され, 2018年に出版された Managi and Kumar (2018)が最新の報告書である.
- 5) たとえば, 最新の両報告書では基本的な資本項目は人工資本, 自然資本, 人的資本の3つは同一であるが, 各資本グループ内に含まれる項目は異なる (Lange et al., 2018; Managi and Kumar, 2018). IWIが人的資本に(長命の価値を捕捉する)健康資本を, また自然資本に水産資源を含めている一方, 世界銀行の富会計は, 人工資本に都市部の土地と海外純資産を含めている.
- 6) Lange et al. (2018)と Managi and Kumar (2018)で計測対象とした国数は, それぞれ, 141か国, 140か国と1か国しか差がなかった点からも計測期間を研究対象の選択要素とした.
- 7) 詳しくは山口 他 (2016)参照.
- 8) 両グループを通じて最も農地資本総額が高く, この結果に影響を与えたと推測される中国(高中所得グループ)が, 1990年度水準から農地資本を伸ばしている点が考えられる. ただし, 1998年を境に, 1990年の水準は超えているものの減少に転じているため, 中国の農地資本の成長を主要因とは断定できない. また, 農地面積が減少している要因は多岐に渡ることが想定される. これらの要因の検討は重要な研究課題であるものの, 本研究の目的を逸脱するため, ここではデータベース上明らかな要因までを検討するにとどめたい.
- 9) Managi and Kumar (2018)は UNU-IHDP and UNEP (2014)と同様に, 温帯・北方林と, 熱帯雨林の2つの森林形態からそれぞれ得られる供給サービス, 調整サービス, 生息・生育地サービス, 文化的サービスの4つの生態系サービスの価値を van der Ploeg and de Groot (2010)から引用し, 森林形態の比率で重みづけして各国の年間の生態系サービスの価値を算出している. そのため, 主に森林形態の比率の違いにより国ごとのシャドウ価格は異なる.
- 10) 13か国は以下の通り. バングラデシュ, ベリーズ, ブラジル, コロンビア, エクアドル, ガイアナ共和国, インドネシア, ジャマイカ, マレーシア, ニカラグア, パナマ, ペルー, トリニダード・トバコ.
- 11) 富会計でのシャドウ価格は, 対象期間を通じて一定として計算するため, 森林面積および体積, そして農地面積の増減が直接的な要因となっている.
- 12) FAOSTAT (<http://www.fao.org/faostat/en/#home>, 2018年6月8日アクセス)に記載されている2014年時点のグロスの農業生産額の上位国のうち, 農地資本, 森林資本のデータが揃っている国を選択した.
- 13) ただし, 耕地資本のシャドウ価格の計算上, 過去の農業生産額が反映されるわけではないが, 将来も続いて生産額が低下するのであれば, 将来のシャドウ価格も低下する. そのためガンビアとオランダなどの国の農業の今後の評価には注意を払うべきだろう.
- 14) 15か国中熱帯雨林グループに該当するのはインドネシア, ブラジル, ペルーの3か国であった.
- 15) 32か国のうち, 1か国はシンガポールであり, 図8には記載されていない点に注意されたい.
- 16) Inclusive Wealth Report 2018と同じく Tol (2009)で示された二酸化炭素排出による損失額を用いた.

謝 辞

本研究は JSPS 科研費(特別推進研究 JP26000001)の助成を受けた。

参 考 文 献

- Arrow, K. J., Dasgupta, P. and Mäler, K.-G. (2003). Evaluating projects and assessing sustainable development in imperfect economies, *Environmental and Resource Economics*, **26**, 647-685.
- Arrow, K. J., Dasgupta, P., Goulder, L. H., Mumford, K. J. and Oleson, K. (2012). Sustainability and the measurement of wealth, *Environment and Development Economics*, **17**(3), 317-353.
- Barbier, E. (2012). Ecosystem services and wealth accounting, *Inclusive Wealth Report 2012: Measuring Progress Towards Sustainability (UNU-IHDP and UNEP)*, 165-194, Cambridge University Press, Cambridge.
- Boyd, J. and Banzhaf, S. (2007). What are ecosystem services? The need for standardized environmental accounting units, *Ecological Economics*, **63**(2-3), 616-626.
- Carlson, K. M., Curran, L. M., Asner, G. P., Pittman, A. M., Trigg, S. N. and Adeney, J. M. (2013). Carbon emissions from forest conversion by Kalimantan oil palm plantations, *Nature Climate Change*, **3**(3), 283.
- Costanza, R., d'Arge, R., de Groot, R., Farber, S., Grasso, M., Hannon, B., Limburg, K., Naeem, S., O'Neill, R. V., Paruelo, J., Raskin, R., Sutton, P. and van den Belt, M. (1997). The value of the world's ecosystem services and natural capital, *Nature*, **387**, 253-260.
- Costanza, R., de Groot, R., Sutton, P., van der Ploeg, S., Anderson, S. J., Kubiszewski, I., Farber, S. and Turner, R. K. (2014). Changes in the global value of ecosystem services, *Global Environmental Change*, **26**, 152-158.
- Dasgupta, P. (2009). The welfare economic theory of green national accounts, *Environmental and Resource Economics*, **42**, 3-38.
- DeFries, R. S., Rudel, T. K., Uriarte, M. and Hansen, M. (2010). Deforestation driven by urban population growth and agricultural trade in the twenty-first century, *Nature Geoscience*, **3**, 178-181.
- de Groot, R. S., Brander, L., van der Ploeg, S., Costanza, R., Bernard, F., Braat, L., Christie, M., Crossman, N., Ghermandi, A. and Hein, L. (2012). Global estimates of the value of ecosystems and their services in monetary units, *Ecosystem Services*, **1**, 50-61.
- Engelbrecht, H. J. (2016). Comprehensive versus inclusive wealth accounting and the assessment of sustainable development: An empirical comparison, *Ecological Economics*, **129**, 12-20.
- Fenichel E. P. and Abbott J. K. (2014). Natural capital: From metaphor to measurement, *Journal of the Association of Environmental and Resource Economists*, **1**(1), 1-27.
- Fenichel, E. P., Abbott, J. K., Bayham, J., Boone, W., Haacker, E. M. and Pfeiffer, L. (2016). Measuring the value of groundwater and other forms of natural capital, *Proceedings of the National Academy of Sciences of the United States of America*, **113**(9), 2382-2387.
- Food and Agriculture Organization of the United Nations (FAO) (2016). Global Forest Resources Assessment 2015: How are the World's Forests Changing?, <http://www.fao.org/3/i4793e/i4793E.pdf>.
- Gibbs, H. K., Ruesch, A. S., Achard, F., Clayton, M. K., Holmgren, P., Ramankutty, N. and Foley, J. A. (2010). Tropical forests were the primary sources of new agricultural land in the 1980s and 1990s, *Proceedings of National Academy of Sciences of the United States of America*, **107**(38), 16732-16737.

- Hamilton, K. (2016). Measuring sustainability in the UN system of environmental-economic accounting, *Environmental and Resource Economics*, **64**(1), 25-36.
- Hamilton, K. and Clemens, M. (1999). Genuine savings rates in developing countries, *World Bank Economic Review*, **13**(2), 333-356.
- Hamilton, K., Ruta, G., Bolt, K., Markandya, A., Pedroso-Galinato, S., Silva, P., Ordoubadi, M. S., Lange, G.-M. and Tajibaeva, L. (2005). *Where is the Wealth of Nations?: Measuring Capital for the 21st Century*, World Bank, Washington, D.C.
- Hassan, R., Scholes, R. and Ash, N. (2005). *Ecosystems and Human Well-being. Current State and Trends*, Island Press, Washington, D.C.
- 林岳, 佐藤真行 (2016). 生態系勘定の開発における諸外国の動向と日本の課題, *環境経済・政策研究*, **9**, 44-47.
- Henle, K., Alard, A., Clitherow, J., Cobb, P., Firbank, L., Kull, T., McCracken, D. I., Moritz, R., Mühle, H., Niemelä, J., Nowicki, P., Rebane, M., Wascher, D., Watt, A. and Young, J. (2008). Identifying and managing the conflicts between agriculture and biodiversity conservation in Europe: A review, *Agriculture, Ecosystems & Environment*, **124**, 60-71.
- Kroeger, T. and Casey, F. (2007). An assessment of market-based approaches to providing ecosystem services on agricultural lands, *Ecological Economics*, **64**, 321-332.
- Lange, G.-M., Wodon, Q. and Carey, K. (2018). *The Changing Wealth of Nations 2018: Building a Sustainable Future*, World Bank, Washington, D.C.
- Managi, S. and Kumar, P. (ed.) (2018). *Inclusive Wealth Report 2018: Measuring Progress Towards Sustainability*, Routledge, UK.
- Millennium Ecosystem Assessment (MEA) (2005). *Ecosystems and Human Well-being: Wetlands and Water. Synthesis*, World Resources Institute, Washington, D.C.
- Nowak, D. J. and Greenfield, E. J. (2018). US urban forest statistics, values and projections, *Journal of Forestry*, **116**(2), 164-177.
- Obst, C., Hein, L. and Edens, B. (2016). National accounting and the valuation of ecosystem assets and their services, *Environmental and Resource Economics*, **64**(1), 1-23.
- Onofri, L., Lange, G.-M., Portela, R. and Nunes, P. A. (2017). Valuing ecosystem services for improved national accounting: A pilot study from Madagascar, *Ecosystem Services*, **23**, 116-126.
- Perry, C. (2012). Accounting for water: Stocks, flows and values, *Inclusive Wealth Report 2012 (UNU-IHDP and UNEP)*, 215-230, Cambridge University Press, Cambridge.
- Peters, C. P., Gentry, A. H. and Mendelsohn, R. (1989). Valuation of an Amazonian rainforest, *Nature*, **339**, 655-656.
- Renwick, A., Jansson, T., Verburg, P. H., Revoredo-Giha, C., Britz, W., Gocht, A. and McCracken, D. (2013). Policy reform and agricultural land abandonment in the EU, *Land Use Policy*, **30**(1), 446-457,
- Sutton, P. C., Anderson, S. J., Costanza, R. and Kubiszewski, I. (2016). The ecological economics of land degradation: Impacts on ecosystem service values, *Ecological Economics*, **129**, 182-192.
- Tol, R. J. (2009). The economic effects of climate change, *Journal of Economic Perspectives*, **23**, 29-51.
- UNU-IHDP and UNEP (2012). *Inclusive Wealth Report 2012: Measuring Progress Towards Sustainability*, Cambridge University Press, Cambridge. (植田和弘, 山口臨太郎 訳, 武内和彦 監修 (2014). 『包括的「富」報告書』, 明石書店, 東京.)
- UNU-IHDP and UNEP (2014). *Inclusive Wealth Report 2014: Measuring Progress Towards Sustainability*, Cambridge University Press, Cambridge.
- van der Ploeg, S. and de Groot, R. (2010). *The TEEB Valuation Database — A Searchable Database of 1310 Estimates of Monetary Values of Ecosystem Services*, Foundation for Sustainable Development, Wageningen, The Netherlands.

World Bank (2011). *The Changing Wealth of Nations: Measuring Sustainable Development in the New Millennium*, World Bank, Washington, D.C.

山口臨太郎, 大久保和宜, 佐藤真行, 籠橋一輝, 馬奈木俊介 (2016). 新しい富の指標計測—持続可能性計測研究の過去と未来—, 環境経済・政策研究, **9**(1), 14-27.

Global Analysis of Ecosystem Services from Agroforestry and Natural Capital

Shinya Ikeda¹, Rintaro Yamaguchi² and Shunsuke Managi^{3,4}

¹College of Agriculture, Ibaraki University

²Center for Social and Environmental Systems Research, National Institute for Environmental Studies

³Graduate School of Engineering, Kyushu University

⁴Urban Institute, Kyushu University

We enjoy tremendous value of ecosystem services from agricultural land and forest. However, it remains challenging to evaluate ecosystem services relevant to human well-being. This study illustrates global trends of the changes in values of ecosystem services from natural capital, based on the dataset in a recent United Nations report (Inclusive Wealth Report 2018), which includes 140 countries' data of ecosystem services from renewable natural capital. Among other results, we show that the value of agricultural land increased in South American countries with high levels of agricultural production, and decreased for some developed countries, e.g., in Europe, although in general natural capital has been decreasing. We also argue that in future work, other aspects of ecosystem services from natural capital should be evaluated, e.g., ecosystem disservices from agricultural land and urban ecosystem services.

外来種駆除の生物多様性保全効果：保全優先地域と脅威動態の関係

楠本 聞太郎^{1,2}・南木 大祐¹・久保田 康裕¹

(受付 2018 年 6 月 28 日；改訂 12 月 17 日；採択 12 月 18 日)

要 旨

限りある保全リソースで、生物多様性を効率的・効果的に保全するには、保全事業の有効性を十分に理解しておく必要がある。本研究では、沖縄島北部のマングース駆除事業を例題として、フイリマングース (*Herpestes auropunctatus*) による脅威と、陸生脊椎動物 (254 種) の保全優先地域の空間的一致性を検証した。脊椎動物種 (哺乳類 29 種, 鳥類 155 種, 爬虫類 47 種, 両生類 23 種) の潜在分布地図を用いて、除去規則に基づく空間的優先順位付け分析を行い、分類群ごとの保全優先地域を明らかにした。捕獲量によるマングース捕獲情報 (2000 年から 2009 年) に基づき、マングースの分布確率の時空間動態を階層ベイズモデルによって推定した。当該地域の駆除事業は、各分類群の保全優先地域におけるマングースの分布確率を低下させた。一方、保全優先地域とマングースの分布確率の空間的一致性には、分類群間で違いが見られた。また、保全優先地域の一部では分布確率が上昇しており、駆除の継続が必要であることが明らかになった。今後は、生物多様性の空間情報を積極的に活用し、他の保全事業との相乗効果や経済的コンフリクトを考慮した上で、保全アクションの戦略的適用を検討していく必要がある。

キーワード：階層ベイズモデル、空間的保全優先順位付け、除去規則、マングース (*Herpestes auropunctatus*)。

1. はじめに

生物多様性損失の防止は、地域から世界規模で取り組むべき重大な課題である (MEA, 2005; CBD, 2011)。投資可能な保全リソース (予算や空間) には限りがあるため、生物多様性保全では社会経済的な負担 (コスト) に対する効率を考慮する必要がある (Weitzman, 1998)。制約条件下での保全効果の最大化問題に対処すべく、この 20 年でシステム化保全計画法が概念的・方法論的に急速に発展してきた (Margules and Sarkar, 2007; Kukkala and Moilanen, 2013; 久保田 他, 2017)。その中でも、“空間的保全優先順位付け” は、保全計画における保護区設置の空間デザインを構築する上での必須ツールになりつつある (Moilanen et al., 2009)。空間的保全優先順位付けは、生物多様性の空間情報を用いて、場所間の相補性を考慮しつつ保全優先地域を特定する (Moilanen et al., 2009)。一方で、たとえ保全優先地域内であっても、その中の生物多様性の存続可能性を脅かす様々な要因が存在するため (Cattarino et al., 2015)、そのような脅威を特定して除去・軽減をする保全努力も重要になる (Auerbach et al., 2015)。

¹ 琉球大学 理学部：〒903-0213 沖縄県中頭郡西原町字千原 1

² 沖縄県環境科学センター 環境科学部：〒901-2111 沖縄県浦添市字経塚 720 番地

侵略的外来種は生物多様性の存続にとって重大な脅威の一つである (Bellard et al., 2016). このため、外来種の駆除は、在来の種または潜在的な生物多様性の存続可能性を改善する有効手段である (Jones et al., 2016). しかし、外来種の駆除事業には膨大な費用と労力が伴うことが一般的なので (Moore et al., 2011), 駆除による生物多様性保全効果と他の保全活動との予算的なコンフリクトを考慮して、効果的かつ効率的な努力配分を検討する必要がある (Carwardine et al., 2012). 外来種駆除の生物多様性保全効果は、駆除によってどの程度生物多様性の存続性が改善されたかによって評価される (Carwardine et al., 2008). したがって、在来の生物多様性に対する脅威である外来種個体群の時空間動態と保全優先地域の空間的な一致性を分析することは、外来種駆除事業の保全効果を評価する有効な方法であり、緊急性に基づく保全意思決定に役立つ情報を提供できる (Dawson et al., 2015).

琉球列島は、日本列島とともに世界的な生物多様性ホットスポットの一つとして知られる (Mittermeier et al., 2011). 琉球列島の中でも、沖縄島北部(通称ヤンバル)の森林は多くの固有動植物が生息しており、保全上の重要地域となっている (Itô, 1997). 沖縄島では、1910 年に南部に移入されたファイリマングース (*Herpestes auro-punctatus*; 以下、マングース) の分布範囲が拡大し、近年では北部地域の生物多様性保全の脅威となっている (Barun et al., 2011). この脅威を緩和あるいは除去するために、環境省と沖縄県では、2000 年から捕獲罟と捕殺罟によるマングース駆除事業を継続して行っている. 駆除努力量や捕獲(殺)量の時空間データが蓄積され、マングース個体群の動態解析が実施されているが、それらと保全優先地域の空間的一致性については検証されていない.

本研究では、沖縄島北部(ヤンバル)において、在来の陸生脊椎動物の保全優先地域と、マングースの駆除努力および駆除成果の空間的一致性を分析する. そのために、1) 陸生脊椎動物 (254 種) の空間分布データを用いて、分類群ごとの保全優先地域を、空間的保全優先順位付けアルゴリズムを使って解析する; 2) 捕獲情報から推定したマングース分布確率の時空間分布と、各分類群の保全優先地域の空間的一致性を解析する. それによって、過去のマングース駆除事業の実効性を生物多様性保全の観点から評価し、今後の課題を特定する.

2. データと方法

2.1 生物多様性空間分布データ

既存の生物分布情報(環境省 <http://www.biodic.go.jp/>) から、沖縄県に生息する陸生脊椎動物の出現情報を抽出した: 哺乳類 (29 種), 鳥類 (155 種), 爬虫類 (47 種), 両生類 (23 種). 種ごとの出現記録と環境要因(気候, 地形, 土壌)を用いて, Maxent ver. 3.3.3 (Phillips et al., 2006) による種分布モデルを行い, 種ごとの日本国内での潜在分布地図を 1 km × 1 km 解像度で整備した (Kubota et al., 2015; Lehtomäki et al., 2019).

2.2 空間的保全優先順位付け

空間的保全優先順位付けには、ZONATION ソフトウェア (バージョン 4.0; Moilanen et al., 2014) を用いた. ZONATION のアルゴリズムでは、対象地域を、任意の大きさの保全ユニット (グリッドセル) に区分する. そして、生物種の分布情報 (在不在情報や、分布確率、個体数など) を使って、各セルを保護しない場合の生物多様性の損失を計算する. 景観全体を保護するという状態から開始し、生物多様性損失が最も小さいサイトから逐次的に除去していくことで、対象地域内での保全優先度の順位付けが生成される. 生物分布情報に基づいて生物多様性損失を計算する関数は “除去規則” と呼ばれる. ZONATION にはいくつかの除去規則が実装されているが、代表的なものの一つは Core-Area Zonation (CAZ) である. CAZ では、セル i にお

ける生物多様性損失 δ_i は以下のように定義される (Moilanen, 2007) :

$$(2.1) \quad \delta_i = \max_j \frac{q_{ij} w_j}{c_i}$$

ここで, q_{ij} は, 種 j の残存分布範囲に対するサイト i の相対分布割合である. すなわち, サイト i に出現する種の中で, 最も残存分布範囲の狭い種の値がサイトの代表値として選択される. このため, CAZ は希少種や局在種を重視したアルゴリズムと言われる. なお, サイトを保護するのに必要な社会経済コスト (c_i) や, 種の保全上の重み (w_j) も明示的に組み込むことができる.

本研究では, 沖縄県全域の $1\text{km} \times 1\text{km}$ グリッドセルの種分布情報 (Maxent の推定値) を用いて, 分類群ごとの CAZ 分析を行った. 優先順位付けにおける種の重み (w_j) は, Pouzols et al. (2014) に従い, 環境省の 2012 年度版レッドリストに基づいて定義した (LC = 1; NT = 2; VU = 4; EN = 6; CR = 8; DD = 2). 保全コストは全サイトで共通と仮定した ($C_i = 1, \forall i$).

人間の生活エリア (居住区, 畑地等) が保全優先地域として選択されるのを避けるために, 各サイトの生息地条件を CAZ 分析に組み込んだ. 生息地条件の指標として, 人口, 交通網, 土地利用の情報から, 人為活動指数 (HII: Human Influence Index; Sanderson et al., 2003) を計算し, $0 \sim 1$ の範囲に正規化した ($0 =$ 生息地が完全に改変されている; $1 =$ 生息地が完全に残っている). この生息地条件情報は, 以下の手順で空間的優先順位付け分析に組み込まれる. 種ごとのオリジナルの分布情報 O_{ij} を $0 \sim 1$ に正規化する ($N_{ij} = O_{ij} / \sum_i O_{ij}$). 規準化した種の分布 N_{ij} を, サイト i の生息条件 H_i で重みづけする ($N_{ij}(\text{conditional}) = N_{ij} * H_i$). この処理によって, より完全性の高い生息場所の保全優先度が高く評価されることになる.

これらの設定の元, 沖縄県レベルで分類群ごとに CAZ 分析を行い, 県内の保全優先順位を $1\text{km} \times 1\text{km}$ グリッドセルレベルで生成した. 尚, 本研究では, 県内の保全優先順位が上位 17% (愛知目標に対応; CBD, 2011) のグリッドセルを保全優先地域と定義した.

2.3 マングースの分布確率の時空間情報

マングースの分布確率の時空間情報は, 久保田・南木 (2011) の推定結果を用いた. 解析の元データは, 沖縄県及び環境省が 2000 年 10 月から 2009 年 3 月に行った, かご罠 (16506 地点) によるマングース捕獲情報である. 一つのかご罠で捕獲できるマングースは一度につき一頭で, 捕獲の有無は毎日確認される. 久保田・南木 (2011) は, $200\text{m} \times 200\text{m}$ メッシュレベルでのマングース分布確率動態を, パッチ占有モデル (Royle and Dorazio, 2008) の枠組みに基づき, 潜在的な捕獲率を考慮した階層ベイズモデルで以下のように定式化した. 各時点 t の罠 i における捕獲頭数 (Y_{it}) は, 罠の稼働日数 J_{it} を試行回数とし, マングースの在/不在 (z_{it}) と捕獲率 p の積を成功確率とする二項分布に従うと仮定する.

$$(2.2) \quad Y_{it} \sim \text{Binomial}(J_{it}, z_{it}p)$$

マングースの在/不在 z_{it} は, 分布確率 ψ_{it} を持つベルヌーイ分布に従うと仮定する.

$$(2.3) \quad z_{it} \sim \text{Bernoulli}(\psi_{it})$$

分布確率 ψ_{it} のロジットは, 時点 t における移入率 (a_t) と, 時点 $t-1$ から t への生存率 (b_t), それぞれに対する場所 i の環境要因 $n(x_i^{(n)})$ の効果 (b_{nt}^1, b_{nt}^2), 空間自己相関項 (ρ_i) の線形結合で表される.

$$(2.4) \quad \text{logit}(\psi_{it}) = a_t + b_t z_{i,t-1} + \beta_t^{1(n)} x_i^{(n)} + \beta_t^{2(n)} x_i^{(n)} z_{i,t-1} + \rho_i$$

移入率と生存率に影響する環境要因として, 各 200m メッシュ内の林道密度, 森林面積, 標高を考慮した. また, 移入率と生存率に関わるパラメータは, 前年度の値に依存してランダム

ウォークすると仮定する.

$$(2.5) \quad a_t \sim \text{Normal}(a_{t-1}, \sigma_a^2)$$

$$(2.6) \quad b_t \sim \text{Normal}(b_{t-1}, \sigma_b^2)$$

$$(2.7) \quad \beta_t^{1(n)} \sim \text{Normal}(\beta_{t-1}^{1(n)}, \sigma_{\beta 1n}^2)$$

$$(2.8) \quad \beta_t^{2(n)} \sim \text{Normal}(\beta_{t-1}^{2(n)}, \sigma_{\beta 2n}^2)$$

なお, 事業開始時点 ($t = 1$) については, これらのパラメータは, 平均 0, 標準偏差 1000 の正規分布に従うと仮定する (無情報事前分布). また, σ は分散パラメータであり, それぞれ平均 0.01, 標準偏差 100 のガンマ分布に従うと仮定する. 空間自己相関項 ρ_i は, 地点 i の近傍の分布確率が与えられた際の条件付き分布確率として与える (Latimer et al., 2006). パラメータの事後分布は, WinBUGS1.43 (Lunn et al., 2000) によるマルコフ連鎖モンテカルロ (MCMC) 法で推定した. 3 つの独立した MCMC 計算をそれぞれ 30000 ステップ行った. 20000 回までのステップはバーニンとして除き, 残りのステップを 10 ステップ間隔でサンプリングし, 合計 3000 個の事後分布推定値を得た. 前述の保全優先地域の分析結果と規格を統一するために, 推定分布確率は, 年毎に, $1 \text{ km} \times 1 \text{ km}$ メッシュレベルに集約 (平均化) した.

2.4 保全優先地域とマングース脅威の関係

CAZ 分析で特定した保全優先地域と, 各年の罝の積算稼働時間, 及びマングースの推定分布確率の関係をグラフ化し, 各分類群の保全優先地域と脅威及び駆除努力の空間的な一致性を調べた. さらに, 期間全体での生物多様性保全効果を明らかにするために, 期間初期 (2001 年) と最終年 (2009 年) のマングース推定分布確率の差をセルごとに計算した. そして, 推定分布確率の増減パターンの地理的分布, 及び分類群を統合した保全優先度との関係性を検証した. なお, 分類群統合の保全優先度は, 分類群ごとの CAZ 分析で求めた相対優先順位 (0~1 の連続値) の合計値として定義した.

3. 結果

保全優先地域の空間的な分布パターンは分類群によって違いが見られた (図 1): 両生類と爬虫類の保全優先地域は, 海岸から内陸にかけて散在していた; 鳥類では, ヤンバル全体が保全優先地域であった; 哺乳類では, ヤンバル南部と北西部に保全優先地域の大きなパッチがみられる. マングース捕獲努力量 (罝の積算稼働時間) には時空間的に偏りが見られた (図 2): 事業初期は南西部に努力量が集中していたが, 後期には南部から中央部を広くカバーするようになった. 捕獲情報に基づく推定分布確率は, 南側で高く, 北側で低かった (図 2). 分布確率は, 事業開始後緩やかに増加していたが, 2007 年以降には分布確率の顕著な低下が見られた.

罝の積算稼働時間は, 両生類, 爬虫類, 哺乳類では事業初期に保全優先地域外で長かった (図 3). 期間全体を通して, 両生類の保全優先地域内外でマングースの分布確率にほとんど差はなかった. 爬虫類では, マングースの分布確率は, 保全優先地域外でやや高かった. 哺乳類では, 保全優先地域内外でマングースの分布確率の差が大きく, 優先地域内で分布確率が低かった. 2009 年と 2000 年の分布確率を比較すると, ヤンバル中南部のほとんどの場所において, マングースの分布確率は低下した (図 4). 一方で, 開始時には分布確率の低かった場所では, わずかではあるが分布確率の上昇が見られた. そのような場所は, ほとんどが北部地域で, 保全優



図 1. 分類群ごとの保全優先地域の分布。

先度が高い地域も含まれていた。

4. 考察

既存の統計数理手法を組み合わせ、実際上の問題を議論することは統計数理の社会還元という観点から見て重要である。本研究では、2つの既存手法(状態空間モデルによる外来種分布の時空間推定と、最適化アルゴリズムによる生物多様性保全重要地域を特定)によって、沖縄島北部のマングース分布の時空間動態と、陸生脊椎群集の保全優先地域との空間的一致性を検証し、保全実務的に重要かつ新規的な結果を提示した。具体的には、全ての分類群の保全優先地域内において、マングースの分布確率は平均的には低下しており、駆除事業が生物多様性の存続可能性の改善に貢献していることが示された。一方で、マングースの分布確率と保全優先地域との空間的一致性は分類群によって異なっており、これまでの駆除事業がもたらした保全効果には分類群間で格差があることも明らかになった。

本研究で取り上げた4つの脊椎動物群は、マングースとの相互作用(捕食や競争)によって、潜在的に存続可能性が脅かされる分類群である(Hays and Conant, 2007)。ヤンバル地域における先行研究で、既に各分類群への被害(食害)が報告されている(小倉 他, 2002)。一方で、種分布データの不足のため、分類群ごとの多様性パターンとマングースの分布の関係は、これまでに検証されてこなかった。本研究では、対象分類群の全種を網羅した種分布情報を用いて保全優先地域の空間分布を明らかにし、マングース駆除努力の各分類群に対する保全効果を明らかにした。2000年から2009年に実施された駆除事業では、開始時点で脅威に曝されていた保全優先地域(特に両生類と爬虫類)で効果的にマングースの分布確率を低下させた。鳥類にとっては、ヤンバル全域が県レベルでの保全優先地域であるため、マングースの駆除努力は、県レベルでの生物多様性保全にも効果的であったと言える。哺乳類の保全優先地域では、期間中のマングース分布確率は低かったが、期間終了時に僅かながら分布確率が上昇する場所もあった。この結果は、保全優先地域における将来的な脅威を示唆している。マングースの脅威を未然に防ぐためにも、生物多様性保全優先度マップに基づく駆除努力の最適空間配分が検討されるべきである。

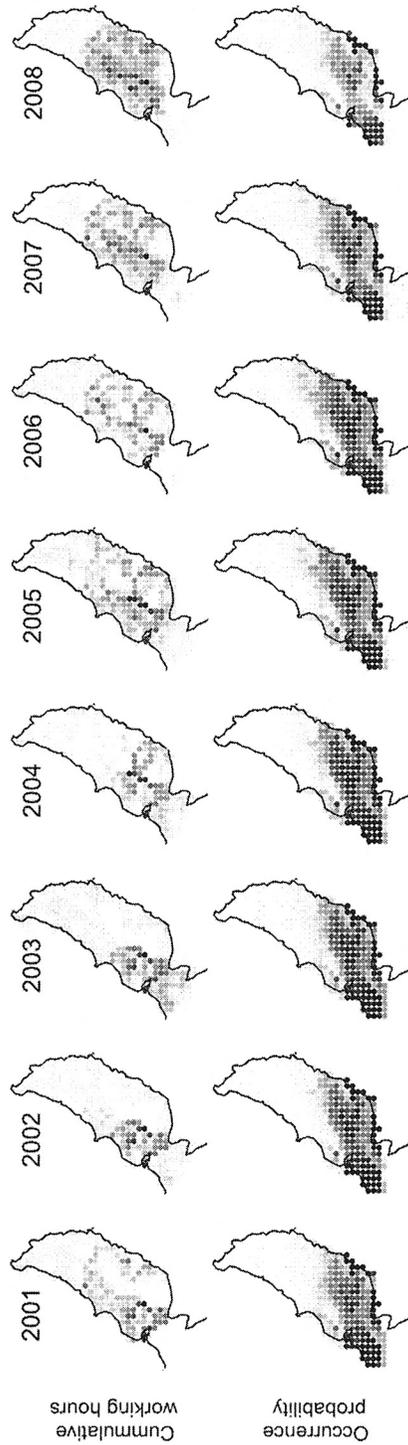


図 2. マングース島の積算稼働時間と推定分布確率の時空間パターン.

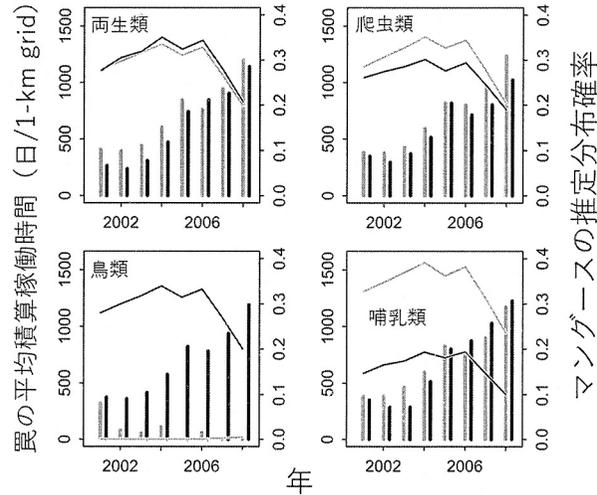


図 3. 保全優先地域内外における農稼働時間(棒)とマングースの推定分布確率(折れ線).

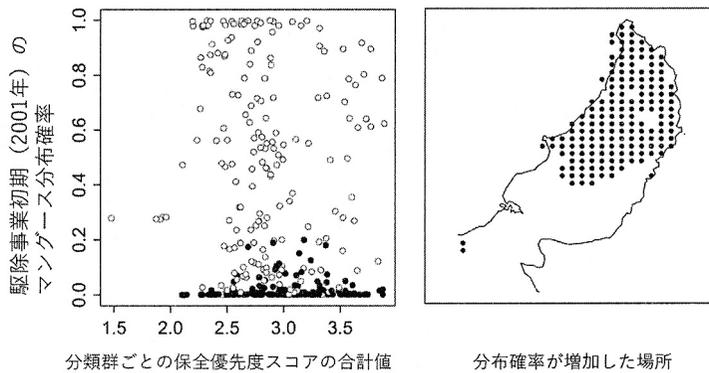


図 4. 保全優先度とマングース分布確率の関係。(左)駆除開始時点(2001年)に比べて分布確率が減少した場所を白, 増加した場所を黒で示した; (右)駆除開始時(2001年)と比べて, 分布確率が上昇した地域の分布.

現在の駆除事業の方針としては、ヤンバル地域を 8つの管理区域に分け、各場所のマングース個体群密度に照らして保全アクション(捕獲や捕殺による低密度化、残存個体の探索、モニタリング)が決定される。マングース個体群の空間分布情報は、事業の実現可能性の評価や経済効率の向上に必須の情報をもたらす(Fukasawa et al., 2013)。一方で、駆除の究極目的は在来の生物多様性の存続可能性の改善なので、生物多様性の空間情報を積極的に取り入れることで、より効率的かつ有効な保全アクションの意思決定が可能になる(Dawson et al., 2015)。本研究で提示した生物多様性空間情報に基づく保全優先順位づけは、各管理ユニットにおけるアクションの緊急性評価に役立つだろう。一方、本研究の解析は、現状の生物分布データの空間解像度の制約により、1 km × 1 km レベルの生物多様性情報に基づいている。管理区域のサイズ(16.9~52.9 km²)を考慮すると、このように粗い空間解像度は管理区域内での保全アクション分析には不適かもしれない(Di Marco et al., 2017)。管理区域内での詳細な保全意思決定を支

援するためには、生物分布情報を拡充し、生物多様性地図の空間解像度を向上させる努力が必要である。

戦略的な保全事業の計画と実施には、明確かつ実現可能な数値目標が不可欠である (Shields et al., 2002)。沖縄島北部のマングース駆除事業では、「固有の生物相を有し、多くの希少種の生息地であるヤンバル地域の生態系を回復し、安定した状態で保全すること」を目的とし、2017年度から2026年度までの10年間で、北部地域からマングースを完全排除することを目標としている。一般的には、対象面積が大きくなるほど、侵略種の個体群間の移動を完全に断ち切るのが困難になるため、完全排除の可能性は低くなる (Barun et al., 2011)。沖縄県では、2005年から2006年にかけて、ヤンバル南部に島を横断するようにマングース防除柵を設置し、北部地域へのマングースの移入を制限している。実際に、マングースの分布確率は2007年付近から低下しており、移入制限と捕獲努力が相乗的にマングースの脅威を緩和したことを示している。一方で、北部地域での分布確率の微小な増加は、偶発的に起こるマングースの移入分散を完全に食い止めることの困難さを示唆している。そのような場所では、保全優先地域へのマングースの侵入を未然に防ぐための予防的措置(例：モニタリングの強化、捕獲罠の重点配備など)が必要であろう。

国や県として保全に投資できるリソース(予算)は限られているので、他の潜在的な保全アクション(例：生物分布データの拡充；保護区設置、希少種の保全など)とのバランスも重要である (Cattarino et al., 2015)。既存研究において、保全アクション間の保全効果の相乗性と経済的なコンフリクト(例えば、外来種駆除への投資が、希少種の保全効果にどの程度貢献し、予算配分的に競合するか)を考慮することにより、全体の保全効率が大きく改善されることが知られている (Carwardine et al., 2012; Adams et al., 2014)。実際、沖縄県では、保護区による経済活動の規制強化 (Kusumoto et al., 2017) や、生態リスクに考慮した森林利用の実施 (Maeshiro et al., 2013; Kusumoto et al., 2016)、希少種の保全増殖事業 (環境省, 2015a; 2015b)、生物分布情報の拡充 (<http://www.biodiversity.okinawa/>) など、様々な保全タスクが掲げられている。これらの保全事業は国や県が協調分担しつつも、個別に行われているのが現状である。今後は、生物多様性保全を大局的な目標として、様々な保全アクションのコストと保全による利益(ベネフィット)を定量し、いつ、どこに、どのような保全アクションを適用するかについて、個別的な保全事業の優先順位付けを分析する必要がある。

5. 結論

空間的保全優先地域の順位付け分析と、外来種による脅威の空間明示的な個体群動態モデリングを組み合わせることで、過去のマングース駆除事業の保全効果を評価した。このような保全事業の有効性評価は、外来種の脅威下での潜在的な生物多様性保全において必須である。全体的には、2001年から2009年までのマングース駆除事業は、保全優先地域におけるマングースの分布確率を低下させた。一部の保全優先地域には分布確率が上昇している場所も見られ、駆除事業の継続が必要であることが明らかになった。今後は、生物多様性の空間情報を積極的に活用し、他の保全事業との相乗効果や経済的コンフリクトを考慮した上で、保全アクションの効率のかつ効果的な適用方法を検討していく必要がある。

謝 辞

本研究は、日本学術振興会の科研費「汎用型離散最適化システムの構築による拡散移動を伴う森林生態系サービスの経済評価」(17H00806)の援助を受け、環境省の環境総合推進費のプロジェクト「生態学的ビッグデータを基盤とした生物多様性パターンの予測と自然公園の実効力

評価」(4-1501)及び「環境変動に対する生物多様性と生態系サービスの応答を考慮した国土の適応的保全計画」(4-1802)の一部として行った。本論文の執筆にあたり、環境省やんばる野生生物保護センターより、マングース駆除事業に関する有益な意見をいただきました。心より感謝申し上げます。

参 考 文 献

- Adams, V. M., Álvarez-Romero, J. G., Carwardine, J., Cattarino, L., Hermoso, V. et al. (2014). Planning across freshwater and terrestrial realms: Cobenefits and tradeoffs between conservation actions, *Conservation Letters*, **7**, 425-440.
- Auerbach, N. A., Wilson, K. A., Tulloch, A. I. T., Rhodes, J. R., Hanson, J. O. and Possingham, H. P. (2015). Effects of threat management interactions on conservation priorities, *Conservation Biology*, **29**, 1626-1635.
- Barun, A., Hanson, C., Campbell, K., Simberloff, D. and Towns, D. (2011). A review of small Indian mongoose management and eradications on islands, *Island Invasives: Eradication and Management* (eds. C. Veitch and M. Clout), 17-25, IUCN, Gland, Switzerland.
- Bellard, C., Genovesi, P. and Jeschke, J. M. (2016). Global patterns in threats to vertebrates by biological invasions, *Proceedings of the Royal Society B: Biological Sciences*, **283**, 1823.
- Carwardine, J., Wilson, K. A., Ceballos, G., Ehrlich, P. R., Naidoo, R. et al. (2008). Cost-effective priorities for global mammal conservation, *Proceedings of the National Academy of Sciences*, **105**, 11446-11450.
- Carwardine, J., O'Connor, T., Legge, S., Mackey, B., Possingham, H. P. and Martin, T. G. (2012). Prioritizing threat management for biodiversity conservation, *Conservation Letters*, **5**, 196-204.
- Cattarino, L., Hermoso, V., Carwardine, J., Kennard, M. J. and Linke, S. (2015). Multi-action planning for threat management: A novel approach for the spatial prioritization of conservation actions, *PloS one*, **10**, e0128027.
- Convention on Biological Diversity (2011). *Conference of the Parties Decision X/2: Strategic Plan for Biodiversity 2011-2020*, <http://www.cbd.int/decision/cop/?id=12268>.
- Dawson, J., Opper, S., Cuthbert, R. J., Holmes, N., Bird, J. P. et al. (2015). Prioritizing islands for the eradication of invasive vertebrates in the United Kingdom overseas territories, *Conservation Biology*, **29**, 143-153.
- Di Marco, M., Watson, J. E., Possingham, H. P. and Venter, O. (2017). Limitations and trade-offs in the use of species distribution maps for protected area planning, *Journal of Applied Ecology*, **54**(2), 402-411.
- Fukasawa, K., Hashimoto, T., Tatara, M. and Abe, S. (2013). Reconstruction and prediction of invasive mongoose population dynamics from history of introduction and management: A Bayesian state-space modelling approach, *Journal of Applied Ecology*, **50**, 469-478.
- Hays, W. S. and Conant, S. (2007). Biology and impacts of Pacific Island invasive species. 1. A worldwide review of effects of the small Indian mongoose, *Herpestes javanicus* (Carnivora: Herpestidae), *Pacific Science*, **61**(1), 3-16.
- Itô, Y. (1997). Diversity of forest tree species in Yanbaru, the northern part of Okinawa Island, *Plant Ecology*, **133**, 125-133.
- Jones, H., Holmes, N., Butchart, S. et al. (2016). Invasive mammal eradication on islands results in substantial conservation gains, *Proceedings of the National Academy of Sciences*, **113**, 4033-4038.
- 環境省 (2015a). ノグチゲラ保護増殖事業計画, 1-3.

- 環境省 (2015b). ヤンバルクイナ保護増殖事業計画, 1-4.
- 久保田康裕, 南木大祐 (2011). 階層ベイズ型の状態空間モデルを用いたマングース個体群の時空間動態の解析(沖縄県文化環境部自然保護課 編), 平成 22 年度沖縄島北部地域生態系保全事業(マングース対策事業)報告書, 45-52.
- Kubota, Y., Shiono, T. and Kusumoto, B. (2015). Role of climate and geohistorical factors in driving plant richness patterns and endemism on the east Asian continental islands, *Ecography*, **38**, 639-648.
- 久保田康裕, 楠本聞太郎, 藤沼潤一, 塩野貴之 (2017). 生物多様性の保全科学: システム化保全計画の概念と手法の概要, 日本生態学会誌, **67**, 267-286.
- Kukkala, A. S. and Moilanen, A. (2013). Core concepts of spatial prioritisation in systematic conservation planning, *Biological Reviews*, **88**, 443-464.
- Kusumoto, B., Baba, A., Fujii, S., Fukasawa, H., Honda, M. et al. (2016). Dispersal process driving subtropical forest reassembly: Evidence from functional and phylogenetic analysis, *Ecological Research*, **31**, 645-654.
- Kusumoto, B., Shiono, T., Konoshima, M., Yoshimoto, A., Tanaka, T. and Kubota, Y. (2017). How well are biodiversity drivers reflected in protected areas? A representativeness assessment of the geohistorical gradients that shaped endemic flora in Japan, *Ecological Research*, **32**, 299-311.
- Latimer, A. M., Wu, S., Gelfand, A. E. and Silander, J. A. Jr. (2006). Building statistical models to analyze species distribution, *Ecological Applications*, **16**, 33-50.
- Lehtomäki, J., Kusumoto, B., Shiono, T., Tanaka, T., Kubota, Y. and Moilanen, A. (2019). Spatial conservation prioritization for the East Asian islands: A balanced representation of multitaxon biogeography in a protected area network, *Diversity and Distributions*, **25**, 414-429.
- Lunn, D. J., Thomas, A., Best, N. and Spiegelhalter, D. (2000). WinBUGS—A Bayesian modelling framework: Concepts, structure, and extensibility, *Statistics and Computing*, **10**, 325-337.
- Maeshiro, R., Kusumoto, B., Fujii, S., Shiono, T. and Kubota, Y. (2013). Using tree functional diversity to evaluate management impacts in a subtropical forest, *Ecosphere*, **4**, art70.
- Margules, C. and Sarkar, S. (2007). *Systematic Conservation Planning*, Cambridge University Press, Cambridge, UK.
- Millennium Ecosystem Assessment (2005). *Ecosystem and Human Well-being: Biodiversity Synthesis*, World Resources Institute, Washington, DC.
- Mittermeier, R. A., Turner, W. R., Larsen, F. W., Brooks, T. M. and Gascon, C. (2011). Global biodiversity conservation: The critical role of hotspots, *Biodiversity Hotspots* (eds. F. E. Zachos and J. C. Habel), 3-22, Springer, Berlin, Heidelberg.
- Moilanen, A. (2007). Landscape zonation, benefit functions and target-based planning: Unifying reserve selection strategies, *Biological Conservation*, **134**, 571-579.
- Moilanen, A., Wilson, K. A. and Possingham, H. P. (2009). *Spatial Conservation Prioritization: Quantitative Methods & Computational Tools*, Oxford University Press, Oxford.
- Moilanen, A., Pouzols, F., Meller, L., Arponen, A., Lppanen, J. and Kujala, H. (2014). *Zonation-spatial Conservation Planning Methods and Software. Version 4. User Manual*, C-BIC, Helsinki.
- Moore, J. L., Runge, M. C., Webber, B. L. and Wilson, J. (2011). Contain or eradicate? Optimizing the management goal for Australian acacia invasions in the face of uncertainty, *Diversity and Distributions*, **17**, 1047-1059.
- 小倉剛, 佐々木健志, 当山昌直, 高原建二, 仲地学 他 (2002). 沖縄島北部に生息するジャワマングース (*Herpestes javanicus*) の食性と在来種への影響, 哺乳類科学, **42**, 53-62.
- Phillips, S. J., Anderson, R. P. and Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions, *Ecological Modelling*, **190**, 231-259.
- Pouzols, F. M., Toivonen, T., Minin, E. D., Kukkala, A. S., Kullberg, P. et al. (2014). Global protected

- area expansion is compromised by projected land-use and parochialism, *Nature*, **516**, 383-386.
- Royle, J. A. and Dorazio, R. M. (2008). Hierarchical modeling and inference in ecology: The analysis of data from populations, metapopulations and communities, Elsevier, Burlington, USA.
- Sanderson, E. W., Jaiteh, M., Levy, M. A., Redford, K. H., Wannebo, A. V. and Woolmer, G. (2003). The human footprint and the last of the wild, *Bioscience*, **52**, 891-904.
- Shields, D. J., Šolar, S. V. and Martin, W. E. (2002). The role of values and objectives in communicating indicators of sustainability, *Ecological Indicators*, **2**, 149-160.
- Weitzman, M. L. (1998). The Noah's Ark problem, *Econometrica*, **66**, 1279-1298.

Effectiveness of Invasive Species Eradication Efforts on Biodiversity Conservation: Spatial Congruence between Conservation Priority Areas and Threat

Buntarou Kusumoto^{1,2}, Daisuke Nanki¹ and Yasuhiro Kubota¹

¹Faculty of Science, University of the Ryukyus

²Department of Environmental Science, Okinawa Prefecture Environment Science Center

To implement biodiversity conservation under limited resource, we should understand the effectiveness of action plans. In this study, we analyzed spatial congruence between distribution of invasive mongoose (*Herpestes auro-punctatus*) as a major threat against biodiversity and conservation priority areas for native terrestrial vertebrate species (254 species) in the northern part of Okinawa island, Japan. Using potential species distribution maps (29, 155, 47, 23 species for mammals, birds, reptiles and amphibians, respectively) estimated at the 1 km grid-cell level, we created spatial conservation priority rankings and detected top 17% priority areas for each taxonomic group within Okinawa prefecture. In the spatial prioritization, we first assumed the perfectly protected state (i.e. 100% of land is protected), and then removed least important cell sequentially following a removal rule where cells include geographically rare species (narrow distribution range) receives higher conservation values. To evaluate the dynamics of mongoose population in space and time, we used catching data obtained by gage traps in the eradication program implemented by Okinawa prefecture and the Ministry of the Environment during 2000 to 2009. Using the catching data (number of mongoose per trapping days in each trap), spatiotemporal pattern of mongoose occurrence was modelled using a hierarchical Bayesian model that estimates survival and immigration rate of mongoose and their environmental dependencies and spatial autocorrelation. The eradication program in the region successfully reduced the occurrence of mongoose within conservation priority areas, whereas the spatial congruence between the priority areas and mongoose occurrence was different between the taxa. Some of the priority areas, which mostly located on the northern part of the region, showed slight increase in the occurrence of mongoose threat, indicating that continuous eradication effort is needed to prevent the mongoose threat. Our findings demonstrate that we should use explicitly spatial biodiversity information and develop strategic conservation plans that account for synergy and conflict between conservation actions and benefits.

インターネット調査におけるモニター情報の利用 による非回答バイアスの補正

—国内草資源を利用した放牧飼養牛肉に対する消費者評価への適用—

楠戸 建¹・後藤 貴文²・高橋 義文³・矢部 光保³

(受付 2018 年 6 月 30 日；改訂 12 月 24 日；採択 12 月 25 日)

要 旨

エビデンスに基づく政策立案などの側面から欠測データに関する研究が進展している。本研究では、逆重み付き推定量(IPWE)を用い、仮想評価法における非回答バイアスについて検証した。また、その際に用いる共変量の入手可能性の問題に対応するため、インターネット調査会社のモニター登録情報を用いた。高所得者を対象に放牧飼養により生産された牛肉(国産放牧牛)に対する消費者評価を行った結果、回答者と非回答者の間には、個人年収、年齢、同居家族、動画サービスや SNS 利用、趣味に関して差異が存在することが明らかになった。他方、回答者のみのサンプルを用いて算出された放牧牛ステーキ肉に対する平均支払意志額は 100g 当たり約 1161.6 円であったのに対し、IPWE を用いて推計を行った場合には 100g 当たり約 1161.7 円と推計され、両者にほとんど差がなかった。以上から本研究においては、回答者・非回答者間に属性的な差異は存在するものの、その差異は平均的な支払意志額の差にはつながっておらず、評価対象の国産放牧牛に対する支払意志に関して非回答バイアスが存在しているとは言えないことが示された。

キーワード：非回答バイアス、仮想評価法、モニター情報、放牧牛、インターネット調査、エビデンスに基づく意志決定。

1. はじめに

1.1 背景

近年、エビデンスに基づく意志決定や政策決定の重要性が認識され、それに対する対応策として、欠測データに関する研究が進展している (Groves and Schoeffel, 2018)。

標本調査における欠測データの問題は不可分の問題として存在しており、非回答者をどのように取り扱うかについては、解決すべき課題として頻繁に取り上げられているところである。特に非回答者が回答者と異なる特徴をもち、それが目的となる変数(目的変数)に影響を与えていれば、サンプルの母集団に対する代表性が損なわれ、非回答バイアスについて適切な対応が求められる。ここで、標本調査における非回答は、ある標本について調査票全体の情報が

¹九州大学大学院 生物資源環境科学府：〒819-0395 福岡県西区元岡 744

²鹿児島大学 農水産獣医学域農学系：〒890-0065 鹿児島県鹿児島市郡元 1 丁目 21-24

³九州大学大学院 農学研究院：〒819-0395 福岡県西区元岡 744

得られない「単位無回答」と調査票の一部について情報が得られない「項目無回答」に大別される (Bethlehem et al., 2011) が、本研究では前者の単位無回答による非回答バイアスについて検討を行う。

上述の問題は、仮想評価法 (CVM: Contingent Valuation Method) や選択実験 (CE: Choice Experiment) を含む、表明選好法による環境評価研究においても同様の課題として認識されている。ここで、仮想評価法を用いた環境評価の手続きは以下のようなものである。「ある環境(状態)に関する支払意志額 (WTP: Willingness To Pay) あるいは受け取り意志額 (WTA: Willingness To Accept) を、関係者あるいはその標本として一部の人々から直接聞きだし、その額を統計的に処理することによって 1 人当たりの金額を計算する。そして、その金額を関係者全体で集計することによって、その環境の価値とするというものである」(鷺田, 1999)。この 1 人当たりの金額を計算した後に、「その金額を関係者全体で集計する」際に回答者・非回答者間の差異に対していかなる対処を行うべきかという問題は、上述の非回答バイアスと同一の問題であると捉えられる。しかしながら、仮想評価法を始めとした表明選好法を用いたほとんどの研究においては、回答者のデータのみを用いるか、または、過大推定を避けるために非回答者の WTP はゼロであるという仮定をして全体価値の推定がなされるという手続きが一般的に行われている (Mitchell and Carson, 1989) もの、そうしてよい根拠については、ほとんどの場合示されていない。

非回答バイアスに適切に対応するためには、調整に用いる共変数をより多く取得することが対応として考えられる。しかし、対象とする母集団をよくカバーする台帳(例えば住民基本台帳や選挙人名簿など)から無作為に抽出された標本を対象として、面接法や郵送法などによって回答を得る既存調査手法では、非回答者の情報を取得することは實際上非常に難しい。星野 (2010) において指摘されているように、このような既存調査手法における非回答者の情報は、目的変数のみならずその他の情報についても取得できない場合がほとんどであり、また取得が可能であったとしても、デモグラフィック変数のうち、性別・年齢別や居住地域などごく一部しか情報が得られないことが多い。一方で、インターネット調査を利用する場合には、あらかじめ登録されているモニター情報を利用することで、より多くの共変量の候補が取得可能であり、この共変量をうまく利用することによってより精度の高い非回答バイアスの補正を行うことができると期待できる。なお、本研究で言うところのインターネット調査とは、調査主体がなんらかの手段で協力意志のある個人を集めて登録し、登録者集団内で対象者を選定し、Web 上に置かれた調査票への回答を求める「リソースタイプ」のインターネット調査 (大隅, 2017) 一般のことを指すものとする。もちろん、このインターネット調査と従来型の確率的標本抽出に基づく既存調査手法(訪問調査や郵送調査法など)との比較から、インターネット調査は母集団との関係が曖昧であり、バイアスを持つことも指摘 (大隅, 2017) されており、その調査法による差異を調整する研究も進展しつつある (星野・前田, 2006; 星野, 2007)。しかしながら本研究ではあくまで非回答バイアスに焦点を当てるため、この点については、ひとまずおいておくこととする。

1.2 先行研究

非回答バイアスは、欠測データによる問題の 1 つとして取り扱われる。欠測データの問題については、これまで近年、医学・疫学の分野をはじめ盛んに研究が行われてきており、欠測データの適切な取り扱いに向けた国際的なガイドラインも National Research Council (2010), Little et al. (2012) をはじめとして、標準化されつつある。

アンケートにおける非回答は、この欠測データの典型例であるが、ここで、アンケートにおける非回答がなぜ起こるかを考える。土屋 (2010) によれば、一般に近年の調査不能の二大理由は拒否と不在であると捉えられる。インターネット調査においては、前者はアンケートを確認

した上で回答を行わなかった場合が対応し、後者は調査期間内にログインを行わなかったり、アンケートに気づかなかったという場合が対応する。しかし、本調査では、この2つを識別することはできないため、調査票を配布したモニターのうち、期間内にアンケートに回答したモニターを回答者、回答が得られなかったモニターを非回答者として取り扱う。

非回答バイアスへの対応としては、星野・前田(2006)に挙げられているように、欠測のメカニズムに応じて様々な方法が取られている。特に国内における社会調査の非回答バイアスに関する先行研究は、ほとんどの場合、事前または事後アンケートにより本調査に回答しなかった標本の共変量情報を取得する方法(土屋, 2005)が中心である。さらにそれに加えて調査訪問時に調査員によって記録された情報を利用したり(松岡・前田, 2015)、二次的な統計情報により回答者の居住する地域の人口密度等を利用するなどして、共変量を取得し、得られた共変量を用いて非回答バイアスの補正が行われてきた。非回答バイアスの補正にあたっては、できるだけ多くの共変量を取得することに加え、調査への協力態度などといったデモグラフィック変数以外の要因も考慮する必要性についても指摘されている(土屋, 2006, 2010)。しかしながら、当然事前・事後アンケートにも回答しなかった標本の情報は得られないという限界は依然として残っている。

他方、本研究で利用する仮想評価をはじめとした表明選好法による環境評価研究においては、非回答バイアスの存在については度々取り上げられている(Mitchell and Carson, 1989)ものの、実際の調整を行った研究は少なく、事前・事後アンケートとCVMの質問を含む本調査を組み合わせることで、データを収集し、サンプルセレクションモデル(Heckman, 1979)の枠組みから分析を行ったもの(Whitehead et al., 1993; Messonnier et al., 2000)など、ごく少数に限られる。また、これらにおいても、上述の内容と同様に事前・事後アンケートにも回答しなかった標本の情報は得られないという課題は残されている。

1.3 研究における課題

本調査では、環境保全型の畜産方法の1つである放牧飼養により生産された牛肉(国産放牧牛)の消費者評価を仮想評価法によりWTPを尋ねることで明らかにする。またその際、評価額の推定における非回答バイアスの補正に利用する共変量の情報として、インターネット調査会社が保有するモニター登録情報を利用し、そこから作成された共変量を用いて回答者・非回答者間の差異を明らかにする。これにより、前で述べた共変量の取得可能性という課題に対処した上で、非回答者を含む母集団全体における平均WTPの逆重み推定量(IPWE: Inverse Probability Weighting Estimate)を求めることで、回答者・非回答者間の差異が母集団全体における平均WTPの推計値にいかなる影響を与えるか明らかにする。

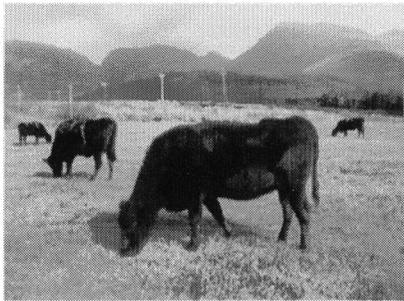
2. データ収集とデータの特徴

2.1 データ収集

本調査は、インターネット調査会社を通して登録モニターに対してアンケート調査を行った。事前調査から、ステーキ肉を主に購入しているのは比較的高所得者であることが確認されたため、本調査もモニターの中でも世帯年収が高い層を対象として行っている。具体的には、インターネット調査会社の登録モニターのうち、世帯年収が1,000万円以上のモニター5,000人をランダムに抽出し、調査依頼を行った。調査期間は、2018年3月1日から3月7日であり、回答者は1,475人(29.5%)であった。この回答者のうち、400サンプルをランダムに抽出する形で購入し、分析の対象とした。アンケート内容には、国産放牧牛ステーキ肉100gに対する支払カード方式による仮想評価の質問(図1)、および日常の牛肉喫食、牛肉生産に対する意識・知

「完全放牧牛」とは、通常の肥育法に比べて以下の特徴を持っている牛肉です。

- ・自給率の低い、トウモロコシやダイズなどの濃厚飼料をほとんど与えず、牛肉生産における国内自給率が向上する
- ・国内の牧草や耕作放棄地を利用することで、飼料の供給が海外の状況に左右されない
- ・速くから飼料を輸送してこる必要がなく、温室効果ガスの削減につながる
- ・飼育期間のほとんどの期間は放牧されているため、牛のストレスが比較的少ない
- ・耕作放棄地(荒れた農地)、中山間地域(使われなくなった山と里の間の地域)などの草資源を利用することで、耕作放棄地や草地を適切に管理できる
- ・適切な管理により牛からの排せつ物は草地に循環する
- ・牧草などを食べさせているため、牛肉のサシはあまり見られず、さっぱりとした赤身の多い肉になる(タンパク質の豊富な食資源となる)
- ・そのため、脂肪は黄色みを帯びる



写真：放牧の様子



写真：放牧牛肉の見た目（イメージ）

Q19

このような国産の「完全放牧牛」のステーキ肉100gに対して、どのくらいまでなら支払ってもよいと思いますか。以下からお選びください。ただし、豪州産ステーキ肉は100g当たり500円とし、購入した分だけ他のものに使えるお金が減ることを念頭に、お答えください。

- 1 500円
- 2 700円
- 3 1,000円
- 4 1,500円
- 5 2,000円
- 6 2,500円
- 7 3,000円
- 8 4,000円
- 9 4,001円以上 円
- 10 豪州産ステーキ肉を購入したい
- 11 どちらの牛肉も購入したくない

次へ

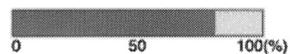


図 1. 提示した仮想状況文と選択肢の提示画面。

表 1. WTP および共変量の候補と基本統計量(連続変数). ***は、1% 水準で Wilcoxon の順位和検定の帰無仮説が棄却されることを示し、表示がない場合は、10% 水準でも帰無仮説が棄却されないことを示す.

変数	回答者		非回答者	
	平均	標準偏差	平均	標準偏差
国産放牧牛ステーキ肉 100g に対する WTP (円; WTP)	1161.61	(703.37)	-	-
年齢 (歳; age) ***	51.33	(9.88)	46.17	(12.59)
個人年収 (100 万円; i_income) ***	1059.60	(722.43)	804.99	(648.50)
世帯年収 (100 万円; h_income)	1661.88	(562.52)	1630.63	(557.07)

識などを含んでいる。仮想評価に関する質問で提示した支払カードは、500 円、700 円、1,000 円、1,500 円、2,000 円、2,500 円、3,000 円、4,000 円、4,001 円以上に加え、「豪州産ステーキ肉を購入したい」、「どちらの牛肉も購入したくない」という選択肢を提示した。

同時に、非回答者の情報を得るためにインターネット調査会社が保持しているモニター情報についても、前述の回答者 400 サンプルに加え、非回答者 800 サンプルについてもランダムに抽出し、データを得た。このモニター情報は、インターネット調査会社に登録されているものであり、年齢、性別、同居者の情報、職業、住居形態や、世帯年収、個人年収、さらには所有している車種や利用している Web サービス、趣味に関する事項など多岐にわたる。これらのモニター情報は、およそ 1 年ごとに情報が更新されている。以降で共変量の候補として利用した変数は、このモニター情報を基にして作成した。本研究で利用したもの以外にも共変量の候補は存在しうるが、解釈が可能ではないものや、著しく低い頻度で選択されるものを除いた結果、表 1、表 2 に示す変数を共変量の候補としている。

2.2 データの特徴

モニター情報として得られたデータのうち、本調査において利用した共変量の候補について示す。比較にあたって利用した検定の方法は、ダミー変数については Fisher の正確検定、連続変数と捉えられるものについては Wilcoxon の順位和検定を用いた。基本統計量に関する情報について述べると、回答者は非回答者に比べて男性の比率が高く、現在婚姻している割合が高いことがわかる。同居者との関係を見ると、回答者は、未就学児や親、祖父母、兄弟姉妹、恋人といった同居者がいる割合が低く、逆にその他学生、配偶者と同居している割合が高いことが確認できる。また、回答者は持ち家の割合が比較的低いことがわかる。逆に回答者は非回答者に比べ SNS や動画サイト、ブログ等の利用がある割合が低い。趣味においては、非回答者の方が音楽、映画や観劇、漫画、ソーシャルゲームやオンラインゲーム、音楽、美容が趣味であると答える割合が高く、逆にスポーツ観戦、株やマネーが趣味と答える割合は、回答者の方が高いことが読み取れる。また、回答者の方が平均的に年齢が高く、個人年収も高いことが確認された。一方で、世帯年収については、両者に統計的に有意な差は見られなかった。

3. 分析

3.1 非回答者が存在する場合の母集団平均の推定

まず、非回答者の存在が、目的となる変数の推定にどのように影響するか概説する。いま、非回答者が存在する場合の、関心のある目的変数 y についての母集団平均 μ の推定において、無作為抽出によって得られた調査対象者 N サンプルのうち、 N_1 人が回答し、 $N_0 (= N - N_1)$ 人が非回答であったとする。また、調査に回答した場合 1、非回答の場合 0 を取る二値変数を

表 2. 共変量の候補と基本統計量(ダミー変数). ***, **, * はそれぞれ 1%, 5%, 10% 水準で Fisher の正確検定の帰無仮説が棄却されることを示す.

属性の内容	変数	回答者	非回答者
性別***	男性 (<i>male</i> = 1)	288 (72%)	472 (59%)
	女性 (<i>male</i> = 0)	112 (28%)	328 (41%)
現在の婚姻状態***	現在婚姻あり (<i>marriage</i> = 1)	340 (85%)	586 (73%)
	現在婚姻なし (<i>marriage</i> = 0)	60 (15%)	214 (27%)
同居家族 (未就学児) ***	あり (<i>with_presch</i> = 1)	26 (6%)	94 (12%)
	なし (<i>with_presch</i> = 0)	374 (94%)	706 (88%)
同居家族 (その他学生) *	あり (<i>with_o_stud</i> = 1)	66 (16%)	100 (12%)
	なし (<i>with_o_stud</i> = 0)	334 (84%)	700 (88%)
同居家族 (配偶者) **	あり (<i>with_spouse</i> = 1)	252 (63%)	445 (56%)
	なし (<i>with_spouse</i> = 0)	148 (27%)	355 (44%)
同居家族 (親) **	あり (<i>with_par</i> = 1)	77 (19%)	197 (25%)
	なし (<i>with_par</i> = 0)	323 (81%)	603 (75%)
同居家族 (祖父母) **	あり (<i>with_granpar</i> = 1)	3 (1%)	33 (4%)
	なし (<i>with_granpar</i> = 0)	397 (99%)	767 (96%)
同居家族 (兄弟姉妹) ***	あり (<i>with_sibl</i> = 1)	16 (4%)	81 (10%)
	なし (<i>with_sibl</i> = 0)	384 (96%)	719 (90%)
同居家族 (恋人) *	あり (<i>with_lover</i> = 1)	3 (1%)	17 (2%)
	なし (<i>with_lover</i> = 0)	397 (99%)	783 (98%)
持ち家が賃貸か*	持ち家 (<i>myhome</i> = 1)	72 (18%)	183 (23%)
	賃貸・寮など (<i>myhome</i> = 0)	328 (82%)	617 (77%)
SNS・動画サービス・ブログ等の利用有無***	利用あり (<i>WebS</i> = 1)	317 (79%)	702 (88%)
	利用なし (<i>WebS</i> = 0)	83 (21%)	98 (12%)
趣味 (音楽) **	あり (<i>hob_music</i> = 1)	166 (42%)	380 (48%)
	なし (<i>hob_music</i> = 0)	234 (58%)	420 (52%)
趣味 (映画・観劇) *	あり (<i>hob_movie</i> = 1)	175 (44%)	396 (50%)
	なし (<i>hob_movie</i> = 0)	225 (56%)	404 (50%)
趣味 (漫画) ***	あり (<i>hob_manga</i> = 1)	52 (13%)	170 (21%)
	なし (<i>hob_manga</i> = 0)	348 (87%)	630 (79%)
趣味 (ソーシャルゲーム・オンラインゲーム) ***	あり (<i>hob_s_on_game</i> = 1)	21 (5%)	85 (11%)
	なし (<i>hob_s_on_game</i> = 0)	379 (95%)	715 (89%)
趣味 (料理) **	あり (<i>hob_cook</i> = 1)	87 (24%)	227 (28%)
	なし (<i>hob_cook</i> = 0)	313 (78%)	573 (72%)
趣味 (美容) **	あり (<i>hob_beauty</i> = 1)	57 (14%)	169 (21%)
	なし (<i>hob_beauty</i> = 0)	343 (86%)	631 (79%)
趣味 (スポーツ観戦) **	あり (<i>hob_sport_w</i> = 1)	111 (28%)	177 (22%)
	なし (<i>hob_sport_w</i> = 0)	289 (72%)	623 (78%)
趣味 (株・マネー) **	あり (<i>hob_finance</i> = 1)	112 (28%)	176 (22%)
	なし (<i>hob_finance</i> = 0)	288 (72%)	624 (78%)

欠測指標 z とする.

ここで, 母集団平均 μ の不偏推定値 $\hat{E}(y)$ は,

$$(3.1) \quad \hat{E}(y) = \frac{1}{N} \sum_{i=1}^N y_i$$

と表される。ただし、添字の i は i 番目の調査対象者を表す。他方、回答者のサンプルから得られる推定値は、

$$(3.2) \quad \bar{y}_{obs} = \frac{\sum_{i=1}^N z_i y_i}{\sum_{i=1}^N z_i}$$

と表される。しかし、 \bar{y}_{obs} は、 y と z が独立でない限り母集団平均 μ の不偏推定量にはならない (星野, 2010)。

3.2 逆重み付き推定量

上述のような非回答者が存在する場合に、偏りを補正する方法は、Little and Rubin (2002)、星野 (2009, 2010) に挙げられているように、欠測のメカニズムに応じた方法が複数提案されている。この中でも本研究では、多くの実証研究で利用されている逆重み付き推定量 (IPWE) を用いる。

いま、強く無視できる割り当て条件 (Rosenbaum and Rubin, 1983) が回答者・非回答者間について成り立つ、すなわち $f(z | y, \mathbf{w}) = f(z | \mathbf{w})$ を仮定する。ただし、 z は調査に回答した場合 1、非回答の場合 0 を取る二値変数で表される欠測指標であり、 y は目的変数 (本研究では、国産放牧牛に対する WTP)、 \mathbf{w} は共変量として用いるモニター情報である。Kang and Schafer (2007) に従えば、このときの興味のある変数 y における母集団平均 μ の逆重み推定量 $\tilde{\mu}_{IPW}$ は、

$$(3.3) \quad \tilde{\mu}_{IPW} = \frac{\sum_{i=1}^N \frac{z_i y_i}{e(\mathbf{w}_i)}}{\sum_{i=1}^N \frac{z_i}{e(\mathbf{w}_i)}}$$

と表され、この $\tilde{\mu}_{IPW}$ は母集団平均の一致推定量となる (星野, 2010)。 $e(\mathbf{w}_i)$ は、 \mathbf{w}_i が与えられたときに y_i が観測される条件付き確率を表し、傾向スコア (propensity score) と呼ばれる。この傾向スコアに関する推計にはロジットモデルやプロビットモデルがよく用いられるが、 $e(\mathbf{w}_i)$ が例えばロジット型のモデルで表されるとすると、以下のように表される。

$$(3.4) \quad e(\mathbf{w}_i) = \Pr[z_i = 1 | \mathbf{w}_i] = \frac{\exp(\mathbf{w}_i' \boldsymbol{\psi})}{1 + \exp(\mathbf{w}_i' \boldsymbol{\psi})} = e(\mathbf{w}_i; \boldsymbol{\psi})$$

ただし、 $\boldsymbol{\psi}$ はモデルのパラメータであり、このパラメータを最尤推定した値 $\hat{\boldsymbol{\psi}}$ を用いて各調査対象者の傾向スコアを推定し、式 (3.3) に代入することで、 μ の IPWE である $\tilde{\mu}_{IPW}$ を得る (Kang and Schafer, 2007)。

3.3 回答・非回答の説明モデル

母集団における国産放牧牛への平均 WTP の IPWE を求めるために、まずアンケートへの回答・非回答を説明するモデルを推定する。推定にあたっては、ロジットモデルを用いた。ここで、回答者へのリワードは一定であるので、モニターがアンケートに回答するか否かは、個人の機会費用と関係すると考えられる。従って、個人の機会費用の代替として個人年収が説明変数の強い候補となる。この際、個人年収を「わからない・答えたくない」と登録しているモニターを削除した結果、回答者のデータは 398 サンプル、非回答者のデータは 786 サンプルとなった。ここで、回答者の割合は 29.5% であることから、分析における回答者のデータは、回答率に合わせ、349 サンプルを用いる。

加えて、消費者の選好の多様性を考慮するため、モニター情報のうち、表 1、表 2 に示した共変量の候補をモデルに組み込み、AIC を基準にした変数増減法によりモデルを決定した。共変

量の候補の選択にあたっては、解釈が可能でないもの、著しく低い頻度で選択される変数については用いず、また、二値変数同士にはテトラコリック相関係数、順序変数間にはポリコリック相関係数、二値変数と連続変数間にはバイシリアル相関係数、順序変数と連続変数間にはポリシリアル相関係数、連続変数間にはピアソンの積率相関係数確認し、高い相関 ($|r| > 0.7$) がみられる場合には、同時にその変数を利用することはなかった。推定にあたっては、統計ソフト R(3.4.3)を用いた。

推計結果を表 3 に示す。推計の結果、個人所得が高く(有意水準: 10%, 以下同様)、年齢が高いモニター(1%)の回答確率が有意に高いことが確認された。同居の有無に関しては、祖父母および兄弟姉妹と同居している場合に、回答率が有意に低いことが確認された。SNS や動画サービスなどのウェブサービスを利用している調査対象者の回答確率は有意に低い(5%)という結果が得られた。趣味については、スポーツの観戦が趣味である(10%)、金融や株式投資等が趣味である(10%)の場合に回答確率が高く、逆に漫画が趣味の場合、回答確率が有意に低いことが明らかになった。以上の内容を解釈するとすれば、インターネット調査への回答は、家庭で時間の空いたときに行うものであり、祖父母や兄弟姉妹等の同居人がいる場合や、家庭内でできる趣味などを楽しんでいる場合には、アンケート回答を行う時間が取れず、回答する確率が低くなっていると考えられる。他方、個人所得が高いモニター、株・マネーが趣味のモニターは、まめに情報をチェックしているため、アンケートにも回答する確率が高くなっていると考えられる。一方、世帯年収は回答の有無には関係しないことも明らかになった。この点は、今回の調査対象者がそもそも世帯年収の高い人々であり、ポイントなどを求めて調査に協力しているわけではないことを示していると考えられる。

3.4 支払意志額の推定

表 3 で求めた回答者・非回答者に関する推定結果を用いて、回答者の国産放牧牛に対する平均 WTP、および、母集団における国産放牧牛への平均 WTP の IPWE をそれぞれ推計する。ここで、一般的な支払カード方式の仮想評価法においては、回答者が J 個の提示された支払カードからある支払カード t_j を選択した場合、回答者の WTP は、 $[t_j, t_{j+1})$ の区間に存在すると仮定し、生存曲線の下部を積分することで平均 WTP を求める (Cameron and Huppert, 1989)。しかし、本研究における関心は、回答者・非回答者間に差があるか否かであるため、単純化のために、回答者がある支払カード t_j を選択したとき、回答者の WTP は t_j であると取り扱い、平均値を求める。このとき「豪州産ステーキ肉を購入したい」、「どちらの牛肉も購入したくない」を選択した回答者の WTP は 0 円、「4,001 円以上」を選択した回答者の WTP は 4,000 円であると仮定した。

式(3.2)を用いて、アンケートへの回答者のみを利用して推計された回答者の国産放牧牛に対する平均 WTP は、100g 当たり約 1,161.6 円であった。他方、式(3.3)を用いて推計された母集団における国産放牧牛への平均 WTP の IPWE は、100g 当たり約 1,161.7 円と推計された。両者の差はほぼゼロであり、回答者と非回答者間の属性的な差異は、評価対象である国産放牧牛への WTP に影響するとは言えないことが示された。

4. おわりに

本研究では、インターネット調査におけるモニター登録情報を用いて、回答者・非回答者間の差異が存在するか、また、その差が評価対象となる財への WTP の推計値にどのような影響を与えるかを検証した。その結果、本研究で使用した回答者・非回答者間の属性には差異があり、回答・非回答に関係する要因ではあるものの、その属性的な差異は、必ずしも評価対象と

表 3. 推定結果(共変量による回答・非回答の推定). ***, **, *, † はそれぞれ, 0.1%, 1%, 5%, 10% 水準でパラメータが 0 と有意に異なることを示す. 括弧内の値は, 標準誤差を表す.

	個人年収のみ	全変数	最終モデル
(Intercept)	-1.903*** (0.304)	-3.661* (1.698)	-1.870*** (0.485)
$\ln(1 + i_income)$	0.174*** (0.047)	0.074 (0.053)	0.087† (0.046)
male		0.136 (0.197)	
age		0.019* (0.008)	0.020** (0.007)
marriage		0.328 (0.278)	
with_presch		-0.461† (0.263)	-0.397 (0.252)
with_o_stud		0.060 (0.191)	
with_spouse		-0.097 (0.171)	
with_par		0.198 (0.213)	
with_granpar		-0.983 (0.627)	-1.025† (0.622)
with_sibl		-0.573 (0.378)	-0.616† (0.345)
with_lover		-0.888 (0.789)	-1.094 (0.773)
myhome		0.028 (0.180)	
WebS		-0.395* (0.182)	-0.443* (0.179)
hob_music		-0.043 (0.155)	
hob_movie		-0.158 (0.152)	
hob_manga		-0.252 (0.207)	-0.322† (0.195)
hob_s_on_game		-0.234 (0.291)	
hob_cook		-0.261 (0.183)	-0.269 (0.167)
hob_beauty		0.236 (0.230)	
hob_sport_w		0.299† (0.163)	0.281† (0.159)
hob_finance		0.287† (0.164)	0.304† (0.159)
$\ln(h_income)$		0.213 (0.223)	
AIC	1388.587	1359.496	1343.671
Log Likelihood	-692.293	-656.748	-659.835
McFadden's pseudo R^2	0.012	0.062	0.058
N	1135	1135	1135

なる財への WTP に影響するとは言えないことが明らかになった。

ただし、本研究で対象とした母集団は、世帯年収 1,000 万円以上の高所得者であり、その他の所得層の集団には必ずしも当てはまらない可能性もある。回答・非回答を分かち要因が各所得階層で同様であるとすれば、子どもや祖父母、兄弟などの同居者が多い場合、また他の SNS や動画サイトの利用がある場合には、インターネットモニターに登録をしても、アンケートに回答する時間が取れず、回答できないという要因が存在することが示唆される。時間がいないからアンケートに回答しないという要因は、環境の価値の評価や環境保全的な農畜産物に対する評価とは関係のないものと捉えられ、それゆえに WTP の推定値にも影響を与えなかったと考えられる。

また、本研究においては、インターネット調査を用いたことについても留意が必要である。前述の通り、インターネット調査によって得られたサンプルは、従来型の確率的標本抽出に基づく既存調査(訪問調査や郵送調査法)によって得られるサンプルと性格が異なる(大隅, 2017)ことが指摘されており、その調査法による差異を調整するという視点からの研究も進展しつつある(星野・前田, 2006; 星野, 2007)。しかし、インターネット調査においては、従来の方法では収集が困難であった共変量について、登録されているモニター情報を用いることが可能であり、このモニター情報は今後もますます情報が蓄積されていくことが期待できる。

今後は、インターネット調査などの多くの共変量を利用可能な調査方法により、非回答バイアスがどのように目的となる変数の推定に影響を及ぼすか考慮しつつ、それを念頭に調査設計を行うことで、より確かな根拠に基づいた意志決定が可能となると考えられる。

謝 辞

本研究は、キヤノン財団研究助成プログラム「理想の追求」の助成を受けて行なった研究「牛肉生産システムの大構造改革：科学と国土をフル活用した大革新」(代表者：後藤貴文)の成果の一部である。記して謝意を表したい。

参 考 文 献

- Bethlehem, J., Cobben, F. and Schouten, B. (2011). *Handbook of Nonresponse in Household Surveys*, John Wiley & Sons, Hoboken.
- Cameron, T. A. and Huppert, D. D. (1989). OLS versus ML estimation of non-market resource values with payment card interval data, *Journal of Environmental Economics and Management*, **17**(3), 230-246.
- Groves, R. M. and Schoeffel, G. J. (2018). Use of administrative records in evidence-based policymaking, *The ANNALS of the American Academy of Political and Social Science*, **678**(1), 71-80.
- Heckman, J. J. (1979). Sample selection bias as a specification error, *Econometrica*, **47**(1), 153-161.
- 星野崇宏 (2007). インターネット調査に対する共変量調整法のマーケティングリサーチへの適用と調整効果の再現性の検討, *行動計量学*, **34**(1), 33-48.
- 星野崇宏 (2009). 『調査観察データの統計科学—因果推論・選択バイアス・データ融合』, 岩波書店, 東京.
- 星野崇宏 (2010). 調査不能がある場合の標本調査におけるセミパラメトリック推定と感度分析：日本人の国民性調査データへの適用, *統計数理*, **58**(1), 3-23.
- 星野崇宏, 前田忠彦 (2006). 傾向スコアを用いた補正法の有意抽出による標本調査への応用と共変量の選択法の提案, *統計数理*, **54**(1), 191-206.
- Kang, J. D. and Schafer, J. L. (2007). Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data, *Statistical Science*, **22**(4), 523-539.

- Little, R. J. and Rubin, D. B. (2002). *Statistical Analysis with Missing Data*, 2nd ed., John Wiley & Sons, New York.
- Little, R. J., D'Agostino, R., Cohen, M. L., Dickersin, K., Emerson, S. S., Farrar, J. T., Frangakis, C., Hogan, J. W., Molenberghs, G., Murphy, S. A., Neaton, J. D., Rotnitzky, A., Scharfstein, D., Shih, W. J., Siegel, J. P. and Stern, H. (2012). The prevention and treatment of missing data in clinical trials, *New England Journal of Medicine*, **367**(14), 1355-1360.
- 松岡亮二, 前田忠彦 (2015). 「日本人の国民性第 13 次全国調査」の欠票分析：個人・地点・調査員の特性と調査回収状況の関連, *統計数理*, **63**(2), 229-242.
- Messonnier, M. L., Bergstrom, J. C., Cornwell, C. M., Teasley, R. J. and Cordell, H. K. (2000). Survey response-related biases in contingent valuation: Concepts, remedies, and empirical application to valuing aquatic plant management, *American Journal of Agricultural Economics*, **82**(2), 438-450.
- Mitchell, R. C. and Carson, R. T. (1989). *Using Surveys to Value Public Goods: The Contingent Valuation Method*, RFF Press, New York.
- National Research Council (2010). *The Prevention and Treatment of Missing Data in Clinical Trials*, National Academies Press, Washington, DC.
- 大隅昇 (2017). インターネット調査, 『社会調査ハンドブック (新装版)』 (林知己夫 編), 第 4 章, 200-240, 朝倉書店, 東京.
- Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects, *Biometrika*, **70**(1), 41-55.
- 土屋隆裕 (2005). 調査不能者の特性に関する一考察—「日本人の国民性 第 11 次全国調査」への協力理由に関する事後調査から—, *統計数理*, **53**(1), 35-56.
- 土屋隆裕 (2006). 「調査への指向性」変数を用いた調査不能バイアス補正の試み：「日本人の国民性調査」データへの適用, *日本統計学会誌*, **36**(1), 1-23.
- 土屋隆裕 (2010). 調査への指向性変数を用いた調査不能バイアスの二段補正—「日本人の国民性第 12 次全国調査」への適用—, *統計数理*, **58**(1), 25-38.
- 鷺田豊明 (1999). 『環境評価入門』, 勁草書房, 東京.
- Whitehead, J. C., Groothuis, P. A. and Blomquist, G. C. (1993). Testing for non-response and sample selection bias in contingent valuation: Analysis of a combination phone/mail survey, *Economics Letters*, **41**(2), 215-220.

Adjusting for Non-response Bias Using Registration Data
from an Internet Research Company
—Application to Consumers' Evaluation of Japanese Grass-fed Beef—

Takeru Kusudo¹, Takafumi Gotoh², Yoshifumi Takahashi³ and Mitsuyasu Yabe³

¹Graduate School of Bioresource and Bioenvironmental Resources, Kyushu University

²Research Field in Agriculture, Agriculture, Fisheries and Veterinary Medicine Area,
Kagoshima University

³Faculty of Agriculture, Kyushu University

Research on missing data is attracting growing attention as a means of improving translation of research into more reliable evidence-based decision making. In this study, we investigate the non-response bias in Contingent Valuation Method (CVM) using the Inverse Probability Weighting Estimator (IPWE). One of the major problems with application of the IPWE is related to the availability of covariates. We dealt with this problem by using individual information registered at an internet research company. Using the CVM format, we evaluated consumers' willingness to pay for a grass-fed beef produced by an environmentally friendly farming system. The results revealed that there were differences between respondents and non-respondents in individual income, age, characteristics of family members, use of SNS or video distribution websites, and hobbies. The sample mean of willingness to pay (WTP) for grass-fed beef was ¥1161.6 per 100 g, and the IPWE of WTP was ¥1161.7 per 100 g. In other words, there was almost no difference between the mean and IPWE of WTP. Based on this survey, we conclude that there are differences in covariates between respondents and non-respondents. However, this difference did not cause a non-response bias in estimation of consumers' WTP for grass-fed beef.

ミツバチの送粉サービスと景観構造との関係解析

—宮崎県綾町における日向夏生産の事例—

光田 靖¹・湯村 昂広^{1,2}・平田 令子¹・伊藤 哲¹

(受付 2018 年 7 月 2 日; 改訂 10 月 2 日; 採択 11 月 20 日)

要 旨

宮崎県綾町において特産品である日向夏(*Citrus tamurana*)に対するニホンミツバチ(*Apis cerana*)による送粉サービスと周囲の景観構造との関係を解析した。綾町内の日向夏農園 16 箇所において対象木を合計 24 本選定し、開花時期にニホンミツバチの訪花数を現地計測した。一方で、オルソフォトを判読して対象地の土地被覆分類図を作成した。各対象木を中心とした半径 2 km 円内において土地被覆タイプ別の面積を計測した。ニホンミツバチの訪花数を送粉サービスの指標として、これを目的変数とした統計モデルを開発した。モデルの説明変数として、景観構造の指標である半径 2 km 円内の天然林面積および農地と草地の合計面積を用いた。また、ニホンミツバチ訪花数の観測においては観測者のミツバチ発見能力に差があることから、ニホンミツバチ訪花数を予測する生態モデルと、観測者がニホンミツバチを発見する確率を推定する観測モデルを組み合わせて、ベイズ推定によりパラメータ推定を行った。推定されたパラメータは天然林面積が多いほど、農地と草地の合計面積が多いほど、ニホンミツバチによる送粉サービスが高くなることが示唆された。

キーワード：エコパーク、GIS、ベイズ推定、天然林再生、有機農業。

1. はじめに

生態系からの恵みである「生態系サービス」の重要性が強く認識されるようになってきた。このような中で、国連が主体となって行った国際規模での生態系サービスの評価であるミレニアム生態系評価(Millennium Ecosystem Assessment, 2005)や、ドイツやその他の国が主体となって行った生態系サービスの経済評価である生態系と生物多様性の経済学(TEEB: The Economics of Ecosystem and Biodiversity)などの生態系サービスに関する国際的な調査結果が報告され、生態系サービスの重要性の認識はさらに高まってきている。

生態系サービスの中で調整サービスに分類されている送粉サービスは、食糧生産の過程において人類に大きく貢献している。これまでに行われてきた研究では、世界の主要農作物の約 75% が送粉サービスに依存しており(Klein et al., 2007)、その経済価値は 2005 年時点において全世界で約 1,530 億ユーロ(Gallai et al., 2009)、また日本においては 2013 年時点で 4,700 億円に達することが明らかにされた(小沼・大久保, 2015)。一方で送粉サービスは食糧生産においてだけでなく野生植物の約 87%(約 308,000 種)の繁殖にも貢献しており、生物多様性の維持に

¹ 宮崎大学農学部：〒 889-2192 宮崎県宮崎市学園木花台西 1-1

² 現 地域環境計画九州支社：〒 814-0006 福岡県福岡市早良区百道 2-9-3

対しても貢献している (Ollerton et al., 2011). しかし送粉サービスの役割を担っている送粉者の減少は世界的な問題となっており (Potts et al., 2010), 送粉者の中でも最も主要な送粉者であると言われているミツバチやその他の野生のハチについては世界各地でその多様性や豊富さが減少していることが報告されている (Biesmeijer et al., 2006; Patiny et al., 2009; Pettis and Delaplane, 2010). このような送粉者の減少要因としては農薬の影響や病気の広まり (たとえば, Kevan et al., 1997), また土地利用の改変によるハビタットの分断化や損失, 劣化などランドスケープ構造の変化による影響が報告されている (Kearns et al., 1998; Goulson et al., 2008; Ricketts et al., 2008). このように送粉者の減少について様々な研究報告がある中で, 2016 年 2 月にマレーシアで開催された生物多様性及び生態系サービスに関する政府間科学政策プラットフォーム (IPBES: Intergovernmental Platform on Biodiversity and Ecosystem Services) では, 送粉サービスに関する報告書 (送粉者と食糧生産に関するアセスメント報告書) が公表され, 送粉者と送粉サービス保全のために配慮すべき 10 の政策と国際的なモニタリングの必要性が示された (IPBES, 2017).

国連により 2015 年に制定された持続可能な開発目標 (Sustainable Development Goals, SDGs) において, 目標 12 として「つくる責任, つかう責任」が挙げられている. SDGs 達成のためには関連する生態系サービスとの関係を明らかにし, これを活用することが重要であると考えられる (Wood and DeClerck, 2015). 先にも述べたように送粉サービスは農業生産に直結しており, SDGs を達成させるためのキーファクターである. これまで, いくつかの事例で送粉サービスが農作物の収量に影響することが明らかにされている. 例えば, コスタリカのコーヒー農園で行われた研究においては, 送粉サービスの担い手である送粉者が豊富な農園ではコーヒーの生産量が 20% 増加することが明らかになった (Ricketts, 2004). 一方で, 送粉サービスの豊富さは農地周辺のランドスケープ構造に影響を受けることが明らかにされている. 例えば, コーヒー農園周辺の 1km 以内に森林がある農園では訪花昆虫の訪問率が増加したことが明らかにされた (Ricketts, 2004). 送粉サービスとランドスケープ構造の関係を定量的に評価できれば, 送粉サービスの恩恵が受けられる農地の適地選択や, 送粉サービスを利用した農業への切り替えを提供できることが期待される. さらに, 送粉サービスの恩恵を向上させられるようなランドスケープデザインの提案へと繋げていくことが期待できる.

送粉サービスを定量的に評価する研究では, 送粉者の豊富さを送粉サービスの指標として, 送粉者の豊富さと景観構造との関係を明らかにしようとするものが多い. 既往の研究において, 送粉者の豊富さと農作物の収量との間に強い関係があることが明らかにされている (Cane, 2005; Bosch et al., 2006; Isaacs and Kirk, 2010). 例えば, コーヒー, マンゴー, ソバ, およびヒマワリなどで送粉者が多いほど収穫量が高くなるという相関関係が報告されている (Ricketts, 2004; Carvalheiro et al., 2010; Taki et al., 2010; Carvalheiro et al., 2011). このように送粉者の豊富さを送粉サービスの指標とすることは妥当であると考えられる. また, スイカ, ソバ, およびサクランボについては農地周辺の森林面積が増加につれ, 送粉者の訪問率が増加する傾向が明らかになっている (Kremen et al., 2002; Taki et al., 2010; Holzschuh et al., 2012). このように送粉者とランドスケープ構造の関係を定量的に明らかにすることで, 送粉サービスの効果を増加させられるような景観管理を行うというような貢献が期待される.

これまで日本においては送粉サービスの重要性について欧米やヨーロッパと比較してあまり注目されてこなかった (小沼・大久保, 2015). そのため, IPBES の報告書でも指摘されたように, 日本では送粉サービスに関する情報が不足している (IPBES, 2017). 国内における送粉サービスの研究蓄積をおこない, 送粉サービスを維持する対策を講じることが求められている.

そこで本研究では, 宮崎県綾町を対象として代表的な農作物の一つである日向夏 (*Citrus tamurana*) の生産に送粉サービスを活用することを想定して, 代表的な送粉者であるニホンミ

ツバチ (*Apis cerana*, 以下, ミツバチと称する)の豊富さと景観構造との関係を明らかにすることを目的とする。

2. 資料および方法

2.1 対象地

本研究の対象地である宮崎県東諸県郡綾町は、国内最大級の照葉樹林と長い歴史をもつ有機農業で知られている。綾町では約半世紀前から人と自然が共生する町づくりを進めてきており、農薬や化学肥料を使用しない自然生態系農業への取り組みや、照葉樹林再生のプロジェクトである「綾の照葉樹林プロジェクト」などの取り組みを行ってきた(朱宮 他, 2013)。2012年にはこれらの取り組みが認められ、日本で初めて移行地域を伴ったユネスコエコパークに登録されている。照葉樹林プロジェクトにおいては国有林、県有林、町有林を合わせた約1万 haの照葉樹林の再生を進めており、このエリア内の人工林を自然林に戻していく計画が進められている。綾町では宮崎県の特産品である日向夏の栽培が県内で盛んであり、県内でも有数の産地として知られている。日向夏は強い自家不和合性(自個体の花粉では授粉できず、別の花粉源からの花粉が必要となる性質)を持っているため、綾町の日向夏農園では人工授粉による花粉の媒介が行われている。しかし人工授粉の作業は大変な時間と労力が必要であり、これが日向夏の栽培コストが高くなる原因となっている。

2.2 調査方法

綾町内において日向夏農園を16箇所選定し、各調査地に対して1から4本の調査対象木の設定を行った。以前の報告(Yumura et al., 2016)では、対象木に調査枠を設置して1時間に枠内に訪れるミツバチの数を計数した。しかし、この方法では観測されるミツバチの数が少なく、時間がかかるという問題があった。そこで本研究では対象木に隣接木2本をあわせた3本について、周囲を歩きながら3分間にわたって全体を観察して存在するミツバチの数を計数した。ミツバチの見落としを考慮して、現地調査においては複数の調査者で同時に調査を行った。現地調査は日向夏の開花期間にあわせて2016年4月29日から5月20日の期間中5日間において行った。また、計測時に対象木の写真を撮影し、開花レベルとして開花の度合いを5段階に分類した。さらに写真において樹冠面積を計測し、個体サイズの指標とした。調査結果の一例を図1に示す。

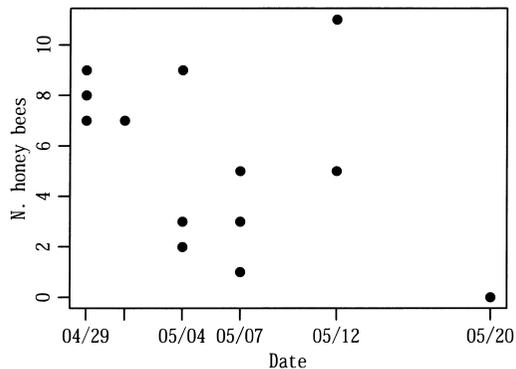


図1. ミツバチ訪花数調査結果の例。

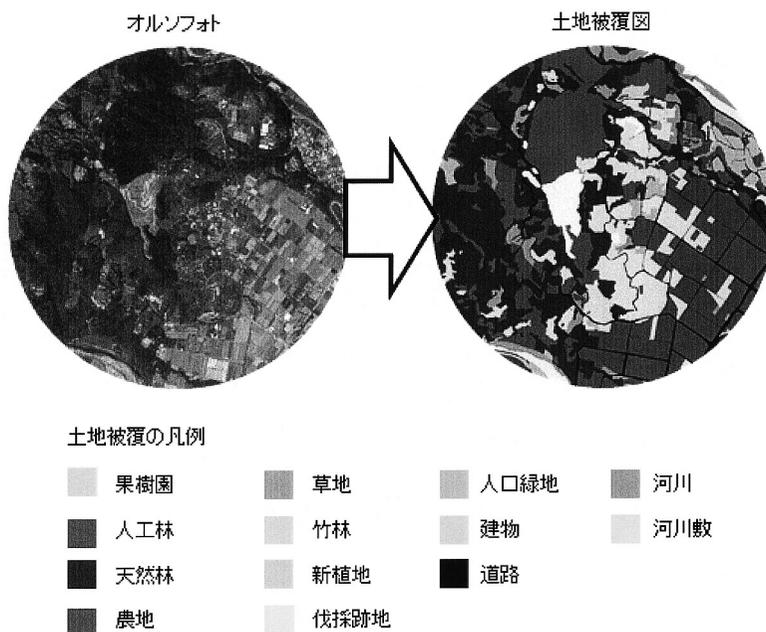


図 2. 対象木を中心とした半径 2 km 円内の土地被覆図の例.

対象日向夏農園周辺の景観構造を指標化するため、平成 25 年撮影の航空写真から作製されたオルソフォトを判読し、土地利用図を作成した(図 2)。本研究で用いた土地被覆タイプは人工林、広葉樹林、農地、住宅地、緑の多い住宅地、竹林、河川、河川敷、伐採地、果樹園、人工物、植林地、道路、ゴルフ場、草地、裸地、および人工緑地の 17 種類である。先行研究(Yumura et al., 2016)および既往の研究(Fujiwara and Washitani, 2017)から、各対象木を中心とした半径 2 km 円について土地被覆タイプ別の面積を計測した。

2.3 解析方法

以前の研究では、ミツバチ訪花数と景観構造との関係について一般化線形混合モデルを用いて解析した(Yumura et al., 2016)。しかし、本研究においては調査者がミツバチを探しながら計数するという方法を採用したため、調査者によってミツバチ発見能力が異なることによる観測データのバラツキが生じる危険性がある。そこで本研究では景観構造とミツバチ訪花数との関係を説明する生態モデルと、観測者の発見能力を推定する観測モデルを結合した統計モデルを用いた。

生態モデルにおいては以前の研究を参考にして、景観構造の指標として半径 2 km 以内の天然林面積(以下、 NF)および農地と草地の合計面積(以下、 AG)を説明変数として用いた。また、既往の研究(Eckhart, 1991)を参考に、開花レベル(以下、 FL)をミツバチへの誘因効果を指標する因子として用いた。気象要因による影響を考慮するため、3段階に区分した調査時刻(10 時~12 時, 12 時~15 時, 15 時~17 時以下、 T)および調査日(以下、 D)をランダム要因として取り扱った。生態モデルのモデル式は一般化線形混合モデルの形式を採用し、リンク関数を対数関数として以下の様に構築した。

$$(2.1) \quad \log(E[N_i]) = b_{0,FL} + b_1 NF + b_2 AG + re_{1,T} + re_{2,D} + \log(S)$$

ここで、 N_t は真のミツバチ訪花数、 S は写真を用いて計測した樹木サイズでありオフセット項として用いた。 $b_{0,FL}$ は開花レベルによって決まる係数、 b_1 および b_2 は各説明変数の回帰係数、および $re_{1,T}$ は調査時刻によって、 $re_{2,D}$ は調査日によって設定されるランダム効果を示す。ランダム効果については平均値を 0 として、分散のパラメータ (s_1 および s_2) をそれぞれ推定する。真のミツバチ訪花数 (N_t) は現実には観測できない値であり、実際に観測されたミツバチ訪花数は各個人の発見能力によって影響を受ける。よって、下記のような観測モデルを導入した。

$$(2.2) \quad N_e = N_t \times P_i$$

ここで P_i は観測者 i ごとに設定される発見率であり、 N_e は真のミツバチ訪花数と個人の発見率から計算されるミツバチ訪花数の推定値である。このような観測モデルを用いることによって、観測者における発見能力の違いを考慮して生態モデルのパラメータを推定することを試みた (Kery and Schaub, 2012)。ミツバチ訪花数はポアソン分布に従うと仮定して、ミツバチ訪花数の推定値と観測値から計算される尤度にもとづきハミルトニアンモンテカルロ法によるパラメータ推定を行った。なお、事前分布として b_0 には -2 から 2 の値をとる一様分布を、 b_1 および b_2 には平均値 0 および標準偏差 100 の正規分布を、および s_1 および s_2 には 0 から 100 の値をとる一様分布を設定した。観測者ごとに設定されるミツバチ発見率である P_i については、0 から 1 の値をとる一様分布を事前分布として設定し、観測者ごとに独立して推定した。ハミルトニアンモンテカルロ法を実行するにあたって、繰り返し回数を 50,000 回とし、初期除外回数を 30,000 回とし、3 つの試行を並列した。

3. 結果および考察

ハミルトニアンモンテカルロ法によって推定されたパラメータを表 1 に示す。パラメータの収束度合いを示す \hat{R} 指標はすべて 1.1 を下回っており、ハミルトニアンモンテカルロ法による

表 1. パラメータ推定の結果.

	Median	\hat{R}
$b_{0,1}$	-1.389	0.998
$b_{0,2}$	-0.869	0.999
$b_{0,3}$	0.064	0.998
$b_{0,4}$	0.779	0.998
$b_{0,5}$	1.064	0.997
b_1	0.381	0.997
b_2	0.255	0.995
s_1	0.364	1.001
s_2	1.717	1.006
p_1	0.708	1.001
p_2	0.635	0.999
p_3	0.626	1.007
p_4	0.568	0.999
p_5	0.796	1.004
p_6	0.945	1.000
p_7	0.647	1.004

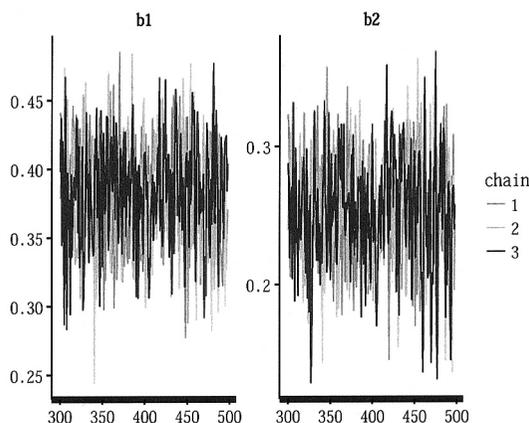


図 3. パラメータ b_1 および b_2 に関するトレースプロット.

パラメータ推定は収束したと判断した．景観構造との関係性をあらわすパラメータ b_1 および b_2 について，ハミルトニアンモンテカルロ法によるサンプリング経過を図 3 に示す．図 3 において 3 回の試行結果(chain 1 から 3 として表示)は同様の傾向を示しており，両パラメータとも適切に収束していることが確認できる．半径 2km 以内の天然林面積に係るパラメータ b_1 は正の値をとっており，周囲に天然林が多いほどミツバチ訪花数は多くなることを示している．天然林はミツバチの営巣場所となっており，既往研究(Taki et al., 2010, 2011)と同様に，ミツバチによる送粉サービスを期待するためには天然林が不可欠であると考えられる．

半径 2km 以内の農地および草地合計面積に係るパラメータ b_2 も正の値をとっており，周囲に農地や草地が多いほどミツバチ訪花数は多くなることを示している．草地在ミツバチの採餌場所として機能することは確認されており(Fujiwara and Washitani, 2017)，本研究の結果も草地在ミツバチの採餌活動に正の影響を与えることを示唆していると考えられる．一方で，既往の研究では農地がミツバチに悪影響を与えるという結果が多く報告されている(Ricketts et al., 2008; Chauzat et al., 2009; Henry et al., 2012)．これは農地における農薬の使用が影響しているものであり，本研究の対象である綾町ではほとんどの農地で有機栽培が行われており，そのことを反映したものであると考えられる．先の研究(Yumura et al., 2016)においても同様の結果となったが，新たなデータでも同様の結果となったことで有機農業の効果が確認された．

推定された個人のミツバチ発見率は 0.626 から 0.945 の値をとっており，ミツバチ発見の能力に大きな差があることが示唆された．動物や昆虫の観測データを取り扱う際には，本研究のような観測モデルを組み合わせた解析のアプローチが重要であると考えられる．

本研究においては，天然林が多く，農地および草地が多い景観構造において，日向夏生産に対する高い送粉サービスが期待できるという結果となった．農地の存在が送粉サービスに正の影響を与えるということは，綾町においては古くから有機農業が推進された結果として，農地が生態系へ与える悪影響が緩和されたことを示唆していると考えられる．さらに調査を継続し，有機農業が生態系サービスに良い影響を及ぼすことが確認できれば，綾町の有機農業生産品をブランド化することが可能になるだろう．ユネスコエコパークの目的は地域の生態系保全と持続可能な経済の両立であり，日向夏や有機農業生産品はまさにその理念の象徴となりうる．ユネスコエコパークにおいては移行地域の活用が課題となっており(比嘉 他, 2012)，このような取り組みを科学的にサポートするような研究が重要である．また，綾町では人工林を天

然林へ再生するプロジェクトが進行中であるが、本研究成果をもとに、ミツバチの送粉サービスを考慮して天然林再生を行う場所を選択するという考え方もあるだろう。このように生態系サービスを対象とした研究は、地域の生態系保全だけでなく、地域社会に対して貢献できる可能性がある。SDGsの達成に向けて、このような研究を推進していく必要があるだろう。

冒頭で述べたように、生態系サービスを定量的に評価することが喫緊の課題となっている。そのために様々な野外調査が実施されるが、特に動物を対象とする野外調査には測定の不確実性が大きくなってしまおうという問題が付きまとう。本研究においてもミツバチ発見率に大きな個人差が存在することが示唆されたが、個人差を考慮しないで解析を行うとミツバチによる送粉サービスと景観構造との関係を誤って理解してしまう危険性がある。観測者間における能力差のような野外調査における測定の不確実性は解析結果を攪乱する要因であるが、野外調査は比較的 low cost で多くの観測点における観測を可能にするという利点がある。そこで本研究のように生態系サービスの評価に関わるモデルと不確実性を伴う観測のモデルをそれぞれ設計して融合することにより、より適正に生態系サービスの評価を行うことが可能であると考えられる。

謝 辞

本研究は JSPS 科研費 (17H00806 および 18H02242) の助成を受けたものです。

参 考 文 献

- Biesmeijer, J. C., Roberts, S. P. M., Reemer, M., Ohlemüller, R., Edwards, M., Peeters, T., Schaffers, A. P., Potts, S. G., Kleukers, R., Thomas, C. D., Settele, J. and Kunin, W. E. (2006). Parallel declines in pollinators and insect-pollinated plants in Britain and the Netherlands, *Science*, **313**(5785), 351-354.
- Bosch, J., Kemp, W. P. and Trostle, G. E. (2006). Bee population returns and cherry yields in an orchard pollinated with *Osmia lignaria* (Hymenoptera: Megachilidae), *Journal of Economic Entomology*, **99**(2), 408-413.
- Cane, J. H. (2005). Pollination potential of the bee *Osmia aglaia* for cultivated red raspberries and blackberries (*Rubus: Rosaceae*), *HortScience*, **40**(6), 1705-1708.
- Carvalho, L. G., Seymour, C. L., Veldtman, R. and Nicolson, S. W. (2010). Pollination services decline with distance from natural habitat even in biodiversity-rich areas, *Journal of Applied Ecology*, **47**(4), 810-820.
- Carvalho, L. G., Veldtman, R., Shenkute, A. G., Tesfay, G. B., Pirk, C. W. W., Donaldson, J. S. and Nicolson, S. W. (2011). Natural and within-farmland biodiversity enhances crop productivity, *Ecology Letters*, **14**(3), 251-259.
- Chauzat, M.-P., Carpentier, P., Martel, A.-C., Bougeard, S., Cougoule, N., Porta, P., Lachaize, J., Madec, F., Aubert, M. and Faucon, J.-P. (2009). Influence of pesticide residues on honey bee (*Hymenoptera: Apidae*) colony health in France, *Environmental Entomology*, **38**(3), 514-523.
- Eckhart, V. M. (1991). The effects of floral display on pollinator visitation vary among populations of *Phacelia linearis* (Hydrophyllaceae), *Evolutionary Ecology*, **5**(4), 370-384.
- Fujiwara, A. and Washitani, I. (2017). Dependence of Asian honeybee on deciduous woody plants for pollen resource during spring to mid-summer in northern Japan, *Entomological Science*, **20**(1), 96-99.
- Gallai, N., Salles, J.-M., Settele, J. and Vaissière, B. E. (2009). Economic valuation of the vulnerability of world agriculture confronted with pollinator decline, *Ecological Economics*, **68**(3), 810-821.
- Goulson, D., Lye, G. and Darvill, B. (2008). Decline and conservation of bumble bees, *Annual Review of Entomology*, **53**(1), 191-208.

- Henry, M., Béguin, M., Requier, F., Rollin, O., Odoux, J.-F., Aupinel, P., Aptel, J., Tchamitchian, S. and Decourtye, A. (2012). A common pesticide decreases foraging success and survival in honey bees, *Science*, **336**(6079), 348-350.
- 比嘉基紀, 若松伸彦, 池田史枝 (2012). ユネスコエコパーク (生物圏保存地域) の世界での活用事例 (<特集 2> ユネスコ MAB (人間と生物圏) 計画—日本発 ユネスコエコパーク制度の構築に向けて), 日本生態学会誌, **62**(3), 365-373.
- Holzschuh, A., Dudenhöffer, J.-H. and Tschardtke, T. (2012). Landscapes with wild bee habitats enhance pollination, fruit set and yield of sweet cherry, *Biological Conservation*, **153**, 101-107.
- IPBES (2017). 花粉媒介者, 花粉媒介及び食料生産に関するアセスメントレポート政策決定者向け要約 (抄訳), p.8, <https://pub.iges.or.jp/pub/ipbes-pollination-keymessages-jp>.
- Isaacs, R. and Kirk, A. K. (2010). Pollination services provided to small and large highbush blueberry fields by wild and managed bees, *Journal of Applied Ecology*, **47**(4), 841-849.
- Kearns, C. A., Inouye, D. W. and Waser, N. M. (1998). Endangered mutualisms: The conservation of plant-pollinator interactions, *Annual Review of Ecology and Systematics*, **29**(1), 83-112.
- Kery, M. and Schaub, M. (2012). *Bayesian Population Analysis Using WinBUGS*, Academic Press, Boston.
- Kevan, P. G., Greco, C. F. and Belaussoff, S. (1997). Log-normality of biodiversity and abundance in diagnosis and measuring of ecosystemic health: Pesticide stress on pollinators on blueberry heaths, *Journal of Applied Ecology*, **34**(5), 1122-1136.
- Klein, A.-M., Vaissière, B. E., Cane, J. H., Steffan-Dewenter, I., Cunningham, S. A., Kremen, C. and Tschardtke, T. (2007). Importance of pollinators in changing landscapes for world crops, *Proceedings of the Royal Society of London B: Biological Sciences*, **274**(1608), 303-313.
- 小沼明弘, 大久保悟 (2015). 日本における送粉サービスの価値評価, 日本生態学会誌, **65**(3), 217-226.
- Kremen, C., Williams, N. M. and Thorp, R. W. (2002). Crop pollination from native bees at risk from agricultural intensification, *Proceedings of the National Academy of Sciences*, **99**(26), 16812-16816.
- Millennium Ecosystem Assessment (2005). *Ecosystems and Human Well-being: Synthesis*, Island Press, Washington, D.C.
- Ollerton, J., Winfree, R. and Tarrant, S. (2011). How many flowering plants are pollinated by animals?, *Oikos*, **120**(3), 321-326.
- Patiny, S., Rasmont, P. and Michez, D. (2009). A survey and review of the status of wild bees in the West-Palaeartic region, *Apidologie*, **40**(3), 313-331.
- Pettis, J. S. and Delaplane, K. S. (2010). Coordinated responses to honey bee decline in the USA, *Apidologie*, **41**(3), 256-263.
- Potts, S. G., Biesmeijer, J. C., Kremen, C., Neumann, P., Schweiger, O. and Kunin, W. E. (2010). Global pollinator declines: Trends, impacts and drivers, *Trends in Ecology & Evolution*, **25**(6), 345-353.
- Ricketts, T. H. (2004). Tropical forest fragments enhance pollinator activity in nearby coffee crops, *Conservation Biology*, **18**(5), 1262-1271.
- Ricketts, T. H., Regetz, J., Steffan-Dewenter, I., Cunningham, S. A., Kremen, C., Bogdanski, A., Gemmill-Herren, B., Greenleaf, S. S., Klein, A. M., Mayfield, M. M., Morandin, L. A., Ochieng', A. and Viana, B. F. (2008). Landscape effects on crop pollination services: Are there general patterns?, *Ecology Letters*, **11**(5), 499-515.
- 朱宮丈晴, 小此木宏明, 河野耕三, 石田達也, 相馬美佐子 (2013). 照葉樹林生態系を地域とともに守る: 宮崎県綾町での取り組みから, 保全生態学研究, **18**(2), 225-238.
- Taki, H., Okabe, K., Yamaura, Y., Matsuura, T., Sueyoshi, M., Makino, S. and Maeto, K. (2010). Effects of landscape metrics on *Apis* and non-*Apis* pollinators and seed set in common buckwheat, *Basic and Applied Ecology*, **11**(7), 594-602.
- Taki, H., Yamaura, Y., Okabe, K. and Maeto, K. (2011). Plantation vs. natural forest: Matrix quality

determines pollinator abundance in crop fields, *Scientific Reports*, **1**, 132.

Wood, S. L. and DeClerck, F. (2015). Ecosystems and human well-being in the sustainable development goals, *Frontiers in Ecology and the Environment*, **13**(3), 123-123.

Yumura, T., Mitsuda, Y., Iwamoto, M., Hirata, R. and Ito, S. (2016). Evaluation of the relationship between abundance of pollinators and landscape structure in Hyuganatsu (*Citrus tamurana*) Orchards in Aya Town, Miyazaki Prefecture, *Journal of Forest Planning*, **21**(2), 23-28.

The Relationship between Pollination Service by the Native Honeybee
(*Apis cerana*) and Landscape Structure
—A Case Study of Hyuganatsu (*Citrus tamurana*) in Aya Town,
Miyazaki Prefecture—

Yasushi Mitsuda¹, Takahiro Yumura^{1,2}, Ryoko Hirata¹ and Satoshi Ito¹

¹Faculty of Agriculture, University of Miyazaki, Miyazaki

²Now at Kyushu Branch, Regional Environmental Planning Inc., Fukuoka

In this study, we examined the relationship between pollination of hyuganatsu (*Citrus tamurana*) (as an ecosystem service) by native honeybees (*Apis cerana*) and the landscape structure in Aya Town, Miyazaki Prefecture. A total of 24 hyuganatsu trees were selected in 16 orchards, and the number of honeybees visiting each tree was counted. A land use map of the study area was developed by photo interruption on the orthophoto. A stochastic model for predicting honeybee visits was developed, using the area of natural forest and the total area of agricultural field and grassland as explanatory variables. The difference in bee-finding ability among observers was also modelled as an observation model. We created an ecological model combining the relationship between the number of honeybee visits, regarded as an indicator of pollination service, landscape structure, and the observation model were combined, and estimated the parameters of these model by Bayesian inference. The estimated parameters suggested that both the area of natural forest and total area of agricultural field and grassland positively affected pollination services by native honeybees.

関数データに基づく統計的モデリング

松井 秀俊^{1,2}

(受付 2018 年 7 月 11 日；改訂 12 月 20 日；採択 12 月 21 日)

要 旨

農業をはじめとする多くの分野では、1つの個体が時間や位置の変化に伴い繰り返して計測されたデータが多く用いられる。本論文では、このような形式で与えられたデータを分析するための方法の1つである、関数データ解析に基づくさまざまな分析手法を紹介する。関数データ解析では、経時測定データを関数化処理し、得られた関数データ集合を対象として分析を行う。特にここでは、関数データ解析の枠組みで回帰分析、時系列解析、空間データ解析を行うための方法を紹介する。また、各手法に対して適用例を示し、どのようなデータに対して、どのような結果が得られるかについて説明する。

キーワード：回帰分析、関数データ解析、空間データ解析、経時測定データ、時系列解析。

1. はじめに

データ科学に基づく意思決定や価値創造が、さまざまな分野で行われるようになってきた。農業分野もその例に漏れず、ICT化による栽培データ管理によって、作物の増産に繋げる動きが強まっている。近年の計測・測定機器の発展に伴い、大量のデータが比較的安価で容易に取得できるようになったことも、この流れを後押ししている。測定されるデータの形式は多岐にわたるが、その中でも多いのが、1つの個体が時間や位置などの変化に応じて繰り返して測定される形式のデータである。例として、日ごとの気温や日照時間、深さに対する土壌成分、波長ごとのスペクトル強度など、さまざまなものが挙げられる。本論文ではこのような形式のデータを経時測定データとよぶ。

農業においては、作物の成長具合は、生育期間中の気温や日照時間といった環境要因が大きく関わっているとされ、これらの関係性を統計モデルによって明らかにすることができれば、収量増産のための栽培管理などに役立つと考えられる。環境要因に関するデータとしては、例えば気象庁のウェブサイト (<http://www.jma.go.jp/jma/index.html>) からダウンロードできる。これらのデータの多くは、各観測地点において経時測定データとして与えられている。このような形式のデータを直接多変量データとして扱い、回帰分析などの統計手法を適用することがまず考えられるが、次のような問題点が生じる。1点目に、観測データは一般的にはノイズが混入されており、経時測定データのばらつきが大きい場合は本質構造を捉えにくくなる。2点目として、観測時点数の増加はそのままデータの次元の増加につながるため、推定量が不安定になる可能性がある。観測時点数が多いデータほどより多くの情報をモデルに取り込むことが

¹ 滋賀大学 データサイエンス学部：〒522-8522 滋賀県彦根市馬場1丁目1-1

² 科学技術振興機構さきがけ：〒332-0012 埼玉県川口市本町4-1-8

できるが、その分解が困難になってしまう恐れがある。また、計測機器の故障などにより、本来計測されるはずの時点で計測が行われない(欠測する)状況が考えられる。あるいは、データの取得環境によっては、個体ごとに観測時点数が不均一であったり、観測時点が不揃いである状況もあり得る。このような場合、欠損値補間(高井 他, 2016)を適用する方法も考えられるが、特に観測時点が不揃いである場合は、一般的には古典的な多変量解析手法を直接適用することは困難となる。これが3点目の問題である。

このような形式のデータに対して、個々のデータを関数化し、得られた関数をデータとして扱う方法が Ramsay and Silverman (2005)によって確立され、関数データ解析(Functional Data Analysis; FDA)とよばれている。関数データ解析では、経時測定データを本論文の2節で述べる方法などを用いて関数化処理することで、観測ノイズを除去し、さらに、観測時点数の増大により引き起こされるデータの高次元化を抑えることができる。また、データを関数化することにより、個体ごとに観測時点や観測時点数が異なっていたとしても、連続時間上でデータを表現できるため、容易に分析が行えるようになる。これにより、上記の3点の問題点が解決される。関数データ解析について詳しく書かれた書籍としては、Ramsay and Silverman (2005)の他に、Ferraty and Vieu (2006)、Horváth and Kokoszka (2012)で全般的な方法についての紹介がある。Hsing and Eubank (2015)は、関数データ解析の理論を詳しく述べている。Shi and Choi (2011)は、ベイズアプローチの観点から関数データ解析におけるモデル推定法を紹介している。Ramsay et al. (2009)、Kokoszka and Reimherr (2017)では、関数データ解析に基づく手法と、それらを実行するためのRなどのプログラムがまとめられている。Ramsay and Silverman (2002)は関数データ解析の応用事例を幅広く紹介している。関数データ解析に関するレビュー論文も数多く報告されており、例えば Cuevas (2014)、Morris (2015)、Shang (2014)、Jacques and Preda (2014)、Wang et al. (2015)、Reiss et al. (2017)が挙げられる。Ullah and Finch (2013)では、さまざまな分野への関数データ解析の応用に関する文献を紹介しているが、医学への応用が多く、農学への応用は僅かであることが分かる。さらに、関数データ解析を行うためのRパッケージも数多く実装されており、その一覧が Scheipl (2018)にまとめられている。

関数データ解析では、スカラーデータに対する統計的分析手法を関数データの枠組みへ拡張したものが数多く提案されている。本論文では、経時測定データを関数データとして扱い、分析するための方法をいくつか紹介する。紹介する手法は農業データ分析に特化したものではなく、汎用的に用いることができるものであるが、農業データへの応用が有効であると考えられるものをまとめた。まず、説明変数と目的変数のいずれか一方、あるいは両方が関数データとして与えられたとき、両者の関係を表す関数回帰モデルについて説明する。次に、関数データに基づく手法として提案された、各個体が関数データとして与えられた時系列データの予測に用いられる関数時系列解析として、Hyndman and Ullah (2007)の方法を紹介する。さらに、複数地点で観測された関数データから、未観測地点の曲線を予測する関数空間データ解析として、Giraldo et al. (2011)の方法についても紹介する。いずれの手法に対しても、適用例を通して、どのようなデータに対して、どのような分析を行うことができるかについて説明する。

本論文の構成は次の通りである。2節では、関数データに基づく分析手法について紹介する前に、取得された経時測定データを関数データ化するための方法について紹介する。続く3節では、関数データに基づく回帰モデルである関数回帰モデルについて、モデルおよび推定方法の概略を紹介する。4節では、関数データの枠組みでの時系列解析の方法について紹介する。そして5節では、空間データ解析に基づき未観測地点の関数データを予測するための方法について解説する。3節から5節については、各手法の適用例についても紹介する。なお、4節と5節の適用例については、Kokoszka and Reimherr (2017)に掲載されているRのプログラムを参

考にした。最後に、6節でまとめと今後の展望について述べる。

2. データの関数化

本節では、1年間の日別平均気温や平均湿度といった経時測定データを、関数データとして表すための方法について述べる。後の節で紹介する、関数回帰分析をはじめとした関数データ解析の多くの方法に対しては、まず第1段階で経時測定データの関数化を行い、続く第2段階で関数データ集合に対して分析を行うという2段階法が用いられている。

いま、第*i*番目の個体が、第 α 時点 $t_{i\alpha}$ で観測値 $x_{i\alpha}$ を得たとする。ここで、個体番号は $i = 1, \dots, n$ 、時点番号は $\alpha = 1, \dots, n_i$ の値を取るものとする。時点 $t_{i\alpha}$ やその数 n_i は、個体ごとに異なってもよい。また、ここでは分かりやすさのために t を時間とみなして説明しているが、 t は位置や深さ、波長など、本来は連続的に変化するようなものであればよい。経時測定データの例として、図1左は、世界のいくつかの都市で観測された月別平均気温のデータを図示したものである。このようなデータを関数化するために、各個体に対して平滑化を適用する。すなわち、第*i*個体のデータ $x_{i\alpha}$ ($\alpha = 1, \dots, n_i$)は、次式のように関数 $x_i(t)$ が時点 $t_{i\alpha}$ において誤差 $e_{i\alpha}$ が加わって得られたものと仮定し、 $x_i(t)$ を推定する方法である。

$$x_{i\alpha} = x_i(t_{i\alpha}) + e_{i\alpha}.$$

関数 $x_i(t)$ を推定する方法の1つとして、 $x_i(t)$ は基底関数展開、すなわち、基底関数とよばれる既知の関数系 $\phi_1(t), \dots, \phi_M(t)$ の線形結合によって表されると仮定する。

$$(2.1) \quad x_i(t) = \sum_{m=1}^M w_{im} \phi_m(t) = \mathbf{w}_i^T \boldsymbol{\phi}(t).$$

ただし $\mathbf{w}_i = (w_{i1}, \dots, w_{iM})^T$ は係数パラメータで、これを推定することで関数データ $x_i(t)$ が得られる。また、 $\boldsymbol{\phi}(t) = (\phi_1(t), \dots, \phi_M(t))^T$ とする。基底関数の種類としては、さまざまなものが提案されている。汎用的に用いられているものとしては、*B*-スプライン (井元・小西, 1999; de Boor, 2001) や動径基底関数 (Bishop, 1995; 安道 他, 2001) などが挙げられる。また、1年間の日別平均気温のデータのように、周期性をもつデータに対してはフーリエ級数が、心電図のようにスパイク状の変動に重要な意味をもつようなデータにはウェーブレット (Donoho and Johnstone, 1994) が有効である。基底関数展開以外の方法としては、局所線形回帰などのノンパラメトリック回帰 (Härdle, 1990) に基づく平滑化によって関数データを得る方法も考えられている (Yao et al., 2003)。平滑化の詳細に関しては Green and Silverman (1994)、小西 (2010) などを、関数データ解析の枠組みでの関数化については Ramsay and Silverman (2005) の4章および5章、荒木・小西 (2004)、Araki et al. (2009b)、Fujii and Konishi (2006) などを参照されたい。図1左における各個体のデータに対して、基底関数展開に基づく平滑化を行うことで、図1右のような関数データ (曲線) が得られる。

関数データを表現するもう1つの方法として、関数主成分分析 (Besse and Ramsay, 1986; 茅野 他, 2006) を用いるものがある。これは、確率的に変動するランダム関数 $X(t)$ が、適当な仮定の下で次のように表されることを利用する。

$$(2.2) \quad X(t) = \mu(t) + \sum_{l=1}^{\infty} \xi_l v_l(t).$$

ただし、 $\mu(t)$ は平均関数とする。ランダム関数の厳密な定義や仮定の詳細については、Ferraty and Vieu (2006)、Horváth and Kokoszka (2012) などを参照されたい。ここで、 ξ_l ($l = 1, 2, \dots$)

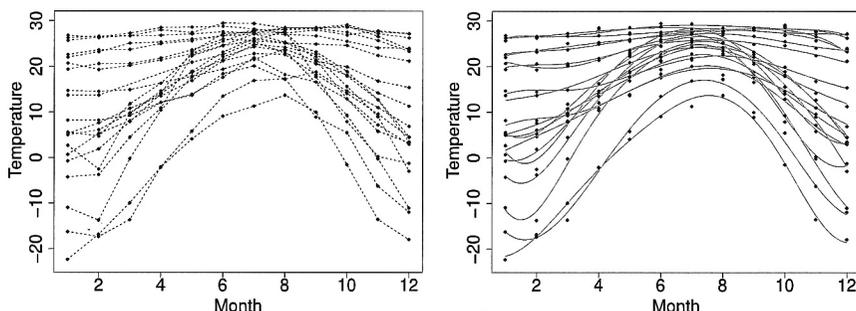


図 1. 左：世界の各都市の月別平均気温データ。破線でつながった点は同じ都市のデータを表す。右：左のデータに対して基底関数展開に基づく平滑化を適用し得られた関数データ。

は平均 0, 分散 λ_l ($\sum_{l=1}^{\infty} \lambda_l < \infty, \lambda_1 > \lambda_2 > \dots$) をもつ確率変数, $v_l(t)$ ($l = 1, 2, \dots$) は正規直交関数とする. この表現は Karhunen-Loève 展開としても知られている. 実際には, 上述した基底関数展開やノンパラメトリック回帰に基づく平滑化によって得られた関数データ $x_i(t)$ を, 次のように $v_l(t)$ の有限個の線形和で表現する.

$$(2.3) \quad x_i(t) = \bar{x}(t) + \sum_{l=1}^L \xi_{il} v_l(t).$$

ここで $\bar{x}(t) = \frac{1}{n} \sum_{i=1}^n x_i(t)$ とし, λ_l (ξ_{il} の分散) と $v_l(t)$ はそれぞれ標本共分散関数 $c(s, t) = \frac{1}{n-1} \sum_{i=1}^n (x_i(s) - \bar{x}(s))(x_i(t) - \bar{x}(t))$ の第 l 固有値, 固有関数として求められる. また, ξ_{il} は $x_i(t) - \bar{x}(t)$ と $v_l(t)$ の内積を計算することで得られる. (2.1) 式で用いた基底関数 $\phi_m(t)$ は汎用的に用いられる形状の関数であったのに対して, 固有関数 $v_l(t)$ は直交性の条件の下で, 与えられたデータを $l = 1, 2, \dots$ の順に最もよく表現する関数を与える. このことから, 一般的に固有関数の個数 L は (2.1) 式の基底関数の個数 M よりも少なくなり, 時点数の多い経時測定データ分析における次元縮小には特に有効である. 固有関数の個数 L の決定に関しては, 累積寄与率や, 交差検証法, AIC などを用いた方法が考えられている (Yao et al., 2005a; Crainiceanu et al., 2009). なお, 関数主成分分析は, 一度関数化したものを主成分によって表現するため, 関数主成分による関数データの近似精度は, 関数化の精度に依存する.

データによっては, 各個体が非常にまばらな時点でのみ観測されるような場合もある. 例えば図 2 左は, 図 1 左のデータの一部だけを人工的に取り出したもので, 各個体が限られた時点でのみ観測されていることがわかる. このような形式のデータはスパース経時測定データとよばれている. スパース経時測定データに対して, 前述のように個体ごとに独立して平滑化を行うと, 観測時点数が少ないために適切な曲線推定ができない可能性が高い. そこで, 個体それぞれではなく, 全個体の観測情報を使ってデータを関数化する方法が考えられている. 1 つは, データ集合 $\{x_{i\alpha}; i = 1, \dots, n, \alpha = 1, \dots, n_i\}$ を局所線形回帰などを用いて平滑化する方法である (Yao et al., 2005a, 2005b). また, 基底関数展開に基づく非線形混合効果モデルを適用することで, 全個体の観測情報を使って個体 1 つ 1 つを関数化することもできる (James et al., 2000; Rice and Wu, 2001; 松井 他, 2016). 非線形混合効果モデルを適用することで, 全ての個体の情報を共有し, 図 2 左のような, 1 つの個体がまばらな時点で観測されたデータであっても, 図 2 右のように, データが観測されている時間の範囲で各個体の曲線を当てはめることができる. Zhang and Wang (2016) は, 時点数がスパースな場合とそうでない場合の経時測定

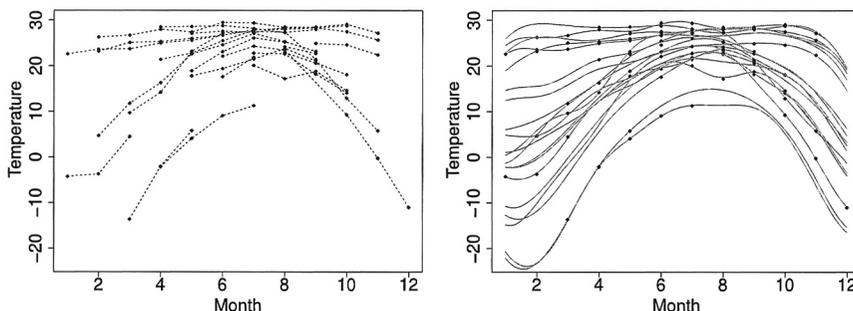


図 2. 左：図 1 左のデータの一部分のみを人工的に抜き取ったもの。右：左のデータに対して非線形混合効果モデルを適用し得られた関数データ。

データについて、局所線形回帰により関数化した際の、推定値の理論的な性質を統一的に示している。

3. 関数回帰モデル

説明変数、あるいは説明変数と目的変数の両方が関数データとして与えられたとき、これらの関係をモデル化する関数回帰モデルの概要を説明する。例えば、作物の生育期間中の気温や日照時間の推移を説明変数、作物の生育状況のデータを目的変数とし、これらの関係をモデル化することを考える。年に 1 度だけ収穫される作物の収量を考えたい場合は、目的変数は一般の回帰モデルと同様にスカラーとして与えられるだろう。一方で、トマトなどのように特定の期間で毎日のように収穫される作物に対しては、目的変数も経時測定データとなるため、関数データとして扱うこともできる。本節では、目的変数がスカラーおよび関数データで与えられた場合のモデルと、その推定方法についてそれぞれ解説する。

3.1 スカラー目的変数に対するモデル

いま、目的変数と説明変数に対して、 n 組の観測値 $\{(y_i, x_i(t)); t \in \mathcal{T} \in \mathbb{R}, i = 1, \dots, n\}$ を得たとする。ここで、 y_i はスカラーとして得られた目的変数、 $x_i(t)$ は関数として表された説明変数とする。また、 \mathcal{T} は関数 $x_i(t)$ の定義域とする。説明変数に関するデータについては、実際には離散時点で観測されたものであるため、2 節で述べた平滑化手法などを用いて関数化されたものとする。このとき、 y_i と $x_i(t)$ との関係を表す関数線形回帰モデルは、次で与えられる (荒木・小西, 2004; Ramsay and Silverman, 2005; Araki et al., 2009b)。

$$(3.1) \quad y_i = \beta_0 + \int_{\mathcal{T}} x_i(t) \beta_1(t) dt + \varepsilon_i.$$

ただし β_0 は切片、 $\beta_1(t)$ は回帰係数関数、 ε_i は平均 0、分散 σ^2 を持つ観測誤差とする。

関数線形回帰モデル (3.1) の推定問題は、未知パラメータである β_0 、 $\beta_1(t)$ 、 σ^2 を推定することに対応する。しかし、 $\beta_1(t)$ は無限次元のパラメータとなるため、これを直接推定することは困難である。これらのパラメータを推定するためのアプローチの 1 つとして、基底関数展開を用いる方法がある。まず、説明変数 $x_i(t)$ が、(2.1) のように基底関数展開によって表されるとする。ここで、 w_i は関数化の段階で推定されたものであり、ここでは既知とする。基底関数の代わりに固有関数を用いてもよい。そして、係数関数 $\beta_1(t)$ も同様に、次のように基底関

数展開によって表されるとする.

$$(3.2) \quad \beta_1(t) = \sum_{m=1}^M b_{1m} \phi_m(t) = \mathbf{b}_1^T \boldsymbol{\phi}(t).$$

ここで, $\mathbf{b}_1 = (b_{11}, \dots, b_{1M})^T$ は未知パラメータとする. また, (2.1)式の基底関数と(3.2)式の基底関数について, ここでは共通のものを用いているが, その個数や種類は異なってもよい. 基底関数展開(2.1), (3.2)の仮定により, 関数線形回帰モデル(3.1)は次のように変形される.

$$(3.3) \quad \begin{aligned} y_i &= \beta_0 + \int_{\mathcal{T}} \mathbf{w}_i^T \boldsymbol{\phi}(t) \boldsymbol{\phi}^T(t) \mathbf{b}_1 dt + \varepsilon_i \\ &= \beta_0 + \sum_{m=1}^M \mathbf{w}_i^T \Phi \mathbf{b}_1 + \varepsilon_i \\ &= \mathbf{z}_i^T \mathbf{b} + \varepsilon_i. \end{aligned}$$

ここで, $\Phi = \int_{\mathcal{T}} \boldsymbol{\phi}(t) \boldsymbol{\phi}^T(t) dt$ は基底関数ベクトル $\boldsymbol{\phi}(t)$ の各成分の積の積分を要素にもつ $M \times M$ 行列, $\mathbf{z}_i = (1, \mathbf{w}_i^T \Phi)^T$ はここでは既知のベクトル, そして $\mathbf{b} = (\beta_0, \mathbf{b}_1^T)^T$ は未知パラメータベクトルとする. 特に, $\boldsymbol{\phi}(t)$ がフーリエ級数や固有関数などのように正規直交基底の場合, Φ は単位行列となる. (3.3)式の形から, 最小 2 乗法や正則化法といった, 古典的な線形回帰モデルに対する推定法と同様の方法を用いて, パラメータを推定できる. 例えば, (3.3)式のモデルのパラメータ β を正則化最小 2 乗法により推定する場合は, 次の正則化誤差 2 乗和を最小にする \mathbf{b} を求める.

$$\sum_{i=1}^n (y_i - \mathbf{z}_i^T \mathbf{b})^2 + n \lambda \mathbf{b}^T \Omega \mathbf{b}.$$

ただし, $\lambda > 0$ は正則化の度合いを調整する正則化パラメータで, Ω は非負値定符号行列とする. 正則化誤差 2 乗和の最小化により, 次の推定量が得られる.

$$\hat{\mathbf{b}} = (Z^T Z + n \lambda \Omega)^{-1} Z^T \mathbf{y}.$$

ただし, $Z = (\mathbf{z}_1, \dots, \mathbf{z}_n)^T$, $\mathbf{y} = (y_1, \dots, y_n)^T$ とする. パラメータの推定値 $\hat{\mathbf{b}} = (\hat{\beta}_0, \hat{\mathbf{b}}_1^T)^T$ を用いて, 新たな観測 $x_0(t)$ が得られたときの目的変数の値を予測したり, 係数関数の推定値 $\hat{\beta}_1(t) = \hat{\mathbf{b}}_1^T \boldsymbol{\phi}(t)$ から, 説明変数の任意の時点 t における寄与を定量化できる. Ogden et al. (2002) は, 水田の俯瞰撮影により得られた画像のデータを関数データとして表現し, イネの倒伏度合いを表す指標を予測するために関数回帰モデルを適用した.

関数線形回帰モデルは, 説明変数が複数の場合へも容易に拡張できる. それだけでなく, 説明変数として関数データだけではなく, スカラーデータを説明変数に加えることもできる. すなわち, p_1 変数のスカラー説明変数 x_{ij} ($j = 1, \dots, p_1$) と, $p - p_1$ 変数の関数説明変数 $x_{ij}(t)$ ($j = p_1 + 1, \dots, p$) が与えられたとき, 関数線形回帰モデルは

$$(3.4) \quad y_i = \beta_0 + \sum_{j=1}^{p_1} x_{ij} \beta_j + \sum_{j=p_1+1}^p \int_{\mathcal{T}} x_{ij}(t) \beta_j(t) dt + \varepsilon_i$$

のように与えられる. なお, 関数データ $x_{ij}(t)$ の係数がスカラー β_j として与えられる, すなわち, 関数説明変数 $x_{ij}(t)$ の目的変数 y_i への寄与が t によらない場合, 対応する項は $\int x_{ij}(t) dt \cdot \beta_j$ となり, $x_{ij}(t)$ の \mathcal{T} 上の積分値 $\int x_{ij}(t) dt$ を説明変数としたものとして(3.4)式右辺第 2 項に含

めることができる。(3.4)式に対して、説明変数が1つのみのモデルに対するものと同様の仮定を置くことで、このモデルも(3.3)式と同様のベクトル表記を行うことができ、やはり一般の線形回帰モデルと同様に推定できる。

複数の説明変数の中から、目的変数に実際に寄与している変数の組み合わせを選びたい場合は、変数選択問題を考える必要がある。例えば、気温、日射量、大気中のCO₂濃度といった複数の環境要因の経時変化のうち、どの組み合わせが作物の収量に寄与しているかを調べるといった問題に対応する。変数選択の方法としては、古典的なステップワイズ法などを関数線形回帰モデルに適用することができる。一方で、スパース推定(Hastie et al., 2015; 川野 他, 2018)を利用することで、ある j について、説明変数 $x_j(t)$ に対する係数関数を $\hat{\beta}_j(t) \equiv 0$ と推定することができる。このとき、 $x_j(t)$ は目的変数に寄与していないものとみなすことができ、変数選択が行われたことになる。関数回帰モデルに対するスパース推定については、Matsui and Konishi (2011), Gertheiss et al. (2013), Lian (2013), Zhao et al. (2015)などを参照されたい。また、James et al. (2009), Lee and Park (2012)は、スパース推定の考えを利用して、係数関数 $\beta(t)$ の定義域 \mathcal{T} の全区間ではなく一部区間のみを0と推定することで、目的変数に寄与する説明変数の区間の選択を行っている。

線形モデルの枠組みを拡張した関数回帰モデルの1つとして、次の関数2次回帰モデルが提案されている(Yao and Müller, 2010; Fuchs et al., 2015; Usset et al., 2016)。

$$(3.5) \quad y_i = \beta_0 + \sum_{j=1}^p \int_{\mathcal{T}} x_{ij}(t) \beta_j(t) dt + \sum_{j,k} \iint_{\mathcal{T} \times \mathcal{T}} x_{ij}(s) x_{ik}(t) \beta_{jk}(s, t) ds dt + \varepsilon_i.$$

ここで、 $\beta_{jk}(s, t)$ は2変数の係数関数である。すなわち、2次回帰モデルは、説明変数 $x_{ij}(t)$ の2次の情報だけでなく、単一および2つの説明変数の、異なる2時点 s, t 間の交互作用まで考慮に入れたモデルとなっている。Wei et al. (2014)は、スカラーデータとして与えられた遺伝型と、経時測定データとして与えられた環境情報の相互作用(遺伝子-環境相互作用)を関数回帰モデルの枠組みで表現している。

本項で述べたモデルはさらに、目的変数がスカラーであることから、一般化線形モデル(McCullagh and Nelder, 1989)の枠組みへも容易に拡張できる(James, 2002)。例えば、気温や日照時間を説明変数としたとき、目的変数として年間の総収量といった連続量ではなく、生育に成功したか否か、あるいは病害の有無といった2値のデータの判別を行いたい場合や、生育に成功した株の割合といった比率のデータの推定および予測を行いたい場合は、関数ロジスティック回帰モデルを用いることが考えられる(Araki et al., 2009a; Matsui et al., 2011; Kayano et al., 2016)。あるいは、病害が発生した回数などのカウントデータを目的変数として扱いたい場合は、関数ポアソン回帰を用いることもできる。関数回帰モデルはこの他にも、適応モデル(James and Silverman, 2005)やニューラルネットワークモデル(Rossi et al., 2005)、加法モデル(Müller and Yao, 2008; Müller et al., 2013; Zhu et al., 2014)、一般加法モデル(Mclean et al., 2014)、ノンパラメトリックモデル(Ferraty et al., 2007; Rachdi and Vieu, 2007)といった非線形モデルなどへ拡張されている。さらに、Reiss and Ogden (2007, 2010)は主成分回帰の枠組みでの関数回帰モデルの推定法を提案している。このように、目的変数がスカラーの場合は、従来の回帰モデルと同様の拡張が行われている。

本項では、関数化を行った後に関数回帰モデルを推定するという、2段階法に基づく方法を紹介した。一方で、Crainiceanu et al. (2009)は、関数回帰モデルの推定において、2段階法ではなく、1段階で同時にモデルを推定する方法を提案している。そして、真の係数関数の値が大きいときや、説明変数に対応する経時測定データの誤差分散が大きいとき、1段階法とは異なり2段階法では係数関数の推定値にバイアスが生じることを示している。しかしCrainiceanu

et al. (2009)は同時に、2段階法はロバストな推定結果を与える上、1段階法と比べて計算コストが少なく、推定結果に大きな差がないことから、必ずしも2段階法を否定しておらず、両手法を適用し結果を比較するのがよいと述べている。

3.2 関数目的変数に対するモデル

3.1項では、目的変数 y_i は古典的な回帰モデルと同様にスカラーを想定した。一方で、トマトなどのように年間の一定期間を通して繰り返して収穫される作物に対して、気象情報と収量との関係を回帰モデルで表現する場合は、目的変数も説明変数と同様に関数データとして扱う方法が考えられる。本項では、説明変数および目的変数が共に関数データで与えられた場合のモデルについて紹介する。

目的変数と説明変数に関する n 組の観測値 $\{(y_i(t), x_i(s)); t \in \mathcal{T} \subset \mathbb{R}, s \in \mathcal{S} \subset \mathbb{R}, i = 1, \dots, n\}$ が得られたとする。なお、 $x_i(s)$, $y_i(t)$ はいずれも中心化されたもの、すなわち、それぞれの標本平均 $\frac{1}{n} \sum_{i=1}^n x_i(s)$, $\frac{1}{n} \sum_{i=1}^n y_i(t)$ を引いたものを扱うこととする。このとき、関数線形回帰モデルは次で与えられる (Ramsay and Dalzell, 1991; 下川 他, 2000; Matsui et al., 2009)。

$$(3.6) \quad y_i(t) = \int_{\mathcal{S}} x_i(s) \beta_1(s, t) ds + \varepsilon_i(t).$$

ここで、 $\beta_1(s, t)$ は説明変数 $x_i(s)$ の時点 s による、目的変数 $y_i(t)$ の時点 t への寄与を表す2変数の係数関数、 $\varepsilon_i(t)$ は平均0の誤差関数とする。

係数関数 $\beta_1(s, t)$ を推定する方法の1つとして、積分誤差2乗和 $\sum_{i=1}^n \int_{\mathcal{T}} \varepsilon_i^2(t) dt$ を最小にするものがある。推定量を求めるにあたり、次の仮定をおく。まず、 $x_i(s)$ および $y_i(t)$ が、次のように基底関数 $\phi(s) = (\phi_1(s), \dots, \phi_{M_x}(s))^T$, $\psi(t) = (\psi_1(t), \dots, \psi_{M_y}(t))^T$ に基づく線形結合によって表されるとする。

$$x_i(s) = \sum_{m=1}^{M_x} w_{im} \phi_m(s) = \mathbf{w}_i^T \phi(s), \quad y_i(t) = \sum_{m=1}^{M_y} v_{im} \psi_m(t) = \mathbf{v}_i^T \psi(t).$$

ここで、 $\mathbf{w}_i = (w_{i1}, \dots, w_{iM_x})^T$, $\mathbf{v}_i = (v_{i1}, \dots, v_{iM_y})^T$ であり、これらは2節で述べた方法であらかじめ関数化の段階で推定されているものとする。さらに、 $\beta_1(s, t)$ は2種類の基底関数 $\phi(s)$, $\psi(t)$ を用いて次のように表されると仮定する。

$$\beta_1(s, t) = \sum_{m=1}^{M_x} \sum_{m'=1}^{M_y} \phi_m(s) b_{mm'} \psi_{m'}(t) = \phi^T(s) B \psi(t).$$

ここで、 $B = (b_{mm'})_{mm'}$ はパラメータからなる行列とする。これらの仮定より、(3.6)式は次のように表される。

$$\begin{aligned} \mathbf{v}_i^T \psi(t) &= \int_{\mathcal{S}} \mathbf{w}_i^T \phi(s) \phi^T(s) B \psi(t) ds + \varepsilon_i(t) \\ &= \mathbf{w}_i^T \Phi B \psi(t) + \varepsilon_i(t). \end{aligned}$$

これを用いて、積分誤差2乗和は次で与えられる。

$$\begin{aligned} \sum_{i=1}^n \int_{\mathcal{T}} \varepsilon_i^2(t) dt &= \sum_{i=1}^n \int_{\mathcal{T}} (\mathbf{v}_i^T - \mathbf{w}_i^T \Phi B) \psi(t) \psi^T(t) (\mathbf{v}_i^T - \mathbf{w}_i^T \Phi B)^T dt \\ &= \text{tr}\{(V - ZB)\Psi(V - ZB)^T\}. \end{aligned}$$

ここで $W = (\mathbf{w}_1, \dots, \mathbf{w}_n)^T$, $V = (\mathbf{v}_1, \dots, \mathbf{v}_n)^T$, $\Psi = \int_{\mathcal{T}} \psi(t)\psi^T(t)dt$, $Z = W\Phi$ とする. これを B について最小化することで, 推定量

$$\text{vec}(\hat{B}) = (\Psi \otimes Z^T Z)^{-1} \text{vec}(Z^T V \Psi)$$

を得る. ただし, vec は行列の列ベクトルを全て 1 列に並べる作用素で, \otimes はクロネッカー積を表す. モデル (3.6) を推定するための他の方法として, 誤差関数 $\varepsilon_i(t)$ にガウス過程を仮定し, バイズアプローチに基づきパラメータを推定する方法 (Shi and Choi, 2011) も考えられる.

ここで, 説明変数および目的変数が時間の関数データである場合に (3.6) 式のモデルを適用する場合, 注意が必要である. (3.6) 式では, $s > t$ を満たす範囲の係数関数 $\beta_1(s, t)$ は, 時点 s の説明変数が, それよりも過去の時点 t の目的変数への寄与を表していることになり, 矛盾した関係性を表す可能性がある (ただし例外として, データに周期性を仮定している場合は, 例えば「12月の気温が翌年1月の収量に影響を及ぼす」というように, 次の周期の時点への寄与を表していると解釈できる). この問題点を解消するために, Malfait and Ramsay (2003) は, 時間の前後関係を考慮に入れた関数線形回帰モデルを構築し, その推定方法を提案した. この他にも, 同様の目的に対する研究が Harezlak et al. (2007), Şentürk and Müller (2010), Kim et al. (2011) によって行われている.

説明変数と目的変数間の関係性を調べるにあたって, 説明変数と同じ時点での目的変数への影響にのみ興味がある場合, すなわち $s = t$ での関係性のみを考える場合は, 回帰モデルは次で与えられる.

$$(3.7) \quad y_i(t) = \beta_0(t) + x_i(t)\beta_1(t) + \varepsilon_i(t).$$

ここで, $\beta_0(t)$ は説明変数に関する定数項に対応するベースライン関数, $\beta_1(t)$ は $x_i(t)$ の係数関数とする. このモデルは関数同時モデルとよばれているほか, Hastie and Tibshirani (1993) によって提案された変化係数モデルの特別な場合とみなすこともできる (Hoover et al., 1998). 関数同時モデルの推定方法については, Fan and Zhang (2008), Manrique et al. (2018) などを参照されたい.

3.3 適用例

3.1 節で紹介した関数線形回帰モデルを用いて, 作物のデータを分析した例を紹介する. ここでは, 兵庫県神戸市にある農場で 2012 年から 2014 年の間の 2 期で計測されたトマトの単位区画あたりの収量を目的変数, 農場で計測された環境情報の経時測定データを関数データの説明変数として扱い, これらの関係をモデル化する.

収量のデータは 10 月から翌 7 月まで日ごとに測定されているが, ここでは各週で総和をとったものを 1 個体, つまりスカラーデータとして扱った. 収量が計測された期間は 1 期目, 2 期目ともに 39 週分であった. つまり, 分析に用いる週の数(サンプルサイズ)は 78 である. トマトの収量については, その実が開花してから結実するまでのおよそ 60 日間の環境要因が関係していると考えられている. そこで, 週次の単位区画あたり収量 [kg] を目的変数, その週から遡った 60 日間の気温 [°C] および日射量 [W/m²] のデータを説明変数とした. 説明変数に対応するデータは毎分測定されているが, ここでは 1 日のデータの平均値をとり, さらに関数データ化したものを扱った. 図 3 は, 各週(個体)から遡った 60 日間の気温および日射量を, 関数データとして表現したものである. また, 図 4 に, 2 期分の単位区画あたりの週次収量を示す. 収量は春から初夏にかけて大きく伸びる時期があるなど, 年間の時期にも依存している. そこで, 上記の変数に加えて, 季節成分も説明変数に加えた. 具体的には, 第 i 週の初日の年間での通算日数 $d_i \in \{1, \dots, 365\}$ を, B -スプライン基底関数展開によって変換した

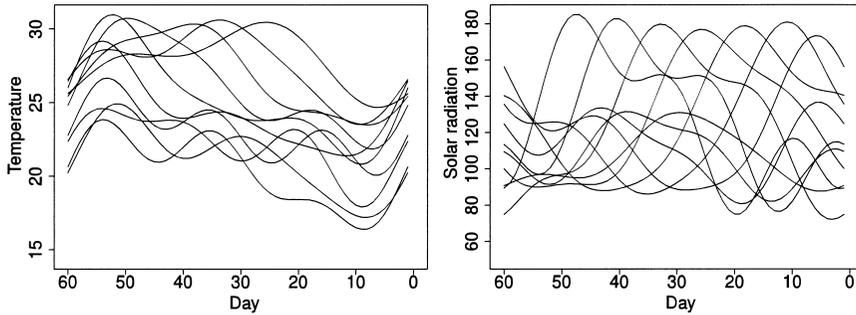


図 3. 週次収量を計測した週から遡った 60 日間の環境情報のデータ。左：気温，右：日射量を関数データ化したもの（一部を抜粋）。各グラフの左側ほど過去の日付を表す。

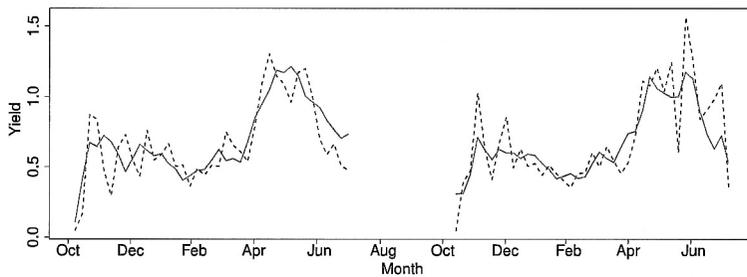


図 4. 2012 年 10 月から 2014 年 7 月の間に計測されたトマトの週次収量(破線)および、関数回帰モデルにより得られた収量の予測値(実線)。

$\psi(d_i) = (\psi_1(d_i), \dots, \psi_{M'}(d_i))^T$ を説明変数としてモデルに加える。基底関数の個数 M' は 30 と固定した。以上をまとめて、収量を推定するためのモデルを次で与える。

$$y_i = \sum_{j=1}^2 \int_{\mathcal{T}} x_{ij}(t) \beta_j(t) dt + \psi(d_i)^T \gamma + \varepsilon_i.$$

ここで、 γ は季節成分 $\psi(d_i)$ の係数、 $x_{i1}(t)$ 、 $x_{i2}(t)$ はそれぞれ気温、日射量の関数データ、 $\beta_1(t)$ 、 $\beta_2(t)$ は対応する係数関数である。2 種類の説明変数については、フーリエ級数による基底関数展開によって表現し、基底関数の個数は共に 7 個とした。また、関数回帰モデルを正則化最小 2 乗法で推定し、推定に伴う正則化パラメータの値を、交差検証法を用いて選択した。

図 4 に、関数回帰モデルによる収量 y_i の予測値を示している。収穫期間を通じて大まかな変動は捉えられていることが分かるが、各期の収穫初めの時期や、春から夏にかけての収量増産期の局所的な変動までは捉えられていない。図 5 左は、季節成分 $\psi(d)^T \gamma$ の推定値である。上述の通り、春から初夏にかけて収量が増えている様子が捉えられていることがわかる。また、図 5 中央と右に、それぞれ係数関数 $\beta_1(t)$ 、 $\beta_2(t)$ の推定曲線を示す。例えば、気温については収穫のおよそ 40 日前に比較的大きな重みがかかっていることから、この期間の気温が高いほど収量が増加すると解釈することができる。しかし、2 つの係数関数の信頼帯がほぼすべての区間で 0 を含んでいることから、今回の方法では、収量・環境要因間の関係性を適切に捉えているとは言い難い。果実の生育には環境情報だけでなく様々な要因が複雑に絡み合っていると

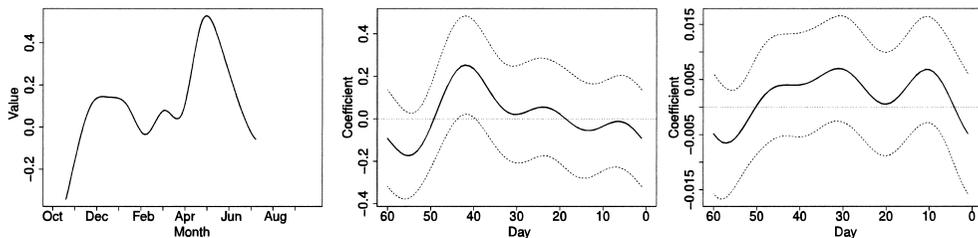


図 5. 左：季節成分 $\psi(t)^T\gamma$ の推定値。横軸は 10 月 1 日から翌 9 月 30 日までの期間に対応している。中央，右：それぞれ気温と日射量の係数関数。左側ほど過去の日付を表し，破線は 95% 信頼帯を表す。

考えられるため，それらのデータを含めた分析を行う必要がある。さらに，今回の分析では説明変数の時間区間を収穫週から遡った 60 日間と固定したが，この区間も不変の事実ではなく，実際には前後している可能性がある。そのため，例えば James et al. (2009) の方法に応用して，何日前までの環境要因が収量に影響を及ぼしているかを推定するといった方法も考えられるが，この点は今後の課題とする。

4. 関数時系列解析

時系列解析は，データが経時的に観測されたとき，その傾向を捉えたり，将来の時点の値を予測することが目的である（たとえば 北川, 2005）。主な時系列解析のための方法は，一連の時系列データに基づいて次の 1 点，または複数の点を予測するものであった。これに対して，時系列解析を関数データの枠組みへ発展させ，将来の点ではなく関数（曲線）を予測するための研究も行われている（Bosq, 2000; Horváth and Kokoszka, 2012; Hörmann and Kokoszka, 2012; 北川 他, 2016）。このような解析は関数時系列解析とよばれている。

関数時系列解析を，気温のデータを例に挙げて説明する。いま，ある地点の 20 年分の月別平均気温のデータが観測されたとする。1 年間のデータを 1 個体の関数データとして捉えると，20 個体の関数データが与えられていることになる。このデータ集合を用いて，次の年の (1 時点ではなく) 年間の気温曲線を予測することが，関数時系列解析の目的となる。

ここでは，Hyndman and Ullah (2007) の方法に基づく関数時系列の予測方法について紹介する。この方法は，アメリカにおける各年齢層の死亡率の毎年の推移のデータから，将来の死亡率を予測することを目的に提案されたものだが，一般的な関数時系列データに対しても適用できる。

4.1 固有関数展開による予測

いま，第 i 期の関数データ $x_i(t)$ からなる関数時系列データ集合 $\{x_i(t); i = 1, \dots, n\}$ が与えられたとする。このデータを用いて， h 期先のデータ $x_{n+h}(t)$ を予測することを考える。 $x_i(t)$ は，各 i に対して (2.2) 式のように，関数主成分分析に基づいて次のように表されると仮定する。

$$x_i(t) = \hat{\mu}(t) + \sum_{l=1}^L \hat{\xi}_{il} \hat{v}_l(t).$$

ここで，平均関数の推定値 $\hat{\mu}(t)$ としては，例えば標本平均 $\frac{1}{n} \sum_{i=1}^n x_i(t)$ などが考えられる。一方で，Hyndman and Ullah (2007) は外れ値にロバストな関数を得るために， $x_1(t), \dots, x_n(t)$ の

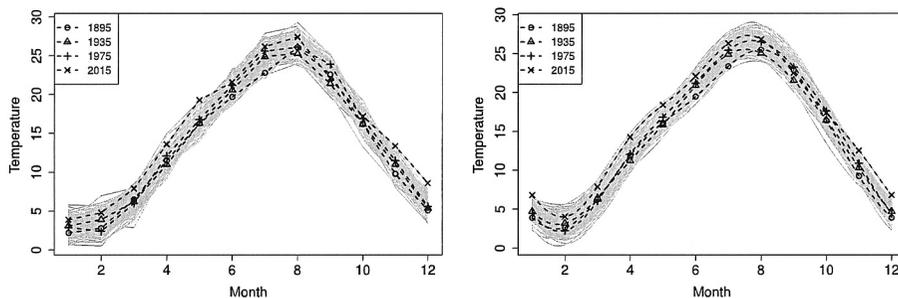


図 6. 左：1894 年から 2017 年までの、彦根市の月別平均気温データ。右：観測データを関数データ化したもの。

中央値曲線を推定値として用いている。続いて、各 l について、 $\hat{\xi}_{1l}, \dots, \hat{\xi}_{nl}$ にスカラーデータに対する時系列解析を適用し、第 n 期に対する h 期先の係数 $\xi_{n+h,l}$ の予測値 $\hat{\xi}_{n+h|n,l}$ を得る。ここで、固有関数 $\hat{v}_1(t), \dots, \hat{v}_L(t)$ は正規直交であることから、 $\hat{\xi}_{i1}, \dots, \hat{\xi}_{iL}$ は無相関となる。したがって、 $l = 1, \dots, L$ それぞれに対して、単変量の時系列解析手法を適用する。こうして得られた予測値 $\hat{\xi}_{n+h|n,1}, \dots, \hat{\xi}_{n+h|n,L}$ を用いて、 h 期先の関数データ $x_{n+h}(t)$ の予測値 $\hat{x}_{n+h|n}(t)$ を次で与える。

$$(4.1) \quad \hat{x}_{n+h|n}(t) = \hat{\mu}(t) + \sum_{l=1}^L \hat{\xi}_{n+h|n,l} \hat{v}_l(t).$$

この方法により、関数時系列データの傾向を、固有関数とその係数を使って表現できる。

4.2 適用例

滋賀県彦根市にある地上気象観測装置で観測された、1894 年から 2017 年までの月別平均気温のデータ（気象庁ウェブサイトより取得）に対して、関数時系列解析を適用する。図 6 左は、1894 年から 2017 年までの年間月別平均気温を図示したものである。このデータを用いて、2018 年の年間月別平均気温曲線の予測を試みる。

まず、R パッケージ `fda` を用いて、月別平均気温のデータを関数データ化した。ここでは各個体の曲線が周期性を持つことから、基底関数としてフーリエ級数を用いた。基底関数の個数は 7 個とした。図 6 右は、各年の気温のデータを関数化したものである。続いて、得られた関数データに対して関数主成分分析を適用し、平均関数 $\hat{\mu}(t)$ および固有関数 $\hat{v}_l(t)$ 、その係数 $\hat{\xi}_{il}$ を導出した。第 1 主成分から第 4 主成分の寄与率はそれぞれ 46.8%, 16.7%, 10.2%, 9.1% で、第 4 主成分までで累積寄与率が 80% を超えたため、ここでは第 4 主成分までを採用した。図 7 は、関数主成分分析により得られた平均関数の推定値および、第 4 主成分までの固有関数と各年の係数を並べたものである。固有関数から、第 1 主成分は年間を通じた気温の高さを表しており、係数の推移を見ると年の経過に伴い緩やかに増加していることが分かる。また、第 2 主成分は冬が寒く（暖かく）、夏が暑い（涼しい）年の特徴を表していると考えられる。

次に、図 7 右下に示す第 1～第 4 主成分の係数をそれぞれ時系列データとみなして、時系列解析を適用した。ここでは R の `forecast` パッケージを利用して自己回帰和分移動平均 (ARIMA) モデルを適用し、2018 年以降に対応する点の予測を行った。その結果を図 7 右下に併せて示している。この結果から、第 1 主成分である「年間を通じた気温の高さ」の係数の予測値としては若干の上昇傾向が見られ、第 2～第 4 主成分の係数の予測値も僅かな変化はあるが、いずれ

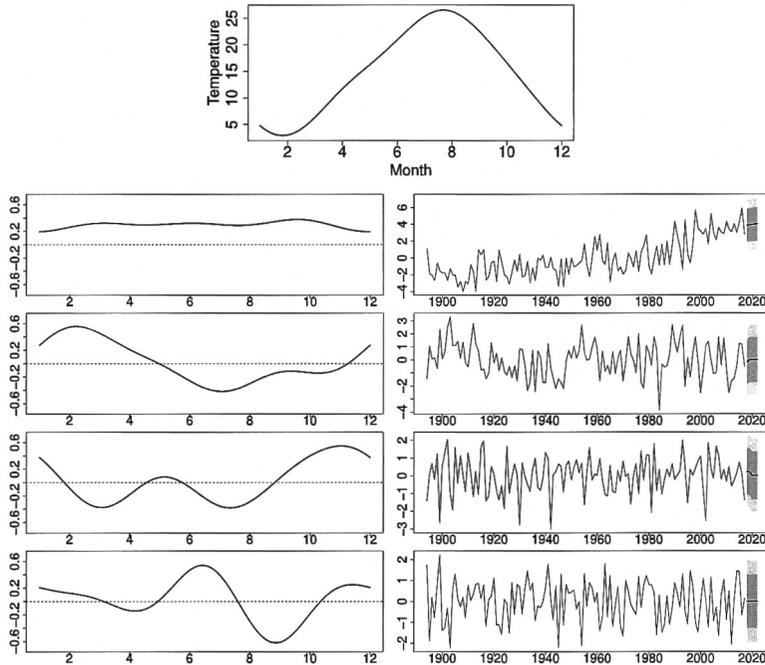


図 7. 気温の関数データ集合に対して関数主成分分析と関数時系列解析を適用し得られた結果. 上: 年間の平均関数. 左下: 上から第 1~第 4 主成分の固有関数. 横軸は時間(月)を表す. 右下: 各主成分に対応する各年の関数主成分. 横軸は個体(年)を表す. 右側にある太線および帯(濃色・淡色)は, それぞれ時系列解析によって得られた将来の係数の予測値および 80%, 95% 信頼帯を表す.

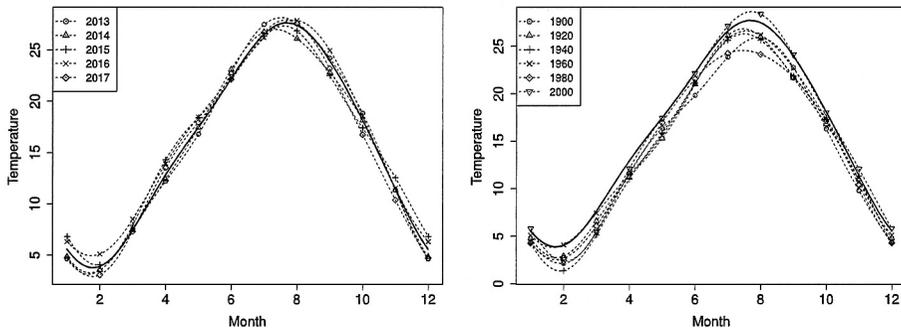


図 8. 左: 2013 年から 2017 年までの月別平均気温曲線(破線)と, 2018 年の予測気温曲線(実線). 右: 1900 年から 2000 年までの 20 年ごとの月別平均気温曲線(破線)と, 2020 年の予測気温曲線(実線).

も信頼帯を考慮に入れると有意な変化であるとは言い難い. 図 8 は, 過去の年の月別平均気温と, 予測した将来の年の月別平均気温を図示したものである. この結果から, 将来の予測値は直近の気温と比べても大きな変化は無いように見える.

4.3 その他の関数時系列解析手法

4.1 節で述べた方法以外にも、関数データの枠組みで自己回帰モデルを構築し、パラメータを推定することで関数時系列データを予測する方法などが考えられている。いま、関数時系列データ $x_i(t) (i \in \mathbb{Z})$ が与えられたとき、1 次の関数自己回帰モデル FAR(1) は、次で与えられる (Bosq, 2000)。

$$x_i(t) = \Phi(x_{i-1}(t)) + \varepsilon_i(t).$$

ここで、 Φ は関数の L^2 空間上への作用素、 $\varepsilon_i(t)$ は平均 0 の誤差関数とする。 Φ としては、実際は積分作用素 $\Phi(x)(t) = \int \varphi(t, s)x(s)ds$ が多く用いられる。すなわち、FAR(1) モデルは次で表される。

$$x_i(t) = \int \varphi(t, s)x_{i-1}(s)ds + \varepsilon_i(t).$$

この他に、関数データに対する ARCH モデルや ARMA モデルなども提案されている (Hörmann et al., 2013; van Delft and Eichler, 2018)。FAR モデルの一致性や収束率といった理論的な性質は、Horváth and Kokoszka (2012) にまとめられている。しかし、FAR モデルについては既存のパッケージでは実装が容易ではないこと (FAR(1) を推定するための R プログラムは Kokoszka and Reimherr (2017) に掲載されている) や、FAR(1) を一般化した p 次の関数自己回帰モデル FAR(p) に関する研究はあまり報告されていないという指摘がある (Aue et al., 2015)。

その一方で、4.1 節で述べた Hyndman and Ullah (2007) の方法は、FAR モデルに比べて定常性や特定の自己相関構造の仮定を必要とせずさまざまなデータに応用できる上、実装も非常に簡単である。しかし、理論的な最適性などの保証はされていない。Aue et al. (2015) は、Hyndman and Ullah (2007) の方法のように係数に対する単変量時系列解析を適用する方法では、ラグを考慮した主成分間の相関を考慮できず情報の損失が起こると指摘している。その上で、係数に対して単変量ではなく多変量時系列解析を適用する方法を提案し、予測量の一致性などの理論的性質を示している。

5. 空間関数データ解析

複数の地点で得られた観測値について、計測された位置情報を保持した形式のデータは空間データとよばれる。空間データ解析は、その位置情報を用いて、距離に応じたデータの関係性の分析や、未観測地点における予測を行うことなどが目的である。空間データ解析に関連する文献としては、例えば Gelfand et al. (2010)、間瀬 (2010)、堤・瀬谷 (2012)、瀬谷・堤 (2014) が挙げられる。また、時間と空間の変化に応じて計測された時空間データの解析法については、矢島・平野 (2012) による詳細な解説がある。

ここでは、複数の地点でそれぞれ経時測定データが観測された状況を考える。例えば、全国のさまざまな都市における年間の月別平均気温の推移や、農場内におけるさまざまな区域で生育されている同一品種の作物の草丈といった生育過程のデータなどが考えられる。このような、各地点で計測された経時測定データを関数データとして扱ったものは空間関数データとよばれ、空間関数データを対象とした分析は空間関数データ解析とよばれている。空間関数データを用いて、データが計測されていない地点の関数データの予測を行う問題などがこれに該当する。空間関数データ解析に関しては、Delicado et al. (2010) によって最近の研究がまとめられているほか、Menafoglio et al. (2013)、Aguilera-Morillo et al. (2017) などによる解析方法の研究がある。応用例として、Giraldo et al. (2011) は土壌の硬度を表す貫入抵抗性を、複数地点それぞれでさまざまな深度で計測したデータに対して適用し、任意の地点での、深度に対する

貫入抵抗性を表す曲線の予測を行っている。また、Hörmann and Kokoszka (2013)は空間関数データの推定量に対する漸近的な性質について述べている。

本節では、時点 t および d 次元空間上の地点 s で観測された空間ランダム関数 $\{X(s; t); s \in \mathcal{D} \subset \mathbb{R}^d, t \in \mathcal{T} \subset \mathbb{R}\}$ を考え、 \mathcal{D} 上の複数の地点それぞれで関数データが与えられたとき、新たな地点における関数データを予測する方法について紹介する。なお、ここでは、空間ランダム関数に2次定常性および等方性を仮定する。すなわち、任意のベクトル $h \in \mathcal{D}$ と、 s に依存しない関数 $\mu(t), \sigma^2(t), C(h; t, u)$ を用いて、

$$\begin{aligned} E(X(s; t)) &= \mu(t), \quad V(X(s; t)) = \sigma^2(t), \\ \text{Cov}(X(s; t), X(s+h; u)) &= C(h; t, u) \end{aligned}$$

と表すことができるとする。つまり、データの平均、分散関数は地点 s に依存せず、共分散関数は2地点間の距離 $r = \|h\|$ のみに依存するものとする。2次定常性の仮定により、 $\gamma(r; t) = \frac{1}{2}E[x(s+h; t) - x(s; t)]^2 = \frac{1}{2}V[x(s+h; t) - x(s; t)]$ も s に依存しない関数となる。これは時点 t における(セミ)バリオグラムとよばれている。空間関数データ解析についても、関数回帰モデルや関数時系列解析と同様に、スカラーデータに対する空間データ解析の枠組みへ帰着させることができる。

5.1 関数クリギング

いま、 n 個の観測地点 s_i ($i = 1, \dots, n$) において関数データ $x(s_i; t)$ を得たとする。このとき、新たな地点 s_0 における関数データの予測値 $\hat{x}(s_0; t)$ は、次のように $x(s_i; t)$ ($i = 1, \dots, n$) の線形結合

$$(5.1) \quad \hat{x}(s_0; t) = \sum_{i=1}^n \lambda_i x(s_i; t)$$

によって表されるとし、未知パラメータ λ_i ($i = 1, \dots, n$) を推定する問題を考える (Goulard and Voltz, 1993; Giraldo et al., 2011)。ここで、 $x_0(s; t)$ の2次定常性を保証するために、 $\sum_{i=1}^n \lambda_i = 1$ という条件が必要となる。この条件を踏まえて、パラメータ λ_i ($i = 1, \dots, n$) は、ラグランジュの未定乗数法を用いて、次の積分予測2乗誤差の制約付き最小化問題によって推定される。

$$(5.2) \quad \min_{\lambda_1, \dots, \lambda_n, \mu} \int_{\mathcal{T}} E[x(s_0; t) - \hat{x}(s_0; t)]^2 dt + \mu \left(\sum_{i=1}^n \lambda_i - 1 \right).$$

ただし μ はラグランジュの未定乗数とする。2次定常性および(5.1)式の仮定を利用することで、パラメータ $\lambda_1, \dots, \lambda_n, \mu$ の推定量は、次の方程式の解として与えられる。

$$\begin{pmatrix} \gamma(\|s_1 - s_1\|) & \cdots & \gamma(\|s_1 - s_n\|) & -1 \\ \vdots & \ddots & \vdots & \vdots \\ \gamma(\|s_n - s_1\|) & \cdots & \gamma(\|s_n - s_n\|) & -1 \\ -1 & \cdots & -1 & 0 \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \\ \mu \end{pmatrix} = \begin{pmatrix} \gamma(\|s_0 - s_1\|) \\ \vdots \\ \gamma(\|s_0 - s_n\|) \\ -1 \end{pmatrix}.$$

ここで、 $\gamma(r) = \int_{\mathcal{T}} \gamma(r; t) dt$ はトレースバリオグラムとよばれる。また、積分予測2乗誤差は、次のように表すことができる。

$$\sigma_{s_0}^2 = \sum_{i=1}^n \lambda_i \gamma(\|s_i - s_0\|) + \mu.$$

パラメータの推定量や積分予測 2 乗誤差の最小値の導出については、スカラーデータに対するクリギングとはほぼ同様の流れで行うことができる。

実際には、トレースバリオグラム $\gamma(r)$ は未知であるため、これを推定する必要がある。トレースバリオグラムは、観測された 2 地点間の距離が r となるようなデータの分散を求めることで得られるが、距離がちょうど r となるようなデータは存在しないか、あったとしてもごく僅かである。そこで、2 点間の距離が r に近い点の集合 $N_r = \{(s_i, s_j) : \|s_i - s_j\| \approx r\}$ を考え、次のようにトレースバリオグラムを推定する。

$$(5.3) \quad \hat{\gamma}(r) = \frac{1}{2|N_r|} \sum_{i,j \in N_r} \int_{\mathcal{T}} [x(s_i; t) - x(s_j; t)]^2 dt.$$

ただし、 $|N_r|$ は集合 N_r の元の数とする。また、(5.3) 式の積分については、関数データ $x(s_i; t)$ が (2.1) 式と同様に基底関数展開 $x(s_i; t) = \mathbf{w}_i^T \phi(t)$ によって表されると仮定することで、次のように表せ、計算できるようになる。

$$\int_{\mathcal{T}} [x(s_i; t) - x(s_j; t)]^2 dt = (\mathbf{w}_i - \mathbf{w}_j)^T \Phi (\mathbf{w}_i - \mathbf{w}_j).$$

2 節で述べた関数化の結果は、関数クリギングの予測に影響を与える（このことは、3 節の回帰分析や 4 節の時系列解析でも同様である）。より具体的には、トレースバリオグラムの推定値 (5.3) は、基底関数展開に基づく関数化における平滑化パラメータや基底関数の個数といった調整パラメータに依存する。Giraldo et al. (2011) は、関数クリギングの予測精度の観点から、関数データ $x(s_i; t)$ を得る際の調整パラメータを選択するための基準として、次の関数交差検証法を提案した。

$$(5.4) \quad FCV = \sum_{i=1}^n \sum_{\alpha=1}^{n_i} \{x(s_i; t_{i\alpha}) - \hat{x}(s_i; t_{i\alpha})^{(-i)}\}^2.$$

ここで $\hat{x}(s_i; t_{i\alpha})^{(-i)}$ は、 n 地点のデータのうち第 i 地点を抜き取った $n-1$ 地点のデータを用いて、第 i 地点のデータを (5.1) 式で予測したものである。(5.4) 式を最小にするような、関数化における調整パラメータの値を、パラメータの最適値とみなして選択する。これにより、関数クリギングの予測の観点から、経時測定データの関数化が行われることになる。

5.2 適用例

地上気象観測装置が設置されている日本国内の観測所のうち、各都道府県で 1 か所の 2017 年の月別平均気温のデータ（気象庁ウェブサイトより取得）を用いる。ただし、沖縄県のデータに関しては、他都道府県に比べて大きく距離が離れているためここでは除外し、46 地点のデータを扱った。観測地点は、図 9 左の地図上に点で示した場所にある。各地点における月別平均気温のデータは、図 9 右で与えられる。ここでは、東京以外の 45 地点のデータに対して関数クリギングを適用することで、東京の年間気温曲線を予測することを試みる。

データの関数化については、R パッケージ `fda` を使い、基底関数としてフーリエ級数を用いた。また、関数クリギングについては R パッケージ `geofd` を適用した。関数データ集合に対して、トレースバリオグラム (5.3) の推定値を求めた結果が図 10 左である。図 10 左中の黒点は、集合 N_r の中心が等間隔に 13 個となるように選択されたものである。この黒点に対して、指数型を仮定してトレースバリオグラムの当てはめを行った。日本の都道府県は位置が細長く分布しているため、距離が大きく離れた 2 地点のデータが少ない。このこともあり、2 点間の距離が離れたデータについては当てはまりがよくない。推定されたトレースバリオグラム $\hat{\gamma}(r)$ を

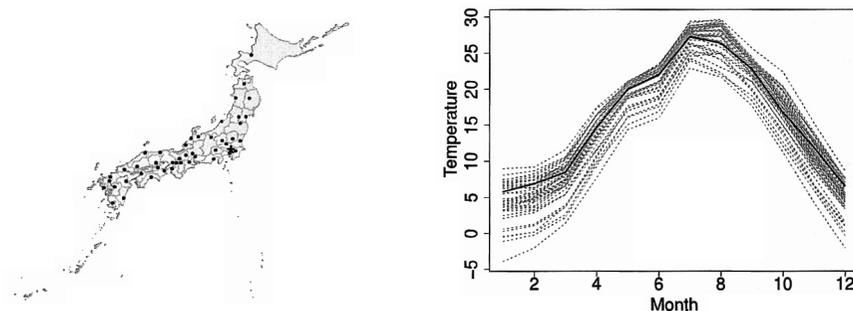


図 9. (左) データが観測された地上気象観測装置の位置。予測地点である東京は「+」で示している。(右) 各地点の月別平均気温のデータ。実線は東京のデータ。

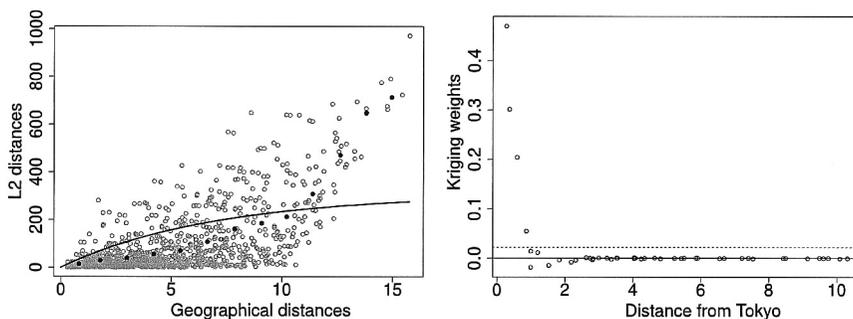


図 10. 関数クリギングの推定。左：各 2 地点間の距離に対するデータの分散(灰丸)および近傍でまとめられた点(黒丸)と、当てはめられたトレースバリオグラム(曲線)。右：東京からの距離(横軸)に対する、重みパラメータ λ_i の推定値(縦軸)。破線は全地点に対する重み λ_i を全て等しくした(1/45 とした)もの。

用いて得られた、(5.1)式の重みパラメータ λ_i の推定値を図 10 右に示す。この図より、東京から近い地点ではより強い重みが与えられている一方で、そこからやや離れた地点の重みはほぼ 0 であり、それらの情報はほぼ加味されていないことが分かる。(5.1)式に基づいて東京の気温 $x(s_0, t)$ を予測した結果を、図 11 に示す。破線で示した、45 地点の気温の平均関数を用いた予測に比べて、よい予測ができていることが分かる。なお、関数クリギングを適用するにあたり、(5.4)式の関数交差検証法を用いて基底関数の個数を選択した結果、3 個が選択された。

6. まとめと今後の展望

本論文では、農業データの分析に有用であると考えられる関数データ解析について、さまざまなデータ形式や分析目的に応じた手法を紹介した。ここでは特に、関数データに基づく回帰分析および時系列解析、空間データ解析について説明した。加えて、それぞれの手法に対して実データを含む適用例を通して、各手法がどのようなデータに対して、どのように分析できるのかについて議論した。

3 節では、気温といった環境情報を時間の関数データとみなし、これらを説明変数として関数回帰モデルを適用することで、トマトの収量の予測を試みた。しかし、作物の成長に関わる

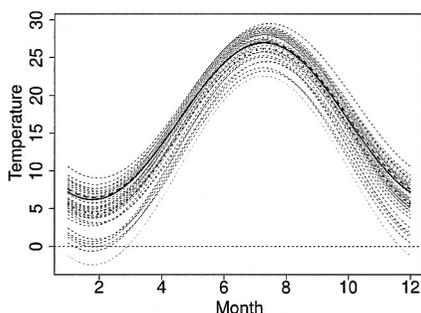


図 11. 関数クリギングによる東京の気温の予測結果. 灰破線は 45 地点の実測値, 黒破線は東京の気温の実測値を関数化したもの, 黒実線は関数クリギングによる東京の予測気温曲線, 黒点線は 45 地点の平均関数を表している.

環境要因は、気温や日射量以外のものも含めて複雑に絡み合っていると考えられるため、作物成長のメカニズムに応じたより適切なモデルを構築する必要があると考えられる。今回用いたモデルに関しても、収量に関連すると考えられる環境情報やその期間を主観で与えたが、これらについてもデータから推定、選択するための方法を適用することが望ましい。4 節では、関数時系列解析について説明し、日本の気温の年間推移の予測に用いた。今回は予測結果だけを示したが、より厳密には予測精度などの検証を行う必要がある。5 節では、日本全国の気温のデータに対して関数クリギングを適用した。しかし、日本の各都道府県の空間データに対して、関数クリギングの適用において用いた 2 次定常性の仮定は強いと考えられる。この仮定を緩めた方法の検討が必要である。

そのほかの課題として、2 節で述べた関数データ化の方法によるその後の関数データ解析への影響や、関数データ解析を見越した関数化の方法について言及された研究はあまり報告されていない。3.1 節の最後でも述べたように、関数回帰モデルに対して、2 段階推定では推定量にバイアスが生じてしまうが、その一方で 1 段階法では計算コストが問題になる。関数クリギングに対する関数交差検証法 (5.4) のように、関数データ化およびその後の分析までを包括したモデル評価を行うことでも、関数データ解析に適した関数データを得られると考えられるが、関数回帰モデルなどに対してはやはり計算コストが問題になることが予想される。また、関数主成分表現 (2.3) における、固有関数の個数の選択についても検討の余地があると考えられる。Yao et al. (2005a) や Crainiceanu et al. (2009) の方法では、固有関数の数を交差検証法や AIC、累積寄与率を使って決定しており、これによって良い推定結果が得られているようであるが、固有関数の個数による、関数回帰モデルなどへの推定精度の影響の理論的評価についても、今後調査すべき課題であると思われる。

本論文で述べた方法以外にも、関数データ解析についてはさまざまなものが提案されている。例えば、関数データに対する判別分析 (Hastie et al., 1995; James and Hastie, 2001)、クラスター分析 (James and Sugar, 2003; Misumi et al., 2019)、統計的仮説検定 (Fan and Lin, 1998; Hall and Van Keilegom, 2007) などの研究がある。さらに、関数データに対する微分を対象とした分析 (Ramsay, 1996) や、微分方程式の解の導出への応用 (Ramsay et al., 2007) など、関数データならではの研究もある。1 節でも述べたように、関数データ解析の農業分野への応用はまだ少ない。今後、農業分野の様々な経時測定データに対する関数データ解析の適用が期待される。

謝 辞

本研究を遂行するにあたり、データを提供していただいた東馬場農園(兵庫県神戸市)に感謝いたします。また、本論文の執筆にあたり、文献調査等に協力いただいた横浜市立大学の三角俊裕氏に感謝いたします。本稿の改訂にあたり、大変貴重なご指摘ならびに適切なご意見をいただいた査読者の方々へ御礼申し上げます。本研究は、若手研究(B)(課題番号:16K16020)および戦略的創造研究推進事業さきがけ(課題番号:JPMJPR16O6)の支援によって行われたものです。

参 考 文 献

- Aguilera-Morillo, M. C., Durbán, M. and Aguilera, A. M. (2017). Prediction of functional data with spatial dependence: A penalized approach, *Stochastic Environmental Research and Risk Assessment*, **31**, 7-22.
- 安道知寛, 井元清哉, 小西貞則 (2001). 動径基底関数ネットワークに基づく非線形回帰モデルとその推定, *応用統計学*, **30**, 19-35.
- 荒木由布子, 小西貞則 (2004). 動径基底関数展開に基づく関数回帰モデリング, *応用統計学*, **33**, 243-256.
- Araki, Y., Konishi, S., Kawano, S. and Matsui, H. (2009a). Functional logistic discrimination via regularized basis expansions, *Communications in Statistics. Theory and Methods*, **38**, 2944-2957.
- Araki, Y., Konishi, S., Kawano, S. and Matsui, H. (2009b). Functional regression modeling via regularized Gaussian basis expansions, *Annals of the Institute of Statistical Mathematics*, **61**, 811-833.
- Aue, A., Norinho, D. D. and Hörmann, S. (2015). On the prediction of stationary functional time series, *Journal of the American Statistical Association*, **110**, 378-392.
- Besse, P. and Ramsay, J. (1986). Principal components analysis of sampled functions, *Psychometrika*, **51**, 285-311.
- Bishop, C. M. (1995). *Neural Networks for Pattern Recognition*, Oxford University Press, New York.
- Bosq, D. (2000). *Linear Processes in Function Spaces: Theory and Applications*, Springer, New York.
- Crainiceanu, C. M., Staicu, A. M. and Di, C. Z. (2009). Generalized multilevel functional regression, *Journal of the American Statistical Association*, **104**, 1550-1561.
- Cuevas, A. (2014). A partial overview of the theory of statistics with functional data, *Journal of Statistical Planning and Inference*, **147**, 1-23.
- de Boor, C. (2001). *A Practical Guide to Splines*, Springer, New York.
- Delicado, P., Giraldo, R., Comas, C. and Mateu, J. (2010). Statistics for spatial functional data: Some recent contributions, *Environmetrics*, **21**, 224-239.
- Donoho, D. and Johnstone, J. (1994). Ideal spatial adaptation by wavelet shrinkage, *Biometrika*, **81**, 425-455.
- Fan, J. and Lin, S.-K. (1998). Test of significance when data are curves, *Journal of the American Statistical Association*, **93**, 1007-1021.
- Fan, J. and Zhang, W. (2008). Statistical methods with varying coefficient models, *Statistics and Its Interface*, **1**, 179-195.
- Ferraty, F. and Vieu, P. (2006). *Nonparametric Functional Data Analysis: Theory and Practice*, Springer, New York.
- Ferraty, F., Mas, A. and Vieu, P. (2007). Nonparametric regression on functional data: Inference and practical aspects, *Australian & New Zealand Journal of Statistics*, **49**, 267-286.
- Fuchs, K., Scheipl, F. and Greven, S. (2015). Penalized scalar-on-functions regression with interaction term, *Computational Statistics & Data Analysis*, **81**, 38-51.
- Fujii, T. and Konishi, S. (2006). Nonlinear regression modeling via regularized wavelets and smoothing parameter selection, *Journal of Multivariate Analysis*, **97**, 2023-2033.

- Gelfand, A. E., Diggle, P., Guttorp, P. and Fuentes, M. (2010). *Handbook of Spatial Statistics*, CRC Press, Boca Raton, Florida.
- Gertheiss, J., Maity, A. and Staicu, A.-M. (2013). Variable selection in generalized functional linear models, *Stat*, **2**, 86-101.
- Giraldo, R., Delicado, P. and Mateu, J. (2011). Ordinary kriging for function-valued spatial data, *Environmental and Ecological Statistics*, **18**, 411-426.
- Goulard, M. and Voltz, M. (1993). Geostatistical interpolation of curves: A case study in soil science, *Geostatistics Tróia '92*, 805-816, Springer, Dordrecht.
- Green, P. and Silverman, B. (1994). *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach*, Chapman & Hall/CRC, London.
- Hall, P. and Van Keilegom, I. (2007). Two-sample tests in functional data analysis starting from discrete data, *Statistica Sinica*, **17**, 1511-1531.
- Härdle, W. (1990). *Applied Nonparametric Regression*, Cambridge University Press, Cambridge.
- Harezlak, J., Coull, B., Laird, N., Magari, S. and Christiani, D. (2007). Penalized solutions to functional regression problems, *Computational Statistics & Data Analysis*, **51**, 4911-4925.
- Hastie, T. and Tibshirani, R. (1993). Varying-coefficient models, *Journal of the Royal Statistical Society Series B*, **55**, 757-796.
- Hastie, T., Buja, A. and Tibshirani, R. (1995). Penalized discriminant analysis, *The Annals of Statistics*, **23**, 73-102.
- Hastie, T., Tibshirani, R. and Wainwright, M. (2015). *Statistical Learning with Sparsity: The Lasso and Generalization*, Chapman & Hall/CRC, Boca Raton, Florida.
- Hoover, D., Rice, J., Wu, C. and Yang, L. (1998). Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data, *Biometrika*, **85**, 809-822.
- Hörmann, S. and Kokoszka, P. (2012). Functional time series, *Handbook of Statistics*, Vol. 30, 157-186, Elsevier, Amsterdam.
- Hörmann, S. and Kokoszka, P. (2013). Consistency of the mean and the principal components of spatially distributed functional data, *Bernoulli*, **19**, 1535-1558.
- Hörmann, S., Horváth, L. and Reeder, R. (2013). A functional version of the ARCH model, *Econometric Theory*, **29**, 267-288.
- Horváth, L. and Kokoszka, P. (2012). *Inference for Functional Data with Applications*, Springer, New York.
- Hsing, T. and Eubank, R. (2015). *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators*, Wiley, Chichester.
- Hyndman, R. J. and Ullah, M. S. (2007). Robust forecasting of mortality and fertility rates: A functional data approach, *Computational Statistics & Data Analysis*, **51**, 4942-4956.
- 井元清哉, 小西貞則 (1999). B-スプラインによる非線形回帰モデルと情報量規準, *統計数理*, **47**, 359-373.
- Jacques, J. and Preda, C. (2014). Functional data clustering: A survey, *Advances in Data Analysis and Classification*, **8**, 231-255.
- James, G. (2002). Generalized linear models with functional predictors, *Journal of the Royal Statistical Society Series B*, **64**, 411-432.
- James, G. and Hastie, T. (2001). Functional linear discriminant analysis for irregularly sampled curves, *Journal of the Royal Statistical Society Series B*, **63**, 533-550.
- James, G. and Silverman, B. (2005). Functional adaptive model estimation, *Journal of the American Statistical Association*, **100**, 565-576.
- James, G. and Sugar, C. (2003). Clustering for sparsely sampled functional data, *Journal of the American Statistical Association*, **98**, 397-408.
- James, G., Hastie, T. and Sugar, C. (2000). Principal component models for sparse functional data, *Biometrika*, **87**, 587-602.
- James, G., Wang, J. and Zhu, J. (2009). Functional linear regression that's interpretable, *The Annals*

- of *Statistics*, **37**, 2083-2108.
- 川野秀一, 松井秀俊, 廣瀬慧 (2018). 『スパース推定法による統計モデリング』, 共立出版, 東京.
- 茅野光範, 小西貞則, 平川英樹, 久原哲 (2006). 正則化基底展開法に基づく関数主成分分析とその応用, *応用統計学*, **35**, 1-17.
- Kayano, M., Matsui, H., Yamaguchi, R., Imoto, S. and Miyano, S. (2016). Gene set differential analysis of time course expression profiles via sparse estimation in functional logistic model with application to time-dependent biomarker detection, *Biostatistics*, **17**, 235-248.
- Kim, K., Sentürk, D. and Li, R. (2011). Recent history functional linear models for sparse longitudinal data, *Journal of Statistical Planning and Inference*, **141**, 1554-1566.
- 北川源四郎 (2005). 『時系列解析入門』, 岩波書店, 東京.
- 北川源四郎, 田中勝人, 川崎能典 (2016). 『時系列分析ハンドブック』, 朝倉書店, 東京.
- Kokoszka, P. and Reimherr, M. (2017). *Introduction to Functional Data Analysis*, CRC Press, Boca Raton, Florida.
- 小西貞則 (2010). 『多変量解析入門』, 岩波書店, 東京.
- Lee, E. R. and Park, B. U. (2012). Sparse estimation in functional linear regression, *Journal of Multivariate Analysis*, **105**, 1-17.
- Lian, H. (2013). Shrinkage estimation and selection for multiple functional regression, *Statistica Sinica*, **23**, 51-74.
- Malfait, N. and Ramsay, J. (2003). The historical functional linear model, *The Canadian Journal of Statistics*, **31**, 115-128.
- Manrique, T., Crambes, C. and Hilgert, N. (2018). Ridge regression for the functional concurrent model, *Electronic Journal of Statistics*, **12**, 985-1018.
- 間瀬茂 (2010). 『地球統計学とクリギング法・R と geoR によるデータ解析』, オーム社, 東京.
- Matsui, H. and Konishi, S. (2011). Variable selection for functional regression models via the L1 regularization, *Computational Statistics & Data Analysis*, **55**, 3304-3310.
- Matsui, H., Kawano, S. and Konishi, S. (2009). Regularized functional regression modeling for functional response and predictors, *Journal of Math-for-Industry*, **1**, 17-25.
- Matsui, H., Araki, T. and Konishi, S. (2011). Multiclass functional discriminant analysis and its application to gesture recognition, *Journal of Classification*, **28**, 227-243.
- 松井秀俊, 三角俊裕, 横溝孝明, 小西貞則 (2016). 非線形混合効果モデルに基づく関数データクラスタリング, *応用統計学*, **45**, 25-45.
- McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Model*, Chapman & Hall/CRC, London.
- McClean, M. W., Hooker, G., Staicu, A.-M., Scheipl, F. and Ruppert, D. R. (2014). Functional generalized additive models, *Journal of Computational and Graphical Statistics*, **23**, 249-269.
- Menafoglio, A., Secchi, P. and Dalla Rosa, M. (2013). A universal kriging predictor for spatially dependent functional data of a Hilbert space, *Electronic Journal of Statistics*, **7**, 2209-2240.
- Misumi, T., Matsui, H. and Konishi, S. (2019). Multivariate functional clustering and its application to typhoon data, *Behaviormetrika* (to appear).
- Morris, J. S. (2015). Functional regression, *Annual Review of Statistics and Its Application*, **2**, 321-359.
- Müller, H. and Yao, F. (2008). Functional additive models, *Journal of the American Statistical Association*, **103**, 1534-1544.
- Müller, H., Wu, Y. and Yao, F. (2013). Continuously additive models for nonlinear functional regression, *Biometrika*, **100**, 607-622.
- Ogden, R. T., Miller, C. E., Takezawa, K. and Ninomiya, S. (2002). Functional regression in crop lodging assessment with digital images, *Journal of Agricultural, Biological, and Environmental Statistics*, **7**, 389-402.
- Rachdi, M. and Vieu, P. (2007). Nonparametric regression for functional data: Automatic smoothing parameter selection, *Journal of Statistical Planning and Inference*, **137**, 2784-2801.
- Ramsay, J. (1996). Principal Differential Analysis, *Journal of the Royal Statistical Society Series B*,

- 58, 495-508.
- Ramsay, J. and Dalzell, C. (1991). Some tools for functional data analysis, *Journal of the Royal Statistical Society Series B*, **53**, 539-572.
- Ramsay, J. and Silverman, B. (2002). *Applied Functional Data Analysis: Methods and Case Studies*, Springer, New York.
- Ramsay, J. and Silverman, B. (2005). *Functional Data Analysis*, 2nd ed., Springer, New York.
- Ramsay, J. O., Hooker, G., Campbell, D. and Cao, J. (2007). Parameter estimation for differential equations: A generalized smoothing approach, *Journal of the Royal Statistical Society Series B*, **69**, 741-796.
- Ramsay, J. O., Hooker, G. and Graves, S. (2009). *Functional Data Analysis with R and MATLAB*, Springer, New York.
- Reiss, P. T. and Ogden, R. T. (2007). Functional principal component regression and functional partial least squares, *Journal of the American Statistical Association*, **102**, 984-996.
- Reiss, P. T. and Ogden, R. T. (2010). Functional generalized linear models with images as predictors, *Biometrics*, **66**, 61-69.
- Reiss, P. T., Goldsmith, J., Shang, H. L. and Ogden, R. T. (2017). Methods for scalar-on-function regression, *International Statistical Review*, **85**, 228-249.
- Rice, J. and Wu, C. (2001). Nonparametric mixed effects models for unequally sampled noisy curves, *Biometrics*, **57**, 253-259.
- Rossi, F., Delannay, N., Conan-Guez, B. and Verleysen, M. (2005). Representation of functional data in neural networks, *Neurocomputing*, **64**, 183-210.
- Scheipl, F. (2018). *CRAN Task View: Functional Data Analysis*, <https://cran.r-project.org/web/views/FunctionalData.html>.
- Şentürk, D. and Müller, H. (2010). Functional varying coefficient models for longitudinal data, *Journal of the American Statistical Association*, **105**, 1256-1264.
- 瀬谷創, 堤盛人 (2014). 『空間統計学：自然科学から人文・社会科学まで』, 朝倉書店, 東京.
- Shang, H. L. (2014). A survey of functional principal component analysis, *Advances in Statistical Analysis*, **98**, 121-142.
- Shi, J. Q. and Choi, T. (2011). *Gaussian Process Regression Analysis for Functional Data*, CRC Press, Boca Raton, Florida.
- 下川真由子, 水田正弘, 佐藤義治 (2000). 関数データ解析における関数回帰分析の拡張, *応用統計学*, **29**, 27-39.
- 高井啓二, 星野崇宏, 野間久史 (2016). 『欠測データの統計科学』, 岩波書店, 東京.
- 堤盛人, 瀬谷創 (2012). 応用空間統計学の二つの潮流：空間統計学と空間計量経済学, *統計数理*, **60**, 3-25.
- Ullah, S. and Finch, C. F. (2013). Applications of functional data analysis: A systematic review, *BMC Medical Research Methodology*, 13:43.
- Usset, J., Staicu, A. M. and Maity, A. (2016). Interaction models for functional regression, *Computational Statistics and Data Analysis*, **94**, 317-330.
- van Delft, A. and Eichler, M. (2018). Locally stationary functional time series, *Electronic Journal of Statistics*, **12**, 107-170.
- Wang, J.-L., Chiou, J.-M. and Mueller, H.-G. (2015). Review of functional data analysis, *arXiv preprint arXiv:1507.05135*.
- Wei, P., Tang, H. and Li, D. (2014). Functional logistic regression approach to detecting gene by longitudinal environmental exposure interaction in a case-control study, *Genetic Epidemiology*, **38**, 638-651.
- Yao, F. and Müller, H.-G. (2010). Functional quadratic regression, *Biometrika*, **97**, 49-64.
- Yao, F., Müller, H.-G., Clifford, A. J., Duiker, S. R., Follett, J., Lin, Y., Buchholz, B. A. and Vogel, J. S. (2003). Shrinkage estimation for functional principal component scores with application to the population kinetics of plasma folate, *Biometrics*, **59**, 676-685.

- Yao, F., Müller, H. and Wang, J. (2005a). Functional data analysis for sparse longitudinal data, *Journal of the American Statistical Association*, **100**, 577-590.
- Yao, F., Müller, H. and Wang, J. (2005b). Functional linear regression analysis for longitudinal data, *The Annals of Statistics*, **33**, 2873-2903.
- 矢島美寛, 平野敏弘 (2012). 時空間大規模データに対する統計的解析法, *統計数理*, **60**, 57-71.
- Zhang, X. and Wang, J. L. (2016). From sparse to dense functional data and beyond, *The Annals of Statistics*, **44**, 2281-2321.
- Zhao, Y., Chen, H. and Ogden, R. T. (2015). Wavelet-based weighted lasso and screening approaches in functional linear regression, *Journal of Computational and Graphical Statistics*, **24**, 655-675.
- Zhu, H., Yao, F. and Zhang, H. (2014). Structured functional additive regression in reproducing kernel Hilbert spaces, *Journal of the Royal Statistical Society Series B*, **76**, 581-603.

Statistical Modeling via Functional Data Analysis

Hidetoshi Matsui^{1,2}

¹Faculty of Data Science, Shiga University

²Japan Science and Technology Agency, PRESTO

The development of technologies for measurement devices enables us to obtain high-throughput data. In agriculture in particular, more and more repeated measured data with respect to time, position, depth, etc. for individual have been obtained. However, sometimes it is difficult to analyze such data for the following three reasons. First, they usually contain observational errors, making it difficult to reveal unknown structures. Second, repeated measurements lead to the high-dimensional data, to which it is hard to apply traditional statistical analyses. Third, the observed numbers and time points may differ for individual. This leads to difficulties in applying the traditional multivariate analysis. Functional data analysis is one of the most useful tools for resolving the aforementioned problems. In this approach, longitudinal data are expressed as a smooth function for each individual, and then information is drawn from the collection of functions. In this paper, we introduce several techniques for functional data analysis that would be useful for analyzing agricultural data. In particular, we describe regression analysis, time series analysis, and spatial data analysis for functional data. For each method, we provide an empirical example in which the technique is used to analyze real data.

農地と森林の生態系サービスの経済評価手法

柘植 隆宏[†]

(受付 2018 年 8 月 7 日; 改訂 12 月 25 日; 採択 12 月 27 日)

要 旨

農地や森林は、農作物や木材の供給以外にも、国土保全や水源涵養、気候の安定化、地球温暖化防止、生物多様性保全、レクリエーション機会の創出などの多様な役割を果たしている。これら農地や森林の生態系サービスの重要性を示すためには、その価値を貨幣単位で評価して可視化することが効果的である。しかし、生態系サービスの多くは市場で取引されることがなく、価格が存在しないため、価格に基づいてその価値を評価することができない。このため、その価値の評価には、環境評価手法と呼ばれる特別な手法が使われる。本稿では、農地や森林の生態系サービスの価値評価に適用可能な環境評価手法として、代替法、ヘドニック価格法、トラベルコスト法、CVM、コンジョイント分析を取り上げ、その経済理論と推定方法について解説を行う。

キーワード：環境評価手法、代替法、ヘドニック価格法、トラベルコスト法、仮想評価法(CVM)、コンジョイント分析。

1. はじめに

農地や森林は、農作物や木材の供給以外にも、国土保全や水源涵養、気候の安定化、地球温暖化防止、生物多様性保全、レクリエーション機会の創出などの多様な役割を果たしている(日本学術会議, 2001)。農地や森林が持つこれらの機能は、多面的機能と呼ばれており、近年は生態系サービスという表現も広く用いられている。生態系サービスとは、生態系が人間にもたらす恩恵のことであり、国連の呼びかけで 2001 年から 2005 年に行われた生態系に関する地球規模の総合評価であるミレニアム生態系評価(Millennium Ecosystem Assessment: MA)で提唱された概念である(Millennium Ecosystem Assessment, 2005)。

農地や森林の生態系サービスは、人間が安全で快適な生活を送る上で重要な役割を果たしている。これからの農林業政策では、農作物や木材の供給者としてだけでなく、生態系サービスを発揮する主体として農林業をとらえ、それらのサービスが適切に発揮されるために必要な対策を実施していくことが重要である。そのためには、農地や森林の生態系サービスの重要性を、受益者であり納税者でもある一般の人々に広く理解してもらう必要がある。農地や森林の生態系サービスの重要性を示すためには、その価値を貨幣単位で評価して可視化することが効果的である。誰もが理解しやすい貨幣を単位としてその価値を示すことで、多くの人々に重要性を理解してもらうことができると考えられる。

農産物や木材などは市場で取引されるため価格が存在する。したがって、それらの供給に関わるサービスについては、市場価格に基づいて価値を評価することが可能な場合がある。しか

[†] 甲南大学 経済学部：〒658-8501 兵庫県神戸市東灘区岡本 8-9-1

し、生態系サービスの多くは市場で取引されないため、価格が存在しない。したがって、価格に基づいてその価値を評価することができない。このため、その価値の評価には、主に環境経済学分野で研究が進められている環境評価手法と呼ばれる特別な手法が使われる(栗山 他, 2013)。

本稿では、農地や森林の生態系サービスの価値評価に適用可能な環境評価手法を紹介する。本稿の構成は以下の通りである。2節では、農地や森林の生態系サービスの価値とその評価手法について概観する。3節では、代表的な環境評価手法のうち、顕示選好法に分類される代替法、ヘドニック価格法、トラベルコスト法について解説する。4節では、同じく表明選好法に分類される仮想評価法(CVM)、コンジョイント分析について解説する。5節では、顕示選好法と表明選好法を統合したRP-SP 結合モデルについて解説する。最後にまとめと今後の課題を提示する。

2. 環境の価値と評価手法

2.1 農地と森林の価値

農地や森林が我々にもたらす便益の観点からそれらの価値を整理してみよう(栗山 他, 2013)。農地からは農産物を、森林からは木材や食料(きのこ・木の実など)をそれぞれ収穫し消費することができる。このように収穫物を消費するなどの形で、環境を直接的に利用することで得られる価値を直接的利用価値という。また、農地や森林では美しい景観を楽しむことができるほか、農地では農業体験、森林ではハイキングなどのレクリエーションを楽しむことができる。このように、その環境が存在することで間接的に得られる価値を間接的利用価値という。さらに、将来レクリエーションに利用する可能性があるから、あるいは将来そこから有用な遺伝資源が発見される可能性があるからといった理由で農地や森林を保全したいと考える人もいるだろう。このように、将来の利用可能性を維持することから得られる価値をオプション価値という。これらの価値は、いずれも利用することで得られるため利用価値と総称される。

一方、農地や森林には利用しなくても得られる価値も存在する。将来世代に貴重な環境を残したいと考える人は、農地や森林を子や孫の世代に引き継ぐことで満足を得ると考えられる。このように、環境を将来世代に残すことから得られる価値を遺産価値という。また、農地や森林が存在すること自体から満足を得る人もいるだろう。このように、貴重な環境が存在するという事実から得られる価値を存在価値という。遺産価値や存在価値は、利用しなくても得られる価値であるため非利用価値、または受動的利用価値と呼ばれる。

2.2 環境評価手法

経済学では、消費者がそれを手に入れることと引き換えに最大限支払ってもいいと考える金額である支払意志額(willingness to pay: WTP)で財やサービスの価値を評価する。

WTPに基づく代表的な環境評価手法の特徴をまとめたものが表1である。環境評価手法は、人々の行動に基づいて分析を行う顕示選好法と、人々の意見に基づいて分析を行う表明選好法に分類される(栗山, 1998; 栗山 他, 2013)。前者の代表的な手法には代替法、ヘドニック価格法、トラベルコスト法があり、後者の代表的な手法には仮想評価法(CVM)、コンジョイント分析がある。

顕示選好法は実際に人々が行った行動をもとに分析を行うため、人々の表明した意見に基づいて分析を行う表明選好法と比較してデータの信頼性が高い。しかしながら、人々の行動に基づいて分析を行う顕示選好法で評価できるのは利用価値だけである。非利用価値は人々の行動に反映されないため、顕示選好法では評価できない。非利用価値を評価するためには、人々の

表 1. 代表的な環境評価手法.

	顕示選好法		
	代替法	ヘドニック価格法	トラベルコスト法
評価に用いる情報	環境が提供するサービスと同等のサービスを人為的に提供するために必要となる費用から評価	環境が住宅価格に及ぼす影響から評価	レクリエーションに費やす費用から評価
農地や森林の生態系サービスの価値評価への適用可能性	農地や森林の多面的機能のうち、代替可能な財やサービスが存在するものの価値を評価可能	農地や森林の景観など、住宅価格に影響するものの価値を評価可能	農業体験、森林浴、ハイキングなど農地や森林を利用したレクリエーションの価値を評価可能
	表明選好法		
	仮想評価法 (CVM)	コンジョイント分析	
評価に用いる情報	アンケート調査で把握した意見から評価	アンケート調査で把握した意見から評価	
農地や森林の生態系サービスの価値評価への適用可能性	あらゆるサービスの価値を評価可能、非利用価値も評価可能	あらゆるサービスの価値を評価可能、非利用価値も評価可能、一度に複数のサービスを評価可能	

意見をもとに評価を行う表明選好法が必要となる。

以下では農地や森林の生態系サービスの価値評価への適用を想定して、代表的な環境評価手法の概要を説明する。

3. 顕示選好法の詳細

3.1 代替法

代替法は、環境が提供するサービスと同等のサービスを人為的に提供するために必要となる費用で環境の価値を評価するものである(栗山 他, 2013)。例えば、ある農地の保水機能の価値は、同等の保水機能を持つダムを建設するのに必要となる費用で評価する。

日本では農業・農村や森林の多面的機能の評価で代替法がしばしば用いられてきた。林野庁は全国の森林の多面的機能の価値を、農林水産省は全国ならびに中山間地域の農地の多面的機能の価値を、それぞれ代替法で評価している(林野庁, 1972; 農業・農村の公益的機能の評価検討チーム, 1998)。また、日本学術会議(2001)は、農林水産大臣の諮問に対する答申として、農業の多面的機能と森林の多面的機能の価値を代替法で評価した結果を発表している(表2)。

代替法は、比較的簡単な計算により分析が可能である。またシンプルな手法であるため直感的に理解しやすい。しかし、代替法にはいくつかの問題がある(栗山 他, 2013)。第一に、代替法を用いるためには、例えば、農地の保水機能がダム何個分に相当するかといった自然科学的な情報が必要であるが、そのような評価は自然科学的にも容易ではなく、信頼性の高いデータが得られない場合がある。第二に、代替法を用いるためには、環境が提供するサービスと全く同じサービスを提供する財やサービス(完全代替の関係にある財やサービス)の価格を用いる

表 2. 代替法で評価した農業の多面的機能と森林の多面的機能の価値.

農業の多面的機能	評価額	代替財
洪水防止	3 兆 4,988 億円/年	治水ダムを代替財として評価
水源涵養（うち河川流況安定機能）	1 兆 4,633 億円/年	利水ダムを代替財として評価
土壌浸食防止	3,318 億円/年	砂防ダムを代替財として評価
水源涵養（うち地下水涵養機能）	537 億円/年	地下水と上水道との利用上の差額によって評価
土砂崩壊防止	4,782 億円/年	土砂崩壊の被害抑止額によって評価
有機性廃棄物分解	123 億円/年	最終処分経費を代替財として評価
気候緩和	87 億円/年	冷房料金の節減額により評価
保健休養・やすらぎ	2 兆 3,758 億円/年	家計支出額の推定により評価
森林の多面的機能	評価額	代替財
二酸化炭素吸収	1 兆 2,391 億円/年	火力発電所の二酸化炭素回収装置を代替財として評価
化石燃料代替	2,261 億円/年	木造住宅が、すべて RC 造・鉄骨プレハブで建設された場合に増加する炭素放出量を上記二酸化炭素回収コストで評価
表面侵食防止	28 兆 2,565 億円/年	砂防ダムを代替財として評価
表層崩壊防止	8 兆 4,421 億円/年	土留工を代替財として評価
洪水緩和	6 兆 4,686 億円/年	治水ダムを代替財として評価
水資源貯留	8 兆 7,407 億円/年	利水ダムを代替財として評価
水質浄化	14 兆 6,361 億円/年	雨水利用施設及び水道施設を代替財として評価
保健・レクリエーション	2 兆 2,546 億円/年	我が国の自然風景を觀賞することを目的とした旅行費用により評価

出典：日本学術会議（2001）、三菱総合研究所（2001）をもとに作成

必要がある。もし環境と財やサービスが完全代替の関係にない場合には、経済学的に正しい評価額が得られない。第三に、代替法で評価可能なものは、市場で取引される財やサービスによって代替できるものだけである。農地や森林が希少種の生息の場の役割を果たしていたとしても、そのような機能を市場で取引される財やサービスによって代替することは困難であるため、その価値を代替法で評価することは困難である。

3.2 ヘドニック価格法

ヘドニック価格法は、環境が住宅価格に及ぼす影響からその価値を評価する方法である（栗山 他, 2013; 庄子 他, 2011）。

図 1 は縦軸に住宅の価格、横軸に農地や森林の美しい景観を楽しむことができる眺望の程度をとっている。ここでは、住宅価格に影響を及ぼす属性のうち、眺望の程度以外はすべて固定し、住宅価格と眺望の程度のみを描いている。他の条件が同じであれば、眺望の程度がよい住宅ほど価格が高くなるので、住宅価格と眺望の程度の間には右上がりの曲線で表される関係がある。この曲線をヘドニック価格曲線と呼ぶ。

例えば、眺望の程度が q_0 で景観を楽しむことができない住宅の価格は p_0 であり、眺望の程度が q_1 で景観を楽しむことができる住宅の価格は p_1 であるとき、この価格の差 $p_1 - p_0$ を景

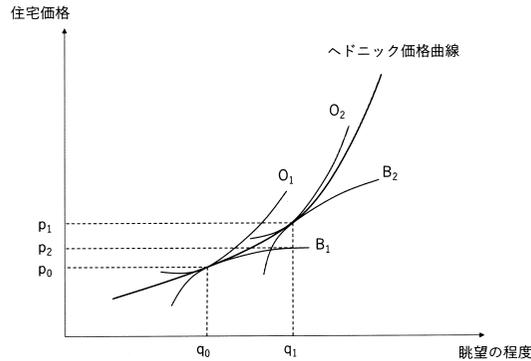


図1. ヘドニック価格曲線. 住宅価格と環境の質(ここでは眺望の程度)の関係を表すヘドニック価格曲線は右上がりとなる. ヘドニック価格曲線は、付け値曲線 B_n とオファー曲線 O_n の包絡線となっている.

観の価値とみなすことができる.

ただし、ヘドニック価格法により環境の価値を評価する場合、過大評価となる場合があることに注意が必要である。ヘドニック価格曲線は、付け値曲線 B_n とオファー曲線 O_n の包絡線となっている (Rosen, 1974)。ここで、付け値曲線 B_n とは、一定の効用を達成するために消費者 n が様々な眺望の程度に対して最大限支払うことができる金額を表し、オファー曲線 O_n とは、一定の利潤を達成するために生産者 n が最低限受け取らなければならない金額を表す。例えば、眺望の程度を q_0 から q_1 に改善することに対する WTP は付け値曲線より $p_2 - p_0$ である。ヘドニック価格曲線を推定した上で、消費者の付け値曲線を推定する二段階推定を行うことができれば、この WTP を正確に評価することができる。しかし、実際にはデータの制約のため、ヘドニック価格曲線の推定しか行われなことが多い。このとき、ヘドニック価格曲線に基づいて評価を行うと、評価額は $p_1 - p_0$ となる。ここから、ヘドニック価格曲線に基づいて評価を行った場合には、 $p_1 - p_2$ だけ過大評価となることがわかる。ただし、眺望の程度の変化が微小であれば、ヘドニック価格曲線に基づく評価により WTP を近似することができるため、過大評価の問題は無視できる。

実際の分析では、様々な住宅のデータを集めて、住宅価格と住宅価格に影響を及ぼすと考えられる様々な属性(部屋数、築年数、交通アクセス、..., 眺望の程度など)の関係を表すヘドニック価格関数を回帰分析により推定することで、眺望の程度が住宅価格に及ぼす影響を特定する。ヘドニック価格関数は以下のように表される。

$$(3.1) \quad p = f(q)$$

ただし、 p は住宅価格、 q は住宅価格に影響を及ぼす属性のベクトルである。

ヘドニック価格関数の関数形としては、線形、両対数、片対数などがよく用いられる。また、Box-Cox 変換によって関数形を決定する方法も用いられる。例えば、 q の要素である変数 q_k (k は属性を表す添え字) に Box-Cox 変換を適用すると、以下ようになる。ただし、 η_k は Box-Cox 変換パラメータであり、 $\eta_k = 0$ のときは対数、 $\eta_k = 1$ のときは線形となる。

$$(3.2) \quad q_k^{(\eta_k)} = \begin{cases} \frac{q_k^{\eta_k} - 1}{\eta_k} & \eta_k \neq 0 \\ \ln q_k & \eta_k = 0 \end{cases}$$

近年は、因果識別を行うために、外生的なショックの影響を受ける地域(処置群)と、ショックの影響を受けない地域(対照群)を比較する準実験的手法とヘドニック価格法を組み合わせた分析も行われている (Parmeter and Pope, 2013; Phaneuf and Requate, 2017)。

そのような準実験的手法の代表的な手法は、差分の差分法(difference-in-difference: DID)である。DID は、環境汚染の発生や環境規制の導入といった外生的なショックの影響を受ける地域のショックの前後での住宅価格の変化と、ショックの影響を受けない地域のショックの前後での住宅価格の変化を比較するものである。このため、DID はショックの前後の住宅価格のデータを使用できる場合に用いることができる。

ショックにより、ある環境の質 q が q_0 から q_1 に変化するとする。ここで、 t 期に販売される住宅 j の価格を p_{jt} と表し、住宅がショックの影響を受ける地域に存在する場合に 1 をとり、そうでないときに 0 をとるダミー変数 D_{jt}^S と、ショックの後($t = 1$)に売買が行われたときに 1 をとり、ショックの前($t = 0$)に売買が行われたときに 0 をとるダミー変数 D_{jt}^T を定義すると、推定式は以下ようになる。

$$(3.3) \quad \ln p_{jt} = \beta' x_{jt} + \delta_1 D_{jt}^S + \delta_2 D_{jt}^T + \delta^{DD} D_{jt}^S D_{jt}^T + \varepsilon_{jt}, \quad j = 1, \dots, J,$$

ここでは、例として片対数型の関数形を仮定した場合の推定式を示している。 x_{jt} は環境の質 q 以外の住宅価格に影響を及ぼす要因のベクトルを表し、 β はそのパラメータのベクトルを表す。なお、「 $'$ 」は転置を意味する。 δ^{DD} はショック後に、ショックの影響を受ける地域で販売されたことによる価格の差をとらえており、環境の質 q の変化が住宅価格に及ぼす影響を表している。

環境経済学分野の適用事例には Davis (2004)がある。Davis (2004)は、ネバダ州のある郡で、ある時期以降にガンの件数が増加したことに注目し、ガンの件数が増加した郡とそうでない郡で住宅価格を比較することで、健康リスクが住宅価格に及ぼす影響を明らかにした。

DID では、処置群と対照群が似ていることを仮定するが、実際に両者が似ているとは限らない。そこで、処置群とできるだけ似た対照群を設定するために、傾向スコアマッチングという手法が用いられることがある。これは、観察可能なデータに基づき各サンプルが処置群に割り当てられる確率を推定し、それが等しい処置群と対照群のペアをマッチングするものである。

因果識別のために用いられるもう 1 つの代表的な方法が回帰不連続デザイン(regression discontinuity design: RDD)である。RDD はクロスセクションデータで分析が可能である。

たとえば、環境の質 q の住宅価格 p_j への影響を計測することが目的であるとしよう。汚染源からそれぞれの住宅までの距離 d_j は連続変数であり、環境の状況が変化する境目となる閾値の距離 \bar{d} が存在すると仮定する。汚染源からの距離が一定の範囲内($d_j < \bar{d}$)の地域では一様に環境の質が悪く($q = q^l$)、汚染源からの距離が一定の範囲外($d_j > \bar{d}$)の地域では一様に環境の質がよい($q = q^h$)とする。ここで、 $H < J$ の住宅は、閾値のすぐ内側またはすぐ外側に立地しているとする。このとき、RDD では、閾値の周辺に存在する H の住宅のデータだけを使用して以下の式を推定する。

$$(3.4) \quad \ln p_j = \beta' x_{jt} + \delta^{RD} D_j + \varepsilon_j, \quad j = 1, \dots, H,$$

ここでは、例として片対数型の関数形を仮定した場合の推定式を示している。 D_j は $q = q^h$ のときに 1 をとり、 $q = q^l$ のときに 0 をとるダミー変数である。 δ^{RD} は環境の質 q の違いが住宅価格に及ぼす影響を表す。ただし、この推定値は閾値周辺でのものであり、すべての住宅でこの推定値が妥当であることは保証されないことに注意が必要である。

環境経済学分野の適用事例には、大気浄化法について分析を行った Chay and Greenstone (2005)がある。大気浄化法では、大気汚染物質の量が基準値を超えると未達成地域に指定さ

れ、既達成地域よりも厳しい規制が行われる。したがって、閾値のすぐ上の地域とすぐ下の地域では当初の汚染の差はわずかであるが、その後の規制の強さが異なることとなる。このことを利用して、閾値のすぐ上の地域とすぐ下の地域の住宅価格を比較することで、住宅価格が大気質の改善にどのように反応するかを調べた。

環境評価の分野では、大気清浄さをはじめとして、騒音の程度や水質の清浄さといった様々な環境の質の評価にヘドニック価格法が用いられている (Smith and Huang, 1995)。

しかし、ヘドニック価格法にはいくつかの問題がある (栗山 他, 2013; 庄子 他, 2011)。第一に、ヘドニック価格法では、住宅市場が完全競争市場であり、取引費用 (経済取引を行う際に発生する費用) が存在しないことが仮定されるが、現実には完全競争市場が想定するように、すべての消費者と生産者 (不動産供給者) が環境の質に関する情報を含むあらゆる情報を把握して行動しているわけではないし、物件を探す手間や引越し費用などの取引費用が存在する。したがって、これらの仮定は非現実的である。第二に、価格が高い住宅は、駅までの距離が近く、市街地までの距離も近いといったように、住宅価格に影響を及ぼす要因間には高い相関が存在することが多いため、ヘドニック価格関数の推定においては多重共線性が発生しやすい。第三に、ヘドニック価格法は住宅価格を用いた分析であるため、住宅価格に影響する環境の価値しか評価できない。例えば、住宅地から離れた場所にある農地や森林の状況は、住宅価格に反映されにくいいため、ヘドニック法でその価値を評価することは困難である。

近年は、空間データ特有の性質を明示的に考慮した計量経済モデルである空間計量経済モデルを用いて分析を行う空間ヘドニック法に関する研究も行われており (星野, 2011)、手法の洗練化が進められている。

3.3 トラベルコスト法

トラベルコスト法は、旅行費用に基づいてレクリエーションの価値を評価する方法である (栗山 他, 2013; 柘植 他, 2011b)。なお、ここでの旅行費用とは、交通費だけではなく、訪問に要した時間の価値を含む、レクリエーションへの参加に要したあらゆる費用の合計である。

トラベルコスト法には、特定のサイトへの旅行費用と訪問回数の関係から、そのサイトにおけるレクリエーションに対する需要曲線 (レクリエーション需要曲線) を推定するシングルサイトモデルと、複数の代替的なサイトの中から訪問するサイトを選択する行動を後述のランダム効用モデルによりモデル化し、効用関数を推定するサイト選択モデルがある。また、近年は、訪問するサイトの選択と訪問回数の決定の双方を扱うことができるモデルの開発も進められている。

3.3.1 シングルサイトモデル

図2は縦軸に旅行費用、横軸に訪問回数をとっている。旅行費用が高いほど訪問回数は減るので、旅行費用と訪問回数の関係を表すレクリエーション需要曲線は右下がりとなる。

レクリエーション需要曲線に基づいて、個人がレクリエーションから得る便益を表す消費者余剰を計算する。例えば、旅行費用が p_1 のとき、消費者余剰は三角形 p_1Ap_C となる。ここで、 p_C はチョークプライスと呼ばれ、訪問回数がゼロになる旅行費用を表す。

個人単位のデータを用いてレクリエーション需要曲線を推定する方法を個人トラベルコスト法と呼ぶ。ある期間中のサイト i への訪問回数を x_i 、サイト i への旅行費用を p_i 、代替的なサイト j への旅行費用を p_j 、所得を M 、年齢や性別などの個人属性のベクトルを z とすると、レクリエーション需要曲線は以下のように表わされる。

$$(3.5) \quad x_i = f(p_i, p_j, M, z)$$

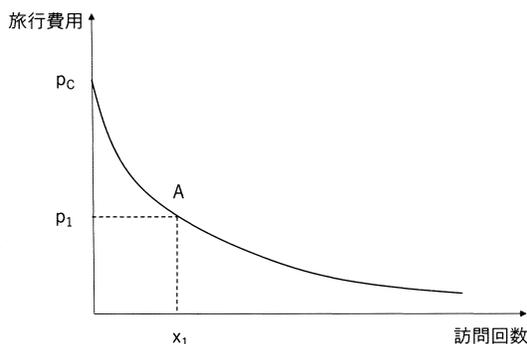


図 2. レクリエーション需要曲線. 旅行費用と訪問回数の関係を表すレクリエーション需要曲線は右下がりとなる.

このとき, 旅行費用が p_1 の場合の消費者余剰 CS は以下のように表される. ただし, p_c は choke price を表す.

$$(3.6) \quad CS = \int_{p_1}^{p_c} f(p_i, p_j, M, z) dp_i$$

訪問回数のデータは非負の整数となるため, レクリエーション需要曲線の推定には, 式(3.7)のポアソン回帰や, 式(3.8)の負の二項分布モデルなどのカウントモデルが用いられる (Shaw, 1988; Haab and McConnell, 2002).

$$(3.7) \quad \Pr(x_{ni}) = \frac{\exp(-\lambda_{ni}) \cdot \lambda_{ni}^{x_{ni}}}{x_{ni}!}$$

$$(3.8) \quad \Pr(x_{ni}) = \frac{\Gamma(\frac{1}{\alpha} + x_{ni})}{\Gamma(\frac{1}{\alpha}) \Gamma(x_{ni} + 1)} \left(\frac{\frac{1}{\alpha}}{\frac{1}{\alpha} + \lambda_{ni}} \right)^{\frac{1}{\alpha}} \left(\frac{\lambda_{ni}}{\frac{1}{\alpha} + \lambda_{ni}} \right)^{x_{ni}}$$

ただし, x_{ni} は個人 n のサイト i への訪問回数を表し, $\Pr(x_{ni})$ はある期間中に個人 n がサイト i を x 回訪問する確率を表す. λ_{ni} は個人 n のサイト i への訪問回数の期待値であり, サイト i への旅行費用 p_{ni} , 代替的なサイト j への旅行費用 p_{nj} , 所得 M_n , 個人属性のベクトル z_n などの関数として, $\ln(\lambda_{ni}) = \beta_{p_i} p_{ni} + \beta_{p_j} p_{nj} + \beta_M M_n + \beta'_z z_n$ のように表される. 左辺を $\ln(\lambda_{ni})$ とするのは, 確率が非負となるようにするためである. 式(3.8)の Γ はガンマ関数を表し, α は過分散 (overdispersion) パラメータを表す. パラメータは最尤法によって推定される. ポアソン回帰では, 訪問回数の期待値と分散が等しいことが仮定されるのに対し, 負の二項分布モデルはこの仮定を必要としない点でより一般的なモデルである.

いずれのモデルでも, 個人 n の期間中の消費者余剰 CS_{ni} は以下のように求められる.

$$(3.9) \quad CS_{ni} = -\frac{\lambda_{ni}}{\beta_{p_i}}$$

分析に必要な訪問回数や居住地などのデータは, アンケート調査により収集する. しかし, 一般市民を対象としたアンケート調査(オフサイトサンプリング)で, あるサイトの訪問者についての十分なサンプル数を確保しようとすると大規模な調査が必要となる. そこで, より効率的に訪問者のデータを収集するため, サイトで訪問者を対象とした調査(オンサイトサンプリング)を行うことが多い. しかし, Shaw (1988)が示したように, オンサイトサンプリング

を行った場合には、回答者は全員訪問者であるため、訪問回数が1以上となる切断(truncation)と呼ばれる現象が発生する。また、訪問回数が多い人ほどサンプルに含まれやすい内生的層化(endogenous stratification)と呼ばれる現象も発生する。これらはパラメータの推定値に影響を及ぼす。これらの問題に対して、Shaw (1988)は、ポアソン回帰の場合には式(3.10)、負の二項分布モデルの場合には式(3.11)のように修正することで、バイアスのない推定値が得られることを示した(Shaw, 1988; Haab and McConnell, 2002)。

$$(3.10) \quad \Pr(x_{ni}) = \frac{\exp(-\lambda_{ni}) \cdot \lambda_{ni}^{x_{ni}-1}}{(x_{ni}-1)!}$$

$$(3.11) \quad \Pr(x_{ni}) = \frac{x_{ni} \Gamma\left(\frac{1}{\alpha} + x_{ni}\right)}{\Gamma\left(\frac{1}{\alpha}\right) \Gamma(x_{ni} + 1)} \left(\frac{\frac{1}{\alpha}}{\frac{1}{\alpha} + \lambda_{ni}}\right)^{\frac{1}{\alpha}} \left(\frac{1}{\frac{1}{\alpha} + \lambda_{ni}}\right)^{x_{ni}} \lambda_{ni}^{x_{ni}-1}$$

シングルサイトモデルでは、ある特定のサイトへの訪問回数を分析することで、そのサイトにおけるレクリエーションの価値を評価する。代替的なサイトが存在しない場合には問題ないが、代替的なサイトが存在する場合には、それを考慮する必要がある。そこで、代替的なサイトへの旅行費用をレクリエーション需要曲線の変数として含めることなどが行われるが、代替的なサイトの影響は限定的にしか分析することができない。代替的なサイトの影響を分析するためには、次節で紹介するサイト選択モデルがより適する。

3.3.2 サイト選択モデル

代替的なサイトの中から訪問するサイトを選択する行動を分析するサイト選択モデルでは、ランダム効用モデルを用いて個人の選択行動をモデル化する。

個人 n がサイト i から得る効用が、効用をもたらす要因が分析者に観察可能なもの(例えば、森林であれば樹木の本数や種類、旅行費用など)と分析者には観察不可能なため確率的に扱わざるを得ないものからなるとする。ランダム効用モデルを用いると、前者を確定項 V_{ni} 、後者を誤差項 ε_{ni} として、サイト i を訪問することから得られる効用を以下のように表すことができる。

$$(3.12) \quad U_{ni} = V_{ni} + \varepsilon_{ni}$$

ここで、確定項 V_{ni} に線形を仮定すると、以下のように表される。

$$(3.13) \quad V_{ni} = \beta'_q q_i + \beta_p p_i$$

ただし、 q_i はサイト i のサイト属性ベクトル、 β_q はそのパラメータのベクトル、 p_i はサイト i の旅行費用、 β_p はそのパラメータである。 β_q はサイト属性の限界効用のベクトルを表す。また、 β_p は旅行費用の限界効用を表し、その絶対値は所得の限界効用を表す。

個人は代替的なサイトの中から最大の効用が得られるサイトを選択すると考えられる。個人 n が、代替的なサイトの集合である選択セット $C = \{1, 2, \dots, J\}$ の中からサイト i を選択する確率 \Pr_{ni} は、サイト i を選択したときの効用 U_{ni} が、その他のサイト $j (j \neq i)$ を選択したときの効用 U_{nj} よりも高くなる確率であるから、以下のように表すことができる。

$$(3.14) \quad \Pr_{ni} = \Pr(U_{ni} > U_{nj} \quad \forall j \in C, j \neq i) = \Pr(V_{ni} - V_{nj} > \varepsilon_{nj} - \varepsilon_{ni} \quad \forall j \in C, j \neq i)$$

ここで、誤差項が独立で同一な第一種極値分布(ガンベル分布)にしたがうと仮定すると、個人 n がサイト i を選択する確率 \Pr_{ni} は、以下の条件付ロジットモデルにより表される(McFadden, 1974)。

$$(3.15) \quad \Pr_{ni} = \frac{\exp(\mu V_{ni})}{\sum_{j \in C} \exp(\mu V_{nj})} = \frac{\exp\{\mu(\beta'_q \mathbf{q}_i + \beta_p p_i)\}}{\sum_{j \in C} \exp\{\mu(\beta'_q \mathbf{q}_j + \beta_p p_j)\}}$$

アンケート調査により、人々がどのサイトの中からどのサイトを選択したかを把握できれば、それらのデータを用いて、最尤法により β_q や β_p を推定することができる。なお、 μ はスケールパラメータを表し、通常は 1 と仮定される (Train, 2009)。

β_q や β_p の推定値を用いることで、サイト属性の変化やサイトの増減に対する WTP を評価することができる。例えば、サイト属性の \mathbf{q}_j^0 から \mathbf{q}_j^1 への変化に対する WTP は以下のように求めることができる (Small and Rosen, 1981; Haab and McConnel, 2002)。

$$(3.16) \quad WTP = -\frac{1}{\beta_p} \left[\ln \left[\sum_{j \in C} \exp[V_{nj}(\mathbf{q}_j^1)] \right] - \ln \left[\sum_{j \in C} \exp[V_{nj}(\mathbf{q}_j^0)] \right] \right]$$

ここで、あるサイト属性 q_k の限界的な変化に対する WTP である限界支払意志額 (marginal willingness to pay: MWTP) は、サイト属性 q_k の限界効用 β_k と所得の限界効用 $-\beta_p$ の比として以下のように求めることができる。

$$(3.17) \quad MWTP_k = -\frac{\beta_k}{\beta_p}$$

また、サイトの新設に対する WTP は以下のように表すことができる。ただし、 C は新設されたサイトを含まない選択セット、 C' は新設されたサイトを含む選択セットを表す。

$$(3.18) \quad WTP = -\frac{1}{\beta_p} \left[\ln \left[\sum_{j \in C'} \exp[V_{nj}(\mathbf{q}_j)] \right] - \ln \left[\sum_{j \in C} \exp[V_{nj}(\mathbf{q}_j)] \right] \right]$$

3.3.3 クーン・タッカーモデル

サイト選択モデルでは、代替的なサイト間での選択行動を分析することはできるが、訪問回数分析することはできない。一方、シングルサイトモデルでは、特定のサイトへの訪問回数を分析することはできるが、代替的なサイトの影響を限定的にしか考慮できない。そこで、両者を扱うことができるモデルの開発が行われている。代表的なものに、訪問するサイトは内点解、訪問しないサイトは端点解として扱うことで、サイト選択と訪問回数選択を 1 つの効用最大化問題としてモデル化するクーン・タッカーモデル (端点解モデル) がある (Phaneuf, et al., 2000; 柘植 他, 2011a; 柘植 他, 2011b)。

クーン・タッカーモデルでは、以下の効用最大化問題を考える。

$$(3.19) \quad \text{Max } U(\mathbf{x}, \mathbf{q}, h, \beta, \varepsilon) \text{ s.t. } \mathbf{p}'\mathbf{x} + h = M, h > 0, x_j \geq 0, j = 1, \dots, J$$

ただし、 U は効用関数、 \mathbf{x} は各サイトへの訪問回数のベクトル、 \mathbf{q} は各サイトの属性行列、 h はニューメールの消費量、 β はパラメータのベクトル、 ε は誤差項のベクトル、 \mathbf{p} は各サイトへの旅行費用のベクトル、 M は所得、 x_j はサイト j の訪問回数である。

この問題を解くと、以下の条件が得られる。

$$(3.20) \quad U_j \leq U_h p_j, x_j^* \geq 0, x_j^* [U_j - U_h p_j] = 0, j = 1, \dots, J$$

ただし、 $U_j = \partial U / \partial x_j$ 、 $U_h = \partial U / \partial h$ であり、 p_j はサイト j への旅行費用、 x_j^* は効用最大化問題の解である。

ここで、 $U_{h\varepsilon} = \partial U_h / \partial \varepsilon = 0$ 、 $\partial U_j / \partial \varepsilon_k = 0 (\forall k \neq j)$ 、 $\partial U_j / \partial \varepsilon_j > 0 (\forall j)$ を仮定すると、式 (3.20) は以下のように書ける。

$$(3.21) \quad \varepsilon_j \leq g_j(\mathbf{x}^*, \mathbf{p}, M, \mathbf{q}, \beta), \quad x_j^* \geq 0, \quad x_j^*[\varepsilon_j - g_j] = 0, \quad j = 1, \dots, J$$

ただし、 \mathbf{x}^* は効用最大化問題を解くことで求められる各サイトへの訪問回数のベクトルであり、 ε_j はサイト j の誤差項である。また、以下の等式の解を g_j とする。

$$(3.22) \quad U_j(\mathbf{x}^*, \mathbf{q}, M - \mathbf{p}'\mathbf{x}^*, \beta, g_j) - U_h(\mathbf{x}^*, \mathbf{q}, M - \mathbf{p}'\mathbf{x}^*, \beta)p_j = 0$$

ここで、訪問回数がゼロとなる確率は $\Pr(\varepsilon_j < g_j)$ であり、訪問回数が 1 以上となる確率は $\Pr(\varepsilon_j = g_j)$ であるため、最初の m 個のサイトを訪問する ($x_j > 0, j = 1, \dots, m$ かつ $x_j = 0, j = m + 1, \dots, J$) 確率は、以下のように表される。

$$(3.23) \quad \Pr(x_1, \dots, x_m, 0_{m+1}, \dots, 0_J) = \Pr[\varepsilon_1 = g_1, \dots, \varepsilon_m = g_m, \varepsilon_{m+1} \leq g_{m+1}, \dots, \varepsilon_J \leq g_J] \\ = \int_{-\infty}^{g_{m+1}} \cdots \int_{-\infty}^{g_J} f_\varepsilon(g_1, \dots, g_m, \varepsilon_{m+1}, \dots, \varepsilon_J) \\ \times |J_m| d\varepsilon_{m+1}, \dots, d\varepsilon_J$$

ただし、 $|J_m|$ は x_1, \dots, x_m から $\varepsilon_1, \dots, \varepsilon_m$ への変換のためのヤコビアンである。

ε_j の分布 f_ε と効用関数の関数形を仮定することで、 β を推定することが可能となる。 ε_j が第一種極値分布にしたがう場合、式 (3.23) の閉形式が存在するため、 ε_j の分布 f_ε には第一種極値分布が仮定されることが多い。また、補償変分の計算を容易にするため、効用関数には加法分離的な関数形が用いられる。代表的な関数形は以下のようなものである (von Haefen et al., 2004; 柘植 他, 2011a)。

$$(3.24) \quad U(\mathbf{z}, \mathbf{q}, h, \beta, \varepsilon) = \sum_{j=1}^J \Psi(\mathbf{z}, \varepsilon_j) \ln(\phi(\mathbf{q}_j)x_j + \pi) + \frac{1}{\rho} h^\rho \\ \Psi(\mathbf{z}, \varepsilon_j) = \exp(\boldsymbol{\delta}'\mathbf{z} + \varepsilon_j) \\ \phi(\mathbf{q}_j) = \exp(\boldsymbol{\gamma}'\mathbf{q}_j)$$

ここで、 \mathbf{z} は個人属性のベクトル、 $\boldsymbol{\delta}$ はそのパラメータのベクトル、 \mathbf{q}_j はサイト j の属性のベクトル、 $\boldsymbol{\gamma}$ はそのパラメータのベクトルである。 π は訪問回数に関係なく効用に影響する部分を表し、トランスレイティングと呼ばれる。 ρ は所得効果を表すパラメータである。

クーン・タッカーモデルでは、二分法などの数値計算により補償変分を探索的に求める。詳しくは、von Haefen et al. (2004) や柘植 他 (2011a) を参照されたい。

3.3.4 トラベルコスト法の課題

トラベルコスト法は、農地や森林の保健・休養機能の評価に適する。しかし、トラベルコスト法にはいくつかの問題がある (栗山 他, 2013)。第一に、トラベルコスト法では、旅行費用を正確に把握することが重要である。しかし、旅行費用には交通費だけでなく、レクリエーションに参加するために費やしたあらゆる費用が含まれる。すべての個人について、それらを正確に把握することは容易ではない。特に、機会費用の正確な把握は課題である。第二に、原始的なサイトほど、自然が豊かで、混雑が少なく、静穏であるといったように、サイト属性のデータ間にはしばしば相関が存在するため、サイト選択モデルでは、サイト属性間に多重共線性が発生しやすい。第三に、トラベルコスト法は旅行費用を用いた分析であるため、レクリエーション行動に反映される環境の価値しか評価できない。したがって、農地や森林の保健・休養機能以外の機能を評価することは困難である。

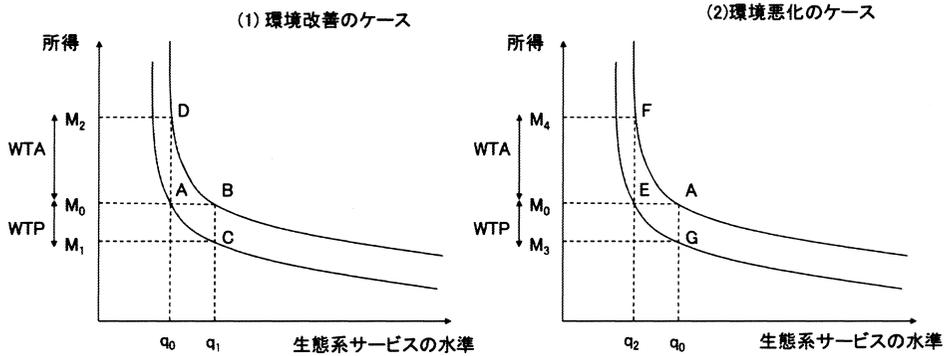


図 3. WTP と WTA. (1)生態系サービスの水準が q_0 から q_1 に改善されることに対する WTP は $M_0 - M_1$ で示され、この生態系サービスの改善が中止されることに対する WTA は $M_2 - M_0$ で示される. (2)生態系サービスの水準が q_0 から q_2 に低下することに対する WTA は $M_4 - M_0$ で示され、この生態系サービスの低下を回避することに対する WTP は $M_0 - M_3$ で示される.

4. 表明選好法の詳細

4.1 仮想評価法 (CVM)

仮想評価法 (contingent valuation method: CVM) は、アンケート調査で把握した人々の意見をもとに環境の価値を評価する方法である (栗山 他, 2013).

CVM では、環境改善に対する WTP、環境悪化に対する受入補償額 (willingness to accept compensation: WTA)、環境改善中止に対する WTA、環境悪化中止に対する WTP のいずれかを尋ねることで環境の価値を評価する.

例えば、ある生態系サービスの水準が q_0 から q_1 に改善することに対する WTP は、間接効用関数 V を用いて次式のように定義される.

$$(4.1) \quad V(q_0, M) = V(q_1, M - WTP)$$

ただし、 M は所得、 q_0 は改善前の生態系サービスの水準、 q_1 は改善後の生態系サービスの水準を表す。つまり、生態系サービスの水準の改善に対する WTP は、生態系サービスの水準が改善された状況で、生態系サービスの水準が改善される前の効用水準を達成するために、消費者から取り去ることのできる最大の金額を表す。一方、生態系サービスの水準の改善が中止されたことに対する WTA は、間接効用関数 V を用いて次式のように定義される.

$$(4.2) \quad V(q_0, M + WTA) = V(q_1, M)$$

つまり、生態系サービスの水準の改善が中止されたことに対する WTA は、生態系サービスの水準が改善される前の状況で、生態系サービスの水準が改善された状況の効用水準を達成するために、消費者に与えなければならない最小の金額を表す.

上記の WTP と WTA を図示したものが図 3 の (1) である。当初、個人が所得 M_0 、生態系サービスの水準が q_0 の点 A にいるとする。ここで、生態系サービスの水準が q_1 に改善する状況を考える。これにより、個人は所得が M_0 、生態系サービスの水準が q_1 の点 B に移動することとなる。このとき、個人は、より右上の無差別曲線上の点に移動することとなる。無差別曲線は、右上方にあるほど効用が高いことを表すので、これは望ましい変化である。したがって、

個人はこの変化を実現したいと考える。個人が、この変化を実現するために支払ってもいいと考える最大の金額、つまり、生態系サービスの水準の改善のための WTP は、 $M_0 - M_1$ で示される。なぜならば、点 A の状態から、最大 $M_0 - M_1$ だけ支払ったとしても、生態系サービスが改善する前の無差別曲線上の点 C に移動することになり、生態系サービスが改善する前の効用水準が維持できるからである。一方、生態系サービスの改善中止に対する WTA は $M_2 - M_0$ で示される。なぜならば、点 A の状態で $M_2 - M_0$ だけ受け取れば、生態系サービスが改善された場合の無差別曲線上の点 D に移動することになり、生態系サービスが改善した場合の効用水準が達成できるからである。

生態系サービスの水準が低下することに対する WTA と、生態系サービスの水準が低下することを回避するための WTP は、図 3 の (2) のように示される。

同じ対象を WTP と WTA で評価した場合、しばしば WTA が WTP を大きく上回る (栗山他, 2013)。CVM ではより控えめな評価額を得ることが推奨されているため、WTP を尋ねることが多いことから、以下では環境改善に対する WTP を尋ねるケースを想定して説明を行う。

CVM には、表 3 に示すように、自由回答形式、競りゲーム (付け値ゲーム) 形式、支払いカード形式、二肢選択形式などの質問形式が開発されているが、最もバイアスが発生しにくく、提示された金額を見て賛成するか否かを決めるという回答形式が日常の購買行動と類似しているため回答しやすい二肢選択形式が広く用いられている (栗山 他, 2013)。また、二肢選択形式の質問の後に、提示額を変更してもう一度質問を繰り返す二段階二肢選択形式も考案されている (Hanemann et al., 1991)。二段階二肢選択形式では、通常の二肢選択形式よりも WTP の存在する区間を特定することができる。以下では二肢選択形式に焦点を当てて説明を行う。

二肢選択形式の分析には様々な方法が用いられるが、ここでは、最も一般的に用いられるランダム効用モデルによる分析について解説を行う (Hanemann, 1984)。

ランダム効用モデルでは、回答者 n が賛成と回答したときの効用 U_{n1} と反対と回答したときの効用 U_{n0} が、それぞれ観察可能な確定項 V と、観察不可能な誤差項 ε からなると仮定する。

$$(4.3) \quad U_{n1} = V(q_1, M_n - p_n) + \varepsilon_{n1}$$

$$(4.4) \quad U_{n0} = V(q_0, M_n) + \varepsilon_{n0}$$

ただし、 M_n は所得、 p_n は賛成と回答した場合の負担額、 q_1 は環境改善がなされる場合の環境の状況、 q_0 は環境改善がなされない場合の環境の状況を示している。このとき、回答者 n が賛成と回答する確率 Pr_{n1} は、賛成と回答したときの効用 U_{n1} が反対と回答したときの効用 U_{n0} よりも高くなる確率であるから、以下のように表すことができる。

$$(4.5) \quad \begin{aligned} \text{Pr}_{n1} &= \text{Pr}[U_{n1} > U_{n0}] = \text{Pr}[V(q_1, M_n - p_n) + \varepsilon_{n1} > V(q_0, M_n) + \varepsilon_{n0}] \\ &= \text{Pr}[V(q_1, M_n - p_n) - V(q_0, M_n) > \varepsilon_{n0} - \varepsilon_{n1}] \\ &= \text{Pr}[\varepsilon_{n1} - \varepsilon_{n0} > -\Delta V_n] \end{aligned}$$

ただし、 ΔV_n は賛成と回答した場合と反対と回答した場合の確定項の差であり、効用差関数と呼ばれる。

効用差関数の関数形を特定し、誤差項に特定の分布を仮定することで、効用差関数のパラメータを推定することが可能となる。例えば、線形の効用差関数 ΔV_n は、以下のように表される。

$$(4.6) \quad \Delta V_n = \alpha + \beta p_n$$

ただし、 α と β は推定されるパラメータである。 α は環境変化に対する評価を表し、 β は所得

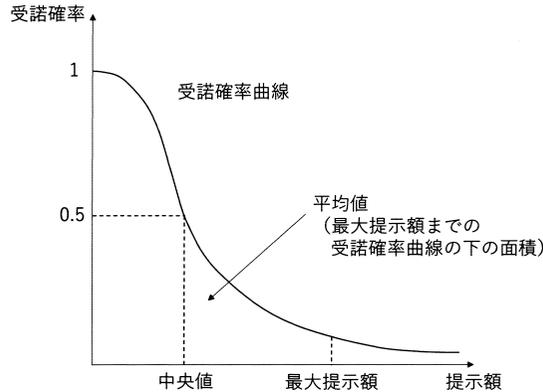


図4. 中央値と平均値. 提示額と支払いに賛成する確率の関係を表す受諾確率曲線は右下がりになる. WTPの中央値は提示額に賛成する確率が0.5となる金額と定義され, 平均値は受諾確率を提示額に関して積分することで求められる.

ただし, Φ は標準正規分布の累積分布関数を表す. パラメータ α と β は最尤法により推定される.

性別, 年齢, 所得といった回答者の社会経済的属性や, 評価対象に関する知識, 環境問題に対する関心の度合いなどを変数化し, 効用差関数の説明変数に加えることも可能である. このようなモデルを推定することで, WTPに影響を及ぼす要因を明らかにすることができる(栗山他, 2013).

縦軸に提示額に賛成する確率(受諾確率), 横軸に提示額をとると, 両者の関係を表す受諾確率曲線は図4のように右下がりになる. 二肢選択形式では, WTPの推計値として中央値と平均値の2種類の金額が得られる. 中央値は, 提示額に賛成する確率が0.5となる金額と定義される. 一方, 平均値は, 回答者が賛成と回答する確率 Pr_{n1} を提示額に関して積分することで求められる. ただし, あまりに高額な提示額まで積分を行うことは現実的でないため, 最大提示額など, 一定の金額までで打ち切ることが多い. 中央値と平均値はそれぞれ図4のように示される. 二項ロジットモデルで線形の効用差関数を仮定した場合の中央値 WTP^* と最大提示額 p_{max} まで積分を行った場合の平均値 WTP^+ は以下のように表される(Hanemann, 1984).

$$(4.9) \quad WTP^* = -\frac{\alpha}{\beta}$$

$$(4.10) \quad WTP^+ = \int_0^{p_{max}} \frac{1}{1 + \exp\{-(\alpha + \beta p)\}} dp$$

4.2 コンジョイント分析

コンジョイント分析もCVMと同様にアンケート調査を用いる方法である. CVMでは農地や森林の生態系サービスのいずれか1つ(例:水源涵養機能)の価値を評価することや, 多様な生態系サービスを発揮する農地や森林の価値全体を評価することができるのに対して, コンジョイント分析では, 一度の調査で複数の生態系サービス(例:水源涵養機能, 保健・休養機能, 生態系保全機能)の価値を個別に評価することができる(栗山他, 2013).

コンジョイント分析では, 森林に生息する生物の種数であれば何種類か, レクリエーション

であれば何が可能かといった、各属性が取り得る具体的な内容や値のことを水準と呼ぶ。また、属性の束として表現される仮想的な政策や製品をプロファイルと呼ぶ。コンジョイント分析では、各属性の水準を直交デザインなどの手法を用いて組み合わせることで作成されるプロファイルに回答者に提示し、それらに対する回答者の評価に基づいてプロファイルを構成する各属性の価値を評価する。たとえば、様々な属性の束として表現される森林整備計画の代替案に対する回答者の評価に基づいて、木材供給が 1 トン増加することに対する WTP、森林に生息する生物が 1 種類増加することに対する WTP といったように、各属性の 1 単位の向上に対する WTP である MWTP を個別に評価することができる。適切なデザイン手法を用いてプロファイルを作成することで、属性間で多重共線性が発生することを回避し、各属性の価値を独立に推定することが可能となる。

コンジョイント分析には、完全プロファイル評定型、ペアワイズ評定型、選択実験、仮想ランキング、ベスト・ワースト・スケーリングなどの質問形式が存在する(表 4)。このうち、環境評価の分野では、選択実験が広く用いられている。選択実験は、最も望ましいものを 1 つ選択するという回答形式が市場での購買行動に近いと回答しやすいといわれている(栗山 他, 2013)。以下では、選択実験に焦点を当てて説明を行う。

選択実験では、トラベルコスト法のサイト選択モデルや CVM の二肢選択形式と同様に、回答者の選択行動をランダム効用モデルのもとでの効用最大化行動としてモデル化し、そこから導出される離散選択モデルを用いて効用関数のパラメータを推定する。

回答者 n が選択肢 i を選択したときの効用を確定項 V_{ni} と誤差項 ε_{ni} の和として $U_{ni} = V_{ni} + \varepsilon_{ni}$ と表し、誤差項 ε_{ni} が独立で同一な第一種極値分布に従うと仮定すると、回答者 n が選択セット C から選択肢 i を選択する確率 Pr_{ni} は以下の条件付ロジットモデルで表される (McFadden, 1974)。

$$(4.11) \quad \text{Pr}_{ni} = \frac{\exp(\mu V_{ni})}{\sum_{j \in C} \exp(\mu V_{nj})}$$

ただし、 μ はスケールパラメータであり、通常は 1 と仮定される。

効用関数のパラメータの推定値を用いることで、各属性に対する MWTP を算出することができる。例えば、効用関数の確定項に $V = \beta'x$ といった線形を仮定した場合、属性 k に対する MWTP ($MWTP_k$) は、属性 k のパラメータ β_k と負担額のパラメータ β_p の比に負号を付けた $MWTP_k = -\beta_k/\beta_p$ として求められる。

近年は選好の多様性を把握することが可能なより洗練されたモデルも用いられている(三谷 他, 2005; 柘植 他, 2011b)。特に広く用いられているのが混合ロジットモデル(ランダムパラメータ・ロジットモデル)である。混合ロジットモデルは、パラメータが確率分布にしたがって個人間で異なることを許容するモデルである (Revelt and Train, 1998; Train, 2009; 三谷 他, 2005)。

個人ごとに異なる選好を持つことを仮定すると、個人 n のパラメータは β_n と表すことができる。ここで、各個人のパラメータベクトル β_n は観察不可能なため、パラメータ β_n の密度に関する条件付ロジットモデルの積分を考える。このとき、個人 n が選択肢 i を選択する確率は以下のように表される。

$$(4.12) \quad \text{Pr}_{ni} = \int \frac{\exp(V_{ni}(\beta_n))}{\sum_{j \in C} \exp(V_{nj}(\beta_n))} f(\beta) d\beta$$

ただし、 $f(\beta)$ は β の確率密度関数を表している。推定にあたっては、 β の確率分布を仮定する必要がある。この確率分布としては正規分布などが用いられる (Revelt and Train, 1998; Train, 2009; 三谷 他, 2005)。また、積分計算が代数的に解けないため、シミュレーションを用いて積

表 4. コンジョイント分析の質問形式.

完全プロフィール評定型: プロファイルを1つ提示して、それがどれくらい望ましいかを回答してもらう
 例: 以下の計画はどれくらい望ましいと思いますか? 点数を付けてください.

森林整備計画	
木材供給	2 トン増加
生息する生物	1 種類増加
負担額	5,000 円

100 点満点で
() 点

ペアワイズ評定型: プロファイルを2つ提示して、どちらがどれくらい望ましいかを回答してもらう
 例: どちらの計画がどのくらい望ましいと思いますか? 番号に○を付けてください.

森林整備計画	1	2
木材供給	2 トン増加	1 トン増加
生息する生物	1 種類増加	2 種類増加
負担額	5,000 円	3,000 円

← 4 3 2 1 0 1 2 3 4 →
 計画1が望ましい どちらともいえない 計画2が望ましい

選択実験: 複数のプロフィールを提示して、その中から最も望ましいと思うものを1つ選択してもらう
 例: どの計画が最も望ましいと思いますか?

森林整備計画	1	2	3
木材供給	2 トン増加	1 トン増加	現状
生息する生物	1 種類増加	2 種類増加	
負担額	5,000 円	3,000 円	0 円

計画 ()

仮想ランキング: 複数のプロフィールを提示して、それらを望ましい順に順位づけてもらう
 例: 以下の計画を望ましいと思う順に並べてください.

森林整備計画	1	2	3
木材供給	2 トン増加	1 トン増加	現状
生息する生物	1 種類増加	2 種類増加	
負担額	5,000 円	3,000 円	0 円

1 位 計画 () 2 位 計画 () 3 位 計画 ()

ベスト・ワースト・スケーリング (case3): 複数のプロフィールを提示して、最も望ましいものと最も望ましくないものを選択してもらう
 例: 最も望ましいと思う計画と最も望ましくないと思う計画を選んでください.

森林整備計画	1	2	3
木材供給	2 トン増加	1 トン増加	現状
生息する生物	1 種類増加	2 種類増加	
負担額	5,000 円	3,000 円	0 円

最も望ましい: 計画 () 最も望ましくない: 計画 ()

分を近似するシミュレートされた最尤法(maximum simulated likelihood)が用いられる。詳細は Train(2009)を参照されたい。

選好の多様性を把握するために用いられるもう 1 つの代表的なモデルが潜在クラスモデルである (Swait, 1994; Boxall and Adamowicz, 2002)。潜在クラスモデルは、個人をいくつかのグループに分類し、各グループの効用関数のパラメータを推定する。このとき、各クラスに所属する確率を、年齢、性別、所得などの個人属性や態度、動機などの心理的な潜在変数で説明するメンバーシップ関数を推定することで、どのような要因で選好の多様性が生じているかを明らかにする。したがって、選好の多様性が生じる要因を解明することに関心がある場合には、特に有益である。

サンプル内に S 個のクラスが存在し、個人 n はあるクラス $s = \{1, 2, \dots, S\}$ に属していると仮定する。このとき、個人 n が選択肢 i を選択する確率 $\text{Pr}_{ns}(i)$ は、以下のように表される。

$$(4.13) \quad \text{Pr}_{ns}(i) = \sum_{s=1}^S \left[\frac{\exp(\zeta \gamma'_s z_n)}{\sum_{s^*=1}^S \exp(\zeta \gamma'_{s^*} z_n)} \right] \left[\frac{\exp(\mu_s \beta'_s x_{ni})}{\sum_{j \in C} \exp(\mu_s \beta'_s x_{nj})} \right]$$

右辺の 1 つ目のロジットモデルの式は、個人がクラス s に振り分けられる確率を表すメンバーシップ関数であり、 z_n は個人属性や心理的な潜在変数のベクトル、 γ_s は推定されるパラメータのベクトル、 ζ はメンバーシップ関数のスケールパラメータを表す。右辺の 2 つ目のロジットモデルの式は、クラス s に属する個人 n が選択肢 i を選択する確率を表しており、 x_{ni} と x_{nj} は選択肢 i と j の属性のベクトル、 β_s はクラス s に固有のパラメータのベクトル、 μ_s はクラス s に固有のスケールパラメータを表す。いずれのロジットモデルの導出においても、誤差項に第一種極値分布を仮定している。

パラメータは最尤法によって推定される。また、クラス数については、AIC や BIC に基づいて決定する。

CVM やコンジョイント分析はアンケートで人々の意見を尋ねるため、どのような対象の価値でも評価できる。また、行動には顕示されない非利用価値も評価できる。しかし、バイアスが発生しやすいという問題がある (栗山, 1998; Mitchell and Carson, 1989)。

表明選好法のバイアスの中でも特に重大なものが、アンケートにおいて表明する仮想的な支払額と実際の支払額の乖離である仮想バイアスである。仮想バイアスについては、List and Gallet (2001), Murphy et al. (2005), Loomis (2011) などがメタ分析を行っている。近年は、実験経済学的手法も援用しつつ仮想バイアスに関する研究が行われている (三谷, 2011)。仮想バイアスは表明選好法の信頼性に根本的に関わる問題であり、さらなる研究が求められる。

5. 顕示選好法と表明選好法の統合

顕示選好法は現実の行動に基づくためデータの信頼性が高いが、分析に使用する変数間の相関が高く多重共線性が発生する場合がある。また、市場に存在しない財についてはデータが存在しないため、分析に含めることができない。一方、表明選好法は、仮想的な質問に対する回答をデータとして用いるため、バイアスの影響を受けやすく、データの信頼性は相対的に低い。しかし、コンジョイント分析において直交デザインを使ってプロファイルを作成すれば属性間の相関をなくすことができる。また、市場に存在しない財についても分析に含めることが可能である。このように、顕示選好法のメリットは表明選好法のデメリットを補い、表明選好法のデメリットは顕示選好法のデメリットを補う。したがって、双方のデータを用いて分析を行うことで、双方の欠点を補完しあい、より信頼性の高い評価が可能になると考えられる。そこで、顕示選好データと表明選好データの両方を使用した分析が行われている。これは、RP-SP 結合

モデルなどと呼ばれる（栗山 他, 2013）。

RP-SP 結合モデルの代表的な手法の1つが、主にレクリエーション分野で使われている仮想行動法である。仮想行動法では、仮にサイトの環境の質が変化したらという仮想的な状況での訪問回数の変化を尋ね、その回答データ（表明選好データ）と実際のレクリエーション行動のデータ（顕示選好データ）の双方を分析に使用する。はじめにトラベルコスト法のシングルサイトモデルで環境変化前（現状）のレクリエーション需要曲線を推定し、次に回答者の回答にしたがってレクリエーション需要曲線をシフトすることで、環境変化後のレクリエーション需要曲線を推定する。そして、2つの需要曲線から、消費者余剰の増減を求める。これにより、シングルサイトモデルの枠組みで環境の質の評価が可能となる。仮想行動法を用いた研究には、Whitehead et al. (2000)や Whitehead et al. (2008)がある。

RP-SP 結合モデルのもう1つの代表的な手法が、共通してランダム効用モデルを用いるトラベルコスト法のサイト選択モデルや離散選択型のヘドニック価格法と選択実験の結合推定である（Ben-Akiva and Morikawa, 1990; 三谷 他, 2005）。この両者を結合して推定することで、仮想的な質問に対する回答をデータとして用いる選択実験の分析を単独で行う場合よりもデータの信頼性を高めることができる。また、属性間の相関がない選択実験のデータを加えることで、サイト選択モデルやヘドニック価格法を単独で用いる場合と比較して多重共線性の影響を緩和できる。さらに、サイト選択モデルやヘドニック価格法では扱うことができない、市場に存在しない財についても分析することができる。たとえば、Adamowicz et al. (1994, 1997)は、レクリエーション行動の分析にこの方法を用いている。また、Earnhart (2001, 2002)は、離散選択型のヘドニック価格法と選択実験を結合して、住宅選択行動の分析を行っている。RP-SP 結合モデルについて、より詳しくは Whitehead et al. (2011)を参照されたい。

6. おわりに

本稿では、農地や森林の生態系サービスの価値評価に適用可能な環境評価手法として、代替法、ヘドニック価格法、トラベルコスト法、CVM、コンジョイント分析を取り上げ、その経済理論と推定方法について解説を行った。

日本では、「生物多様性国家戦略 2012-2020」や「生物多様性及び生態系サービスの総合評価報告書」において、生態系サービスの経済評価を推進していくことが述べられている（環境省, 2012, 2016）。また、近年、自然の働きを利用したインフラや土地利用である「グリーンインフラストラクチャー（グリーンインフラ）」に対する関心が高まっている（西田・岩浅, 2015）。なかでも注目を集めているのが、自然を防災・減災に活用する「生態系を活用した防災・減災（Ecosystem-based Disaster Risk Reduction: Eco-DRR）」と呼ばれる取り組みである。生物多様性条約や気候変動枠組条約の締約国会議などでもグリーンインフラの推進が推奨されているほか、国内では「生物多様性国家戦略 2012-2020」、「森林・林業基本計画」、「第五次環境基本計画」などでグリーンインフラや Eco-DRR の活用が明記されている（環境省, 2012, 2018; 林野庁, 2016）。気候変動の進行により自然災害が増加する中、農地や森林をグリーンインフラや Eco-DRR として活用することに対する期待は今後ますます高まるものと考えられる。それに伴い、費用対効果の検討などのために、農地や森林の生態系サービスの経済評価に対するニーズも高まるものと考えられる。

今後は、拡大する社会的ニーズに応えるため、本稿で取り上げた環境評価手法を適用して、農地や森林の生態系サービスの経済評価を推進することが必要であると考えられる。また、本稿で述べた通り、いずれの環境評価手法にも課題があり、いまだ開発の途上である。今後は、信頼性の向上に向けたさらなる研究が求められる。

謝 辞

本研究は JSPS 科研費 16K03686, 16H03004, 16KT0029, 第 IV 期環境経済の政策研究「我が国における自然環境施策への効果的な資源動員に向けた研究」(研究代表者: 栗山浩一)の助成を受けたものです。

参 考 文 献

- Adamowicz, W., Louviere, J. and Williams, M. (1994). Combining revealed and stated preference methods for valuing environmental amenities, *Journal of Environmental Economics and Management*, **26**(3), 271-292.
- Adamowicz, W., Swait, J., Boxall, P., Louviere, J. and Williams, M. (1997). Perceptions versus objective measures of environmental quality in combined revealed and stated preference models of environmental valuation, *Journal of Environmental Economics and Management*, **32**(1), 65-84.
- Ben-Akiva, M. and Morikawa, T. (1990). Estimation of switching models from revealed preferences and stated intentions, *Transportation Research Part A: General*, **24**(6), 485-495.
- Boxall, P. C. and Adamowicz, W. L. (2002). Understanding heterogeneous preferences in random utility models: A latent class approach, *Environmental and Resource Economics*, **23**, 421-446.
- Chay, K. Y. and Greenstone, M. (2005). Does air quality matter? Evidence from the housing market, *Journal of Political Economy*, **113**(2), 376-424.
- Davis, L. W. (2004). The effect of health risk on housing values: Evidence from a cancer cluster, *American Economic Review*, **94**(5), 1693-1704.
- Earnhart, D. (2001). Combining revealed and stated preference methods to value environmental amenities at residential locations, *Land Economics*, **77**(1), 12-29.
- Earnhart, D. (2002). Combining revealed and stated data to examine housing decisions using discrete choice analysis, *Journal of Urban Economics*, **51**(1), 143-169.
- Haab, T. C. and McConnell, K. E. (2002). *Valuing Environmental and Natural Resources: The Econometrics of Non-market Valuation*, Edward Elgar Publishing, Cheltenham, UK.
- Hanemann, W. M. (1984). Welfare evaluations in contingent valuation experiments with discrete responses, *American Journal of Agricultural Economics*, **66**(3), 332-341.
- Hanemann, M., Loomis, J. and Kanninen, B. (1991). Statistical efficiency of double-bounded dichotomous choice contingent valuation, *American Journal of Agricultural Economics*, **73**(4), 1255-1263.
- 星野匡郎 (2011). 顕示選好法の最新テクニック 2: 空間ヘドニック法, 『環境評価の最新テクニック: 表明選好法・顕示選好法・実験経済学』(柘植隆宏, 栗山浩一, 三谷羊平 編著), 勁草書房, 東京.
- 環境省 (2012). 生物多様性国家戦略 2012-2020~豊かな自然共生社会の実現に向けたロードマップ~.
- 環境省 (2016). 生物多様性及び生態系サービスの総合評価報告書.
- 環境省 (2018). 第五次環境基本計画.
- 栗山浩一 (1998). 『環境の価値と評価手法—CVM による経済評価』, 北海道大学出版会, 北海道.
- 栗山浩一, 柘植隆宏, 庄子康 (2013). 『初心者のための環境評価入門』, 勁草書房, 東京.
- List, J. A. and Gallet, C. A. (2001). What experimental protocol influence disparities between actual and hypothetical stated values?, *Environmental and Resource Economics*, **20**(3), 241-254.
- Loomis, J. (2011). What's to know about hypothetical bias in stated preference valuation studies?, *Journal of Economic Surveys*, **25**(2), 363-370.
- McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior, *Frontiers in Econometrics* (ed. P. Zarembka), 105-142, Academic Press, New York.

- Millennium Ecosystem Assessment (2005). *Ecosystem and Human Well-being: Synthesis*, Island Press, Washington, DC.
- 三谷羊平 (2011). 実験経済学アプローチの新展開, 『環境評価の最新テクニック: 表明選好法・顕示選好法・実験経済学』(柘植隆宏, 栗山浩一, 三谷羊平 編著), 勁草書房, 東京.
- 三谷羊平, 栗山浩一, 庄子康 (2005). 離散選択モデルによる推定方法の新たな展開, 『環境と観光の経済評価 国立公園の維持と管理』(栗山浩一, 庄子康 編著), 勁草書房, 東京.
- Mitchell, R. C. and Carson, R. T. (1989). *Using Surveys to Value Public Goods: The Contingent Valuation Method*, Resources for the Future, Washington, DC.
- 三菱総合研究所 (2001). 地球環境・人間生活にかかわる農業及び森林の多面的な機能の評価に関する調査研究報告書.
- Murphy, J. J., Allen, P. G., Stevens, T. H. and Weatherhead, D. (2005). A meta-analysis of hypothetical bias in stated preference valuation, *Environmental and Resource Economics*, **30**(3), 313-325.
- 日本学術会議 (2001). 地球環境・人間生活にかかわる農業及び森林の多面的な機能の評価について (答申).
- 西田貴明, 岩浅有記 (2015). わが国のグリーンインフラストラクチャーの展開に向けて: 生態系を活用した防災・減災, 社会資本整備, 国土管理, 季刊 政策・経営研究, 2015, vol. 1, 46-55.
- 農業・農村の公益的機能の評価検討チーム (1998). 代替法による農業・農村の公益的機能評価, 農業総合研究, **52**(4), 113-138.
- Parmeter, C. F. and Pope, J. C. (2013). Quasi-experiments and hedonic property value methods, *Handbook of Experimental Economics and the Environment* (eds. J. A. List and M. K. Price), 3-66, Edward Elgar Publishing, Cheltenham, UK.
- Phaneuf, D. J. and Requate, T. (2017). *A Course in Environmental Economics: Theory, Policy, and Practice*, Cambridge University Press, New York.
- Phaneuf, D. J., Kling, C. L. and Herriges, J. A. (2000). Estimation and welfare calculations in a generalized corner solution model with an application to recreation demand, *Review of Economics and Statistics*, **82**(1), 83-92.
- Revelt, D. and Train, K. (1998). Mixed logit with repeated choices: Households' choices of appliance efficiency level, *Review of Economics and Statistics*, **80**(4), 647-657.
- 林野庁 (1972). 森林の公益的機能に関する費用分担及び公益的機能の計量, 評価ならびに多面的機能の高度発揮の上から望ましい森林について (中間報告).
- 林野庁 (2016). 森林・林業基本計画.
- Rosen, S. (1974). Hedonic prices and implicit markets: Product differentiation in pure competition, *Journal of Political Economy*, **82**(1), 34-55.
- Shaw, D. (1988). On-site samples' regression: Problems of non-negative integers, truncation, and endogenous stratification, *Journal of Econometrics*, **37**(2), 211-223.
- 庄子康, 星野匡郎, 柘植隆宏 (2011). 顕示選好法の新展開, 『環境評価の最新テクニック: 表明選好法・顕示選好法・実験経済学』(柘植隆宏, 栗山浩一, 三谷羊平 編著), 勁草書房, 東京.
- Small, K. A. and Rosen, H. S. (1981). Applied welfare economics with discrete choice models, *Econometrica*, **49**(1), 105-130.
- Smith, V. K. and Huang, J. C. (1995). Can markets value air quality? A meta-analysis of hedonic property value models, *Journal of Political Economy*, **103**(1), 209-227.
- Swait, J. (1994). A structural equation model of latent segmentation and product choice for cross-sectional revealed preference choice data, *Journal of Retailing and Consumer Services*, **1**, 77-89.
- 柘植隆宏, 庄子康, 栗山浩一 (2011a). 顕示選好法の最新テクニック 1: 端点解モデル, 『環境評価の最新テクニック: 表明選好法・顕示選好法・実験経済学』(柘植隆宏, 栗山浩一, 三谷羊平 編著), 勁

草書房, 東京.

- 柘植隆宏, 庄子康, 栗山浩一 (2011b). トラベルコスト法の研究動向, *環境経済・政策研究*, **4**(2), 46-68.
- Train, K. E. (2009). *Discrete Choice Methods with Simulation*, 2nd ed., Cambridge University Press, Cambridge, UK.
- von Haefen, R. H., Phaneuf, D. J. and Parsons, G. R. (2004). Estimation and welfare analysis with large demand systems, *Journal of Business & Economic Statistics*, **22**(2), 194-205.
- Whitehead, J. C., Haab, T. C. and Huang, J. C. (2000). Measuring recreation benefits of quality improvements with revealed and stated behavior data, *Resource and Energy Economics*, **22**(4), 339-354.
- Whitehead, J. C., Dumas, C. F., Herstine, J., Hill, J. and Buerger, B. (2008). Valuing beach access and width with revealed and stated preference data, *Marine Resource Economics*, **23**(2), 119-135.
- Whitehead, J. C., Haab, T. C. and Huang, J.-C. (eds.) (2011). *Preference Data for Environmental Valuation: Combining Revealed and Stated Approaches*, Routledge, New York.

Economic Valuation Methods for Ecosystem Services Provided by Agricultural Land and Forest

Takahiro Tsuge

Faculty of Economics, Konan University

Agricultural land and forests not only supply agricultural crops and timber, but also play various roles in land conservation, cultivation of water sources, climate stabilization, prevention of global warming, conservation of biodiversity, and opportunities for recreation. To recognize the importance of these ecosystem services, it is effective to visualize their value by evaluating them on a monetary term. However, most ecosystem services are not traded on the market, and real prices do not exist; therefore, they cannot be evaluated based on price. Instead, specialized methods called environmental valuation methods are used to determine their value. In this paper, we explain the economic theory and estimation methods of the replacement cost method, hedonic price method, travel cost method, contingent valuation method (CVM), and conjoint analysis. These methods are applicable to the valuation of ecosystem services provided by agricultural land and forests.

個人企業向けクレジットスコアリングモデルに おける業歴の有効性

尾木 研三¹・内海 裕一¹・枇々木 規雄²

(受付 2018 年 9 月 9 日；改訂 2019 年 3 月 16 日；採択 3 月 25 日)

要 旨

日本の多くの銀行は、融資審査業務の一部を自動化して審査コストを削減するために、クレジットスコアリングモデルを用いている。モデルにはさまざまなタイプがあるが、従業員数 20 人以下の小企業の信用力評価は、財務指標とデフォルトとの相関を利用したロジスティック回帰モデルの利用が一般的である。小企業には法人企業だけでなく個人企業も含まれるが、個人企業向けモデルの精度は期待するほど高くはない。主な理由は、個人企業は財務諸表を作成する義務がないため、資産や負債といった B/S に関する情報が不足しているからである。精度を上げるには、資産や負債に関する客観的な説明変数の使用が有効と思われるが、個人企業の大半は、事業主世帯の私的資産と事業の会計が混同している場合が多く、正確なデータを得ることが難しい。

尾木 他 (2016) は、小規模な法人企業向けモデルに業歴を用いることを提案し、業歴が重要な変数であることを示した。その理由は次の 3 つである。第 1 に、小企業の評価では、経営者の私的資産を企業の資産に追加することが効果的であるが、そのような情報は容易かつ正確に入手できない。第 2 に、小企業の実績デフォルト率が、業歴が長いほど低いのは、年々蓄積される経営者の私的資産が経営を支えるからと考えられる。第 3 に、経験豊富な銀行員は、融資審査で業歴を債務者の信用リスク評価の重要な要素に位置付けている。

本論文では、個人企業向けモデルにおいても業歴を用いることを提案する。そして、2007 年から 2014 年の間に日本政策金融公庫が融資した 68 万以上の個人企業のデータセットを利用して、業歴とデフォルト率の相関関係を分析する。本論文では、測定値として一般的な AR 値でモデルを評価する。分析の結果、業歴別デフォルト率はワイブル分布の密度関数もしくは「0-3 年」「4-25 年」「26 年以上」の 3 つの区分線形関数で定式化できることがわかった。いずれの関数を用いても、AR 値を約 9% 改善させることができ、実務的観点からモデルが有効であることを確認した。

キーワード：クレジットスコアリングモデル，クレジットリスク，個人企業，業歴，ロジスティック回帰モデル，ワイブル分布。

1. はじめに

金融機関は信用リスク管理だけでなく、融資審査業務の一部を自動化して審査コストを削

¹ 株式会社日本政策金融公庫 国民生活事業本部：〒100-0004 千代田区大手町 1-9-4

² 慶應義塾大学 理工学部：〒223-8522 横浜市港北区日吉 3-14-1

減するために、デフォルト確率(Probability of Default : PD)を推定するクレジットスコアリングモデルを使用している。モデルには、企業の財務諸表データを用いて PD を推定する統計モデル、株価を用いて推定する構造型モデル、社債価格を用いて推定する誘導型モデルなど、さまざまなタイプがある。しかし、中小企業のほとんどは株式や社債を発行していないため、中小企業を対象とする場合は、統計モデルを用いることが多い。なかでも金融機関の間では、ロジスティック回帰モデルが普及している。173 の金融機関(2018 年 4 月現在)が加入する一般社団法人 CRD 協会や、22 の銀行等が共同出資して設立した日本リスクデータバンク株式会社もロジスティック回帰モデルを提供している。

日本政策金融公庫国民生活事業本部(以下、公庫という)は、主に従業者数が 20 人未満の小企業に対する融資を行っている。公庫もロジスティック回帰モデルを使用しているが、融資先の半数近くは個人企業である。個人企業とは、法人としての登記をしていない個人が経営する経営体のことであり、税務申告においては、青色申告もしくは白色申告している個人事業者のことである¹⁾(なお、本論文では個人企業と法人企業の経営者の混同を避けるため、個人企業に対しては「事業主」、法人企業に対しては「経営者」と呼ぶことにする)。

個人企業は、法人が作成するような決算書はなく、その代わりとなる税務申告書においても、貸借対照表(B/S)を作成しているケースが少ないうえ、資産や負債に関する情報の開示にも消極的である。したがって、融資審査にあたっては、金融機関同士が情報を共有するデータベースから負債に関する情報を取得し、審査担当者による事業主からのヒアリングや提出書類、帳簿、通帳などから資産に関する情報を把握して、簡便的に B/S を作成して評価している。このように、個人企業は審査の手間とコストがかかるうえ、法人企業に比べて融資金額が小さく採算ベースに乗りにくいいため、スコアリングモデルの精度向上による審査業務の一層の効率化は金融機関にとって長年の課題となっている。

先行研究をみると、尾木 他(2016)は小規模な法人企業向けモデルに業歴を用いることを提案し、業歴が重要な変数であることを示した。その理由は次の 3 つである。第 1 に、小企業の評価では、経営者の私的資産を加味することが効果的であるが、そのような情報は容易かつ正確に入手できない。第 2 に、小企業の実績デフォルト率が、業歴が長いほど低いのは、年々蓄積される経営者の私的資産の大きさが影響していると考えられる。第 3 に、経験豊富な銀行員は、業歴を債務者の信用リスク評価の重要な要素として扱っている。

さらに、業歴別デフォルト率は、決算書の数値には表れない小企業のパフォーマンスを示す指標と考えられるが、その一つとして経営者の私的資産の大きさが含まれ、大きなウェイトを占めている可能性があることも示した。具体的には、個人企業の業歴別デフォルト率と事業主の保有する預貯金額あるいは不動産評価額との相関係数がいずれも約 0.8 と高いことを明らかにしている²⁾。分析の目的は法人企業といえども、小企業の場合には経営者の私的資産額が重要なファクターと考えられる一方で、法人経営者の私的資産額のデータが存在しないため、個人企業のデータを用いて検証したとしている。

一連の分析の結果、尾木 他(2016)は、財務指標に加えて、経営者の私的資産を含む小企業のパフォーマンスを示す指標として、業歴別デフォルト率を定式化した 3 次関数を変数として投入すると、法人企業向けモデルの精度を示す AR 値³⁾が改善することを示した。

個人企業の場合、事業資産と事業主の私的資産が混在しており、それを明確に区分することは難しいが、個人企業の業歴別デフォルト率を、非開示となっている事業主の私的資産を含む個人企業のパフォーマンスを示す指標として利用できれば、それを定式化した業歴関数をモデルの変数に含めることによって、法人企業と同様に個人企業向けモデルの精度を改善できる可能性があると考えた。そこで、本研究では尾木 他(2016)の研究を参考にし、個人企業の業歴別デフォルト率を定式化した業歴関数の有効性について検証することを試みる。分析の結果、

以下の点が明らかになった。

- (1) 個人企業の業歴別デフォルト率の形状は、法人企業とは異なることがわかった。具体的には、法人企業と同様に業歴 25 年頃まではデフォルト率が単調減少するが、業歴 35 年頃からデフォルト率が上昇するという特徴は個人企業ではみられず、業歴 25 年以降は概ね一定であった。
- (2) 業歴関数は法人企業で有効性が確認された 3 次関数よりも、ワイブル分布の密度関数もしくは、業歴が「0-3 年」「4-25 年」「26 年以上」で 3 区分した区分線形関数のいずれかを用いることが有効であり、どちらを用いてもパフォーマンスに大きな差はないことを確認した。
- (3) 業歴関数をモデルの変数に投入すると、医療・福祉業、不動産業、運輸業を除く業種で AR 値が改善し、特に飲食・宿泊業においては約 9% ポイント改善することを確認した。
- (4) 個人企業全体の AR 値の改善幅は 2.3% ポイントから 2.6% ポイントと、法人の 7.6% ポイントに比べて小さいことがわかった。要因として、開示されている事業資産に事業主の私的資産が相応に含まれている（非開示の資産が少ない）、もしくは、非開示の私的資産は事業に投入しない傾向にあるなどの理由により、業歴関数の効果が期待したほど高くならなかったと考えられる。

なお、本研究では、デフォルトの定義は破綻懸念先以下へのランクダウンおよび 3 か月以上延滞先とし、デフォルト率は融資企業数のうち融資した翌年度末までにデフォルトした企業の割合とする⁴⁾。ただし、廃業に伴い融資残高を一括返済した企業については成功企業と考え、デフォルト企業に含めていない。

本論文の構成は以下の通りである。2 章では、関連する先行研究を概観する。3 章では、個人企業の概要と特徴を述べ、4 章はモデルおよび使用データの概要と業歴別デフォルト率の特徴を確認する。5 章では、個人企業の業歴別デフォルト率の定式化を試みる。6 章では業歴関数のモデル投入によるパフォーマンスを検証し、7 章では結論と今後の課題を述べる。

2. 先行研究

デフォルト確率や倒産確率を推計するモデルの研究にはきわめて多くのものがあり、さまざまな手法が提案されている。山下 他 (2003) によれば、これらは、貸出先企業の財務データから判別分析や回帰分析などの統計手法を用いて推定する統計モデルと株価や社債金利などの市場データからオプション理論を用いて推定するオプションアプローチに大別される。この他にも、サポートベクターマシーン (SVM) やニューラルネットワークといった機械学習的な手法を用いて推定する方法がある。

本論文は、個人企業を対象にし、金融機関の間で最も普及しているロジスティック回帰モデルにおいて、その変数として業歴別デフォルト率を定式化した業歴関数の投入により AR 値が向上するかどうかを検証することを目的としている。そこで本章では、2.1 節でロジスティック回帰モデル、2.2 節では個人企業、2.3 節では業歴とデフォルトとの関係に関する先行研究について概観する。

2.1 ロジスティック回帰モデルに関する研究

ロジスティック回帰モデルに関する研究は、非常に盛んに行われている。Ohlson (1980) が行ったロジスティック回帰モデルを使って倒産確率を推定する研究に始まり、Altman and Sabato (2007) の中小企業の 1 年間のデフォルト確率を推定する研究など、国外には数多くの研

究がある。国内においても数多くの研究が行われている。柳澤 他 (2007) は、2000 年度から 7 年間の約 35,000 件の RDB (Risk Data Bank) データを用いて、複数のロジスティック回帰モデルを構築し、AR 値の推移について検証している。森平・岡崎 (2009) は上場企業の財務データと景気動向指数や日経平均株価、原油先物価格指数などを用いて、マクロファクターを加味した期間構造型のモデルの有効性を提案している。山下 他 (2003) は、スコアリングモデルの評価手法の特徴などについて詳細な分析を行っている。この他、変数選択とデータ量に関する研究 (山下 他, 2003)、財務変数に加えて非財務変数として業歴の有効性を検証した研究 (枇々木 他, 2010)、創業企業の信用リスクを評価するスコアリングモデルに関する研究 (尾木 他, 2017) などもある。

多数の研究に加えて、データの量と質の向上によって、モデルの精度は着実に上がってきている。今やほとんどの金融機関が統計モデルを使って中小企業のクレジットリスクを定量化し、審査やリスク管理などの実務に活用している。最近では、機械学習の手法を用いた人工知能 (AI) 型スコアリングモデルを使った審査に取り組む金融機関も増えており、実務におけるスコアリングモデルの活用範囲は着実に広がっている。ただ、尾木 他 (2017) を除いて、いずれの研究も個人企業ではなく、法人企業を対象にしている。

2.2 個人企業に関する研究

前述のように、個人企業のスコアリングモデルに関する先行研究はほとんど見当たらない。ただ、個人企業に関する調査・研究はいくつか存在する。国民生活金融公庫総合研究所 (2002) は、「自己雇用者 (≡ 自営業者) に関する実態調査」を行い、個人企業の事業主と法人企業の経営者の違いについて分析している。上村 (2006) は、個人企業世帯の家計資産について、国勢調査や貯蓄動向調査をもとに分析し、個人企業世帯の家計は複雑・不安定で把握しにくいことを述べている。高川・亀田 (2008) は、事業所・企業統計調査や国勢調査などをもとに個人企業の企業数が減少している理由について、高齢化や規制緩和による法人との競争激化にあることを明らかにした。

Schreiner (2000) は、個人企業の信用評価は、定性情報の依存度が高い一方で、スコアリングは定量情報に依存しているため、貸し手がスコアリングモデルを使って業務の効率化を図ることは容易ではないことを述べている。Kneiding and Kritikos (2011) は、ドイツの個人企業世帯の資金移転について所得と消費に関する調査データを使って分析し、個人企業世帯はビジネスの目的で消費者ローンを使用しており、家計と事業の資金移転が行われている実態を明らかにした。Gweyi (2013) は、個人企業はデフォルトリスクが非常に高く、ケニアの銀行は、マイクロファイナンスの分野に進出することが困難である現状を解説している。以上のように、個人企業向けスコアリングモデルに関する研究はほとんどない。個人企業は家計と事業の会計が混在していて、実態が掴みにくいことがわかる。本研究は、個人企業向けのスコアリングモデルの変数として、業歴別デフォルト率を定式化した業歴関数を利用する点に特徴がある。

2.3 業歴と倒産およびデフォルトとの関係に関する研究

業歴に関する研究は近年盛んに行われ、さまざまな研究成果が発表されている。まず、業歴と企業のパフォーマンスとの関係を分析した研究を紹介する。Cabral and Mata (2003) は、ポルトガルの製造業のデータを用いて、業歴が長いほど企業規模が大きくなることを示した。Sakai et al. (2010) は、一般社団法人 CRD 協会が保有する 20 万社以上のパネルデータを使用して中小企業の業歴と借入コストとの関係を分析し、業歴が長くなるほどパフォーマンスが上がり、借入コストが下がることを明らかにした。一方で、森川 (2008) は、従業員数 50 人以上の法人企業約 5 千社のデータを用いて同族企業の生産性について分析し、その中で企業年齢と生

産性が有意に負の関係にあるという結果を示した。Loderer and Waelchli (2010)は、米国の上場企業においては、業歴が長いほど組織の硬直化や研究開発活動の低下が進行し、ROA(総資産利益率)やTobinのqのパフォーマンスが悪化することを述べている。また、Akben-Selcuk (2016)は、302社のトルコの企業を対象に、業歴とROAやROEといった収益性との関係を分析し、業歴が長くなるほど収益性が低下することを明らかにした。このように、業歴が長いほど企業のパフォーマンスが上がるという研究がある一方で、パフォーマンスが下がるとする研究もある。

次に、業歴とデフォルトもしくは倒産との関係を分析した研究をいくつか紹介する。清水(1985)は、企業年齢と倒産について分析し、倒産する企業数だけを見ると、企業年齢が3~10年未満の企業が倒産しやすく、10年以上の企業は倒産し難いものの、生存する企業との比較では単純にそうとも言えないことを指摘している。Evans(1987)は米国の中小企業は、業歴が長いほどデフォルトが減ると述べている。一方、中小企業庁の中小企業白書(2002)は、東京商工リサーチのデータを用いて業歴とデフォルトとの関係についてプロビット分析し、業歴はデフォルトを説明する有意な変数にはならないとしている。ただ、白書のなかに記載されている生存企業と倒産企業の業歴別構成比を用いて、その比率(倒産企業のうち業歴a年の企業の割合÷生存企業のうち業歴a年の企業の割合)を計算すると業歴と倒産率との関係は、業歴が長いほど倒産率が低くなるが、この点は指摘されていない。

また、鶴田(2005)はノンバンクの融資先を対象に、業歴が長いほど経営は安定しているという仮説を検証した。債務超過企業をデフォルト企業とした場合は有意ではないが、インタレストカバレッジレシオが1未満の企業をデフォルト企業とした場合には有意になることを示している。東京商工リサーチの友田(2008)は業歴別倒産件数構成比の推移を示し、業歴30年以上の老舗企業の構成比が年々上昇し、業歴10年未満の新興企業の構成比が下がっていることを示している。枇々木他(2010)は、日本政策金融公庫が保有する約48万件のデータを用いて、小企業の業歴とデフォルトとの関係は3次関数で表現できることを明らかにした。このように、業歴とデフォルトとの間に関係があるかどうかについては、必ずしも明確になっていない。さらに、個人企業の業歴とデフォルトとの関係を分析した研究は、筆者たちの知る限り存在しない。

3. 個人企業の概要と特徴

本章では、個人企業の概要と特徴について、国内の個人企業の実態に関する先行研究をもとに、法人企業との比較を交えながら概観する。

3.1 個人企業の概要

個人企業の一般的なイメージは、商店主や工場主など、夫が事業主で妻が家族従業者という形である。具体的な数値でみてみよう。

平成26年の経済センサス(総務省統計局, 2016)をみると、個人企業の事業所数の割合は、38.4%と全体の約4割を占める。業種構成を表1に示す。卸・小売業の割合が最も高く21.8%となっている。法人企業と比較すると、飲食業の割合が、法人企業が6.1%であるのに対し、個人企業が19.4%と高くなっているほか、理美容業の割合が、法人企業が3.6%であるのに対し、個人企業が14.4%と高いという点が特徴的である。製造業の減少とともに、工場主は減り、現在では商店主、あるいは、飲食店や理美容店の店主が多くなっている。

次に、個人企業の規模について確認する。2002年に実施したものであることには注意を要するが、表2に国民生活金融公庫総合研究所(現:日本政策金融公庫総合研究所)(2002)「自己雇

表 1. 個人企業の業種構成比。数値は事業所数に対する割合。値の小さな業種は除いている。したがって、合計は 100% にならない。出典：総務省統計局 (2016) 「平成 26 年経済センサス-基礎調査」。

業種	個人企業	法人企業
建設業	7.2%	10.7%
製造業	6.9%	10.0%
情報通信業	0.1%	1.9%
運輸業, 郵便業	0.9%	3.4%
卸売業, 小売業	21.8%	27.8%
金融業, 保険業	0.3%	2.4%
不動産業, 物品賃貸業	7.0%	6.9%
学術研究, 専門・技術サービス業	4.9%	3.7%
宿泊業, 飲食サービス業	20.8%	8.4%
(飲食店)	(19.4%)	(6.1%)
生活関連サービス業, 娯楽業	15.1%	4.9%
(理美容業)	(14.4%)	(3.2%)
教育, 学習支援業	4.4%	2.2%
医療, 福祉	8.2%	7.2%
サービス業 (他に分類されないもの)	2.4%	8.5%

表 2. 個人企業の概要(法人企業との比較)。出典：国民生活金融公庫総合研究所 (2002) 「自己雇用者(≒自営業者)に関する実態調査」(以下, 図 1~図 3 も同じ)。

	平均 従業員数	平均売上高 (年間)	創業者の 割合	経営者の 平均年齢
個人企業 (n=2,380)	1.9人	2,116万円	84.5%	53.8歳
法人企業 (n=1,426)	36.9人	8億6,314万円	55.0%	56.9歳

用者(≒自営業者)に関する実態調査」(以下, 公庫調査という)の結果を示す。平均従業員数は本人を含めて 1.9 人, 配偶者が家族従業員となっている割合は 57.2% で, 夫婦で事業を営んでいる企業が大半である。平均売上高は年間 2,116 万円と, 法人企業の 8 億 6,314 万円に比べて規模はかなり小さい。また, 経営者または事業主が創業者である割合は, 法人企業が 55.0% であるのに対し, 個人企業が 84.5% となっており, 事業承継が少ないことがわかる。2014 年版中小企業白書(中小企業庁, 2014)によると, 廃業した企業のうち個人企業が約 9 割を占めている。日本政策金融公庫総合研究所が 2016 年に行った「中小企業の事業承継に関するインターネット調査」(日本政策金融公庫総合研究所, 2016)でも廃業予定企業の 68.4% が個人企業であることから, 法人企業に比べて一代限りで廃業する企業が多いことがわかる。

3.2 個人企業の特徴

3.1 節のとおり, 個人企業の平均像は小売店, 飲食店, 理美容店などを営む経営者と配偶者で構成される年商 2 千万円程度の小規模な企業であることがわかった。本節では, 個人企業の特徴についてさらに 4 つの視点から確認する。

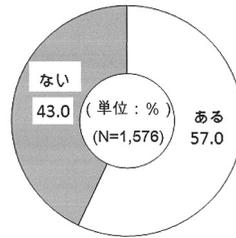


図 1. 個人企業世帯の別収入。自己雇用者(個人事業主)本人、配偶者、同居家族のそれぞれについて、別収入の有無を尋ね、集計したものである。出典：表 2 に同じ。

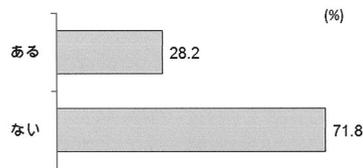


図 2. 個人企業経営者の副業の有無。出典：表 2 に同じ。

(1) 家族経営で別収入がある

内閣府(2011)「年次経済財政報告」によると、労働政策研究・研修機構「副業者の就労に関する調査」を用いて、副業の日数を本業の就業形態や性別などの要因で回帰分析した結果、個人企業の事業主とその家族従業者は、正社員に比べて、1カ月あたり3日程度副業従事日数が多いということがわかった。公庫調査でも、図1のとおり、本人も含めた家族に別収入がある割合は57%となっている。

事業主自身が他に本業を持っていることもある。図2のとおり、個人企業の事業主自身に別収入がある割合は28.2%となっている。たとえば、プロゴルファーが経営するゴルフショップや勤務者が経営する不動産賃貸業など、別収入の方が多い場合もある。個人企業は規模が小さいため、事業の赤字を別収入で補てんする企業は少なくない。

(2) 家計と事業の会計が混在している

上村(2006)は「個人企業(自営業)の家計は事業経営との関係が緊密で、勤労者の家計に比べて複雑・不安定で把握がしにくく、その内部構造を明らかにした研究は少ない」「全くの井勘定で家族員の生活に必要な費用がその日の家業の売上から支出されてしまうタイプもあろう」と述べている。前(1)の特徴も踏まえると、個人企業は、別収入があるうえ帳簿類が未整備で、家計と事業の会計が混在していることがわかる。

(3) 業種や経営形態が多様である

平成23年の年次経済財政報告に、個人企業は「日銭を稼いで暮らしていくための「生業」がある一方、イノベーションのけん引役としての「ベンチャービジネス」がある」という記述がある。一口に個人企業といっても、大宗を占める商店や飲食店、理美容店だけでなく、開業医、弁護士などの専門職業、在宅ワークやネット関連ビジネスといったものまで、多様な業種や形態がある。

(4) 法人成りする

個人企業はデフォルトや廃業以外に、法人化することによっても企業数が減る。企業規模を維持したまま個人企業として事業を継続する企業も少なくないが、一般に、規模が大きくなる

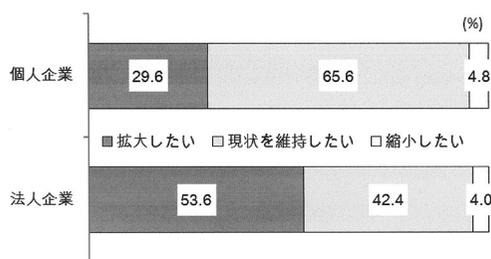


図 3. 事業規模の拡大意欲. 出典: 表 2 に同じ.

と法人化する傾向がある. 図 3 のとおり, 公庫調査によると, 個人企業の拡大志向は法人企業より低いものの, 3 割程度は事業規模の拡大に意欲的であることがわかる.

4. モデル構築手順の概要とデータの特徴

枇々木 他 (2010) や尾木 他 (2016) は, 小規模な法人企業向けモデルにおける業歴の有効性について示した. 本章では, 尾木 他 (2016) の研究をもとに, モデル構築手順およびデータの概要と業歴別デフォルト率の特徴について述べる.

4.1 モデル構築手順の概要

尾木 他 (2016) が示した業歴を考慮したロジスティック回帰モデル (以下, 業歴モデルという) は, 財務指標を用いた一般的なモデル⁵⁾ をベースに, 説明変数として業歴別デフォルト率の形状を定式化した業歴関数 (3 次関数) を加えたものである.

尾木 他 (2016) が示した業歴関数 $\sum_{k=1}^3 \beta_k (g_i)^k (i = 1, \dots, N)$ を加えた法人企業向けの業歴モデルを (4.1) 式で示す. ここで g_i は企業 i の業歴 (61 年以上は 61 年) である.

$$(4.1) \quad p_i = \frac{1}{1 + e^{z_i}}, \quad z_i = \ln \left(\frac{1 - p_i}{p_i} \right) = \alpha_0 + \sum_{j=1}^J \alpha_j f_{ij} + \sum_{k=1}^3 \beta_k (g_i)^k \quad (i = 1, \dots, N)$$

尾木 他 (2016) が示した業歴モデルの特徴は, (4.1) 式の右辺第 3 項の業歴関数にある. すなわち, 個人企業の業歴モデル構築のポイントは, 業歴別デフォルト率をどのように定式化するかにある. 本研究におけるモデルは, (4.1) 式のモデルをベースに, 右辺第 3 項の業歴関数にさまざまなタイプの関数の当てはめを行い, 修正 R^2 によるフィッティングの良さ, AR 値を使ったモデルの精度を比較検討しながら, モデルを構築する.

4.2 使用データの概要

表 3 に本研究で使用する個人企業データの概要を示す. まず, 年度別の企業数については, 対象となる融資年度は異なるが, 特徴を比較するために, 法人企業データを用いた尾木 他 (2016) の分析データおよび結果も示す⁶⁾. 本研究で使用するデータは公庫が 2007 年度～2014 年度の 8 年間に融資した個人企業の延べ 684,994 社の年度別データである. また, 尾木 他 (2016) の分析データは, 2004 年度～2011 年度の 8 年間に融資した法人企業, 延べ 1,089,362 社⁷⁾ である.

年度ごとの企業数は重複を許している. 例えば, 業歴 n 年の企業が x 年度と $x+3$ 年度に 1 回ずつ合計 2 回の融資を受けた場合, x 年度は業歴 n 年の企業, $x+3$ 年度は業歴 $n+3$ 年の企業としてカウントしている. 延べサンプルを使用しているのは, 同一の企業でも, 時点が異なれば, 企業の財務内容, 経営者もしくは事業主の資質や私的資産額などが異なるためである.

表 3. 使用データの概要.

(年度別データ数)

年度	2007	2008	2009	2010	2011	2012	2013	2014	合計
個人企業数	88,053	93,779	99,999	95,642	83,210	77,590	74,762	71,959	684,994
年度	2004	2005	2006	2007	2008	2009	2010	2011	合計
法人企業数	133,643	127,496	123,174	130,821	140,181	162,616	147,718	123,713	1,089,362

(業種別データ数)

	製造	建設	卸小売	飲食宿泊	医療福祉	サービス	不動産	運輸	その他
個人企業	44,305	111,876	145,022	85,801	43,979	123,725	47,148	14,054	35,527

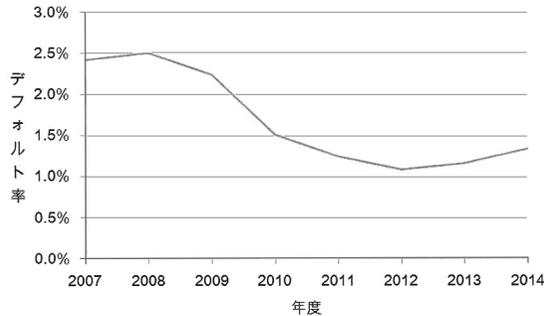


図 4. 年度別デフォルト率.

次に、業種別のデータ数をみると、卸・小売業が約 14 万 5 千件と最も多く、次にサービス業が約 12 万 3 千件、建設業が約 11 万 1 千件と続く。製造業と医療・福祉業は 4 万件台、運輸業は約 1 万 4 千件と業種間の差が大きくなっている。

最後に、図 4 で年度別のデフォルト率を見てみると、リーマンショックがあった 2008 年度前後のデフォルト率が 2% 台と高くなっている。2010 年度以降は、おおむね 1%~1.5% の間で推移している。デフォルトや倒産は、企業個別の要因だけではなく、マクロ経済要因の影響も受けるため、時系列での分析も重要であるが、データの制約もあり、この点については今後の課題と認識している。

4.2.1 業歴別企業数の特徴

業歴別企業数を図 5 に示す。業歴が長くなるほど個人企業数は減少する。業歴 15 年頃から 30 年頃にかけてやや減少のペースが遅くなるものの、業歴 15 年ぐらいまでと、業歴 30 年頃から 50 年頃にかけて、おおむね一定のペースで減少する。創業時に個人企業であっても、業歴が長くなるにつれて法人化する企業が増えるため、個人企業数は一貫して減少する傾向にある。業歴 15 年頃から 30 年頃にかけては、法人化する企業が一段落するため、減少のペースが緩やかになるが、事業承継期もしくは廃業時期を迎えて業歴 30 年以降は再び減少のペースが上がると考えられる。一方、法人企業は、尾木 他 (2016) が示したとおり、業歴 10 年ぐらいまで低下したあと、しばらく安定した期間が続くが、業歴 40 年ぐらいから大きく減少する。なお、業歴が 60 年を超えると企業数が大幅に減少するため、図 5 では示していないが、業歴 100 年以上の個人企業は 3,500 社程度存在する。

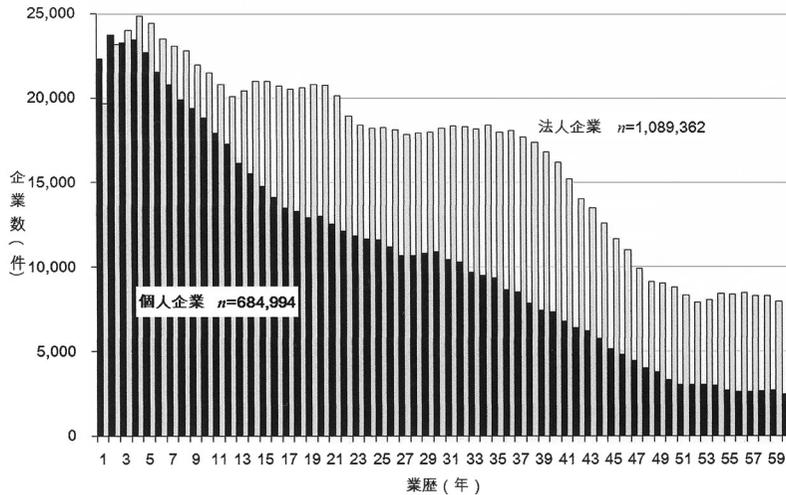


図 5. 業歴別企業数.

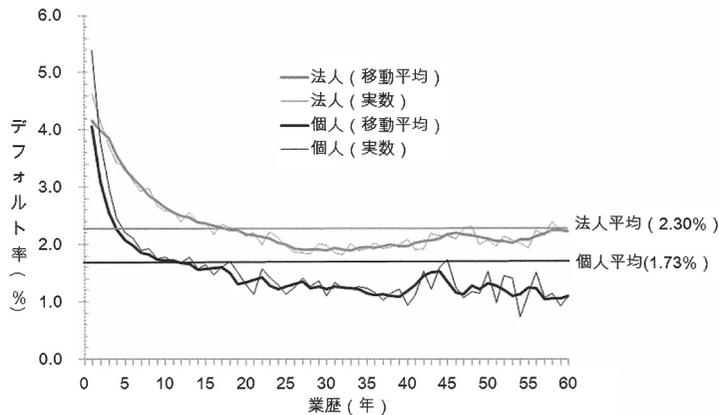


図 6. 業歴別デフォルト率.

4.2.2 業歴別デフォルト率の特徴

尾木 他 (2016)の分析した法人企業は 2004 年度から 2011 年度までのサンプルをそれぞれプールした状態でデフォルト率を計算している。本研究でも同様の条件で、2007 年度から 2014 年度までの個人企業のサンプルを使って計算した。

デフォルト率を業歴別に計算した結果を図 6 に示す。デフォルト率の傾向を示すために、前後 2 年を含む 5 年間 (-2, -1, 0, +1, +2 年)のデフォルト率を平均した移動平均(太線)⁸⁾も示す。平均値をみると、個人企業が 1.73% であるのに対して法人企業は 2.30% と、個人企業の方が低い。観測期間が異なる影響を除去するため、共通する年度(2007~2011 年度)の結果も表 4 に示す。同じ年度で比べても、個人企業は法人企業に比べてデフォルト率が低い傾向にある。

業歴とデフォルト率との関係を見ると、個人企業、法人企業とも業歴 5 年ぐらいまでの企業のデフォルト率は高く、業歴 15 年ではほぼ平均まで低下(個人 1.64%, 法人 2.38%), その後も低下を続けて業歴 25 年から 35 年ぐらいまではおおむね横ばいとなる。

表 4. 個人・法人企業別年度別デフォルト率.

年度	2007	2008	2009	2010	2011
個人企業	2.42%	2.50%	2.24%	1.51%	1.25%
法人企業	3.75%	3.20%	2.22%	1.54%	1.57%

35年頃(個人1.23%, 法人1.93%)から法人企業のデフォルト率は徐々に上昇しはじめ、45年(2.18%)を過ぎると再び低下する。一方、個人企業は44年と45年に高い時期もあるが、これを除くと40年を過ぎてもおおむね横ばいか、やや低下傾向にある。業歴別デフォルト率の動きにおける個人企業と法人企業との違いについて、現場感覚や中小企業研究における指摘などと整合的かという視点を交えながら、業歴区間ごとの特性を詳しくみていくことにする。

(1) 業歴5年未満

業歴1年のデフォルト率は、個人企業が5.39%、法人企業が4.63%といずれも高い水準にある。開業後5年を過ぎるまではデフォルト率が高い時期がしばらく続くことがわかる。安田他(2007)は、開業したての経営者には資質に欠く者が一定数含まれており、開業後間もない時期にこうした企業が自然淘汰されるため退出率が高くなると指摘している。

個人企業を法人企業と比較すると、法人企業は業歴5年に3.37%に低下するのに対し、個人企業は2.20%と半分以下の水準まで急激に低下する。低下のペースは個人企業の方が早いことがわかる。この理由はいくつか考えられる。一つ目は、3.2節(4)のとおり、事業が拡大すると法人化するが、拡大期はデフォルトしやすい傾向があるため、法人成りした後でデフォルトしている企業があると考えられる。図5をみると、業歴5年未満の企業数は、個人企業が横ばいであるのに対し、法人企業は5千社程度増加しており、個人企業が法人成りしている可能性がある。二つ目に、個人企業は規模が小さく小回りが利くため、事業が一旦軌道に乗るとデフォルトしにくい傾向にあるということである。3.1節のとおり、飲食店や理美容店のような規模が小さく現金商売の業種が多いうえ、3.2節(1)(2)のとおり、事業収入の減少を家族の別収入で補ったり、経営者自らが副業を行ったりするケースは珍しくない。個人企業のデフォルト率が法人企業のデフォルト率より低い理由もこの点にあると考えられる。

(2) 業歴5年以上15年未満

徐々にデフォルト率が低下していく。生き残った企業の経営者は経営経験を積むことによって経営ノウハウを身につけるため、業歴10年を過ぎるころからデフォルト率が平均に近づき、15年ではほぼ平均の水準になる。中小企業白書(2002)では、誕生期の危機を乗り越えた開業者は、十数年で既存事業者との差がさほど見られなくなるという分析がある。

個人企業もデフォルト率が低下していく傾向は法人企業と同じであるが、デフォルト率低下のペースが緩やかである。拡大志向のある個人企業は業歴10年ぐらいいままでに法人化し、規模の小さな個人企業が残るため、先述のとおり、小回りの利く経営でデフォルトしにくいと考えられる。

(3) 業歴15年以上35年未満

デフォルト率が平均を下回り、個人企業は業歴25年頃まで、法人企業は業歴28年頃までデフォルト率が低下して、それ以降はしばらく横ばいとなる。

(4) 業歴35年以上50年未満

この業歴区間の個人企業は法人企業と大きな違いがある。その違いを明らかにするためにまず、法人企業の特徴およびその理由を説明する。

法人企業は業歴35年を過ぎたころからデフォルト率が少しずつ上昇し始め、45年ごろにピークを迎え、その後、また下降する。日本政策金融公庫総合研究所の新規開業実態調査(2017)で

は、開業時の経営者の平均年齢は 42.6 歳となっており、業歴 35 年の経営者の年齢は 70 歳代後半になる。事業承継を考える時期であるが、小企業は後継者難という構造問題を抱えており、事業承継の失敗によるデフォルトが増えている可能性がある。

一方、個人企業のデフォルト率は 44 年(1.61%)と 45 年(1.73%)にやや高い時期もあるが、これを除くと 40 年を過ぎてもおおむね横ばいである。その理由としては、個人企業の業歴別企業数をみると、業歴が長くなるほど企業数が減少していることから、①表 2 のとおり、大半の個人企業においては事業承継が行われていない、②事業承継する企業は拡大志向があり、法人化する、ことが背景にあると思われる。つまり、個人企業の組織形態を維持しているということは、そもそも事業承継を考えていないと思われる。

(5) 業歴 50 年以上

法人企業は業歴 45 年ごろから再びデフォルト率が低下に転じたあと、55 年ごろをボトムに再び上昇するが、事業承継が少ない個人企業にはこの傾向はみられない。もっとも、個人企業の企業数はかなり少なくなるため評価が難しい。

5. 業歴別デフォルト率の定式化

4 章で議論した個人企業の業歴別デフォルト率の特徴を踏まえて、本章では業歴別デフォルト率の定式化を試みる。枇々木 他 (2010)と尾木 他 (2016)において行われた法人企業に対する定式化と一部、同じ枠組みで分析を行う。そこで、その分析内容を最初にやや詳しく考察したあとで、個人企業に関して追加的に行った新たな定式化および分析について説明する。

5.1 先行研究における法人企業の業歴別デフォルト率の定式化の考察

枇々木 他 (2010)と尾木 他 (2016)は、法人企業のデフォルト率の形状と業歴区間ごとの特性から、多項式関数(n 次関数)を業歴関数として定式化を試みた。手順は以下のとおりである。

まず、(5.1)式の 1~8 次関数のパラメータを最小二乗法により推定している。

$$(5.1) \quad p_g = \beta_0 + \sum_{k=1}^n \beta_k g^k \quad (g = 1, \dots, G)$$

ここで、 g は融資を受けた企業の融資年度時点の業歴、 p_g は業歴 g 年で融資を受けた翌年度時点のデフォルト率、 β_k は業歴の k 乗に対する回帰係数、 β_0 は定数項を表している。(5.1)式において業歴関数の次数 n を 1~8 まで試したときの回帰係数($\beta_0 \sim \beta_8$)の p 値と修正 R^2 が示されている。尾木 他 (2016)の表 3 をみると、1 次から 8 次へと次数が上がるほど修正 R^2 は上昇しているが、7 次関数は 5 乗以上、8 次関数は 2 乗以上の p 値が 5% を超えている。また、同表に示されている(4.1)式の業歴モデル⁹⁾(右辺第 3 項の業歴関数の次数 n を 1~8 まで試した場合)の AR 値も合わせて見てみると、6 次関数まではすべての p 値が 1% 以下であるものの、AR 値は 3 次関数で 42.7%、8 次で 43.0% と 3 次以上ではほとんど上昇していない。

加えて、業歴が長くなるにつれて、デフォルト率が低下する傾向があるという大きな特徴もあるので、尾木 他 (2016)は指数関数についても検討した。分析の結果は、どちらの関数も修正 R^2 は遜色ないが、時系列での頑健性が若干劣るほか、現場感覚や中小企業研究における指摘などを踏まえると業歴とともにデフォルト率が低下し続ける指数関数は選択しにくいなど、総合的に判断して 3 次関数を選択している。

以上のような先行研究の結果を踏まえて、個人企業についても、まず法人企業と同様に多項式関数と指数関数の当てはめを行い、次に個人企業に対する新たな業歴関数として、ワイブル分布の密度関数および区分線形関数を使ったアプローチを試みる。パラメータの推定は

表 5. 個人企業の業歴別デフォルト率の多項式関数のモデル化.

	$n=1$	$n=2$	$n=3$	$n=4$	$n=5$	$n=6$	$n=7$	$n=8$
β_0	0.023 (<0.001)	0.031 (<0.001)	0.039 (<0.001)	0.044 (<0.001)	0.052 (0.010)	0.059 (<0.001)	0.065 (<0.001)	0.069 (<0.001)
β_1 [$\times 10^{-3}$]	-0.251 (<0.001)	-1.015 (<0.001)	-2.534 (<0.001)	-4.214 (<0.001)	-7.423 (<0.001)	-11.901 (<0.001)	-15.891 (<0.001)	-20.123 (<0.001)
β_2 [$\times 10^{-5}$]		1.253 (<0.001)	7.428 (<0.001)	19.581 (<0.001)	55.198 (<0.001)	125.002 (<0.001)	207.303 (<0.001)	317.899 (0.166)
β_3 [$\times 10^{-6}$]			-0.675 (<0.001)	-3.757 (<0.001)	-19.130 (<0.001)	-63.925 (<0.001)	-136.506 (0.004)	-263.656 (0.800)
β_4 [$\times 10^{-8}$]				2.526 (<0.001)	30.757 (<0.002)	167.128 (<0.003)	489.113 (<0.004)	1252.392 (<0.005)
β_5 [$\times 10^{-9}$]					-1.851 (<0.000)	-21.455 (<0.001)	-96.872 (<0.002)	-353.780 (0.001)
β_6 [$\times 10^{-10}$]						1.071 (<0.001)	9.973 (<0.002)	58.782 (0.003)
β_7 [$\times 10^{-12}$]							-4.170 (0.001)	-53.026 (0.007)
β_8 [$\times 10^{-14}$]								20.023 (0.013)
AR値	39.7%	42.0%	42.7%	42.8%	42.9%	43.0%	43.0%	43.0%
修正 R^2	0.383	0.612	0.765	0.814	0.873	0.918	0.933	0.941

SAS/STAT を用いている。

5.2 多項式関数と指数関数の当てはめ

個人企業の業歴別デフォルト率をみると、法人企業と異なる点はあるものの、業歴5年前後までデフォルト率が大きく低下し、業歴15年以降は平均値を下回って安定するという傾向は共通している。定式化にあたっては、まず、尾木 他 (2016) の分析と同様に多項式関数と指数関数による当てはめを行った。

5.2.1 多項式関数

多項式関数を当てはめた結果を表5に示す。個人企業は法人企業と異なり、業歴35年以降にデフォルト率が上昇する傾向がみられないことから、3次関数の修正 R^2 は0.765となり、法人企業の0.932を下回る結果となった。次数が上がるにつれて修正 R^2 は上昇するが、AR値は2次以降ほとんど上昇しない。デフォルト企業数は業歴が長くなるほど減少し、業歴20年を過ぎると、業歴1年～2年の1/8程度の数になるため精度が安定しないことが背景にあると思われる。また、尾木 他 (2016) が示したような審査担当者の経験やノウハウとの整合性、中小企業研究における指摘、オーバーフィッティングにより経年劣化が大きくなる可能性などを考慮すると、高次の関数は採用しにくい。

5.2.2 指数関数

次に、業歴35年以降にデフォルト率が上昇しない傾向を踏まえて、指数関数による当てはめを行った。最小二乗法を用いてパラメータを推定すると(5.2)式が得られる。

$$(5.2) \quad p_g = 0.0511e^{-0.3158g} + 0.0130 \quad (g = 1, \dots, G)$$

指数関数の R^2 は0.898となった。図7のとおり、開業から業歴5年までにデフォルト率が大きく低下したあと、業歴25年ぐらゐまで緩やかに低下する期間があるため、指数関数では

この期間がうまく当てはまっていない可能性がある。

5.3 新たな業歴関数の提案

法人企業では有効であった多項式関数は、個人企業においては当てはまりが良いとはいえない結果となった。また、指数関数の修正 R^2 は 0.898 と、比較的良好な結果となったが、業歴 5 年から 25 年あたりのフィッティングが良くない。

そこで、図 8 の灰色の実線の指数関数に近いワイブル分布の密度関数、さらには、業歴 5 年まで急激に低下、業歴 25 年ぐらいまで緩やかに低下、それ以降はほぼ横ばいという業歴区間ごとの特徴を踏まえて、新たに図 7 の黒色の実線のような全体を 3 つに区分(急激な低下区間、緩やかな低下区間、横ばい区間)した区分線形関数による当てはめを行った。

5.3.1 ワイブル分布の密度関数

信頼性工学の分野では、機械などの故障率は、縦軸に故障率、横軸に時間をとると、初期不良に起因する故障が多く発生し、時間とともに減少する時期、偶発的な事故が発生するのみで故障率が一定で安定する時期、寿命が近づいて時間と共に故障率が増加する時期の 3 つの時期

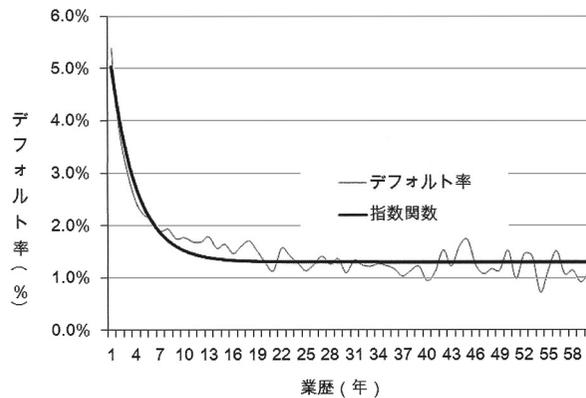


図 7. 指数関数の当てはめ.

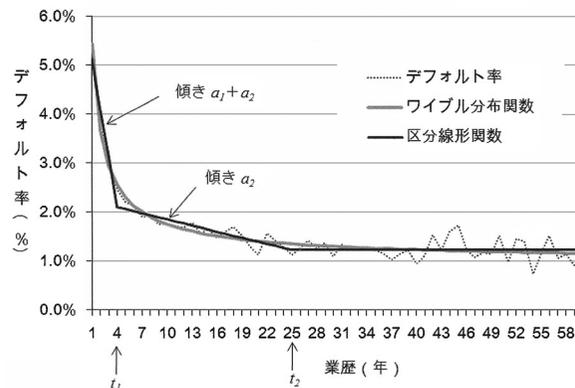


図 8. ワイブル分布の密度関数および区分線形関数の当てはめ.

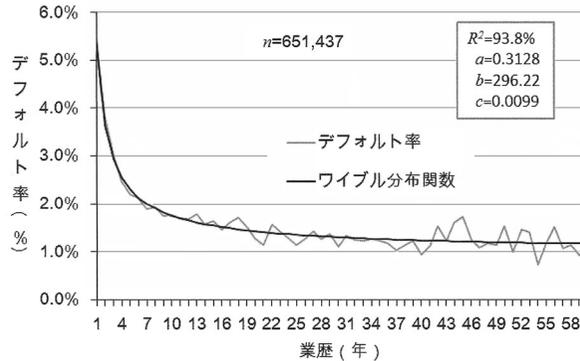


図 9. ワイブル分布の密度関数.

があり、バスタブのような曲線を描くと言われている。個人企業の多くが一代限りで事業承継しないとすれば、個人企業には寿命があると考えられることもできる。とすれば、デフォルト率もバスタブ曲線を描くと仮定することは可能である。このバスタブ曲線にはワイブル分布を仮定することが多い。ワイブル分布の密度関数を(5.3)式に示す。

$$(5.3) \quad f(x) = \frac{a}{b} \left(\frac{x}{b}\right)^{a-1} \exp\left[-\left(\frac{x}{b}\right)^a\right]$$

ここで、 a は分布の形状、 b は分布の幅を決めるパラメータ、 x は経過時間を入力する。次に、(5.3)式に定数項を加えた(5.4)式を用いてフィッティングする。

$$(5.4) \quad f(x) = \frac{a}{b} \left(\frac{g}{b}\right)^{a-1} \exp\left[-\left(\frac{g}{b}\right)^a\right] + c \quad (g = 1, \dots, G)$$

ここで、 g は業歴年数である。

結果を図9に示す。修正 R^2 は 0.938 と、法人企業の業歴関数と同水準となった。

5.3.2 区分線形関数

次に、区分線形関数による当てはめを行う。再度、図8を見てほしい。区分数を考えるにあたって、業歴25年以降の区分数については、法人企業と同様に35年から45年にかけて上昇し、45年以降に減少するという傾向に見えなくもない。ただ、4.2.2項(4)で述べたように、個人企業は事業承継せず1代限りの企業が多いこと、5.2.1項で述べたように、業歴20年以降の業歴別デフォルト率はデフォルト企業数が少ないため不安定なことを踏まえて3区分とした。

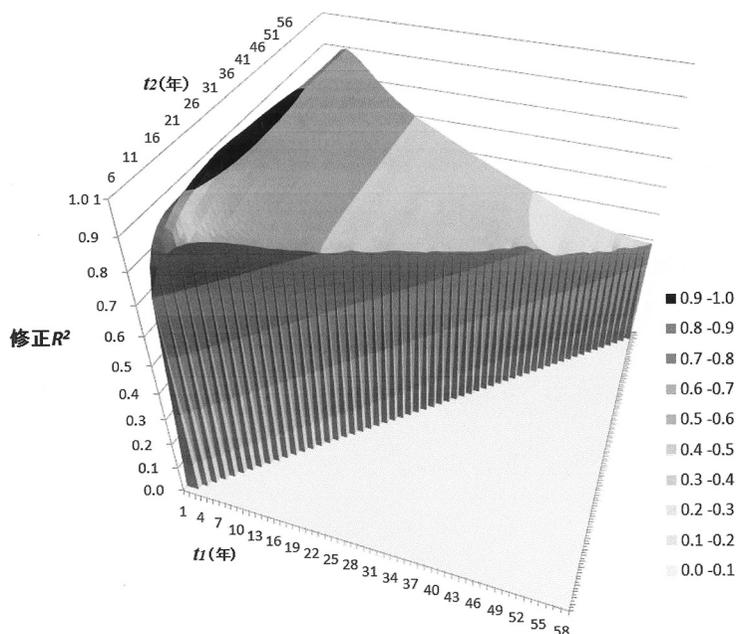
まず、デフォルト率を区分する業歴は、急激な低下が続く創業から t_1 年までと、緩やかな低下が続く t_2 年までの2点とし、 t_2 年以降は定数とする。次に、 $t_1 = 2, \dots, 59$, $t_2 = 3, \dots, 60$ の1,711通りについて、 y_t を $\min(t - t_1, 0)$ と $\min(t - t_2, 0)$ で回帰し、最も決定係数の高い組み合わせを求める。具体的には以下のとおり、デフォルト率 y_t を区分線形関数で近似した \hat{y}_t を求める。なお、業歴60年超については、業歴60年として扱う。

$$\begin{aligned} 1 \leq \text{業歴 } t \leq t_1 & \quad \hat{y}_t = b + a_2(t - t_2) + a_1(t - t_1) \\ t_1 \leq \text{業歴 } t \leq t_2 & \quad \hat{y}_t = b + a_2(t - t_2) \\ t_2 \leq \text{業歴 } t \leq 60 & \quad \hat{y}_t = b \end{aligned}$$

分析の結果を以下に示す。表6のとおり、 t_1 年は4年が、 t_2 年は25年のパターンでの修正 R^2 が 0.929 と最も高くなり、ワイブル分布の密度関数よりもやや低いものの、こちらも法人企業

表 6. 区分線形関数の t_1, t_2 の分析結果(上位 20 位の組み合わせ).

順位	修正 R^2	t_1 年	t_2 年	順位	修正 R^2	t_1 年	t_2 年
1	0.9289	4	25	11	0.924	4	30
2	0.9288	4	24	12	0.923	4	31
3	0.928	4	26	13	0.922	4	19
4	0.928	4	23	14	0.921	4	32
5	0.927	4	22	15	0.920	4	33
6	0.927	4	27	16	0.919	4	18
7	0.927	4	21	17	0.918	4 </td <td>34</td>	34
8	0.926	4	28	18	0.918	3	21
9	0.925	4	29	19	0.918	3	20
10	0.925	4	20	20	0.918	6	26

図 10. 区分線形関数の t_1, t_2 の修正 R^2 の分布.

の 3 次関数とほぼ同水準となった。3 区分以上にすれば、 R^2 がさらに上昇する可能性もあるが、多項式関数の結果を踏まえるとオーバーフィッティングの可能性があるため、これ以上の検証は行わない。上位 20 位の組み合わせをみると、 t_1 年は 4 年が上位を占め安定している。 t_2 年は 25 年の前後が多いが、図 10 のとおり、修正 R^2 が 0.9 以上の部分(黒色部分)の t_2 年は 15 年ぐらいから 40 年ぐらいまで幅がある。

5.4 業種別業歴関数の当てはめ

業歴別デフォルト率は業種別に異なる。そこで、ワイブル分布の密度関数と区分線形関数について業種別に定式化し、修正 R^2 の比較を行う。

表 7. 業種別ワイブル分布の密度関数の分析結果. b は 10^6 を上限値としてパラメータ推定している.

	製造	建設	卸小売	飲食宿泊	医療福祉	サービス	不動産	運輸	その他
n	44,305	111,876	145,022	85,801	43,979	123,725	47,148	14,054	35,527
a	0.7819	0.4791	0.4533	0.4360	0.2137	0.2996	13.0272	3.3426	0.1423
b	55.50	75.28	88.58	19.40	1,000,000	3,082	81.41	90.44	786.7
c	0.0088	0.0116	0.0114	0.0101	0.0059	0.0071	0.0036	0.0126	0.0078
修正 R^2	55.7%	82.4%	84.8%	86.9%	5.3%	65.0%	6.1%	3.1%	60.2%

5.4.1 ワイブル分布の密度関数(業種別)

まずワイブル分布の密度関数を業種ごとに当てはめた結果を表 7 に示す。「医療・福祉業」「不動産業」「運輸業」については、修正 R^2 の値が低くなっている。図 11 の業種別業歴別デフォルト率をみると、この 3 業種は「全体」に比べて終期に上昇しており、グラフの形状が異なっているが、業歴別デフォルト率がバスタブ曲線になると考えれば、その他の業種については初期で減少、中期で一定の 2 段階目までを示し、終期の上昇は何らかの理由で確認できていないと考えることもできる。

ただ、業種別になると、この 3 業種の業歴 50 年以降の業歴別企業数は件数が 100 件程度と少なくなり、デフォルト率の変動が大きくなることから評価が難しい。さらに、「医療・福祉業」および「不動産業」のデフォルト率は、ほぼ 1.0% 以下の低位水準での変動であり、「医療・福祉業」の大宗が医者であること、「不動産業」には不動産賃貸業が多く含まれていることを考えると、他の業種に比べれば、おおむね横ばいで安定推移していると見ることができる。なお、3 業種以外の業種についてもデータ数が少ないため、55.7%~86.9% と「全体」に比べて低い値となったが、指数関数的に単調減少するというグラフの基本的な形状は「全体」の形状と類似している。

5.4.2 区分線形関数(業種別)

次に、区分線形関数について業種別に値を求めた結果を表 8 に示す。5.4.1 項のワイブル分布の密度関数と同様に、「医療・福祉業」「不動産業」「運輸業」については、修正 R^2 の値が低くなっている。それ以外の業種についてもワイブル分布の密度関数と同水準となった。「全体」に比べてサンプル数が少ないため、57.5%~86.7% と低い値となったが、図 12 をみると、 t_1 年にかけて急激に低下し、 t_2 年にかけて緩やかに低下したあと、横ばいになるという基本的な形状は類似している。

次に、「医療・福祉業」「不動産業」「運輸業」を除いた業種の t_1 の値をみると、製造業が 8 年である以外は 4 年前後が多い一方で、 t_2 の値は、20 年~47 年とやや幅がある。「全体」の t_2 を 15 年くらいから 40 年くらいまでどの値にしても修正 R^2 値に差がないと同様に、 t_1 年から t_2 年にかけては低下が緩やかなため、どこから横ばいになるのかは微妙なところがあり、時系列的にやや不安定となる可能性がある。

そこで、 t_1 、 t_2 の値を 4 年と 25 年で固定し、業種ごとに係数を変えることにした。再び表 8 で修正 R^2 を比較した結果をみると、業種別に t_1 、 t_2 の年数を 4 年と 25 年で固定しても、修正 R^2 は大きく低下しないことがわかる。ワイブル分布の密度関数との比較では、やや低下するものの、頑健性を考慮すれば、 t_1 、 t_2 の年数は業種別にしない方が好ましいといえる。そこで、 t_1 、 t_2 の値は 4 年と 25 年で固定し、業種ごとに係数を変えた業歴関数を採用することにした。

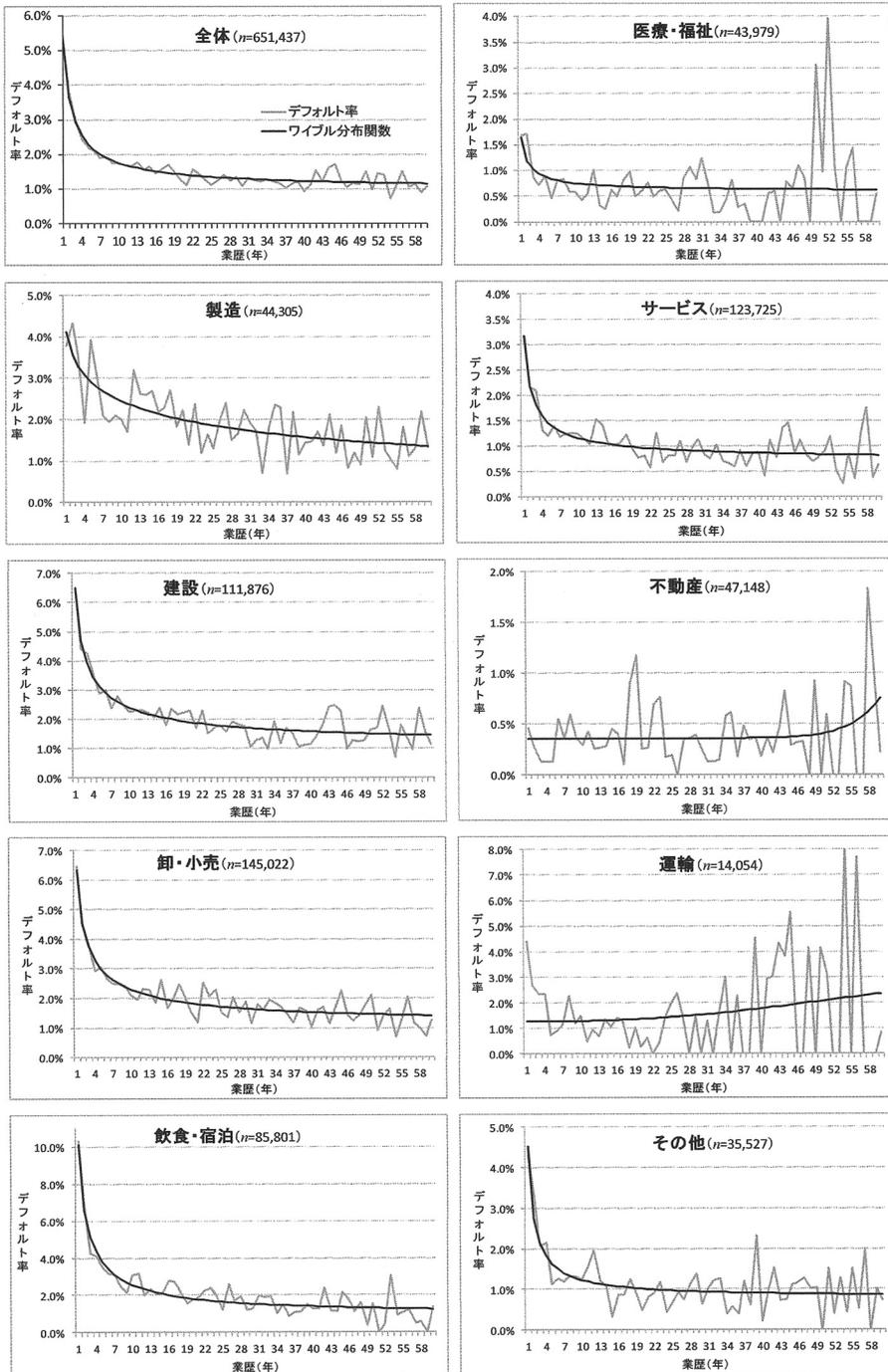


図 11. 業種別業歴別デフォルト率のワイブル分布の密度関数.

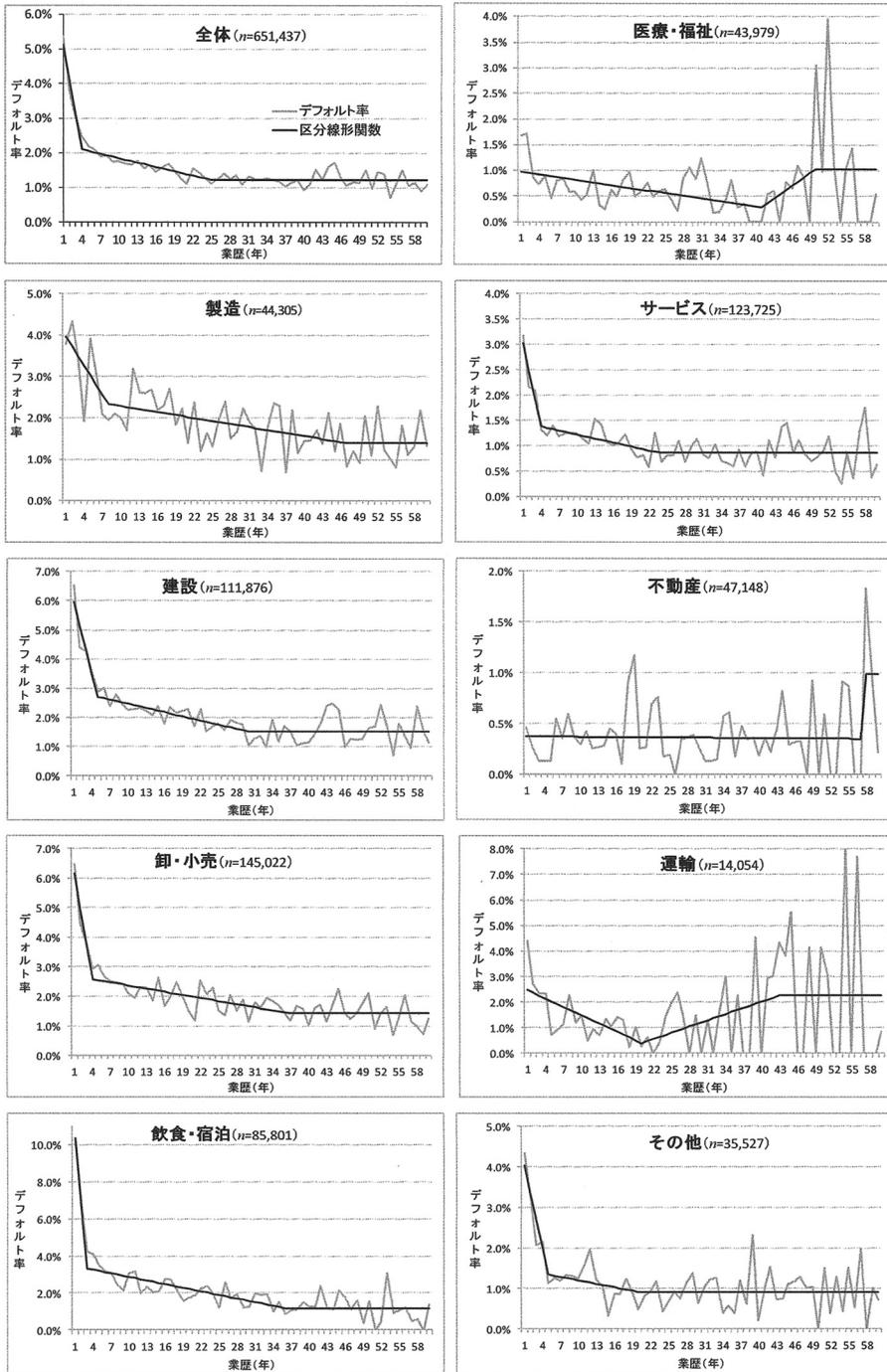


図 12. 業種別業歴別デフォルト率の区分線形関数。

表 8. 業種別区分線形関数の t_1, t_2 の分析結果.

	製造	建設	卸小売	飲食宿泊	医療福祉	サービス	不動産	運輸	その他
n	44,305	111,876	145,022	85,801	43,979	123,725	47,148	14,054	35,527
t_1	8	5	4	3	41	4	57	20	5
t_2	47	30	37	36	50	24	58	43	20
R^2	57.5%	82.4%	83.4%	86.7%	12.7%	66.6%	17.2%	11.6%	62.0%
$t_1=4, t_2=25$ 修正 R^2	54.9%	81.3%	82.3%	84.5%	-	66.6%	-	-	60.8%
差異	-2.6%	-1.1%	-1.1%	-2.2%	-	0.0%	-	-	-1.2%

6. 個人企業向けモデルにおける業歴関数の有効性

財務モデルに業歴関数を変数に加えた業歴モデルを構築し、財務モデルと業歴モデルのパフォーマンスを AR 値により比較する。業歴関数については、ワイブル分布の密度関数と区分線形関数の修正 R^2 の値がほぼ同水準であったことを踏まえて、両者を用いた業歴モデルを構築し、そのパフォーマンスを比較する。なお、「医療・福祉業」「不動産業」「運輸業」については、修正 R^2 が低水準にあることから、本研究においては業歴関数を不採用とした。

6.1 個人企業向け財務モデルの AR 値

2007 年度～2014 年度の 8 年間に融資した 651,437 件のデータを使用して財務モデルを構築する。モデル構築に必要な財務データ等が揃っていない企業は除外しているため、図 5 の企業数と一致しない。モデルは(4.1)式の第 3 項を除いたものである。

分析の結果、財務モデルの AR 値は 46.7% と、法人企業向けに比べてインサンプルで 11.9% ポイント高くなった。業種別の結果を表 9 に示す。ここで、「飲食・宿泊業」「医療・福祉業」「サービス業」については、年商規模によって特徴が異なることから規模の大小によって区分している。業種別の AR 値をみると、「サービス」「不動産」が高く、「飲食・宿泊 s」「医療・福祉 s」など規模の小さな企業の AR 値が低くなっている。

6.2 個人企業向け業歴モデルの構築

財務モデルと業歴モデルのパフォーマンスを AR 値により比較するため、財務モデルに、第 3 項に業歴関数を加えた業歴モデルを構築する。業歴関数には、修正 R^2 が同水準であったワイブル分布の密度関数を使ったモデル(以下、ワイブル分布関数モデルという)と区分線形関数(以下、区分線形関数モデルという)を用いる。

(1) ワイブル分布関数モデル

ワイブル分布の密度関数のパラメータも含めて、ロジットモデルの中で推定することは難しいので、業歴関数(5.4)式に基づき、(6.1)式のパラメータ a_h, b_h, c_h を最小二乗法で業種 h ごとに推定する。

$$(6.1) \quad p_g = \frac{a_h}{b_h} \left(\frac{g}{b_h} \right)^{a_h-1} \exp \left[- \left(\frac{g}{b_h} \right)^{a_h} \right] + c_h, \quad (g = 1, \dots, G)$$

推定されたパラメータ a_h^*, b_h^*, c_h^* と個人企業 $i (i = 1, \dots, I)$ の業歴 g_i を使用して、業歴関数によるデフォルト率の推定値 q_i^* を業種 $h (h = 1, \dots, H)$ ごとに(6.2)式で計算する。

表 9. 財務モデルの推計結果.

業種	n	定数項 (α_0)	申告形態 (α_1)	流動性指標 (α_2)	安全性指標 (α_3)	効率性指標 (α_4)	AR値
製造	44,597	3.68 (<0.001)	-0.47 (<0.001)	0.49 (<0.001)	0.06 (0.028)	-0.12 (<0.001)	0.39
建設	112,653	3.74 (<0.001)	-0.47 (<0.001)	0.74 (<0.001)	0.05 (<0.001)	-0.36 (<0.001)	0.37
卸小売	145,141	3.64 (<0.001)	-0.47 (<0.001)	0.49 (<0.001)	0.07 (<0.001)	-0.12 (<0.001)	0.40
飲食宿泊s	47,911	3.24 (<0.001)	-0.47 (<0.001)	0.49 (<0.001)	0.05 (<0.001)	-0.12 (<0.001)	0.36
飲食宿泊	37,510	3.67 (<0.001)	-0.89 (<0.001)	0.89 (<0.001)	0.05 (<0.001)	-0.12 (<0.001)	0.39
医療福祉s	16,087	4.02 (<0.001)	-0.47 (<0.001)	0.49 (<0.001)	0.05 (<0.001)	-0.12 (<0.001)	0.36
医療福祉	27,759	5.28 (<0.001)	-2.17 (<0.001)	0.49 (<0.001)	0.05 (<0.001)	-0.12 (<0.001)	0.43
サービスs	85,221	4.16 (0.003)	-0.68 (<0.001)	0.49 (<0.001)	0.06 (0.002)	-0.37 (<0.001)	0.41
サービス	39,786	4.48 (0.003)	-0.68 (<0.001)	0.94 (0.002)	0.06 (<0.001)	-0.37 (<0.001)	0.50
不動産	48,925	4.63 (<0.001)	-0.81 (0.045)	0.49 (<0.001)	0.07 (0.041)	-0.64 (<0.001)	0.51
運輸	14,073	4.02 (<0.001)	-0.47 (<0.001)	0.49 (<0.001)	0.05 (<0.001)	-0.59 (<0.001)	0.44
その他	31,774	4.02 (<0.001)	-0.47 (<0.001)	0.49 (<0.001)	0.05 (<0.001)	-0.12 (<0.001)	0.37

注 1：業種の後部に s が付いているものは、年商規模が「小」のものである。

注 2：その他業種は、全体モデルの係数を使用している。

注 3：カッコ内は Wald カイ 2 乗検定に対する p 値。業種モデルの係数は、全体モデルの係数に p 値 5% 水準で有意になった業種ダミー（定数項ダミー、係数ダミー）の値を加えたものである。したがって、業種別モデルの係数の p 値は係数ダミーの値に対する結果である。

注 4：実際に使用しているモデルであり、具体的な財務指標名については、実務での影響を考慮し、秘匿している点をご理解いただきたい。

$$(6.2) \quad q_i^* = \frac{a_h^*}{b_h^*} \left(\frac{g_i}{b_h^*} \right)^{a_h^* - 1} \exp \left[- \left(\frac{g_i}{b_h^*} \right)^{a_h^*} \right] + c_h^*, \quad (i \in F_h; h = 1, \dots, H)$$

ここで、 F_h を業種 h に属する企業の集合とする。ワイブル分布関数モデルを (6.3) 式に示す。個人企業 i ($i = 1, \dots, I$) について、財務変数 $f_{i,j}$ ($j = 1, \dots, J$) と業歴関数によって計算されたデフォルト率 q_i^* を使用してロジットモデルを構築し、最尤法によってパラメータ a_h , $\beta_{j,h}$, w_h を業種 h ごとに推定する。

$$(6.3) \quad p_i = \frac{1}{1 + e^{z_i}},$$

$$z_i = \ln \left(\frac{1 - p_i}{p_i} \right) = \alpha_h + \sum_{j=1}^J \beta_{j,h} f_{i,j} + w_h q_i^*, \quad (i \in F_h; h = 1, \dots, H)$$

(2) 区分線形関数モデル

区分線形関数モデルを (6.2) 式に示す。個人企業 i ($i = 1, \dots, I$) について、財務変数 $f_{i,j}$ ($j = 1, \dots, J$) と業歴 g_i を使用して、ロジットモデルを構築し、最尤法によってパラ

メータ $a_h, \beta_{j,h}, \gamma_{h,1}, \gamma_{h,2}$ を推定する.

$$(6.4) \quad p_i = \frac{1}{1 + e^{z_i}},$$

$$z_i = \ln \left(\frac{1 - p_i}{p_i} \right)$$

$$= \alpha_h + \sum_{j=1}^J \beta_{j,h} f_{i,j} + \gamma_{h,1} \times \min(g_i - 4, 0) + \gamma_{h,2} \times \min(g_i - 25, 0),$$

$$(i \in F_h; h = 1, \dots, H)$$

6.3 個人企業向け業歴モデルの AR 値

業歴モデルの推計結果を表 10 に示す. 業歴関数を投入すると全体の AR 値はワイブル分布関数モデルが 49.0%, 区分線形関数モデルが 49.3% となり, ほぼ同水準で, それぞれ法人企業モデルに比べて 2.3% ポイント, 2.6% ポイント上昇する. 法人企業に比べて AR 値の水準は高いが, 業歴関数を加えたことによる改善幅はいずれも小さい. 仮説では, 法人企業と同様に, 業歴別デフォルト率は, 非開示となっている事業主の私的資産が多くを占める小企業のパフォーマンスを示す指標として期待したが, 本分析結果を見る限り, 開示されている事業資産に事業主の私的資産が相応に含まれている(非開示の資産が少ない), もしくは, 非開示の私的資産は事業に投入しない傾向にあるなどの理由で期待したほどの効果が得られなかった可能性が考えられる.

表 11 と表 12 で業種別に AR 値の改善幅をみると「飲食宿泊」はワイブル分布関数モデルが 9.6% ポイント, 区分線形関数モデルが 9.4% ポイント, 「その他」が同 6.4% ポイント, 同 6.5% ポイント, 「飲食宿泊 s」が同 6.1% ポイント, 同 5.9% ポイントと大きくなっている. それ以外は改善幅が小さく, サービス業はほとんど改善していない. ただ, AR 値の水準の低い業種ほど改善幅が大きい傾向にあり, 業歴関数を投入したことにより, 業種間の差が改善されている.

二つの業歴関数を比較すると, 水準および改善幅に大きな違いはないが, ワイブル分布関数モデルでは, 「製造」「飲食宿泊」「サービス s」「サービス」「その他」の業種のいくつかの財務変数の p 値が 5% 水準で有意にならなかった.

6.4 モデルの頑健性の検証

個人企業向け業歴モデルの頑健性を確認するために, アウトオブサンプルテストを行った. アウトオブサンプルは, 2015 年度に融資を行った債権 70,383 件で, 2007 年度~2014 年度に融資した 651,437 件のインサンプルに対して 10.8% のサンプルサイズである. 結果を表 13 に示す. 全体の AR 値は, いずれのモデルでもほぼ横ばいで良好な結果となっている. ワイブル分布関数モデルと区分線形関数モデルの差はほとんど確認できない.

表 10. 業歴モデルの AR 値(インサンプル). 法人企業の数値は, 尾木 他 (2016) の値.

	個人企業(a)	法人企業(b)	(a) - (b)
財務モデル	46.7	34.8	11.9
ワイブル分布関数モデル (改善幅)	49.0 (2.3)	42.4 (7.6)	6.6 (-5.3)
区分線形関数モデル (改善幅)	49.3 (2.6)	42.4 (7.6)	6.9 (-5.0)

表 11. 区分線形関数業歴モデルの推計結果((6.3)式).

業種	定数項 (α)	申告形態 ($\beta 1$)	流動性指標 ($\beta 2$)	安全性指標 ($\beta 3$)	効率性指標 ($\beta 4$)	$min(ki-4.0)$ ($\gamma 1$)	$min(ki-25.0)$ ($\gamma 2$)	AR値	改善幅
製造	3.76 (<0.001)	-0.38 (<0.001)	0.50 (<0.001)	0.06 (0.004)	-0.16 (<0.001)	0.24 (<0.001)	0.01 (<0.001)	40.2%	1.6%
建設	3.93 (<0.001)	-0.38 (<0.001)	0.72 (<0.001)	0.05 (<0.001)	-0.32 (<0.004)	0.24 (<0.001)	0.02 (0.001)	40.5%	3.5%
卸小売	3.87 (<0.001)	-0.38 (<0.001)	0.50 (<0.001)	0.06 (<0.001)	-0.16 (<0.001)	0.24 (<0.001)	0.02 (0.002)	42.6%	2.7%
飲食宿泊s	3.63 (<0.001)	-0.38 (<0.001)	0.50 (<0.001)	0.05 (<0.001)	-0.16 (<0.001)	0.24 (<0.001)	0.03 (<0.001)	42.0%	5.9%
飲食宿泊	4.18 (<0.001)	-0.69 (<0.001)	0.88 (<0.001)	0.05 (<0.001)	-0.16 (<0.001)	0.35 (0.004)	0.03 (<0.001)	48.6%	9.4%
医療福祉s	4.08 (<0.001)	-0.38 (<0.001)	0.50 (<0.001)	0.05 (<0.001)	-0.16 (<0.001)	-	-	36.2%	-
医療福祉	5.31 (<0.001)	-2.17 (<0.001)	0.50 (<0.001)	0.05 (<0.001)	-0.16 (<0.001)	-	-	42.7%	-
サービスs	4.29 (0.008)	-0.63 (<0.001)	0.50 (<0.001)	0.06 (<0.001)	-0.36 (<0.001)	0.24 (<0.001)	0.01 (<0.001)	41.8%	0.4%
サービス	4.58 (<0.001)	-0.63 (<0.001)	0.93 (0.003)	0.06 (<0.001)	-0.36 (<0.001)	0.24 (<0.001)	0.01 (<0.001)	50.0%	-0.1%
不動産	4.62 (<0.001)	-0.81 (0.010)	0.50 (<0.001)	0.07 (<0.001)	-0.63 (0.010)	-	-	50.8%	-
運輸	4.08 (<0.001)	-0.38 (<0.001)	0.50 (<0.001)	0.05 (<0.001)	-0.69 (<0.001)	-	-	43.7%	-
その他	4.08 (<0.001)	-0.38 (<0.001)	0.50 (<0.001)	0.05 (<0.001)	-0.16 (<0.001)	0.24 (<0.001)	0.01 (<0.001)	43.3%	6.5%

注 1: その他業種は、全体モデルの係数を使用している。

注 2: カッコ内は Wald カイ 2 乗検定に対する p 値。業種モデルの係数は、全体モデルの係数に p 値 5% 水準で有意になった業種ダミー(定数項ダミー, 係数ダミー)の値を加えたものである。したがって、業種別モデルの係数の p 値は係数ダミーの値に対する結果である。

業種別にみると、「サービス業」が 10% ポイント以上の上昇と変動幅が大きくなっている。サービス業は年商規模がやや大きな「サービス業」と年商規模がやや小さな「サービス業_s」の二つに分類している。「サービス業」は、①サンプル数が少ない、②「学術研究、専門・技術サービス業」「生活関連サービス業、娯楽業」「複合サービス事業」「サービス業(他に分類されないもの)」など大分類ベースでも 4 つに分類されるなど多種多様である、③年商規模の上限がなく規模のバラつきが大きい、といった複数の要因が影響していると考えられる。

また、「製造業」「医療・福祉業_s」「医療・福祉業」「運輸業」とサンプル数の少ない業種でやや変動幅が大きくなる傾向にあるほか、景気変動の影響を受けやすい「建設業」が上昇している。ワイブル分布関数モデルと区分線形関数モデルは、アウトオブサンプルテストの結果においても大きな差はみられない。

以上のように、業種別にみると、サンプル数や景気変動、業種特性などの影響を受けるため、AR 値の変動幅はやや大きくなる傾向があるものの、今回のアウトオブサンプルテストの結果を見る限り、実務での使用には大きな問題はなさそうである。ただ、通常、アウトオブサンプルの AR 値はインサンプルより低い傾向にあるが、本検証では上回っているケースもあり、データ数が少なく不安定な結果となっている点是否めない。頑健性については、データの蓄積を待って、引き続き検証する必要があると考えている。

7. まとめと今後の課題

本研究では、個人企業向けスコアリングモデルの精度を改善するために、業歴別デフォルト

表 12. ワイブル分布密度関数業歴モデルの推計結果((6.4)式).

業種	定数項 (α)	申告形態 (β_1)	流動性指標 (β_2)	安全性指標 (β_3)	効率性指標 (β_4)	ワイブル分布 (w, g)	AR値	改善幅
製造	4.21 (<0.001)	-0.27 (0.379)	0.55 (0.030)	0.06 (<0.001)	-0.15 (0.529)	-27.94 (0.081)	39.3%	0.6%
建設	4.29 (<0.001)	-0.41 (<0.001)	0.71 (<0.001)	0.05 (<0.001)	-0.32 (0.038)	-24.78 (0.001)	40.6%	3.6%
卸小売	4.21 (<0.001)	-0.39 (0.007)	0.57 (<0.001)	0.06 (<0.001)	-0.18 (0.007)	-25.78 (0.002)	42.8%	2.9%
飲食宿泊s	3.74 (<0.001)	-0.39 (0.020)	0.46 (<0.001)	0.05 (0.021)	0.00 (<0.001)	-17.19 (<0.001)	42.3%	6.1%
飲食宿泊	4.39 (<0.001)	-0.67 (0.023)	0.95 (0.002)	0.04 (0.006)	-0.41 (0.436)	-21.57 (<0.001)	48.7%	9.6%
医療福祉s	4.22 (<0.001)	-0.38 (<0.001)	0.58 (<0.001)	0.05 (<0.001)	-0.20 (<0.001)	-	36.2%	-
医療福祉	5.28 (<0.001)	-2.17 (<0.001)	0.58 (<0.001)	0.05 (<0.001)	-0.20 (<0.001)	-	42.7%	-
サービスs	4.66 (0.008)	-0.62 (0.001)	0.31 (0.018)	0.07 (<0.001)	-0.33 (0.050)	-25.64 (0.045)	41.8%	0.4%
サービス	5.36 (<0.001)	-0.85 (0.128)	0.89 (0.051)	0.06 (0.007)	-0.57 (0.210)	-48.58 (0.1294)	50.8%	0.6%
不動産	4.65 (<0.001)	-0.84 (0.005)	0.58 (<0.001)	0.05 (<0.001)	-0.64 (0.010)	-	50.8%	-
運輸	4.22 (<0.001)	-0.38 (<0.001)	0.58 (<0.001)	0.05 (<0.001)	-0.89 (<0.001)	-	43.7%	-
その他	4.74 (<0.001)	-0.29 (0.337)	0.57 (<0.001)	0.05 (<0.001)	-0.19 (0.322)	-51.31 (<0.001)	43.1%	6.4%

注 1: その他業種は, 全体モデルの係数を使用している。

注 2: カッコ内は Wald カイ 2 乗検定に対する p 値. 業種モデルの係数は, 全体モデルの係数に p 値 5% 水準で有意になった業種ダミー(定数項ダミー, 係数ダミー)の値を加えたものである. したがって, 業種別モデルの係数の p 値は係数ダミーの値に対する結果である. 網掛けは, p 値 5% 水準で有意にならなかったもの。

表 13. アウトオブサンプルテストの結果(AR 値). 医療福祉, 不動産業, 運輸業は業歴関数を用いていない(シャドー部分).

業種	n	ワイブル分布関数モデル			区分線形関数モデル		
		イン(A)	アウト(B)	差分(A-B)	イン(C)	アウト(D)	差分(C-D)
製造	3,629	39.8%	36.8%	-2.4% [†] イト	40.2%	36.3%	-3.9% [†] イト
建設	11,998	40.7%	46.1%	5.5% [†] イト	40.5%	45.9%	5.4% [†] イト
卸小売	13,707	42.7%	40.7%	-2.0% [†] イト	42.6%	40.3%	-2.3% [†] イト
飲食宿泊s	5,006	42.2%	44.5%	2.2% [†] イト	42.0%	43.5%	1.5% [†] イト
飲食宿泊	4,200	48.4%	51.2%	2.5% [†] イト	48.6%	51.9%	3.3% [†] イト
医療福祉s	2,251	36.2%	42.2%	6.0% [†] イト	36.2%	42.2%	6.0% [†] イト
医療福祉	2,832	42.7%	37.8%	-5.0% [†] イト	42.7%	37.8%	-5.0% [†] イト
サービスs	9,767	41.8%	39.5%	-2.3% [†] イト	41.8%	40.2%	-1.7% [†] イト
サービス	4,111	50.7%	62.0%	11.2% [†] イト	50.0%	63.1%	13.1% [†] イト
不動産	7,046	50.8%	49.1%	-1.7% [†] イト	50.8%	49.1%	-1.7% [†] イト
運輸	1,566	43.7%	39.1%	-4.7% [†] イト	43.7%	39.1%	-4.7% [†] イト
その他	4,270	43.4%	43.5%	0.3% [†] イト	43.3%	45.6%	2.3% [†] イト
全体	70,383	49.0%	50.2%	1.2% [†] イト	49.1%	49.7%	0.6% [†] イト

率を定式化した業歴関数を投入した. 先行研究をみると, 尾木 他 (2016) は, 小規模な法人企業向けモデルに業歴関数を用いることを提案し, 業歴別デフォルト率が重要な変数であることを示した. さらに, 業歴別デフォルト率は, 決算書の数値には表れない小企業のパフォーマンス

スを示す指標と考えられるが、主要な要素として経営者の個人資産の大きさが含まれている可能性があることも示した。

個人企業の場合、事業資産と事業主の私的資産が混在しており、それを明確に区分することは難しいが、個人企業の業歴別デフォルト率を事業主の私的資産が主要な要素とする個人企業のパフォーマンスを示す指標として利用できれば、それを定式化した業歴関数をモデルの変数に含めることによって、個人企業向けモデルの精度を改善できる可能性があると考えた。そこで、業歴別デフォルト率を定式化した業歴関数をモデルの変数に投入して、個人企業向けモデルの精度改善を図ることにした。公庫が保有する約 68 万件のデータを使用して分析した結果、以下のことがわかった。

- (1) 個人企業の業歴別デフォルト率は、業歴 4 年まで急激に低下し、その後 25 年まで緩やかに低下、25 年を過ぎるとおおむね一定になるという法人企業と異なる特徴を有していることがわかった。法人企業で有効であった 3 次関数の R^2 は 0.765 となり、法人企業の 0.932 を大きく下回る結果となり、個人企業に対しては 3 次関数を利用することは適切でないことがわかった。
- (2) 個人企業の特徴を踏まえて、ワイブル分布の密度関数および業歴が「0-3 年」「4-25 年」「26 年以上」の 3 つの区分に分けた区分線形関数を用いた結果、修正 R^2 はいずれも 93% 程度となり、法人企業の 3 次関数と同程度の水準を確保することができた。
- (3) ワイブル分布関数の密度関数および区分線形関数をモデルの変数に投入すると、業種別では AR 値を最大でワイブルは 9.6% ポイント、区分線形で 9.4% ポイント、全体では、ワイブルで 2.3% ポイント、区分線形で 2.6% ポイント向上させることができた。
- (4) ワイブル分布関数の密度関数と区分線形関数は、ほぼ同水準のパフォーマンスを示しているが、ワイブル分布関数の密度関数を投入するといくつかの財務変数が有意にならなくなること、実務では線形の方がオーバーフィッティングしにくい経験則を考慮すると、区分線形関数が望ましい可能性がある。

本研究の成果において、実務上の観点からは(3)の貢献が大きい。個人企業は決算書がなく、その代わりとなる税務申告書に貸借対照表(B/S)を添付していない企業も多く、B/Sに関する情報は法人企業に比べて不足している。モデルの精度を改善するには、個人企業の資産負債に関連する客観的な変数を加えることが有効であるが、個人企業は家計と事業経営の関係が緊密で、会計が混在しているため、資産や負債の把握が極めて困難かつ手間暇がかかる。

本研究によって、個人企業においても業歴別デフォルト率を定式化した業歴関数があることがわかり、信用スコアリングモデルの精度を手間とコストをかけずに高めることが可能になった。ただ、改善幅は法人企業の 7.6% ポイントと比べると低い水準にとどまった。仮説では、個人企業も法人企業と同様に、業歴別デフォルト率が事業主の私的資産の大きさも含めた非開示の小企業のパフォーマンスを示す指標として有効となることを期待したが、開示されている事業資産の中に事業主の私的資産が相応に含まれている(非開示の資産が少ない)、もしくは、非開示の私的資産は事業に投入しない可能性があるため、効果が小さかったと考えられる。

本研究の結果を踏まえた今後の課題は以下のとおりである。

- (1) 個人企業向けモデルの精度をさらに改善するため、入手の手間とコストがかからない資産負債関連の変数を追加することを検討する。
- (2) 副業の収入が個人企業の資産形成に少なからぬ影響を与えていると思われるため、副業に関するファクターの導入について検討する。

- (3) ワイブル分布関数の密度関数と区分線形関数の精度比較を行うため、データの蓄積を待つて再度検証を行う。

なお、他の金融機関に対する利用可能性に関して、本分析で使用したデータは公庫が融資した企業のみであり、サンプルバイアスを持つ可能性があることには注意が必要である。現在、民間金融機関は取引先拡大のために、個人企業をはじめとした小企業への融資を積極化しており、これらのデータを含む本論文の結果は、リスクテイクを積極化する民間金融機関に多くの示唆を与えるものと信じている。本研究の成果が小企業向けの融資を行っている他の金融機関における信用リスク管理の高度化の参考になれば幸いである。

注.

- 1) 個人企業とは法人格を持たない個人で事業を行う個人事業者のことであるが、個人企業の定義は明確にされておらず、個人事業主、自営業者、自己雇用者といった表現も用いられる。本論文では、いずれも表記を個人企業に統一して用いる。総務省統計局「統計 Today No.82」(総務省統計局, 2014)によれば「個人企業とは、法人としての登記をしていない個人が経営する経営体を意味します。例えば、街中の法人登記をしていない個人営業の工場、工務店、小売店、飲食店、美容室、学習塾、税理士などは個人企業ですし、医療法人となっていない医療機関も 1 つの個人企業です。また、例えば、複数の店舗を持つ美容のチェーン企業であっても法人登記していなければ個人企業になります。」と記されている。
- 2) 個人資産額以外のファクターとして経営者の資質や経営ノウハウなどが考えられるが検証は困難である。
- 3) AR 値 (Accuracy Ratio) はスコアリングモデルの精度を示す代表的な指標として普及している。所得の集中度や不平等度を測る「Gini 係数」の考え方を応用したもので、推計されたデフォルト確率を順序性の尺度とし、モデルの精度を評価する。0 から 1 の値をとり、1 に近いほど精度が高い。詳しくは、山下・三浦 (2011) を参照されたい。
- 4) デフォルトに統一的な定義はなく、金融機関やモデルによって定義や算出方法が異なるので、比較には注意が必要である。たとえば、RDB は、過去 12 カ月以内に、3 カ月以上延滞先もしくは破綻懸念先以下の債務者区分に初めて該当した債務者をデフォルトと定義している。また、CRD 協会は①3 カ月以上延滞先、②実質破綻先、③破綻先、④信用保証協会による代位弁済先と定義している。
- 5) 企業 i の決算書 1 期分の J 個の財務指標を用いた変数 $f_{ij}(i = 1, \dots, N; j = 1, \dots, J)$ を使用して、ロジットモデルを構築し、最尤法によってパラメータ $\alpha_j(j = 1, \dots, J)$ を推定する。ここで、 p_i は融資した企業が翌年にデフォルトする確率、 N は企業数を表す。 z_i が大きいほどデフォルト確率は低くなる。

$$p_i = \frac{1}{1 + e^{z_i}}, z_i = \ln \left(\frac{1 - p_i}{p_i} \right) = \alpha_0 + \sum_{j=1}^J \alpha_j f_{ij}$$

次に、推定されたパラメータを用いて計算された z_i から企業 i の信用スコア CS_i を計算する。

$$CS_i = \eta_0 + (\eta_1 - \eta_0) \left(\frac{z_i^* - Z(1\%)}{Z(99\%) - Z(1\%)} \right)$$

ここで、 $Z(1\%)$, $Z(99\%)$ はモデル構築時のインサンプルデータにおける z_i^* の 1 パーセント点、99 パーセント点を表す。これは信用スコアが $z_i^* = Z(1\%)$ ならば η_0 点、 $z_i^* = Z(99\%)$ ならば η_1 点となるように基準化している。本研究では、 $\eta_0 = 10$, $\eta_1 = 90$ としている。

- z_i^* を直接用いても結果に影響を与えない。
- 6) 法人企業も同じ融資年度で分析を行えば、観測期間の影響も受けずに比較が可能である。ただし、本研究の目的は個人企業向けのモデルを構築し、その特徴を考察することであるため、この点は今後の課題としたい。
 - 7) 日本政策金融公庫国民生活事業本部は従業者数 20 人(商業・サービス業は 5 人)未満の小企業への融資が約 8 割を占める。融資先の業種や地域構成比はわが国の事業所の構成比と比べて大きな差はない。公庫の融資ポートフォリオと事業所・企業統計調査との比較、融資先の特徴等については、枇々木 他 (2010)および公庫の HP を参照されたい。
 - 8) データ数で加重平均しているため、5 年間のデフォルト率と等しい。ただし、業歴 1 年は 3 年間(1~3 年)、業歴 2 年は 4 年間(1~4 年)のデフォルト率を計算している。
 - 9) 業歴以外に、財務指標 15 変数(全業種共通 9 変数、業種別 6 変数)を使用している。モデルの概要は枇々木 他 (2010)も参照されたい。

参 考 文 献

- Akben-Selcuk, E. (2016). Does firm age affect profitability? Evidence from Turkey, *International Journal of Economic Sciences*, **5**(3), 1-9.
- Altman, E. I. and Sabato, G. (2007). Modelling credit risk for SMEs: Evidence from the U.S. market, *ABACUS*, **43**, 332-357.
- Cabral, L. M. B. and Mata, J. (2003). On the evolution of the firm size distribution: Facts and theory, *American Economic Review*, **93**(4), 1075-1090.
- 中小企業庁 (2002). 2002 年版中小企業白書, 中小企業庁, 東京.
- 中小企業庁 (2014). 2014 年版中小企業白書, 中小企業庁, 東京.
- Evans, D. S. (1987). The relationship between firm growth, size, and age: Estimates for 100 manufacturing industries, *Journal of Industrial Economics*, **35**(4), 567-581.
- Gweyi, M. O. (2013). Credit risk mitigation strategies adopted by commercial banks in Kenya, *International Journal of Business and Social Science*, **4**, 71-87.
- 枇々木規雄, 尾木研三, 戸城正浩 (2010). 小企業向けスコアリングモデルにおける業歴の有効性, 『定量的信用リスク評価とその応用』, ジャファイア・ジャーナル: 金融工学と市場計量分析, 83-116, 朝倉書店, 東京.
- Kneiding, C. and Kritikos, S. A. (2011). Funding self-employment: The role of consumer credit, IZA Discussion Paper, 6093 IZA, Germany.
- 国民生活金融公庫総合研究所 (2002). 自己雇用者(≡ 自営業者)に関する実態調査～アンケート結果の概要～, 国民生活金融公庫レポート, 国民生活金融公庫総合研究所, 東京.
- Loderer, C. F. and Waelchli, U. (2010). Firm age and performance, available at SSRN: <http://ssrn.com/abstract=1342248>.
- 森平爽一郎, 岡崎貴治 (2009). マクロファクターを考慮したデフォルト確率の期間構造推定, 早稲田大学大学院ファイナンス総合研究所ワーキングペーパーシリーズ, WIF-09-004, 早稲田大学ファイナンス総合研究所, 東京.
- 森川正之 (2008). 同族企業の生産性—日本企業のマイクロデータによる実証分析—, RIETI Discussion Paper Series, 08-j-029, 独立行政法人経済産業研究所, 東京.
- 内閣府 (2011). 平成 23 年 年次経済財政報告, 内閣府, 東京.
- 日本政策金融公庫総合研究所 (2016). 中小企業の事業承継に関するインターネット調査～アンケート結果の概要～, 日本政策金融公庫総合研究所, 東京.
- 日本政策金融公庫総合研究所 (2017). 2017 年度新規開業実態調査～アンケート結果の概要～, 日本政策

- 金融公庫総合研究所, 東京.
- 尾木研三, 戸城正浩, 枇々木規雄 (2016). 小規模企業向け信用スコアリングモデルにおける業歴の頑健性と経営者の個人資産との関係性, *Transactions of the Operations Research Society of Japan*, **59**, 134-159.
- 尾木研三, 内海裕一, 枇々木規雄 (2017). 創業企業の信用リスクモデル, 『リスク管理・保険とヘッジ』, ジャフイー・ジャーナル: 金融工学と市場計量分析, 133-156, 朝倉書店, 東京.
- Ohlson, J. A. (1980). Financial ratios and the probabilistic prediction of bankruptcy, *Journal of Accounting Research*, **18**, 109-131.
- Sakai, K., Uesugi, I. and Watanabe, T. (2010). Firm age and the evolution of borrowing costs: Evidence from Japanese small firms, *Journal of Banking and Finance*, **34**, 1970-1981.
- Schreiner, M. (2000). Credit scoring for microfinance: Can it work?, *Journal of Microfinance*, **2**, 105-111.
- 清水龍宝 (1985). 中小企業倒産の構造的要因の分析, 三田商学研究, **28**(2), 1-24.
- 総務省統計局 (2014). 平成 26 年度統計 Today, No.82, 総務省統計局, 東京, <https://www.stat.go.jp/info/today/082.html>.
- 総務省統計局 (2016). 平成 26 年経済センサス—基礎調査, 総務省統計局, 東京.
- 高川泉, 亀田制作 (2008). わが国における個人企業の動向とその背景, 日銀レビュー, 2008-J-12, 日本銀行, 東京.
- 友田信男 (2008). 中小企業の倒産動向, 日本政策金融公庫調査月報, **2**, 4-15.
- 鶴田大輔 (2005). ノンバンク融資と中小企業のモラルハザード問題, RIETI Discussion Paper Series, 05-J-035, 独立行政法人経済産業研究所, 東京.
- 上村協子 (2006). 自営業世帯の家計資産, 信金中金月報, **397**, 55-74.
- 山下智志, 三浦翔 (2011). 『信用リスクモデルの予測精度—AR 値と評価指標—』, 朝倉書店, 東京.
- 山下智志, 川口昇, 敦賀智裕 (2003). 信用リスクモデルの評価方法に関する考察と比較, Financial Research and Training Center Discussion Paper Series, **11**, 金融庁金融研究研修センター, 東京.
- 安田武彦, 高橋徳行, 忽那憲治, 本庄裕司 (2007). 『テキストライフサイクルからみた中小企業論』, 同友館, 東京.
- 柳澤健太郎, 下田啓, 岡田絵理, 清水信宏, 野口雅之 (2007). RDB データベースにおける信用リスクモデルの説明力の年度間推移に関する分析, 日本金融・証券計量・工学会 2007 年夏季大会予稿集, 249-263.

Effect of Firm Age in Credit Scoring Model for Loans to the Self-employed

Kenzo Ogi¹, Yuichi Utsumi¹ and Norio Hibiki²¹Micro Business and Individual Unit, Japan Finance Corporation²Faculty of Science and Technology, Keio University

A number of Japanese banks have utilized credit scoring models to reduce screening costs by automating part of the screening process. Several types of credit scoring models exist. To calculate the credit scores of small firms with twenty or fewer employees, it is common to utilize the logistic regression model, linked to the correlations between financial indicators and the occurrence of default.

Small firms include not only corporate businesses, but also self-employed businesses. The accuracy of the credit scoring model for the self-employed is not as high as might be expected. The main reason for this is that we lack adequate information regarding financial accounting variables on the balance sheet, such as assets or debts, because self-employed firms are not obligated to prepare a financial statement. We are convinced that it is effective to use explanatory variables associated with assets or debts. However, it is difficult for banks to obtain accurate data, because most self-employed households confuse the owner's private assets with business finances.

Ogi et al. (2016) suggest the use of "firm age" in the credit scoring model for small corporate business, and we have found this to be a statistically significant variable. There are at least three reasons for this. First, for loans to small firms, it is effective to evaluate the private assets of the owner together with the assets of the firm, but accurate information on this subject is not easily available. Second, the default rates of small firms decrease as the firms get older because the private assets of owners are expected to accumulate year by year, and this can help with management activities. Third, in the traditional screening process of loans by experienced bankers, firm age is an important factor for assessing the credit risk to debtors. In this paper, we propose to utilize "firm age" in credit scoring models for self-employed as well as small corporate businesses, and to analyze the correlation between firm age and default rate using a dataset for the more than 680,000 self-employed in Japan from 2007 to 2014. This dataset is owned by the Japan Finance Corporation. We will evaluate the credit scoring model using the accuracy ratio, which is commonly used as a measure.

We find that the default rate can be expressed by the density function of the Weibull distribution or a piecewise-linear function of firm ages with three ranges: up to 3 years, 4–25 years, and 26 or more years. We also show that the accuracy ratios increase by about nine percent at the maximum in both approaches, and that the model can be effectively used in practice.