

令和 2 年 7 月 8 日現在

機関番号：13601

研究種目：若手研究(B)

研究期間：2016～2019

課題番号：16K17241

研究課題名(和文)元留学生の帰国と職歴パターン

研究課題名(英文)On the return and job history of ex-international students

研究代表者

前田 豊 (Maeda, Yutaka)

信州大学・学術研究院人文科学系・助教

研究者番号：50637303

交付決定額(研究期間全体)：(直接経費) 2,200,000円

研究成果の概要(和文)：本研究では、元留学生の帰国に影響する要因の解明に向けた方法論の構築を目的とした。具体的には、在留の元留学生を対象とする調査データの利用を想定し、疑似尤度にもとづく2つの推定方法(SC推定・TLR推定)の比較検討、および、回答者ネットワークによる標本抽出の観点からTLR推定を拡張した推定方法の提唱と検討を試みた。結果として、サンプルサイズ、および従属変数の値のシェアが既知であるか否かが、SC推定・TLR推定の選択の基準になりうることを示した。また、TLR推定を拡張した推定方法が回答者ネットワークの構造に頑健であることが示し、ネットワークベースの抽出方法にも対応可能であることを明らかにした。

研究成果の学術的意義や社会的意義

本研究の学術的意義としては、既存の従属変数の値による抽出データに対応した推定方法の議論と、回答者ネットワークを通じた抽出方法の議論を接合する新たな推定方法の提唱により、双方の領域での発展に寄与した点を挙げることができる。また、本研究の関心であった元留学生のみならず、無作為抽出での捕捉が難しい、いわゆる隠れた母集団への適用が可能となっており、広範な社会現象に適切に対応する推定方法を提唱したという点においても、学術的・社会的意義が認められる。

研究成果の概要(英文)：The purpose of this study is to consider a methodology for explaining the factors that influence the return of ex-international students. Specifically, assuming the use of survey data of ex-international students in Japan, this study conducts a comparative study of two estimation methods (SC estimation and TLR estimation) based on pseudo-likelihood, and attempts to propose and examine a new estimation that extends TLR estimation from a perspective of respondent's network based sampling. As a result, this study shows that the sample size and whether the shares of the values of the dependent variables are known can be the criteria for the selection of SC estimation and TLR estimation. In addition, this study shows that the estimation method that extends the TLR estimation is robust to the structure of the respondent network, and clarifies that this estimation method can be applied to network based sampling method.

研究分野：社会学

キーワード：帰国 留学生 選択バイアス ネットワーク

## 様式 C - 19、F - 19 - 1、Z - 19 (共通)

### 1. 研究開始当初の背景

グローバル化の進展、少子高齢化に伴う日本の労働力人口減少を背景に、高度な専門知識・技術を有する高度外国人労働者の受入が積極的に推進されてきた。なかでも日本の教育機関を修了した留学生は「高度外国人材の卵」として位置付けられ、緩和的な在留資格(例えば、就職活動中の「特定活動」)など、日本での就労を促進する政策が整備されてきた。日本は国際的にみても主要な留学生受け入れ国の一つであり(OECD 2014)、日本の教育機関を経た元留学生の存在は、日本労働市場において無視できない存在になりつつある。

元留学生の日本での就労に関する特徴として、留学生の多くが日本人とは区別されない採用過程を経て、その後も同様の処遇を受けていることが明らかになっている(守屋 2012)。この事実は、「定住者」や「研修」・「技能実習」資格による「サイドドア」からの外国人労働者が、日本人労働者とは分断された外部労働市場を構成しているのに対して(梶田ら 2005, 上林 2015)、留学生が参入する労働市場が日本人新卒者も参入する内部労働市場に近いことを示唆する。

しかし、労働市場参入後にも日本人新卒者と同様の職歴を歩むとは限らない。というのも、元留学生の意思決定が日本での期待効用最大化(Todaro and Harris 1970)に向いているのではなく、第三国、もしくは母国への帰国を念頭に置いている可能性が指摘されてきたためである(e.g. Mahmud 2014, 坪谷 2008)。それゆえ、たとえ日本人新卒者と同様の労働市場に参入したとしても、国家間移動のオプションがあることで、日本人新卒者とは異なる職歴パターンになり得る可能性がある。しかし、移民の帰国は伝統的に社会学で問われてきた理論的・実証的問い(Bovenkerk 1975)だが、日本における元留学生の帰国そのものを実証的に検討した試みは数が少ない。どのような要因が元留学生の帰国や第三国への移動、翻っては残留を促進するのか、そして、その要因が日本の内部労働市場の特徴とどのように絡んでいるのか、などの移動に関連する問題はいまだ未解決のままである。

### 2. 研究の目的

国家間の移動に関わる要因の解明にアプローチする計量的な方法としては、「移動」「在留」の二値変数を従属変数とする離散選択モデルの援用が一般的であると理解できる。しかし、比較的容易に実施が可能な1時点の日本在留の元留学生を対象とする調査では、従属変数の一方の値に相当する「移動」は観察されていないため、一般的な尤度関数を用いた推定は難しい。また、「移動」した回答者の情報がすべて欠測しているため、形式的には従属変数で観察されていない値によって欠測が決まるというNMARに相当していると理解でき、この点においても推定の困難さが理解できる。

これらの点は、1時点の日本在留の元留学生を対象とする調査データが、とりわけ在留を従属変数とする関心に対しては必ずしも適当ではないことを示唆する。しかし、継時的に国家を移動した回答者までを追跡する調査の実施は現実的に難しい。

そこで、本研究では、在留を促進する要因に関する知見を経験的に妥当する形で析出するための基礎的な方法論の確立を目的とし、1時点の日本在留の元留学生を対象とする調査データを用いた離散選択モデルの推定方法を検討した。

### 3. 研究の方法

1時点の日本在留の元留学生を対象とする調査データは、従属変数の値「在留」に限定される形で共変量が測定されている調査データと理解することができる。こうした従属変数の値で限定されている調査データを用いた推定については、経済学・生物学分野を中心に蓄積されている。

本研究では、まず、これらの先行研究で提唱されている種々の推定方法のなかでも、とりわけ実装の容易さの観点から、疑似尤度関数を用いた推定方法であるSteinberg and Cardell(1992)の方法と(SC推定)とTang, Little and Raghunathan(2003)による方法(TLR推定)を取り上げ、モンテカルロシミュレーションから、この2つの方法の相違を検討した。

また、日本在留の元留学生を対象とする調査データの抽出方法に関するヴァリエーションに関心とし、とりわけ回答者ネットワークを利用した抽出方法であるRespondent Driven Sampling(Heckathorn 1997. RDS)を取り上げ、TLR推定との接合から新たな推定方法の提唱、および推定量のネットワーク構造への頑健性を、同じくモンテカルロシミュレーションから検討を試みた。

### 4. 研究成果

#### (1) SC推定とTLR推定の検討

SC推定とTLR推定は、いずれも片方の従属変数のみの値に制限して共変量が測定されているデータ(完全データ)と、従属変数が完全に欠測しているが、共変量は完全に測定されたデータ(補足データ)の2つのデータを用いる方法である。以下、完全データのサンプルサイズを $N_{comp}$ とし、補足データのサンプルサイズを $N_{sup}$ とする。また、推定に使用する離散選択モデルを $\Pr(Y|X; \beta)$ で表し、完全データで観察可能な従属変数 $Y$ の値を1とする。

SC推定は、基本的には二項ロジスティックモデルを離散選択モデルに想定した推定方法で、以下の式で定義される対数尤度関数 $LL_{SC}$ を用いる。

$$LL_{sc} = \sum_{R_i=0} \log Pr(Y=0|X_i, \beta) + s \frac{N_{sup}}{N_{comp}} \sum_{R_i=1} \log \frac{Pr(Y=1|X_i, \beta)}{Pr(Y=0|X_i, \beta)}$$

ここで $s$ は、母集団における $Y = 1$ のシェアを示し、 $R_i = 0, 1$ のサメーションは、それぞれ補足データ、完全データでのサメーションを示す。

対して、TLR 推定は完全データの直接尤度関数のベイズの公式による変形から導出される、以下の式で定義される対数尤度関数 $LL_{TLR}$ を用いる（式の積分部分は補足データを使用）。

$$LL_{TLR} = \sum_{R_i=1} \log Pr(Y=1|X_i, \beta) - \sum_{R_i=1} \log \int Pr(Y=1|X, \beta) dF(X)$$

SC 推定と TLR 推定の対数尤度関数上の違いは、完全データ / 補足データのサンプルサイズ比、および  $Y = 1$  のシェア  $s$  を用いているか否かの点に見出される。こうした相違が生み出す推定値の違いを理解するため、Lancaster and Imbens (1996) のセッティングを参考にし、二つの共変量を想定する二項ロジスティックモデルを想定したモンテカルロシミュレーションを行った。

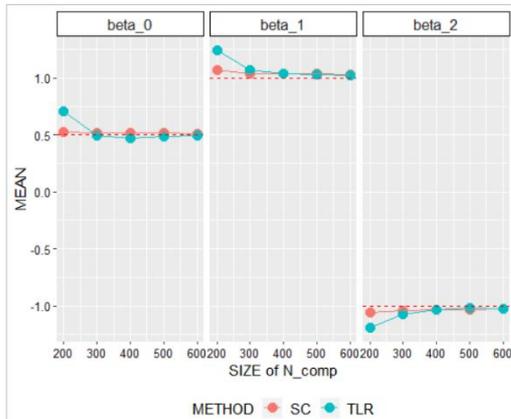


図 1 推定方法による推定値平均（赤破線：真値）

結果として、シェア  $s$  が既知であるという条件のもと、補足データのサンプルサイズを固定とした場合に、完全データのサンプルサイズが相対的に小さい場合には、TLR 推定では比較的大きなバイアスが生じる、完全データのサンプルサイズが相対的に大きくなれば、TLR / SC 推定ともにバイアスがほとんど確認されないことが明らかとなり、完全データのサンプルサイズに頑健な SC 推定の有用性が示された（図 1）。しかし、真のシェア  $s$  が既知でなく、真のシェア  $s$  からかい離した値を誤って用いた場合には、翻っては大きなバイアスを生み出すことも同時に明らかとなり（図 2）、完全データのサンプルサイズと同時にシェアの情報の有無も推定方法を選択する基準であることが明らかとなった。

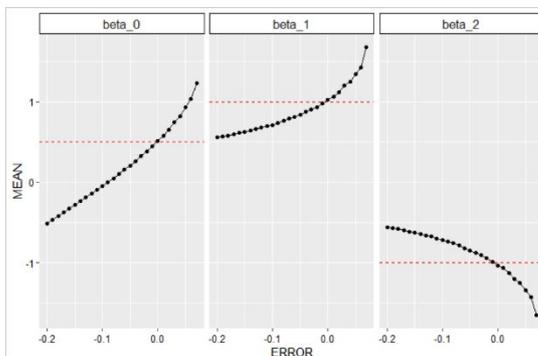


図 2 真の  $s$  とのかい離 (ERROR) と SC 推定による推定値平均（赤破線：真値）

## (2) RDS データに対応する推定方法

RDS は、各回答者が保有するパーソナルネットワークを利用した一種の有意抽出法で、回答者から次の回答者を紹介してもらうことで標本データが生成される。本課題では、完全データが従属変数の値に基づく RDS から収集された状況を想定し、有意抽出における TLR 推定の発展を試みた。

RDS の基本的な仮定を所与とし、十分な紹介プロセス数を仮定すれば、それぞれの回答者が紹介可能な回答者数  $d_i$  を用いて、各回答者の抽出確率  $\pi_i$  は以下の式で得ることができ (Salganik

and Heckathorn 2004 ),

$$\pi_i = \frac{d_i}{\sum_{R_1=1} d_j}$$

この抽出確率 $\pi_i$ を用いた Hajek タイプの重み $w_i$ は、各標本の抽出確率のばらつきを補正する重みとして利用することが可能である (Volz and Heckathorn 2008). 以上を踏まえ、本研究課題では、完全データが RDS で得られた場合に対応する TLR 推定の発展として、 $w_i$ と TLR 推定で使用された疑似尤度関数 $LL_{TLR}$ を M-estimator の形で接合した疑似尤度関数 $LL_{wTLR}$ を提唱した。

加えて、本研究課題では、回答者ネットワーク構造、とりわけ共変量による同類性結合に対する $LL_{wTLR}$ の頑健性についてモンテカルロシミュレーションから検討を加えた。

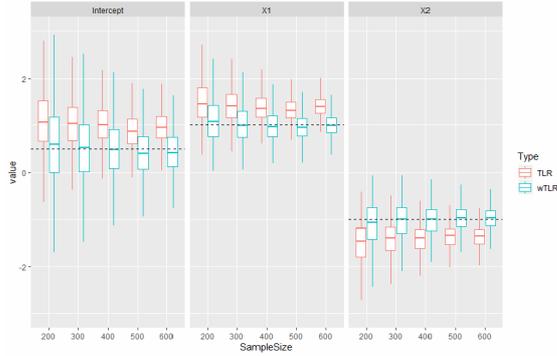


図 3 共変量による同類性結合が存在する場合の推定値 (破線: 真値)

結果として、回答者のネットワークに共変量による同類性結合が見られる場合、従来の $LL_{TLR}$ では完全データのサンプルサイズに依らない形で一貫したバイアスが生じているのに対して、今回提唱した $LL_{wTLR}$ では相対的にバイアスが少ない安定した推定結果を与えていることが明らかとなった (図 3). また、同類性結合だけではなく、共変量によるネットワークサイズに対しても同様の結果となった。

以上の結果は、日本在留の元留学生を対象とする調査データに適用する場合、そのデータが必ずしも無作為抽出を前提とする必要がない点、そして、例えば国籍や学校歴といった移動と回答者ネットワークに同時に影響を及ぼす共変量が存在したとしても、共変量の効果の推定が可能であることを示唆する。

5. 主な発表論文等

〔雑誌論文〕 計2件（うち査読付論文 1件/うち国際共著 0件/うちオープンアクセス 0件）

1. 著者名 前田豊・鎌田拓馬	4. 巻 34(1)
2. 論文標題 Synthetic Control Method を用いた個別事例の因果効果の識別	5. 発行年 2019年
3. 雑誌名 理論と方法	6. 最初と最後の頁 78-96
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 多田はるみ・前田豊	4. 巻 18
2. 論文標題 ポジティブ・アクションによる不満 相対的剥奪理論を拡張したモデルによる考察から	5. 発行年 2017年
3. 雑誌名 立教大学ジェンダーフォーラム年報	6. 最初と最後の頁 27-41
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計2件（うち招待講演 1件/うち国際学会 0件）

1. 発表者名 前田豊・朝岡誠
2. 発表標題 「隠れた人口」に対応する推定方法について
3. 学会等名 第64回大会数理社会学会大会
4. 発表年 2017年

1. 発表者名 Yutaka Maeda
2. 発表標題 Who became Sex worker? The case of Japan in Post-war Period
3. 学会等名 1st Korea-Japan Joint Workshop on Frontiers of Social Survey Research（招待講演）
4. 発表年 2018年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
--	---------------------------	-----------------------	----