

Covariate Balancing Propensity Score を 用いた, スクイズ作戦の有効性の解析

中村 知繁¹・南 美穂子²

(受付 2017 年 1 月 15 日; 改訂 4 月 30 日; 採択 8 月 28 日)

要 旨

近年, スポーツではデータの取得が盛んに行われており, 特にアメリカの大リーグの試合データは全打席一球ごとの記録が公開されている. 本論文では, このデータを用いて野球におけるスクイズ作戦が得点する確率にあたる影響について Covariate Balancing Propensity Score (CBPS, Imai and Ratkovic, 2014) を用いて解析を行った結果を報告する. 特に, 本解析においては, どうしても 1 点が欲しい状況である得点差が 0 点または 1 点の場合のスクイズ作戦の有効性に焦点を当てる (ただし満塁の場合は除く). スクイズ作戦が得点する確率にあたる影響を推定するためには, スクイズ作戦をとった場面と, とらなかった場面を比較するというのが簡単な方法である. しかし, スクイズ作戦をとるかとらないかはランダムではなく, 打者や投手の情報といった共変量に依存して決定されるのが一般的であるため, 単純な 2 群の比較を行ってしまうと, 推定されたスクイズ作戦の影響に共変量の影響が含まれてしまい, 適切に推定を行うことができない. そこで, 本論文では傾向スコアを用いて共変量調整を行った上で, スクイズ作戦が得点する確率に与える影響を推定した. さらに傾向スコアの推定をロジスティック回帰モデルを用いてスクイズ作戦の影響を推定した場合と, Covariate Balancing Propensity Score を用いてスクイズ作戦の影響を推定した場合の結果を比較した. その結果, CBPS を用いて影響を推定する方が, ロジスティック回帰モデルを用いて傾向スコアを推定した場合に比べて, 比較する 2 つの群の共変量の分布の平均がよく釣り合い, スクイズ作戦が得点する確率に与える影響を安定して推定できることがわかった. また, CBPS を用いた解析からスクイズ作戦は平均の意味で 18.2% 得点する確率を上昇させることがわかった.

キーワード: 野球, スクイズ作戦, 因果推論, 共変量調整, Covariate Balancing Propensity Score.

1. はじめに

近年スポーツの世界では, プレイデータの取得が盛んに行われるようになり, データを用いた作戦や戦術の評価を行うことができるようになりつつある. 特に, 大リーグのデータは, 過去の試合記録が 20 年分以上公開されており, これらのデータを用いて野球における戦略や戦術について, 統計的に評価をすることが可能になりつつある.

野球における戦略には, 手段の有効性の検証が十分になされていないものも多く, 野球ファ

¹ 慶應義塾大学大学院 理工学研究科: 〒 223-8522 神奈川県横浜市港北区日吉 3-14-1

² 慶應義塾大学 数理科学科: 〒 223-8522 神奈川県横浜市港北区日吉 3-14-1

ンや、解説者などの間でも意見が割れる戦術が存在する。そのうちの代表的な作戦が、バント作戦である。バント作戦にはいくつかの種類があり、よく知られたものでは、ノーアウト、または1アウトかつランナー1塁の状況において、バントを行い、打者はアウトになるが、ランナーを得点圏に進めるといった犠牲バント作戦や、ノーアウト・1アウトでランナーが3塁にいる状況で、バントをし、ランナーを生還させるスクイズバント作戦(以下、スクイズ作戦)などがある。これらの作戦は、成功した時の華やかさがあり、とくにスクイズ作戦については、成功すると歓声上がる。一方で、そのような華やかな側面があるために、実際には失敗しているケースも少なからずあるが、スクイズ作戦の有効性について過剰に評価している可能性もある。そこで、本論文では、スクイズ作戦に焦点を絞り、その有効性について解析した結果を報告する。

野球におけるバント作戦の評価を行った文献には、Albert and Bennet (2001)や、仁志・鳥越(2012)などがある。これらの文献では、選手の打率と犠牲バントの成功確率を計算した上で、得点する確率の期待値を計算し、犠牲バントの有効性について解析している。また、Lee (2011)では確率的フロンティア分析の手法を用いて、犠牲バントや盗塁などを包括したスモールベースボール戦略の有効性を、得点数を犠牲バントや盗塁数で回帰することによって推定している。この他にも、論文という形式は取られていないが、日本経済新聞が「日本でも「脱バント」が浸透するか」(丹羽, 2011)という連載を行っていたり、アメリカにおいても、Turner Sportsというメディアが「Explaining Why the Bunt Is Foolish in Today's MLB」(Zachary, 2013)という記事を掲載していることから、バント作戦については興味関心が高いことがうかがえる。これらの記事で話題にしているのは、犠牲バント作戦であり、バント作戦をとった場面と、取らなかった場面での進塁率や、生還率について単純な標本平均を用いて比較し、その結果、犠牲バントは有効な作戦とは言えないという結論を導いている。

これらの手法におけるバント作戦の有効性の評価は、単純な2群比較をもとにした期待値の比較や、回帰係数を因果効果として解釈し行われている。では、果たしてこれらの解析によって犠牲バント作戦や、今回の対象であるスクイズ作戦の有効性を解析することはできるだろうか？

バントの有効性について議論するためには、有効性とは何かを定義する必要がある。そのために以下の記号を導入する。サンプル(場面) i に対して、 T_i をバント作戦を行ったかどうかの処置変数、 Y_i を得点したかどうかの2値の結果変数、 X_i をバント作戦及び、結果変数に関係する状況を表す共変量ベクトルとする。例えば、共変量としては、打者や投手の情報が挙げられる。さらに、 $Y_i(1)$ をバントをした時に観測される結果変数、 $Y_i(0)$ をバントをしなかった時に観測される結果変数とすると、結果変数 Y_i は

$$(1.1) \quad Y_i = T_i Y_i(1) + (1 - T_i) Y_i(0)$$

と表すことができる。実際、同一の標本 i から $Y_i(1), Y_i(0)$ の両方が観測されることはなく、バント作戦をとったかどうかで、いずれか一方が観測される。ここで、バント作戦の有効性を、バントをした場合と、バントをしなかった場合に観測された結果の差の期待値であると考えらることにすれば、バント作戦の有効性は以下のように定義することができる。

$$(1.2) \quad \mu = E[Y_i(1) - Y_i(0)]$$

この μ を平均処置効果(Average Treatment Effect; ATE)と呼ぶ。この考え方は統計的因果推論の枠組みの1つであり、Rosenbaum and Rubin (1983), Hirano et al. (2003), 星野 (2009)などが詳しい。

ここで、先ほどの単純な標本平均を用いた比較は、上記の記号を用いて以下のように表さ

れる。

$$(1.3) \quad \frac{1}{N_1} \sum_{i=1}^N T_i Y_i - \frac{1}{N_0} \sum_{i=1}^N (1 - T_i) Y_i$$

ここで、 N は観測されたデータのサンプルサイズである。また、 N_1 はスクイズ作戦をとった群のサンプルサイズであり、 N_0 はスクイズ作戦をとらなかった群のサンプルサイズである（ただし、 $N = N_1 + N_0$ である）。この推定値は漸近的 ($N_0, N_1 \rightarrow \infty$) に、 $E[Y_i(1)|T_i = 1] - E[Y_i(0)|T_i = 0]$ へと収束するため、これはバント作戦の有効性を、平均処置効果の意味で表していると考えられることはできない。このような理由から、単純平均を用いた比較では、バントの有効性を正しく推定できていない。

そこで、中村・南 (2016) では、このような問題へ対処するために、傾向スコア (Rosenbaum and Rubin, 1983) を用いて、大リーグのデータから犠牲バントの因果効果の推定を行った。傾向スコア $\pi(X_i)$ とは、共変量が与えられた元で処置を受ける確率であり以下で定義される。

$$(1.4) \quad \pi(X_i) = P(T_i = 1|X_i)$$

本論文では、投手や打者の情報(共変量)が与えられたもとで、スクイズ作戦をとる確率に対応する。

さて、Rosenbaum and Rubin (1983) では、以下の強く無視可能な割り付けの仮定(1.5)が成立するとき、傾向スコアを用いることで、 μ の推定が可能であることを示した。

$$(1.5) \quad (Y_i(1), Y_i(0)) \perp\!\!\!\perp T_i | X_i$$

これは、スクイズ作戦をとるか、とらないかに関係する変数 X_i を条件付けたもとで、得点のインジケータである潜在結果変数と、バント作戦を取るか取らないかの割り付け変数が独立になることを意味する。強く無視可能な割り付けの仮定が成立するもとは、傾向スコアの真値が既知であるとき、式(1.6)で定義される推定量は μ に対する一致推定量となる。式(1.6)で定義される推定量を、傾向スコアによる逆重み付け推定量 (Inverse Probability Weighted estimator, IPW 推定量) と呼ぶ。

$$(1.6) \quad \hat{\mu} = \frac{\sum_{i=1}^N \frac{T_i Y_i}{\pi_i}}{\sum_{i=1}^N \frac{T_i}{\pi_i}} - \frac{\sum_{i=1}^N \frac{(1 - T_i) Y_i}{1 - \pi_i}}{\sum_{i=1}^N \frac{1 - T_i}{1 - \pi_i}}$$

ただし、この推定量は傾向スコアが既知であるもとで定義されているが、解析では傾向スコアは未知であることが多い。そのため、実際にはデータから推定された傾向スコアを用いて推定量を計算する。一般的には、傾向スコア π_i を p 次元のパラメータベクトル $\beta \in \mathbb{R}^p$ を用いてパラメータ化し、データからパラメータを推定することで傾向スコアの推定値を得る。ここで、 β でパラメータ化された傾向スコアを $\pi_\beta(X_i)$ と表すことにする。

$\pi_\beta(X_i)$ を推定する際には、 $\pi_\beta(X_i)$ に対して式(1.7)を仮定したロジスティック回帰モデルが用いられることが多い。以降、ロジスティック回帰モデルを用いた傾向スコアの推定値とは、最尤法を用いて得られたパラメータの推定値を $\pi_\beta(X_i)$ に代入したものを意味することとする。

$$(1.7) \quad \pi_\beta(X_i) = \frac{\exp(X_i^T \beta)}{1 + \exp(X_i^T \beta)}$$

中村・南 (2016) においても、傾向スコアをロジスティック回帰モデルを用いて推定し、犠牲バントが得点する確率に与える影響について解析を行っている。しかしながら、傾向スコアを用いた解析においては、傾向スコアに対するモデルを誤特定した場合には、IPW 推定量の一致

性が失われるという点に注意が必要である。特に、Kang and Schafer (2007)では、傾向スコアに対するモデルの誤特定が、推定値に大きな影響を与えることをシミュレーションによって確認している。実際、野球の試合においてスクイズ作戦をとるかどうかを決定する際、投手や打者の情報に依存していることは先ほども指摘したが、どのような関数の形で依存しているかを正確に特定するのは困難である。そのため、傾向スコアに対するモデルは多少の誤特定を含むと考えるのが自然であり、傾向スコアに対するモデルを誤特定をしているならば IPW 推定量の一致性が損なわれる可能性がある。

このような問題に対処するため、本論文では傾向スコアに対するモデルが、真のモデルからそれほど大きくずれていない場合に、結果を頑健に推定することができる Covariate Balancing Propensity Score (Imai and Ratkovic, 2014; CBPS)を用いる。CBPS は、傾向スコアの推定で用いられる方程式が、両群の共変量分布を釣り合わせる重みであると解釈できる点に着目して傾向スコアを推定する手法である。CBPS の頑健な性質は、シミュレーションによっても確かめられている。

本論文における解析では、CBPS の釣り合いの関数として、Imai and Ratkovic (2014)においても使用される、ロジスティック回帰モデルのスコア方程式から導出される釣り合いの関数と、共変量の 1 次のモーメントの関数の両方を用いて推定を行った。その結果、ロジスティック回帰モデルを用いて傾向スコアを推定した場合に比べて、CBPS を用いた場合のほうが、推定された傾向スコアで重み付けされた 2 群の共変量の分布が平均の意味で釣り合った。よって、今回の解析で推定されたスクイズ作戦が得点する確率に与える影響は CBPS を用いる方が、バイアスが小さく推定されていると考えることができる。

本論文の構成は以下の通りである。まず、第 2 節でスクイズの有効性を推定するために用いたデータと、その加工手順について説明したうえで、スクイズ作戦をとった場面と、とらなかった場面の単純な比較を行った結果について述べる。第 3 節では、ロジスティック回帰モデルを用いて傾向スコアを推定し、スクイズ作戦の有効性を推定した結果と、その問題点について指摘する。続く、第 4 節では第 3 節での問題点を踏まえて、CBPS を用いてスクイズ作戦の有効性を解析した結果と、第 3 節の結果との比較を行い、第 5 節で議論をまとめる。

2. 解析対象のデータと、単純な標本平均による比較

本節では、解析に用いたデータの加工手順について述べた後で、スクイズ作戦の得点する確率に与える影響を、スクイズ作戦をとった場面と、とらなかった場面の単純な 2 群の比較を行った結果について述べる。

2.1 解析対象となるデータの前処理

まず本論文で扱うデータは、2006 年～2014 年の大リーグの試合データである(全サンプルサイズは 15 万 8491 であった)。このデータは、<http://www.retrosheet.org/>から取得することが可能である。

特に本論文ではスクイズ作戦の得点する確率に与える影響を解析する場面を、対戦チーム同士の点差が僅差であるような場合、すなわち得点差が $-1, 0, 1$ の状況のもとでのスクイズの有効性を解析するため、まずこのような場面にデータを絞り込んだ(このような場面は 7 万 7096 であった)。

スクイズ作戦には、状況の異なる 2 種類の作戦がある。それは、満塁におけるスクイズ作戦と、満塁ではない場面におけるスクイズ作戦である。これらの 2 つは、作戦としてのリスクが大きく異なるため、今回の解析においては、後者を対象とする。具体的には、「ランナー三塁」

表 1. スクイズをした場合としなかった場合で、得点した場面と、しなかった場面を集計した結果.

	得点が入らなかった	得点があつた	得点率
スクイズをした群	189	222	0.544
スクイズをしなかった群	12846	15343	0.540

表 2. スクイズをした群と、しなかった群の単純平均の差のばらつきをブートストラップ法で計算した結果.

	2.5%	25%	50%	75%	97.5%
スクイズをした群と、しなかった群の得点する確率の差	-0.054	-0.021	-0.004	0.013	0.043

「ランナー一塁・三塁」「ランナー二塁・三塁」の3つの場面を解析の対象とする(このような場面は5万9640であった).

ここからは、解析対象としたデータの加工の手順について具体的に述べる. 本解析で対象としたデータでは、前年に試合に出場していない選手(例えば、新人選手)が含まれている. このような選手に対しては、前年の試合記録を利用することができないため、解析の対象とすることが難しい. そこで、今回はこのようなデータは解析の対象からは取り除くこととした. 続けて、解析の対象となるデータに含まれている選手の中でも、前年度の出場記録が少ない選手を解析の対象から除外した. 具体的には、打席数が30打席に満たない打者や、投球イニング数が10イニングに満たない投手が含まれているデータを解析の対象から取り除いた(このような場面は3万8458であった).

さらに、野球の試合においては延長戦があるが、延長戦は1点で勝負が決まるため、通常のイニングとは性質が異なっている. そのため、延長戦のデータも解析の対象から取り除いている(このような場面は3万3718であった).

最後に、ボールカウントが3の場合には、打者はスクイズ(フォアボールを狙うほうが良い)をしないため、このような場面も解析対象から取り除いた. 以上のようなデータの加工を行った結果、解析の対象となったサンプルサイズは2万8600であった.

2.2 単純な2群の比較による、スクイズ作戦の得点する確率に対する影響の解析

ここでは、スクイズ作戦が得点する確率に与える影響を、単純な2群の比較を行って推定した結果について述べる. まず、スクイズ作戦をとった場合と、とらなかった場合のそれぞれでデータを分割し、その中で得点が入った場面数と、入らなかった場面数をまとめたものが表1である.

表1から、スクイズ作戦をとった場合の方が、スクイズ作戦をとらなかった場合に比べて、得点する確率は0.4%上昇することがわかる. 次に、ブートストラップ法を用いて、この推定値のパーセントイル点を計算した結果が、表2である.

表2より、スクイズ作戦が得点する確率に与える影響の95%信頼区間は $[-0.054, 0.043]$ となり、信頼区間は0を含むことがわかる. この結果、単純な2群の比較を用いた解析では、スクイズ作戦は得点する確率には有意水準5%で影響を与えていないということになる.

ところで、ランナーが3塁にいる場面では、得点する確率に影響を与える変数は、スクイズ作戦をとるかどうかだけではない. 例えば、ランナー3塁の状況での打者の打率や、長打率は、得点する確率(結果変数)に影響を与えることは野球経験者なら想像がつく. また、打率の高い打者であれば、打者にスクイズ作戦を取らせるよりも、バッティングを行わせたほうが成功確

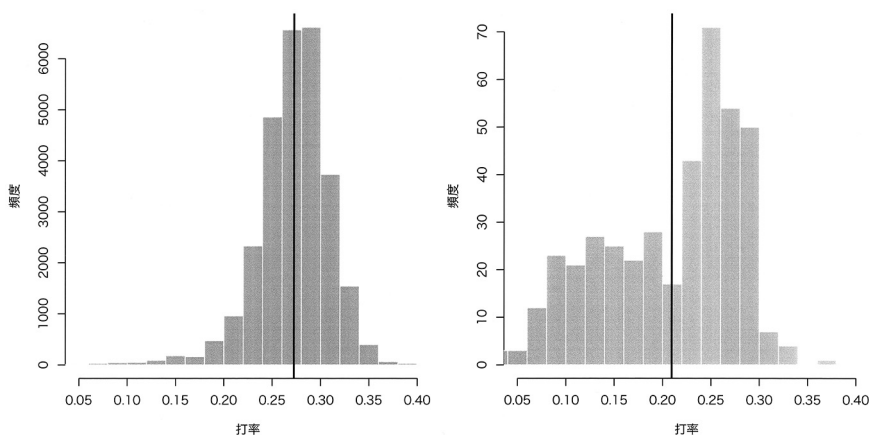


図 1. スクイズ作戦をとった群と、とらなかった群の打率の分布。右：スクイズ作戦をとった群，左：スクイズ作戦をとらなかった群。実線はそれぞれの群の打率の平均値。

率が高くなると考え、スクイズ作戦を取らせないようにするかもしれない。つまり、打率の変数は処置にも関連している可能性がある。ここで、例に上げた打率と長打率のスクイズ作戦を取った群と取らなかった群のヒストグラムは図 1 である。

図 1 から、スクイズした場面と、スクイズを行っていない場面では、打率の分布が異なっていることが視覚的に確認でき、平均での差を計算すると 0.063 であった。よって、単純な 2 群の比較によるスクイズ作戦の影響の推定では、スクイズ作戦の選択と、得点する確率の両方に影響を与えていると考えられる変数が存在しているにもかかわらず、その影響を無視して推定が行われている可能性が高く、スクイズ作戦のみの影響を推定できていないと考えられる(ここでは、スクイズ作戦の選択に関係する代表的な変数の 1 つである打率の分布について紹介したが、これら以外の変数のヒストグラムについては、補足資料として本論文の最後に記載した URL に掲載している)。

これらの問題に対処するためには、スクイズ作戦をとるかとらないかと、得点する確率の両方に影響を与える変数の影響を調整する(共変量調整を行う)必要がある。次節以降の解析では、傾向スコアを用いて共変量の影響を取り除くことで、スクイズ作戦の影響を推定した結果を述べる。

3. ロジスティック回帰モデルを用いた傾向スコアの推定と、その解析結果

本節では、まず傾向スコアで用いた共変量の選択について述べる。そのあと選択した共変量を用いて傾向スコアをロジスティック回帰モデルで推定した結果について述べ、最後に推定した傾向スコアを用いてスクイズ作戦の因果効果を推定した結果と、ロジスティック回帰モデルを用いた傾向スコアの推定の問題点を指摘する。

3.1 傾向スコアの推定に用いた共変量の選択

ここでは、傾向スコアを推定する際にモデルに含めた共変量について述べる。第 1 節で述べたようにモデルに含める共変量は、強く無視可能な割り付けの仮定 (1.5) が成立するように選択する必要がある。解析では、観測されている変数のうち、以下の変数を調整する変数として選択した。

共変量の選択においては、まず得点する確率を上昇させる(もしくはは下降させる)可能性のあ

表 3. 傾向スコアの推定に用いた共変量.

基本情報	打者に関する共変量	投手に関する共変量
イニング	代打かどうか	被安打率
イニングの表・裏	打順	セーブ率
アウトカウント	打率	失点率
走者の位置	長打率	奪三振数
得点差	次の打者の打率	
打者と投手の利き腕	次の次の打者の打率	
	ホームラン率	
	打点率	
	四死球率	
	敬遠率	
	三振率	
	犠打率	
	犠飛率	
	盗塁率	
	死球率	

る変数を選択した。例えば打者の情報に関しては、打点や長打率といった変数は、高いほど打席でヒットを打つ確率が上昇するため、結果変数に関係する変数であると考えられる。また、打率や長打率は結果変数に対して影響を与えることはよく知られた事実であり、スクイズ作戦を指示する監督もよく知っているはずであるから、この変数はスクイズ作戦を取るかどうかを決定する際に考慮されていると考えるのが自然である。よって、打率と長打率はスクイズ作戦を取るかどうかの処置へも影響を与える変数である。

また対戦投手に関する情報も重要である。被安打率が高い打者であれば、打率が低い打者でも打てる可能性が高まり、一方で奪三振率が高い投手では打率が高い打者でもヒットを打てる可能性が小さくなると思えることができる。よって、結果変数に影響を与える変数である。一方で、このような情報も監督は考慮してスクイズ作戦を取るかどうかを選択するため、これはスクイズ作戦をとるかどうかの処置変数へも影響を与えると思えることができる。

上記で上げていない表 3 に掲載した共変量も、結果変数へと影響を与えられ得る変数であり、その情報をもとにスクイズ作戦をとるかどうかを選択していると思えると、これらの変数についても調整するのが妥当であると判断し、解析においては調整することとした。また、これらの変数は昨年度のプレイ成績に基づいているため、処置後変数ではない点をここで指摘しておく。

最後に、因果推論における共変量の選択は、強く無視可能な割り付けを成立させるように選択することが必要であるが、この仮定が潜在結果変数を含むため検証することはできない。よって、ここで述べた共変量を調整することによって交絡を完全に除去することができると思えるのは考えにくい。しかしながら、上記で述べた変数は結果変数にも、処置変数へも影響を与えていると思えられる変数であり、これらが比較する 2 群で異なることが原因で平均処置効果がバイアスを持って推定されてしまう (Greenland et al., 1999) ため、これらについて調整することは重要であると思え、解析では表 3 にあげる共変量について調整を行うこととした。

3.2 ロジスティック回帰モデルを用いた傾向スコアの推定

ここでは、選択した共変量を用いたロジスティック回帰モデルをデータに当てはめて推定された傾向スコアの推定結果について述べる。一般的に傾向スコアを推定するにはロジスティック回帰モデルがよく用いられる。ここでも、同様にロジスティック回帰モデルを用いて

表 4. ロジスティック回帰モデルを用いて推定した、各群の傾向スコアの逆数(重み)のパーセンタイル点. 基準化した重みは、重みの和が各群サンプル数になるように調整をかけている.

パーセンタイル点	5%	25%	50%	75%	95%	99%	99.9%
スクイズした群	1.78	3.15	12.29	36.66	216.85	637.28	9277.42
スクイズしなかった群	1.00	1.00	1.00	1.01	1.06	1.19	6.24
スクイズした群 (基準化)	0.03	0.05	0.18	0.53	3.12	9.17	133.54
スクイズしなかった群 (基準化)	0.99	0.99	0.99	0.99	1.04	1.18	6.15

表 5. ロジスティック回帰モデルを用いて傾向スコアを推定した場合の因果効果の推定量のブートストラップ法によるパーセンタイル点の計算結果.

	2.5%	25%	50%	75%	97.5%
平均処置効果 (ATE)	-0.293	-0.092	0.043	0.261	0.330

傾向スコアの推定を行った. すなわち, 傾向スコア $\pi(X_i)$ に対して, 以下のモデルを仮定した.

$$(3.1) \quad \pi(X_i) = \pi_{\beta}(X_i) = \frac{\exp(X_i\beta)}{1 + \exp(X_i\beta)}$$

ここで, X_i は選択した共変量のベクトルである.

ここで, データにモデル(3.1)を当てはめて得られた β の最尤推定値を $\hat{\beta}$ とする. このとき, $\hat{\beta}$ を代入して得られる傾向スコアの推定値 $\pi_{\hat{\beta}}(X_i)$ の逆数の, パーセンタイル点をまとめたものが表 4 である.

ここで, スクイズを行わなかった群の傾向スコアの推定値の逆数が, スクイズしなかった群の傾向スコアの推定値の逆数より全体として大きくなっているが, これは 2 つの群のサンプルサイズの差によるものである. また, 傾向スコアとは与えられた共変量 X_i のもとでスクイズ作戦をとる確率であるから, その推定値の逆数が大きな値をとるとするのは「ほとんどスクイズ作戦を取らないところでスクイズ作戦をとった」と解釈することができる. 例えば, スクイズ作戦をとった群で, 99.9 パーセンタイル点の標本の傾向スコアの推定値の逆数が 9000 を超えているが, ほとんどスクイズ作戦をとらない場面で, スクイズ作戦をとった標本である. 本節の最後に, 少数の標本の傾向スコアの逆数が, 大きな値をとる場合の問題点について述べる.

3.3 推定した傾向スコアを用いたスクイズ作戦の影響の推定

ここでは, 推定した傾向スコアを用いてスクイズ作戦が得点する確率に与える因果効果を推定した結果について述べる. まず, 第 1 節で述べたように, IPW 推定量は強く無視可能な割り付けが成立するもとの, バントの有効性, 式(1.2), に対する一致推定量となる. 推定した傾向スコアに基づいて IPW 推定量を計算した結果, スクイズ作戦の平均処置効果は -0.007 となった. この推定量はすべての場面でスクイズ作戦をとった場合と, とらなかった場合に得点する確率の差の推定値であるから, 全場面でスクイズ作戦をとったとすると平均して約 0.7% 得点する確率が低下すると解釈することができる. また, この推定値に対する信頼区間を求めるためブートストラップ法を用いたところ, 結果は表 5 のようになった.

この結果から, 95% 信頼区間が $[-0.293, 0.330]$ となり信頼区間が 0 を含むため, 5% の有意水準では影響があるとはいえないことがわかった.

さて統計的因果推論を行う際には, 平均処置効果の他にも, 処置群における平均処置効

表 6. ロジスティック回帰モデルを用いて傾向スコアを推定した場合の、スクイズ作戦の ATT 及び、ATU のブートストラップによるパーセンタイル点の計算結果.

	2.5%	25%	50%	75%	97.5%
スクイズした群における平均処置効果	0.065	0.102	0.120	0.140	0.174
スクイズしなかった群における平均処置効果	-0.297	-0.094	0.042	0.264	0.334

果 (Average Treatment Effect on the Treated; ATT) と、対照群における平均処置効果 (Average Treatment Effect on the Untreated; ATU) にも興味があることが多い. ATT と ATU はそれぞれ式 (3.2), (3.3) で定義される.

$$(3.2) \quad \mathbb{E}[Y_1 - Y_0 | T = 1]$$

$$(3.3) \quad \mathbb{E}[Y_1 - Y_0 | T = 0]$$

ATT は、処置が行われたケースで、もし処置を行っていなかったら、どの程度結果に変化があったかに対応する推定量であり、ATU は処置が行われなかったケースで、もし処置を行っていなかったらどの程度結果に変化があったかに対応する推定量である. つまり、ATT はスクイズを行った場面で、もし行っていなかったら、得点する確率がどの程度変化するかに対応する. ここで、傾向スコアの真値が既知で、強く無視可能な割り付けの仮定が成立するとき式 (3.4) と式 (3.5) はそれぞれ、ATT 及び ATU に対する一致推定量となる.

$$(3.4) \quad \mathbb{E}[Y_1 - Y_0 | T = 1] = \bar{y}_1 - \frac{\sum_{i=1}^N \frac{(1 - T_i)\pi_i Y_i}{1 - \pi_i}}{\sum_{j=1}^N \frac{(1 - T_j)\pi_j}{1 - \pi_j}}$$

$$(3.5) \quad \mathbb{E}[Y_1 - Y_0 | T = 0] = \frac{\sum_{i=1}^N \frac{T_i(1 - \pi_i)Y_i}{\pi_i}}{\sum_{j=1}^N \frac{T_j(1 - \pi_j)}{\pi_j}} - \bar{y}_0$$

ここで、 \bar{y}_1, \bar{y}_0 は、それぞれ処置群 $T = 1$ と対照群 $T = 0$ における結果変数の標本平均である. これらの推定量を傾向スコアの推定値を用いて計算した結果、スクイズ作戦をとった場面で、スクイズをとらなかった場合の平均処置効果は 0.119 であり、逆にスクイズ作戦をとらなかった場面で、スクイズ作戦をとった場合の平均処置効果は 0.005 であった.

この結果から、もしスクイズ作戦をとった場面でとらなかったとすると、平均の意味で 11.9% の得点する確率の低下となることがわかる. 一方で、もしスクイズ作戦をとらなかった場面でとったとすると得点する確率は平均の意味で 0.5% 上昇することがわかる. さらに、先ほどと同様に、これらの推定量の信頼区間を 1500 回のブートストラップ法によって求めたものをまとめたのが表 6 である.

この結果から、スクイズ作戦をとった場面で、もしスクイズ作戦をとらなかった場合には、スクイズ作戦が得点する確率に与える影響の 95% 信頼区間は [0.065, 0.174] となり、5% の有意水準で影響があるといえる. 一方で、スクイズ作戦をとらなかった場面で、もしスクイズ作戦をとった場合には、スクイズ作戦が得点する確率に与える影響の 95% 信頼区間は [-0.297, 0.334] となり、5% の有意水準では影響があると結論付けることはできないという結果となる.

ここで、傾向スコアを用いた解析を行ったモチベーションは、比較する 2 群の共変量が得点する確率に与える影響を、傾向スコアの逆数で各標本に重み付けすることで取り除くことであった. 図 1 で確認した通り、スクイズ作戦をとった群と、スクイズ作戦を取らなかった群の打率の分布は視覚的に確認可能な程度には異なっている. 傾向スコアによって、この分布の差が調整されたと考えると、傾向スコアの逆数で重み付けした打率の分布は比較する 2 群で同じ

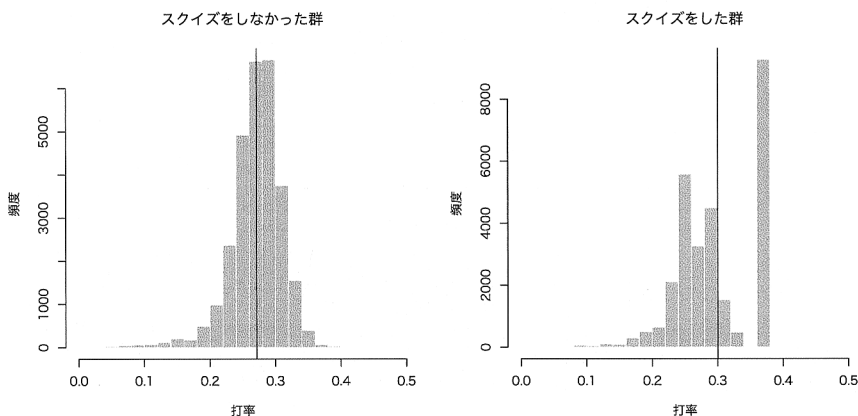


図 2. スクイズ作戦をとった群と、とらなかった群の打率の分布を傾向スコアの推定値で重み付けしたもの. 右: スクイズ作戦をとった群, 左: スクイズ作戦をとらなかった群. 黒の実線は重み付けられた打率の平均値である.

ような形状となるのが望ましい. 実際, 重みをつけたヒストグラムは図 2 となる(ここでは, スクイズ作戦の選択に関する代表的な変数の 1 つである打率の分布について紹介したが, これら以外の変数の重み付けヒストグラムについては, 補足資料として本論文の最後に記載した URL に掲載している).

図 2 より, ロジスティック回帰モデルによって推定された傾向スコアを用いて重み付けされた打率の分布は平均の意味で釣り合っているとは言い難い. また, スクイズ作戦をとった群のヒストグラムでは右端が大きく跳ねており, この周辺の標本に非常に大きな重みがかかっているが, スクイズ作戦をとらなかった群のヒストグラムの右端はなだらかに減少していることが確認できる. 実際, スクイズ作戦をとった群の右端で大きく跳ねているのは 1 つの標本に大きな重みがかかっているためである. この 1 つの標本にかかる重みは, スクイズを行った群の傾向スコアの逆数の和のうち 32.5% に対応する重みとなっている. このように 1 つの標本に大きな重みがかかるとき, 式 (1.6) より IPW 推定量の推定値は 1 つの標本の影響を強く受けることになる. ここでのスクイズの解析では, スクイズ作戦をとった場合の結果の 32.5% を 1 つの標本の結果によって説明されることになり, 小さな重みのかかった場面の結果が過少に評価されて結果が推定されることになる. 以上の議論から, ロジスティック回帰モデルを用いて傾向スコアを推定した上で, 推定した傾向スコアを用いてスクイズ作戦の因果効果を推定した場合には, 少なくとも打率の影響が平均の意味で取り除かれていない可能性が示唆された. また, ロジスティック回帰モデルを用いた場合には, 稀なケースに大きな重みがかかりすぎるという問題があることがわかった.

ロジスティック回帰モデルを用いた場合に, 稀なケースに大きな重みをかけすぎるという問題はブートストラップ法を用いて推定量の信頼区間を計算した際にもわかる.

図 3 は, ブートストラップ標本から計算された ATE の推定値のヒストグラムである. このヒストグラムからは 2 つの峰があることが確認できる. これはブートストラップ標本に, 稀にしかスクイズしない場面でスクイズしたかしなかったかで, 傾向スコアの推定値が変化し, 結果として得られるスクイズ作戦の因果効果が大きく変化することを示している. すなわち, ロジスティック回帰モデルで傾向スコアを推定するというのは, スクイズ作戦の因果効果を推定する場合には結果が安定的に得られないということを示唆している.

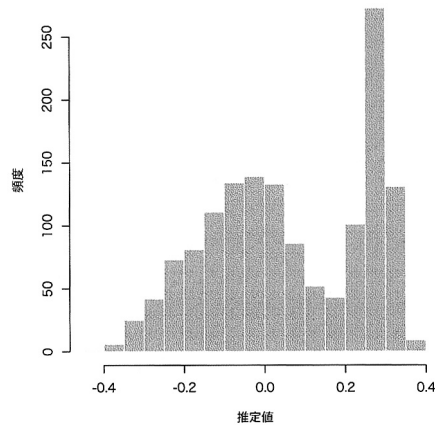


図 3. 傾向スコアの推定にロジスティック回帰モデルを用いた場合の平均処置効果の推定値のブートストラップ標本のヒストグラム。

ここまでの議論から、因果効果を推定する際には、スクイズ作戦及び得点する確率に影響する共変量の分布の平均は揃っている方が望ましく、結果はリサンプリングに対して安定的である方が望ましい。そこで、これらのロジスティック回帰モデルを用いた傾向スコアの推定から生じる問題点を解決するために、次節では Covariate Balancing Propensity Score を用いることで、これらの問題を解消・緩和して因果効果を推定した結果について述べる。

4. CBPS を用いた解析

本節では、Covariate Balancing Propensity Score (以下、CBPS) を用いてスクイズ作戦の因果効果を推定した結果について述べる。傾向スコアをロジスティック回帰モデルで推定した場合には、少数の標本に大きな重みが与えられ過ぎるという問題と、平均の意味での両群の共変量の釣り合いが取れない問題などがわかった。本節では CBPS について述べた後で、CBPS を用いることでロジスティック回帰モデルを用いて傾向スコアを推定した場合に指摘した問題点を解消しつつ、スクイズ作戦が得点する確率へ与える影響を推定することができることを指摘する。

4.1 Covariate Balancing Propensity Score とは

CBPS (Imai and Ratkovic, 2014) とは、式 (4.1) を満たす $\pi_\beta(X_i)$ として定義される。

$$(4.1) \quad \mathbb{E} \left[\left(\frac{T_i}{\pi_\beta(X_i)} - \frac{(1-T_i)}{1-\pi_\beta(X_i)} \right) f(X_i) \right] = 0$$

ここで関数 f は解析者が選択する共変量の関数である。この $\pi_\beta(X_i)$ が Covariate Balancing (共変量釣り合い) と呼ばれる理由は、式 (4.1) を式 (4.2) のように変形することにより $\pi_\beta(X_i)$ が、比較する 2 つの群の共変量の分布のモーメントを釣り合わせていると解釈できるためである。

$$(4.2) \quad \mathbb{E} \left[\frac{T_i}{\pi_\beta(X_i)} f(X_i) \right] = \mathbb{E} \left[\frac{1-T_i}{1-\pi_\beta(X_i)} f(X_i) \right]$$

実際、ロジスティック回帰モデルによって傾向スコアを推定する際には、 $\pi_\beta(X_i)$ に対して式 (1.7) を仮定し、 f に対して $\partial \pi_\beta(X_i) / \partial \beta$ をおくことに対応する。すなわち、ロジスティック回

帰モデルを用いて傾向スコアを推定した場合には、推定された傾向スコアは、 $\partial\pi_\beta(X_i)/\partial\beta$ を2つの群で平均の意味で釣り合わせる重みであると解釈できる。

次に、式(4.1)を満たす β の推定について述べる。パラメータ β をデータから推定する際には、式(4.3)を β について解く必要がある。

$$(4.3) \quad g(\beta) = \sum_{i=1}^N \left[\left(\frac{T_i}{\pi_\beta(X_i)} - \frac{(1-T_i)}{1-\pi_\beta(X_i)} \right) f(X_i) \right] = 0$$

しかしながら、この方程式(4.3)は β の次元と方程式の数が一致するかどうかは f の選択に依存し、一般的には解が存在しない。また、CBPSによる傾向スコアの推定では、関数 f の選択によって、パラメータの次元よりも、方程式の数が多くなることが多い。そこで、この方程式を直接的に解くのではなく、式(4.4)によって β の推定値を計算する。

$$(4.4) \quad \hat{\beta} = \underset{\beta}{\operatorname{argmin}} g(\beta)^T \Sigma(\beta) g(\beta)$$

ここで、 $\Sigma(\beta)$ は共変量 X に依存する正定値行列であり、解析ではImai and Ratkovic (2014)の式(14)を用いた。このようにすることで、 β をデータから推定することが可能となり、傾向スコアの推定値を計算することができる。このようにして得られる推定量の性質については、Hansen (1982)やHansen et al. (1996)が詳しい。

4.2 CBPSを用いたスクイズ作戦の解析

ここまで、CBPSが共変量の関数 $f(X_i)$ を比較する2つの群で釣り合わせるという性質を持つことを述べた。この性質を利用して、第3節で傾向スコアの逆数で重み付けされた打率の分布の平均が釣り合わなかった問題を解消する。すなわち、共変量 X_i が与えられた元でスクイズ作戦をとる確率 $\pi_\beta(X_i)$ は、共変量釣り合いの関数 f として式(4.5)を選んでモデル化されると仮定する。

$$(4.5) \quad f(X_i) = \begin{pmatrix} s_\beta(X_i) \\ X_i \end{pmatrix}$$

ここで、

$$(4.6) \quad s_\beta(X_i) = \frac{1}{1 + \exp(X_i^T \beta)} \left(1 - \frac{1}{1 + \exp(X_i^T \beta)} \right)$$

である。 $s_\beta(X_i)$ はロジスティック回帰モデルを用いて傾向スコアを推定する際に2つの群で釣り合わせる関数である。ここで、 f として X_i を加えることは、スクイズ作戦をとるかとならないかと、得点する確率の両方に影響をあたえる共変量の分布の平均を釣り合わせる重みとして、傾向スコアを推定していることに対応する。 f として、式(4.5)を選択したもとの、データから傾向スコアを計算した結果を表7にまとめた。

この結果(表7)と、ロジスティック回帰モデルを用いて推定された傾向スコアの表4を比較すると、スクイズしなかった群の傾向スコアの逆数にはそれほど大きな違いが見取れない。一方で、スクイズした群の傾向スコアの逆数については、CBPSの方が5%点は大きく、50%点以降ではロジスティック回帰モデルを用いた方が大きく推定されていることがわかる。すなわち、CBPSの方が傾向スコアのばらつきが小さいことがわかる。

推定したCBPSを用いた平均処置効果(ATE)の推定結果は0.182となった。また第3節と同様に、スクイズ作戦をとった場面でスクイズ作戦をとらなかった場合の因果効果(ATT)を推定した結果は0.106、逆にスクイズ作戦をとらなかった場面でスクイズ作戦をとった場合の因果

表 7. 推定された、各群の CBPS の逆数(重み)のパーセンタイル点. 基準化した重みは、重みの和が各群サンプル数になるように調整をかけている.

パーセンタイル	5%	25%	50%	75%	95%	99%	99.9%
スクイズした群	1.85	3.16	14.10	34.19	149.51	409.61	1133.13
スクイズしなかった群	1.00	1.00	1.00	1.01	1.05	1.16	6.14
スクイズした群 (基準化)	0.05	0.09	0.36	0.89	3.85	11.79	28.01
スクイズしなかった群 (基準化)	0.98	0.99	0.99	1.00	1.04	1.16	5.89

表 8. CBPS を用いた場合の、スクイズ作戦の因果効果(ATE, ATT, ATU)のブートストラップによるパーセンタイル点の計算結果.

	2.5%	25%	50%	75%	97.5%
平均処置効果 (ATE)	0.092	0.162	0.203	0.240	0.300
処置群における平均処置効果 (ATT)	0.006	0.077	0.101	0.123	0.172
対照群における平均処置効果 (ATU)	0.093	0.165	0.206	0.244	0.303

効果(ATU)を推定した結果は0.185であった. さらに、それぞれの推定量の信頼区間を計算するために、1500回のブートストラップを行った結果を表8にまとめた.

表8より、平均処置効果の推定値の95%信頼区間は[0.092, 0.300]であることから、有意水準5%で得点する確率に与える影響は有意であるとわかる. よって、すべての場面でスクイズ作戦を行った場合にも得点する確率を上昇させることがわかる. 同様に、スクイズ作戦をとった場面でとらなかった場合の平均処置効果の95%信頼区間は[0.006, 0.172]であり、スクイズ作戦をとらなかった場面でスクイズ作戦をとった場合の平均処置効果の95%信頼区間は[0.093, 0.303]で、ともに5%の有意水準で有意であり、得点する確率を上昇させることがわかる. この結果は、第3節でロジスティック回帰モデルを用いた場合とは異なっている.

次に共変量の分布を傾向スコアの逆数で重み付けした分布について考える. 第3節での傾向スコアで重み付けされた共変量(打率)の分布は平均の意味で釣り合っていないかった. 図4は推

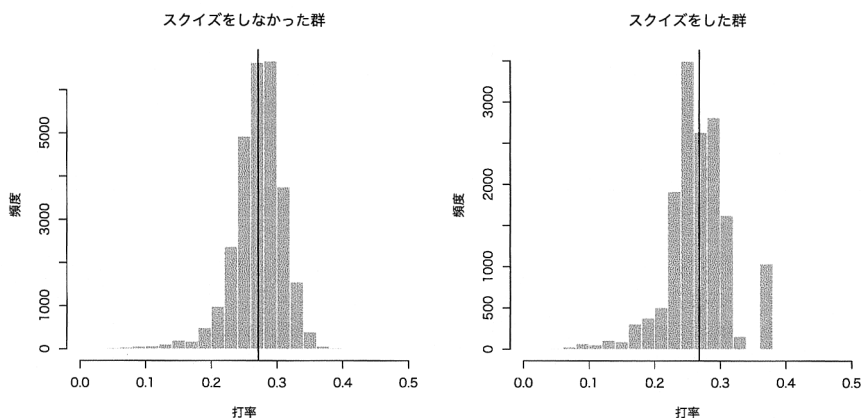


図 4. スクイズ作戦をとった群と、とらなかった群の打率の分布を CBPS の推定値で重み付けしたもの. 右:スクイズ作戦をとった群, 左:スクイズ作戦をとらなかった群. 黒の実線は重み付けられた打率の平均値である.

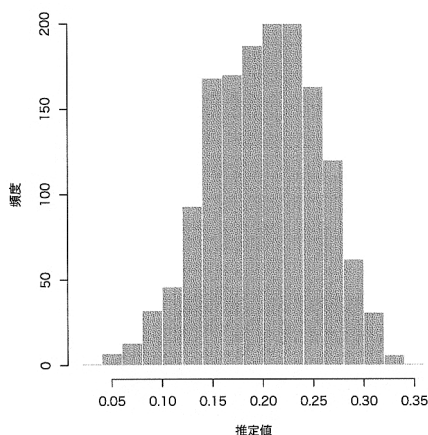


図 5. CBPS を用いた場合の平均処置効果の推定値のブートストラップ標本のヒストグラム。

定した CBPS で重み付けした共変量(打率)の分布である(ここでは、スクイズ作戦の選択に関係する代表的な変数の 1 つである打率の分布について紹介したが、これら以外の変数の CBPS による重み付けヒストグラムについては、補足資料として本論文の最後に記載した URL に掲載している)。

実際、スクイズ作戦をとった群と、とらなかった群の打率の重み付け分布の標本平均の差の絶対値は 0.003 であり、これは第 3 節で重み付けした分布の標本平均の差の絶対値 0.029 よりも小さいことがわかる。実際、視覚的にも図 2 よりも、図 4 の方が分布の形状が似ていることが確認できる。

ロジスティック回帰モデルを用いた場合には、ブートストラップ標本に基づいて計算されるスクイズの平均処置効果の推定値の分布が 2 つの峰を持つことが図 3 から確認できた。一方、CBPS を用いた場合のブートストラップ標本に基づく平均処置効果の推定量の分布は図 5 である。

2 つの図を比較すると、CBPS を用いた場合の分布の峰は 1 つになっているのがわかる。実際、ロジスティック回帰モデルを用いた場合、ブートストラップ標本に、スクイズ作戦をほとんどとらない場面をとったケースが含まれると 1 つの標本に大きな重みがかかり推定結果がブレやすくなる。しかし、CBPS では、スクイズ作戦をほとんどとらない場面をとったケースがブートストラップ標本に含まれたとしても、傾向スコアの値を大きくしすぎないように推定できていた。

以上の議論からスクイズ作戦の解析では、ロジスティック回帰モデルを用いると推定結果が不安定になるが、CBPS を用いることで影響を安定的に推定できることがわかった。さらに、ロジスティック回帰モデルを用いて傾向スコアを推定するとスクイズ作戦の有効性が得られなかったが、一方で CBPS を用いた場合には有効性が示された。また、スクイズ作戦をとらなかった場面で、もしスクイズ作戦をとった場合の有効性についてもロジスティック回帰モデルを用いて傾向スコアを推定した場合には示されなかったが、一方で CBPS を用いると有効性が示された。

5. まとめ

スクイズ作戦が得点する確率に与える影響は、得点差が 0 点または 1 点というどうしても 1

点が欲しい場面において、CBPSを用いた解析から得点する確率を上昇させる効果があることがわかった(ただし、満塁の場面を除く)。スクイズ作戦は、作戦の性質上、相手にスクイズ作戦をとることを察知されてしまうと、アウトカウントを増やしランナーを失うリスクのある作戦である。そのため、やらないであろうと想定される場面でやることで効果が生まれる。例えば、打率の高い打者が打席に立った際には「打つだろう」と想定するし、逆に打率が低い打者であれば1点を取るために「スクイズも警戒しよう」となると考えることができる。このような実際、スクイズを行った場面で傾向スコアが小さい、すなわちスクイズをほとんどしない場面での標本ではスクイズ作戦の成功率が高かった。このように考えれば、CBPSによる解析の結果は直感的にも誤った結果ではない。ただし、この結果からすべての場面でスクイズをすれば良いという結論は早計である。なぜなら、スクイズ作戦は相手の意表をつく作戦であるから、すべての場面でスクイズ作戦をとるのは、もはやスクイズ作戦にはならないため、あくまでこの結果は「現状で行われているスクイズ作戦の選択」に対する評価であると考えるのが適切である。

次に、スクイズ作戦の解析において、第3節の結果と第4節の結果を比較することでCBPSを用いた方がよいと考えられる理由についてまとめる。まず、CBPSを用いて得られた結果は、ロジスティック回帰モデルを用いて傾向スコアを推定した場合には得られなかった結果である。CBPSの特に重要な性質は、第4節で述べたように、比較する2群で結果変数と処置の両方に影響を与える共変量の関数の重み付き分布のモーメントを釣り合わせる性質があることである。今回の解析では、共変量分布の1次モーメントが釣り合うように制約をつけることで、比較する2つの群の共変量の1次モーメントの差が得点する確率に与える影響を取り除いた。さらに、1次モーメントの釣り合いをとることで、ほとんどスクイズ作戦をとらない場面でスクイズ作戦をとった標本に大きすぎる重みを与えないようにすることができた。

また、ロジスティック回帰モデルを用いた共変量分布の調整結果である図2と、CBPSを用いた共変量分布の調整結果である図4を確認すると、CBPSを用いた場合のほうが2群の共変量の分布の形状が似ていることが視覚的に確認できる。Greenland et al. (1999)でも指摘されたように交絡が生じる原因は、比較する2群の背後にある共変量の分布が異なっていることである。このような理由からも、2群の共変量の分布をよく釣り合わせているCBPSを用いた場合のほうが、交絡の影響を取り除いてスクイズ作戦の効果を推定できていると考えられることができる。

さらに、傾向スコアを用いてデータからスクイズ作戦の因果効果を推定する際の課題として、共変量が観測されたもとスクイズ作戦をとる確率に対するモデルは正確に特定できないという点である。実際、Kang and Schafer (2007)において、ロジスティック回帰モデルを用いて傾向スコアを推定した場合には、傾向スコアに対するモデルを誤特定すると、それが些細なモデルの誤特定であっても結果が安定しないことが指摘されている。一方で、CBPSはImai and Ratkovic (2014)において小さなモデルの誤特定に対して結果を頑健に推定できることがシミュレーションで示されており、このような理由からも、スクイズ作戦の因果効果の推定にはCBPSを用いた方がよいことがわかる。

最後に、CBPSを用いて共変量を調整した上で推定した、処置群での平均処置効果(ATT)と対照群における平均処置効果(ATU)が異なっていることから、スクイズ作戦の効果は、共変量に依存する不均一性がある(すなわちスクイズ作戦の効果は共変量に依存している)可能性が示唆される。本論文で扱いきれなかったが、実際の戦術に因果推論の結果を反映させるためには、どのような状況のもとでスクイズ作戦がより効果的なのかについて解析を行う必要があると考える。例えば、周辺パラメトリックモデル(Rotnitzky and Robins, 1995; 星野, 2009)などを用いて、潜在結果変数の周辺構造を指定した上でスクイズの効果は推定すれば、どのような

状況のもとでスクイズ作戦が有効なのかについて更に深く解析できると考える。

また、本論文で交絡の調整に用いた共変量である表3について、これらが調整すべきすべての交絡変数であるかの検討も行う必要がある。また、未観測の交絡が存在する場合の感度分析についても行う必要があるが、本論文では扱いきれていないので、これらの点については今後の課題としたい。

謝 辞

本稿を作成するに当たり、査読者の方から多くの詳細なコメントを頂いた。ここに記して感謝の意を表したい。なお本研究の一部は、2016年度KLL後期博士課程研究助成金(慶應義塾大学先端科学技術センター)及び、先端研究拠点事業(日本学術振興会: JSPS Core-to-Core Program)の助成を受けたものである。

付記

本論文におけるスクイズ作戦を行った場合と、行わなかった場合のそれぞれの共変量の分布、及び推定した傾向スコアの逆数によって重みづけし、調整したそれぞれの場合の共変量の分布についての詳細を<http://www.stat.math.keio.ac.jp/labs/mminami/research/>にこれを記載した。

参 考 文 献

- Albert, J. and Bennet, J. (2001). *Curve Ball: Baseball, Statistics, and the Role of Chance in the Game*, Springer, New York.
- Greenland, S., Pearl, J. and Robins, J. (1999). Confounding and collapsibility in causal inference, *Statistical Science*, **16**(1), 29–46.
- Hansen, P. (1982). Large sample properties of generalized method of moments estimators, *Econometrica*, **50**, 1029–1054.
- Hansen, P., Heaton, J. and Yaron, A. (1996). Finite-sample properties of some alternative GMM estimators, *Journal of Business & Economic Statistics*, **14**, 262–280.
- Hirano, K., Imbens, G. and Ridder, G. (2003). Efficient estimation of average treatment effects using the estimated propensity score, *Econometrica*, **71**, 1161–1189.
- 星野 崇宏 (2009). 『調査観察データの統計科学』, 岩波書店, 東京.
- Imai, K. and Ratkovic, M. (2014). Covariate balancing propensity score, *Journal of the Royal Statistical Society, Series B*, **76**, 243–263.
- Kang, Y. and Schafer, L. (2007). Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data, *Statistical Science*, **22**(4), 523–539.
- Lee, H. (2011). Is the small-ball strategy effective in winning games? A stochastic frontier production approach, *Journal of Productivity Analysis*, **35**, 51–59.
- 中村知繁, 南美穂子 (2016). 『傾向スコアを用いたバント効果の推定』, 岩波データサイエンス, Vol.3, 岩波書店, 東京.
- 仁志敏久, 鳥越規央 (2012). 『プロ野球のセオリー』, KKベストセラーズ, 東京.
- 丹羽政善 (2011). 日本でも「脱バント」が浸透するか, 日本経済新聞, 東京, <http://www.nikkei.com/article/DGXZZO23324240T10C11A2000000/>.
- Rosenbaum, P. and Rubin, D. (1983). The central role of propensity score in observational studies for causal effects, *Biometrika*, **70**, 41–55.

Rotnitzky, A. and Robins, M. (1995). Semiparametric regression estimation in the presence of dependent censoring, *Biometrika*, **82**, 805–820.

Zachary, R. (2013). Explaining Why the Bunt Is Foolish in Today's MLB, Bleacher Report, <http://bleacherreport.com/articles/1639658-explaining-why-the-bunt-is-foolish-in-todays-mlb>.

Effectiveness of the Squeeze Play Using Covariate Balancing Propensity Scores

Tomoshige Nakamura¹ and Mihoko Minami²

¹Graduate School of Science and Technology, Keio University

²Department of Mathematics, Keio University

Major League Baseball (MLB) has collected play-by-play data for the past 20 years. This data is available to the public. In this paper, we estimate the effect of a squeeze play on scoring using the covariate balancing propensity score (CBPS, Imai and Ratkovic, 2014) method. We focus on the case where the score difference is 0 or 1, except when the bases are loaded. A simple method is used to estimate the effect of a squeeze play on scoring. Specifically, sample averages are compared between two groups (attempting and not attempting a squeeze play). However, the decision to attempt a squeeze play is not random; it depends on the batter, pitcher, inning, etc. If these confounding variables are not considered, the estimated result will not represent the true effect of a squeeze play. In this paper, we estimate the effect of a squeeze play using a propensity score approach to adjust the effect of other variables. In the analysis, two types of estimation procedures for the propensity score are compared: the logistic regression model and the CBPS method. CBPS produces more balanced distributions of the covariates and the estimated effect of a squeeze play becomes more stable than using the logistic regression model to estimate the propensity score. CBPS indicates that a squeeze play has a positive effect on the scoring probability and increases the probability of scoring by 18.2%.