

# CNS アスリートとの雇用に伴うリスクの研究

---

CNS から NPB へのトランザクションと次年度パフォーマンス推定

宮崎浩一<sup>†</sup>・野村周平<sup>††</sup>

2014 年 11 月

## 要旨

2013 年 9 月から、キューバ政府は自国アスリートの国外でのプロ契約を条件付で承諾するプログラムの施行を開始した。このプログラムを受けて、2014 年はキューバにてプレーしていた野球選手が日本にてプロ契約の下プレーを行うケースが見受けられるようになり、このような日玖間を結ぶ新たな人材ルートは 2015 年以降も用いられる可能性が高い。しかし、キューバに限定せず、他国にてプレーを行っていた選手との雇用契約は、福山 (2014) が述べているように、大きなリスク要因である。同一な環境でプレーを続行する場合と比較して、別リーグからのトランザクションを経験した選手の将来予測は、データが示す価値の変動等を考慮する点で困難な問題である。現在日本にて、スポーツにおけるパフォーマンスの将来予測は積極的に取り上げられるテーマでは無く、上記の問題に対する有効な対処のヒントとなる研究成果は存在していない。以上の状況を踏まえ、本論ではキューバから日本へのトランザクションが行われた際のパフォーマンスの将来予測を実現し、雇用契約リスクに対する対処の 1 ケースと成り得る統計手法を利用したモデルを設定した。トランザクションを伴った将来予測を、パネルデータ分析、時系列分析の 2 点を組み合わせることで、アルゴリズム的に実証分析を進め、検証を行った。<sup>1</sup>

---

<sup>†</sup> 慶應義塾大学経済学部

<sup>††</sup> 慶應義塾大学経済学部

<sup>1</sup> 本論の執筆にあたって、慶應義塾大学の長倉大輔准教授から熱心かつ的確なコメントを頂戴した。ここに記して、感謝の意を示したい。

## 目次

- 第 1 章 はじめに
- 第 2 章 先行研究
  - 第 1 節 Equivalent Average
  - 第 2 節 Davenport Translation
  - 第 3 節 Davenport (1995) に見られる課題
- 第 3 章 実証分析モデル
  - 第 1 節 モデルのコンセプト
  - 第 2 節 実証分析モデルの詳細設定
  - 第 3 節 実証分析の基本統計量と手順
- 第 4 章 分析結果
  - 第 1 節 MLB
  - 第 2 節 NPB
  - 第 3 節 CNS
  - 第 4 節 検証
- 第 5 章 インプリケーション・課題

## 参考文献・データ出典

## 補論 キューバ野球の概況

- 補論 A キューバ革命に伴う対外交流の変化
- 補論 B キューバ野球のシステム

## 第1章 はじめに

2013年9月付の在日キューバ大使館のプレスリリースにて、キューバ政府が自国のトップアスリートに対し、国外チームとのプロ契約を条件付で許可するプログラムを施行することが明らかとなった。名目上は全てのスポーツに対し有効とされるプログラムだが、特に政府は国技である野球の輸出に奔走し、このプレスリリース以前からスポーツ体育リクリエーション庁長官のフリオ・クリスチャン・ヒメネスが日本の外務省を来訪、自国リーグ『セリエ・ナシオナル・デ・ベイスボル』（Serie Nacional de Beisbol 以下、CNS）でプレーする選手の日本プロ野球（以下、NPB）への派遣について会談を行っている。

キューバ革命成立から54年間、キューバは「ステート・アマ・システム」という、垂直統合的システム下で自国の競技レベルの向上を図った（詳細は付論を参照）。しかし、このシステム下では以下2点の問題点を露呈したことで、これらの解消に向けて2013年9月のプログラム施行に至っている。

まず1つ目の問題点として、「亡命人口の傾向的増加」があげられる。「ステート・アマ・システム」の下、キューバではトップアスリートの亡命が傾向的な増加を辿っている。図-1は、1991~2013年の23年間でのキューバの野球選手の亡命人口を示したグラフである。図-1によると、2007年を起点にアップダウンを経て、亡命人口は傾向的に増加を辿っていることを示している。特に2013年はこの期間内で最多の12人が亡命しており、この中でJose Abreu, Erisbel Arruebaruenaa, Odrisamer Despaigne, Rusney Castillo, Raisel Iglesiasの5人は2013年のWBCでナショナル・チーム入りを果たしており、「ステート・アマ・システム」下での低い報酬、そして乏しい対外交流に対し、トップアスリートが国外の、金銭、競技レベル両面で高いレベルを示すリーグでのプレーを望んでいることを明らかとしている。<sup>2</sup> 施行プログラムは、亡命を経て国内アスリートが流出する状況に対する予防策であり、政府管理下でアスリートが持つ国外へのプレー意欲を満足させる効果が期待されている。

第2の問題点として国内競技レベルの低下があげられる。亡命人口の増加を経て、キューバの野球は同時に競技レベルの低下を国際大会にて露呈をしている。1951年のIBAF<sup>3</sup>ワールドカップで3位に入ってから、キューバは2009年まで58年間、オリンピック大会、IBAFワールドカップ、インターコンチネンタルカップ、パンアメリカン大会、WBC全て

---

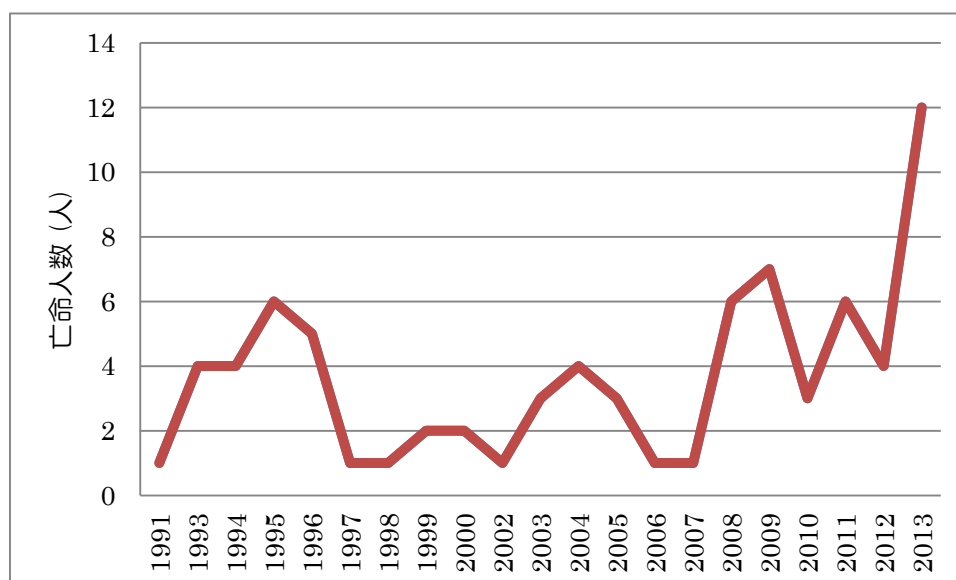
<sup>2</sup> Gourriel は2013年9月のプログラム施行について、2014年2月3日付の“MLB.com”によるコラム“Gourriel eyes baseball career outside of native Cuba”にて、国外でのプレーを熱望している旨を明らかにしている。

"I've played 12 years in Cuba and now I want to play outside. I want to see what it's like and improve," Gourriel said. "It's always been my dream to play the best baseball that I can, be that in the Major Leagues or Japan or wherever." (Sanchez, 2014)

<sup>3</sup> 国際野球連盟 (International Baseball Federation) の略称

の主要な国際大会全てで2位以上を逃すことは無かった。また、1997年のインターコンチネンタルカップで日本に11-2で敗れるまで、キューバは国際大会で151連勝をマークしていた。しかし、ここ5年のキューバは特にWBCで平凡なパフォーマンスに甘んじており、2009年の第2回、2013年の第3回のWBCではいずれもベスト8入りに止まり、オリンピック大会、IBAFワールドカップ、インターコンチネンタルカップが全て2011年時点では開催が行われなくなった現在、国際的なポジションを大きく低下させている要因となっている。<sup>4</sup> 施行プログラムは、亡命人口の減少を実現すると同時に、海外への選手派遣による技術の吸収も目的としている。<sup>5</sup>

図-1 キューバから亡命した野球選手の人口



1959年以降、実に54年にも及んだキューバのスポーツの閉鎖的な対外交流が上記の変換点を迎えたことで、NPBに限らず、メキシコ、韓国、ヨーロッパ諸国等、キューバとの政治的な対立が続くアメリカ合衆国以外の各国内リーグでは、CNSとの直接的な結び付きを持つ新しい人材ルートを得ることが出来る可能性が生まれた。<sup>6</sup> 2014年9月現在、既に

<sup>4</sup> IBAFが作成するナショナル・チームのランキングでも、2009年にランキングが制定されて以来、キューバは常に1位を維持していたが、2013年にその座をアメリカ合衆国に譲ると、2014年4月付のランキングでは3位へとランキングを下げている。

<sup>5</sup> 全日本野球協会（以下、BFJ）事務局長の柴田穰はこのプログラムに対し「キューバには、代表を強くするには海外に学ばないといけない、との危機感がある」と述べている。2014年4月にNPB読売ジャイアンツ入りしたCepedaも「日本野球を学んでキューバに伝えることも使命」と口にし（鷺津，前掲，2014）、プログラムを通じた競技レベルの向上を目指していることがうかがえる。

<sup>6</sup> フリーランス・スポーツライターの鷺田康はキューバ政府が施行するこのプログラムについて「両国の球界にとってメリットを生む契機となる」と論じている（鷺田康，「プロ野球亭日乗 日本球界に乗り込むキューバの英雄。門戸開放の陰に『亡命』と『裏開催』」（2014年5月24日），

CNS から NPB への直接のトランザクションとしては Frederick Cepeda (読売ジャイアンツ)、Yulieski Gourriel (横浜 DeNA ベイスターズ)、Alfredo Despaigne (千葉ロッテマリーンズ)、Hector Mendoza (読売ジャイアンツ) の 4 人が来日、プレーを行っている。この人材ルートの利用は、NPB 各チームの戦力補強にて新たに重要な意味を占めると考えることが出来る。

このような状況を受け、本論では、パネルデータ分析における固定効果モデルと、自己回帰平均モデル (以下、AR モデル) の手法を主に用いることで、CNS のパフォーマンスを NPB での内容に変換を行った上で、次年度のパフォーマンスを推定する。

CNS と NPB で行われている野球のルールは同一だが、プレー環境やリーグ内の技術レベルの分布等の要素の差異から、CNS で示されたパフォーマンスを工夫無く NPB での物差しに当てはめて評価を行うことは妥当では無い。CNS から人材を得る上で、異なるリーグでのパフォーマンスを評価する視点を用いることは、NPB 各チームが負わなければならない雇用リスクの最小化において重要な意味を持つであろう。実証分析の結果から、NPB 各チームが CNS とのパイプを利用する上での留意点について整理し、彼らが負う雇用リスクを可視化出来る形にすることが、本論の目的である。

本論の構成は以下の通りである。次章では、パフォーマンスの将来予測に関する Davenport (1995) の先行研究を参照した上で、本論で用いるモデル式のコンセプト、そして式詳細の設定を行う。第 3 章では、第 2 章で設定したモデルを用いた実証分析、検証の結果を整理する。最後に第 4 章は、第 2～3 章の内容を総括し、得られた結果に対する考察を行った後、本論全体の課題点を明らかにする。付録では、19 世紀後半から現在に至るまでのキューバにおける野球史を整理することで、本論内容が持つ意味の重大性の再認識を行った。

## **第 2 章 先行研究**

ここでは実証分析に先立ち、本論が取り上げる「別リーグ間の内容の変換」について扱った先行研究を参照する。本論では Davenport (1995) で提示された “Davenport Translation” (以下、DT) と、DT に付随する “Equivalent Average” (以下、EqA) の 2 手法について参照する。

### **第 1 節 Equivalent Average**

EqA は Davenport が提示したバッティング評価の手法であり、野球データ分析のシンク

---

<http://number.bunshun.jp/articles/-/820844>, (検索日: 2014 年 8 月 25 日)),

タンクである “Baseball Prospectus” には以下の説明がされている。

“ Equivalent Average. A measure of total offensive value per out, with corrections for league offensive level, home park, and team pitching. EqA considers batting as well as baserunning, but not the value of a position player's defense. The EqA adjusted for all-time also has a correction for league difficulty. The scale is deliberately set to approximate that of batting average. League average EqA is always equal to .260.<sup>7</sup> “

ここでは、EqA とは以下式により算出される RawEqA をリーグ、本拠とする球場等プレー環境の差異の補正を行ったものとする。

$$RawEqA = \frac{H + TB + 1.5 \cdot (BB + HBP + SB) + SH + SF - \frac{IBB}{2}}{AB + BB + HBP + SH + SF + CS + SB}$$

上式の説明変数については、以下表 1 を参照して頂きたい。

表 1 RawEqA の説明変数

打者データ名	意味	式
H	安打	—
1B	単打	—
2B	2 塁打	—
3B	3 塁打	—
HR	本塁打	—
TB	塁打	$1B + 2 \cdot 2B + 3 \cdot 3B + 4 \cdot HR$
BB	四球	—
HBP	死球	—
SB	盗塁	—
SH	犠打	—
SF	犠飛	—
CS	盗塁死	—
AB	打数	$PA - BB - HBP$

<sup>7</sup> Baseball Prospectus, “Glossary EqA”,  
<http://www.baseballprospectus.com/glossary/index.php?mode=viewstat&stat=61>

RawEqA を EqA に変換する具体的な過程については、“Baseball Prospectus” を参照頂きたい。この変換過程の特徴としては、計 10 式を用いてプレー環境、DH システムの有無、チーム間のパワーバランス等の部分における補正を試みている点にある。また上記サイトには以下の説明が掲載されている。

“ Convert RawEqA into EqR, taking into account the league EqA LgEqA, league runs per plate appearance, the park factor PF, an adjustment pitadj for not having to face your own team's pitchers, and the difficulty rating.<sup>8</sup> ”

RawEqA から EqA への変換過程は非常に煩雑であり、また一般的に入手が困難なデータの利用も見受けられるが、この点については本章 3 節にて詳細を述べる。

## 第 2 節 Davenport Translation

DT は EqA をベースに、別リーグへのトランザクションが行われた時のパフォーマンス推定を行う手法だ。リーグ間の差異を補正する場合に限らず、時代別で補正を行い、パフォーマンスを比較する上でも扱われる。Davenport は DT について、その目的意識を以下のように述べている。

“While these Translations look like player stats, they are NOT the players' actual statistics. The Translations are an attempt to show how well the player would have performed in a standard league (the American League of 1992), knowing how well he played in his actual league. We know that some leagues are tougher than others; that's why we have the majors, AAA, AA, and so on. We know that some leagues are easier to hit in; we know that some parks favor the pitchers; and we know that these effects are not constant from one year to the next. We can estimate how big a difference each of those makes and correct for them, and that is what the Translations try to do. How well they work I shall leave for you to judge.<sup>9</sup> “

引用内において、Davenport は異なるリーグ、および時代間の成績を加工無しに比較検討

---

<sup>8</sup> Baseball Prospectus, 前掲

<sup>9</sup> Davenport, “Davenport Translation”, [https://groups.google.com/forum/embed/#!topicsearchin/rec.sport.baseball.analysis/after\\$3A1995\\$2F01\\$2F01\\$20before\\$3A1995\\$2F01\\$2F31](https://groups.google.com/forum/embed/#!topicsearchin/rec.sport.baseball.analysis/after$3A1995$2F01$2F01$20before$3A1995$2F01$2F31), 1995

することの危うさ、そしてこれらの中に存在する差異を解消したパフォーマンス変換モデルの必要性について論じている。

DTの詳細なモデル式は Davenport (1995) 内では公表されておらず、打者評価では上述の EqA、投手では防御率を奪三振率 (以下、K/9)、与四球率 (以下、BB/9)、被打率、被本塁打率等で回帰した Normalized Earned Runs Average (以下、NERA) や、防御率を EqA と同じ過程で補正を行った Pitchers Earned Runs Average (以下、PERA) 等を用いることのみが明らかとされている。

DTに至るまでの Davenport の実証分析上の姿勢は、分析が持つ信頼性の根拠を詳細な補正に持たせている点に特徴を見出すことが出来る。特に EqA ではその傾向が強く出ており、RawEqA を EqA に変換する上で、Davenport は補正に計 10 式を用いている<sup>10</sup>。リーグ間の差異を均し、同じ土俵でパフォーマンスを比較することが目的だ。

DTによるパフォーマンス推定を、Davenport は自身のサイト“claydavenport.com”で公表している。1例として、彼は MLB へ移籍した井口資仁 (シカゴ・ホワイトソックス) と城島健司 (シアトル・マリナーズ) の1年目のパフォーマンスを以下表-2のように推定している。

表-2 井口資仁と城島健司の前年、DT推定、MLBでの実測値

➤ 井口資仁

年度	打率 (AVG)	出塁率 (OBP)	長打率 (SLG)	OPS (OBP+SLG)
前年値 (NPB)	.333	.394	.549	.943
DT推定値	.269	.337	.446	.783
実測値 (MLB)	.278	.342	.438	.780

➤ 城島健司

年度	打率 (AVG)	出塁率 (OBP)	長打率 (SLG)	OPS (OBP+SLG)
前年値 (NPB)	.309	.381	.557	.938
DT推定値	.292	.345	.435	.789
実測値 (MLB)	.291	.332	.451	.783

DTは決して完成された評価では無く、例えば2013年シーズンよりオークランド・アスレチックスでプレーしている中島裕之のパフォーマンスについては、その予見に失敗している<sup>11</sup>。しかし、上の推定は、DTによる予測が一定の分散を伴いながらも、非常に高い精

<sup>10</sup> Baseball Prospectus, 前掲

<sup>11</sup> Davenport は2013年シーズンのパフォーマンスを.264/.332/.384を推定したが、実際はメジャーで1回もプレー無



度の下で行われていることを示している。

### 第3項 Davenport (1995) に見られる課題

ここでは、本論にて実証分析を行う上で、Davenport (1995) が抱えている2点の課題について整理する。1つ目の課題として「プレー環境補正に要するデータ入手が困難」が挙げられる。前項では、Davenport (1995) の研究は、その信憑性をプレー環境の補正によって担保している特徴を指摘した。しかし、上で示した図-2のように、EqAにて補正の完了には計10式を要することが明らかとされている。ここで課題として浮上する点は、算出過程の煩雑さでは無く、算出過程にて必要になるデータの入手面である。Davenport は補正を行う上で、いくつか算出、もしくは入手が困難であるデータを使用しているが、以下表3にそれらを整理する。

表3 図-2内式にて入手・算出が困難なデータ

データ名	概要
PF (Park Factor)	各球場が固有に有する投、もしくは打に対する偏り
Adjustment Pitadj	リーグ全体から自チーム投手の内容を差し引いた投手成績
Difficulty Rating	チーム別に設定されるプレー難易度
DH Adjustment	DHの有無に沿う補正

Davenportはこの課題に対し、DH Adjustment以外のデータについて、代替として定数を利用出来ることを示している。しかし、定数の利用によって、推定値の精度を少なからず犠牲となっていることは明らかであり、実証分析を行う上でこのような措置は避ける必要がある。また、入手・算出が出来ないデータの利用も、当然ながら高い精度の下での推定を妨げる要因となる。

2つめの課題は「Davenport Translationにおけるブラックボックス」が挙げられる。Davenport (1995) では、DTがEqA, NERA, PERAを利用したスタッツの変換であることが示されている。しかし、DTの具体的な推定過程は明らかでは無く、EqA以下の指標の利用、次年度パフォーマンス推定での時系列分析の手法についてはブラックボックスである。すなわち、EqA以下の指標がDTに至るまでの過程が曖昧であり、本論でDTを体系的に参照することが出来ない最大の要因となっている。

---

くシーズンを終えている

### 第3章 実証分析モデル

以上を踏まえ、本論では Davenport (1995) とは異なるモデルとして、パネルデータ分析、時系列分析を複合した実証分析モデルを組み立てる。

#### 第1節 モデルのコンセプト

本論は Davenport (1995) の手法を体系的に参照が出来ない点から、モデル立式では手法の意味合い、目的について整理をした上で、モデル立式に及ぶ必要がある。そこで、ここでは Lewis (2004) の以下の言及を参照した上で、全てのプレーが選手固有の「実力」（以下、Ability）と、それとは独立した「運」（以下、Luck）の2要素で構成されていることを仮定し、前提とする。

この前提を土台にし、パネルデータモデルを組むと、パフォーマンス評価の土台となる項目のデータ（以下、Overall）は Ability と Luck により、

$$Overall = Ability + Luck$$

と表現することが出来る。この時、具体的にはパネルデータモデルのうち、個別効果として選手毎にダミー変数を設定する「固定効果モデル」を利用することで、Ability はダミー変数として抽出が可能であると仮定する。

また、次年度パフォーマンス推定では、Ability はリーグ、プレー年度間での変動は 0、対して Luck は Ability とは独立しランダムな変動を行い、かつリーグ間で異なる水準を持つため、その修正を要すると仮定する。すなわち、Ability は固定効果モデルによる算出値を次年度にも引き継ぐが、Luck として設定した説明変数については、各データにて時系列分析を行い、次年度の値を詳細に推定する。ただし、上記はモデル設定時点でのコンセプトであるので、4章「分析結果」では、これら仮説についても検証を行う。

#### 第2項 実証分析モデルの詳細設定

前項を踏まえ、ここではコンセプトに沿い具体的な実証分析モデルを定式化する。特に前項では Overall と Ability, Luck と大別をするに止めた説明、被説明変数について、詳細な設定を行う。

##### 1. 投手について

分析に用いるモデル式は以下を設定する。

$$ERA_{it} = a_i + b_1 \cdot BABIP_{it} + b_2 \cdot LOB\%_{it} + b_3 \cdot HR/9_{it} + b_4 \cdot H/9_{it} + \varepsilon_{it}$$

説明・被説明変数、及びこれらの導出に必要なデータ項目は、以下表 4 に整理する。

表 4 投手モデルの説明・被説明変数、及びこれらの導出に必要なデータ項目

投手データ名	意味	式
IP	投球イニング	—
H	被安打	—
HR	被本塁打	—
K	奪三振	—
BB	与四球	—
HBP	与死球	—
R	失点	—
ER	自責点	—
ERA	防御率	$9 \cdot \frac{ER}{IP}$

式内 4 変数の導出式は以下に示す。

$$BABIP = \frac{(H - HR)}{(2.8 \cdot IP + H - HR - K)}$$

$$LOB\% = \frac{(H + BB + HBP - R)}{(H + BB + HBP - 1.4 \cdot HR)}$$

$$HR/9 = 9 \cdot \frac{HR}{IP}$$

$$H/9 = 9 \cdot \frac{H}{IP}$$

ここでは ERA を最終的なパフォーマンス評価に用いるとして被説明変数に設定した。a で Ability を示す個別ダミーとして固定効果モデルにより抽出を行い、BABIP 以下 4 つの説明変数が Luck を示す要素である。BABIP は HR を除く H で算出した打率、LOB% は投手が背負った走者のうち、生還を許さなかった割合を示す。これらが Luck を示す変数として使用出来る理由については、前項の引用部分を再度参照して頂きたい。HR/9 と H/9 は、BABIP と LOB% の値を決定する要因を誤差項から除外し、説明変数と誤差項の相関を解消する上で設定する必要がある。

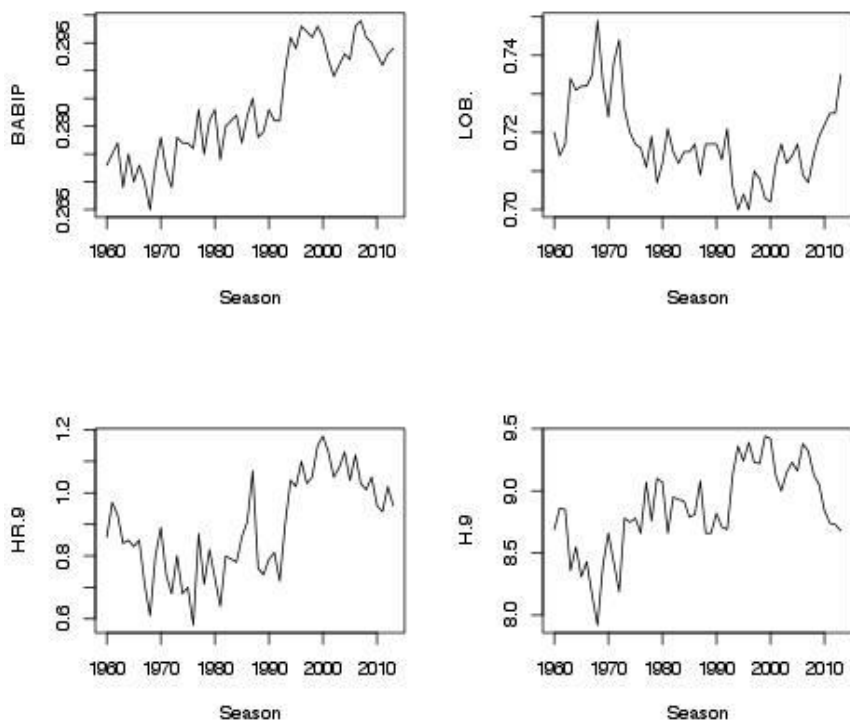
また、BABIP 以下 4 つの説明変数の次年度推定では、各変数の前年比変化%で自己回帰モデル (以下、AR モデル) を設定する。AR モデルは時系列モデルでも「過程が自己の過

去に回帰された形で表現される」という特徴を持ち、例えば 1 次の AR 過程は以下式で示される。

$$y_t = c + \phi_1 \cdot y_{t-1} + \varepsilon_t \quad \varepsilon_t \sim W.N.(\sigma^2)$$

y の「前年比変化%」への換算は、AR モデルによる分析を行う上で、自己共分散が時点 t に左右されず、時点差 k (AR モデルの次数) にのみ依存する「弱定常性」(以下、定常性) を満たす上で必要な設定である。下図-3 は、1960~2013 年の MLB にて観測された BABIP 以下 4 変数の原系列データであるが、本論ではこれらをより分析が容易な図 4 の形式に加工し、推定を行う。

図 3 4 変数の原系列時系列データ



## 2. 打者について

打者に用いるモデル式は以下を設定する。

$$\log(TA)_{it} = a_i + b_{11} \cdot \log(BABIP)_{it} + b_2 \cdot \log(ISO)_{it} + b_3 \cdot \log(K\%)_{it} + \varepsilon_{it}$$

$$TA = \frac{(TB + BB + HBP + SB)}{(AB - H + CS + GDP)}$$

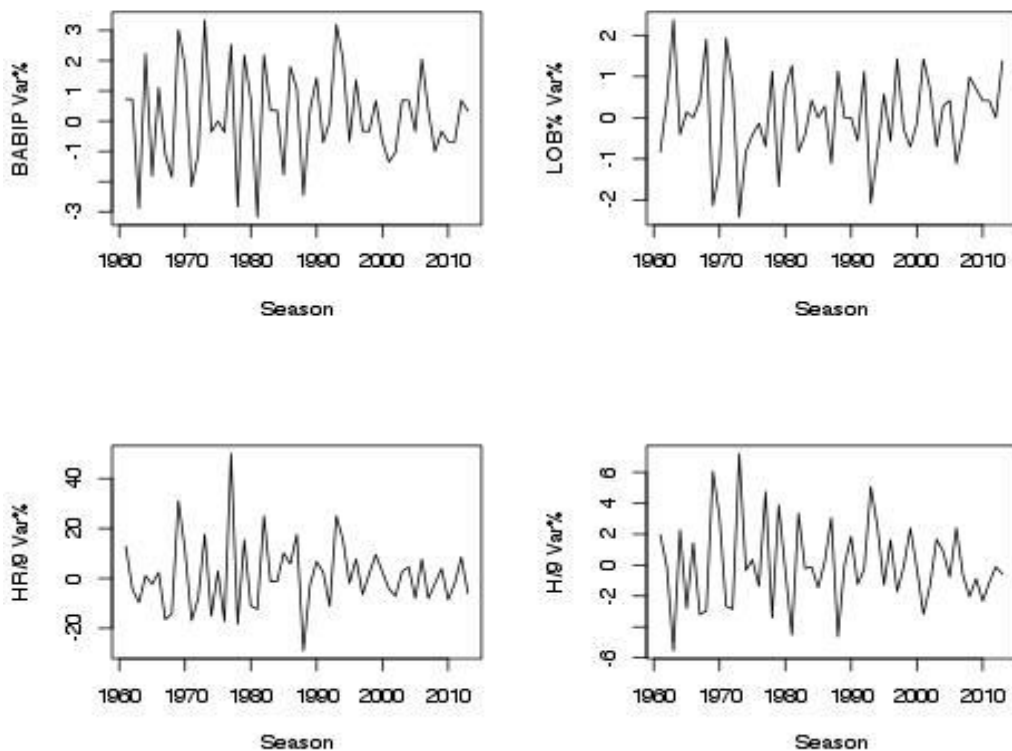
$$BABIP = \frac{(H - HR)}{(AB - HR - K)}$$

$$ISO = \frac{TB}{AB} - AVG$$

$$BB\% = \frac{BB}{PA}$$

ここでも投手同様に、説明・被説明変数、及びこれら上記式中の変数の項目を表5に示す。ここではパフォーマンス評価の項目として、TA (Total Average) の対数を被説明変数に用いる。これは打者が1アウトを献上するまでに得た単位量あたりの塁打数を示す。また、

図-4 4変数の変化%時系列データ



表・5 打者モデルの説明・被説明変数、及びこれらの導出に必要なデータ項目

打者データ名	意味	式
1B	単打	—
2B	2 塁打	—
3B	3 塁打	—
HR	本塁打	—
TB	塁打	$1B + 2 \cdot 2B + 3 \cdot 3B + 4 \cdot HR$
BB	四球	—
HBP	死球	—
SB	盗塁	—
CS	盗塁死	—
PA	打席	—
AB	打数	$PA - BB - HBP$
DP	併殺打	—

説明変数としては BABIP, K%, ISO の対数を設定した。BABIP は投手モデルに使用したものと同一意味合いであり、K%は三振の打席に対する比率、ISO は長打率と打率の差より、パワーの有無を判別するデータであるが、これらの値は投手の説明変数と比べ個体差が大きい点から、Ability としての性質が強い特徴を持つ。そのため、このモデルにおける a は「式内変数の Ability では示されていない Ability 要因」を抽出していると仮定する。モデル設定の時点では、a には打席でのアプローチ、つまりは BB を多く選ぶことが出来るかの否かを示していると仮定するが、詳細な検証結果については次項目で述べる。

時系列推定においても、投手と同様に各変数の変化%を算出し、AR モデルに沿って分析を行う。

### 第 3 節 実証分析の基本統計量と手順

前項のモデルを扱い、本論では以下の手順に沿って実証分析を行った。

1. 実証分析を行う上で、ここでは NPB, CNS, MLB の公式データからパネルデータを作成した。パネルデータの作成にあたり、CNS は “Serie Nacional de Beisbol (<http://www.inder.cu/beta/competitions/baseball/Campeonato Nacional/Serie Nacional/>, 8/19)から 2006～2013 年、NPB は 「NPB 公式ホームページ」

(<http://www.npb.or.jp>, 8/20)より 2009～2013 年、MLB は ”FanGraphs.com” (<http://www.fangraphs.com>, 8/18) より 2009～2013 年のデータを、それぞれ入手した。パネルデータの作成においては、各年度単位で打者は 80 打席、投手は 20 イニング以上のプレーを行った年度のみを対象に、最も実働期間が多かった連続する 2～4 年のデータを抜き出した。その際にデータの擬似的なパネルデータ化を行うべく、年度は古い順から 1, 2, ... と設定し、欠損したシーズンのデータに対しても処理を行った。この手法を取り入れることで、各リーグのデータを収集する上で設定した上記期間のデータ以外にも過去に MLB 入りをしている選手が存在していた場合に限定し、その選手のデータも疑似パネル化して組み込むことが可能となる。

各リーグのデータのサンプル・主体数は表 6 に示す。上記の通り、ここでのサンプル・主体数は当該期間データに加え、その期間外にて MLB でプレーした選手のデータも含まれていることを留意頂きたい。

表 6 本論データの主体・サンプル数

リーグ	投手サンプル数	投手主体数	打者サンプル数	打者主体数
MLB	974	195	1505	302
CNS	1092	294	1096	290
NPB	2081	420	1627	336

- 上記で作成したパネルデータを、前項で設定したモデル式に組み込み、実証分析を行った。ここではまず、MLB のデータにてモデル式がコンセプト通りに機能するかの検証を行う。次いで、NPB, CNS のデータでも同様の操作を行った上で、Ability, Luck いずれの項目に対しても、時系列分析を用いながら MLB との相関関係を確認し、NPB, CNS のデータを MLB でのパフォーマンスに変換し、同環境の下で比較検討が出来るように加工を行った。時系列推定では、各リーグ全体のデータから、リーグ別で説明変数変動の AR (1) モデルを設定し、その式を推定を行いたい各選手に対し当てはめる手順を踏んでいる。
- NPB, CNS データの MLB への変換が完了した後、両者間の関係を確認し、CNS から NPB へのパフォーマンス変換の関係式を設定する。その際、Luck で行う時系列分析は既に CNS データで完了し、次年度予測に変換を加えることで、CNS から NPB へのトランザクションに伴う次年度パフォーマンス推定を行った。

## 第4章 分析結果

設定したモデル式に沿って実証分析を行った結果を、以下リーグ毎に整理する。

### 第1節 MLB

ここでは、モデル式が設定したコンセプト通りに機能するかを確認すべく、MLBのデータを用い、説明変数の有意度合、モデル式全体が有する説明力の把握を行う。

まず、各変数の係数を整理する。表7と8は、MLBパネルデータにおける投打モデルの係数と、その決定係数を示している。

表7 MLBデータにおける投手モデルの係数

説明変数	係数	p値
BABIP	1.45203	0.07165
LOB%	-8.33281	2e-16
HR/9	0.19887	2e-16
H/9	1.20530	2e-16
R <sup>2</sup>	0.90489	

表8 MLBデータにおける打者モデルの係数

説明変数	係数	p値
log(BABIP)	0.2892432	2.2e-16
log(ISO)	0.3455545	2.2e-16
log(K%)	-0.3176203	2.2e-16
R <sup>2</sup>	0.61844	

投手BABIP項、打者モデルの決定係数が脆弱だ。しかし、モデル式内の符号条件は設定した理論と合致しており、いずれの式もMLB内で運用が可能なモデル式であることが明らかとなった。この結果を受けて、我々は本論の目的である「時系列分析を踏まえたパフォーマンス推定」においては、このモデル式を用いることが出来る。

また、図5と6には、設定モデルにより導出されたAbility項のヒストグラム、表9、10では各ポジションでAbility項のトップ/ワースト5人を示す。主体別ダミー(個別効果)はモデル式内の被説明変数に対して影響を与える要因であるため、投手はダミーが低い値、打者は高い値を得ることが高い評価となる。

各ポジションのダミーが持つ意味については、まず投手ダミーだが、これはパネルデータ分析の性質上、ERAに影響を与え、かつ説明変数外であるK、BB、ゴロ%(以下、GB%



図5 投手 Ability のヒストグラム

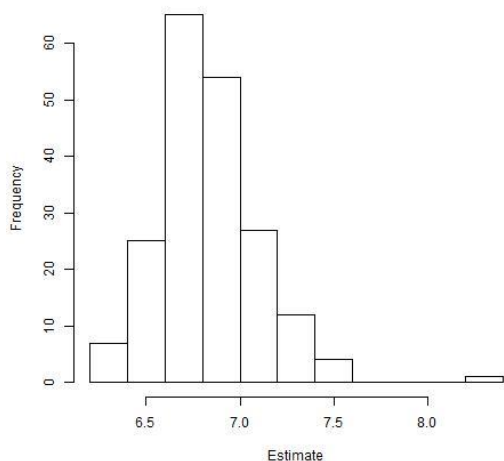


図6 打者 Ability のヒストグラム

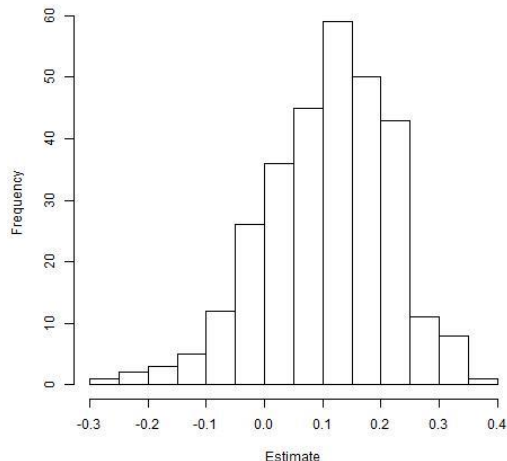


表9 MLB 投手 Ability

順位	選手名	主体別ダミー	p 値
1	Jaime Garcia	5.99746	2.2e-16
2	Brandon McCarthy	6.06212	2.2e-16
3	Francisley Bueno	6.09681	2.2e-16
4	Tommy Milone	6.09787	2.2e-16
5	Brett Anderson	6.11529	2.2e-16
...			
190	Kyle Davies	7.26802	2.2e-16
191	Yunesky Maya	7.34541	2.2e-16
192	Edinson Volquez	7.3849	2.2e-16
193	Kei Igawa	7.38597	2.2e-16
194	Ryota Igarashi	7.98297	2.2e-16

インプレー打球のうちゴロが占める割合) の3要因を複合した総合評価が ERA ベースで算出されたものと考えられる。上位5人はいずれも3種要因いずれかのリーダーでは無く、これはバランス良く高いレベルのパフォーマンスを示すことでダミー値が向上することを示している。一方で下位5人を見ると Volquez は BB、他4人は GB% で低い値に甘んじており、3種要因を高いレベルで両立することがここでは求められていると結論出来る。

表 10 MLB 打者 Ability

順位	選手名	主体別ダミー	p 値
1	Chipper Jones	0.35103	2.2e-16
2	Kurt Suzuki	0.336601	2.2e-16
3	Matt Holiday	0.3284917	2.2e-16
4	Chris Coghlan	0.3236007	2.2e-16
5	Randy Winn	0.322978	2.2e-16
...			
298	Daric Burton	-0.1729384	1.68e-6
299	Jed Lawrie	-0.1950424	4.32e-7
300	Yoenis Cespedes	-0.2402975	1.47e-10
301	Andre Ethier	-0.2461811	5.22e-11
302	Tsuyoshi Shinjo	-0.2589721	2.2e-16

打者ダミーも同じく、説明変数外で TA に影響を与える要因である BB、走塁の 2 要因が総合的に評価されていると考えられる。ここでの BB とはアプローチの良さを示す尺度であり、走塁は SB に限らず、ベースランニングのスキル等によって得られた進塁も加味された評価である。また図 7 は打者モデルの説明変数の変化%の時系列データである。投手モデルの変化%については図 4 を参照頂きたい。

以下の式は、各説明変数の変化%の AR (1) モデルの推定式である。これら式からは、K% の変化%以外では符号変化を伴うアップダウンを描いていることを示している。このモデルは、以降で CNS から NPB へのトランザクションに伴う将来推定を行う際に利用する形である。表 11 は推定結果である。

(投手)

$$BABIPRate_t = 0.1489 - 0.3148 \cdot BABIPRate_{t-1} + \varepsilon_t$$

$$LOB\%Rate_t = 0.0430 - 0.1605 \cdot LOB\%Rate_{t-1} + \varepsilon_t$$

$$H/9Rate_t = 0.0286 - 0.2529 \cdot H/9Rate_{t-1} + \varepsilon_t$$

$$HR/9Rate_t = 1.0734 - 0.3468 \cdot HR/9Rate_{t-1} + \varepsilon_t$$

(打者)

$$BABIPRate_t = 0.1408 - 0.2833 \cdot BABIPRate_{t-1} + \varepsilon_t$$

$$ISORate_t = 0.4763 - 0.3337 \cdot ISORate_{t-1} + \varepsilon_t$$

$$K\%Rate_t = 0.7962 + 0.0299 K\%Rate_{t-1} + \varepsilon_t$$

図 7 打者モデルの説明変数の変化%時系列データ

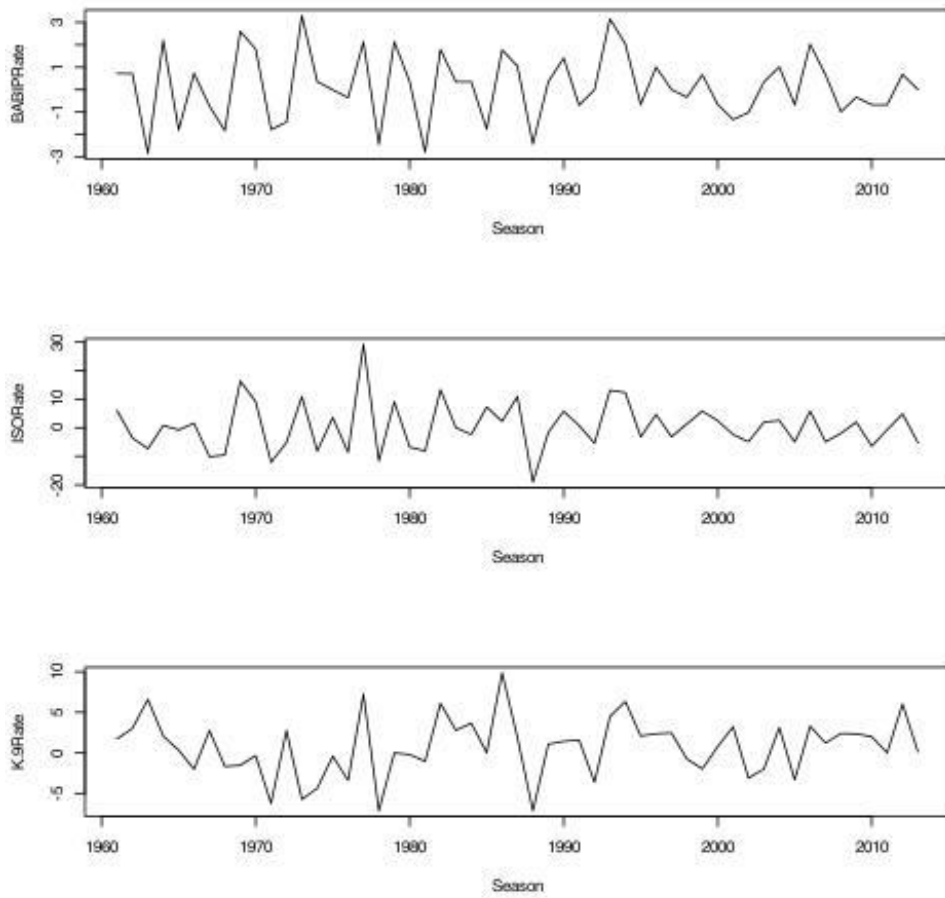


表 11 MLB 投手モデルにおける説明変数の変化%の AR(1)モデル推定結果

説明変数	係数	切片
BABIP	0.1489	-0.3148
LOB%	0.0430	-0.1605
H/9	0.0286	-0.2529
HR/9	0.10734	-0.3468

表 12 MLB 打者モデルにおける説明変数の変化%の AR(1)モデル推定結果

説明変数	係数	切片
log (BABIP)	0.1408	-0.2833
log (ISO)	0.4763	-0.3337
log (K%)	0.7962	0.0299

## 第 2 節 NPB

ここでは、MLB と同様に NPB のデータを用い、説明変数の有意度合、モデル式全体が有する説明力の把握を行う。

まず、各変数の係数を整理する。表 13 と 14 は、NPB パネルデータにおける投打モデルの係数と、その決定係数を示している。

表 13 NPB データにおける投手モデルの係数

説明変数	係数	p 値
BABIP	-2.154718	0.1620
LOB%	0.022315	0.9565
HR/9	-0.086427	0.2466
H/9	0.037523	0.3453
R-Squared	0.001845	

表 14 NPB データにおける打者モデルの係数

説明変数	係数	p 値
log (BABIP)	0.5950690	2.2e-16
log (ISO)	0.2581001	2.2e-16
log (K%)	-0.1913356	2.2e-16
R-Squared	0.803	

投手における全ての項、並びに決定係数が非常に脆弱だ。符号関係も H/9 項以外は設定したモデルと合致していない。一方で、打者モデルに関しては符号関係、決定係数共に問題ない値となった。

図 8 と 9 には、設定モデルにより導出された主体別 Ability 項のヒストグラム、表 15、16 では各ポジション別で Ability 項のトップ/ワースト 5 人を示す。

図8 投手 Ability のヒストグラム

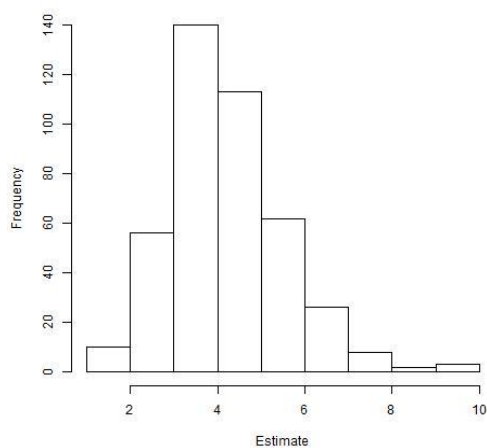


図9 打者 Ability のヒストグラム

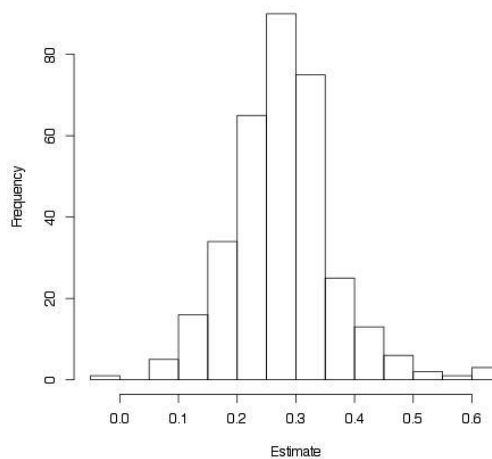


表 15 NPB 投手 Ability

順位	選手名	主体別ダミー	p 値
1	Masato Kobayashi	1.44824	0.0174896
2	Scott Mathieson	1.54543	0.0129636
3	Kyuji Fujikawa	1.66653	0.006946
4	Masa Kobayashi	1.70009	0.0072306
5	Yoshitaka Hashimoto	1.76271	0.0024638
...			
416	Yasuhiro Ichiba	8.24061	2.2e-16
417	Fumikazu Kimura	8.90338	2.2e-16
418	Yi-Cheng Wang	9.24522	2.2e-16
419	Chris Leroux	9.39006	2.2e-16
420	Ken Takahashi	9.63046	2.2e-16

投手ダミーにおけるトップと、打者におけるワーストの値があまり有意ではないという結果が出た。また、投手はメジャーで算出した主体ダミーに比べ個体差に幅が出た。打者は最下位の G.G Sato 以外全員が正の値をとった。

表 16 NPB 打者 Ability

順位	選手名	主体別ダミー	p 値
1	Yuhei Takai	0.6385328	2.2e-16
2	Yasushi Ihara	0.6124318	2.2e-16
3	Akihisa Makida	0.6069484	2.2e-16
4	Takahiro Suzuki	0.5986472	2.2e-16
5	Koichi Hori	0.5130294	2.2e-16
...			
332	Francisco Caraballo	0.0952106	0.028705
333	Rickey Guttormson	0.0844383	0.069637
334	Kazuo Matsui	0.0776887	0.1016389
335	Ryosuke Hirata	0.0549422	0.2261606
336	G.G Sato	-0.0016591	0.970486

次に、NPB のデータを MLB に変換する際に用いる回帰式について表 17 に整理する。図 10、11 はそれぞれ投手、打者の回帰直線を示している。尚、この回帰直線は原点を通過するように定めた。投手に関しては非常に有意な結果が得られたが、反対に打者の値は説明力の低いものとなった。次に、NPB のデータを MLB に変換する際の時系列分析で用いる、変換式の説明変数について表 18、19 に整理する。尚、図 12～18 は各説明変数の回帰直線を示している。投手において、設定した理論と符号条件が一致したのは LOB%のみであった。切片は良い値が取れているが、係数を含め全体的に脆弱な値となった。打者は K%のみ設定した理論と符号が合致しなかったが、決定係数が 3つの項目の中で一番高い値をとった。BABIP と ISO に関しては、決定係数は低いものの、まずまずの p 値が算出された。

表-17 NPB データにおける投手・打者の Ability 変換式の係数

	係数	p 値	R-Squared
投手	1.9932	1.95e-10	0.9595
打者	0.1799	0.333	0.1042

表-18 NPBデータにおける投手変換式の係数

説明変数	係数	p 値	切片	p 値	R-Squared
BABIP	-0.34953	0.136	37.577	3.3e-06	0.05603
LOB%	0.08621	0.78871	71.590	0.00225	0.001863
HR/9	0.2688	0.34246	74.39	0.00115	0.02313
H/9	-0.1171	0.607	851.75	3.74e-05	0.006829

表-19 NPBデータにおける打者変換式の係数

説明変数	係数	p 値	切片	p 値	R-Squared
log(BABIP)	0.62010	8.18e-05	10.574	0.0237	0.4687
log(ISO)	0.26329	0.00408	6.848	0.00213	0.2856
log(K%)	0.89109	3.83e-08	1.862	0.253	0.7081

図 10 投手 Ability の回帰直線

図 11 打者 Ability の回帰直線

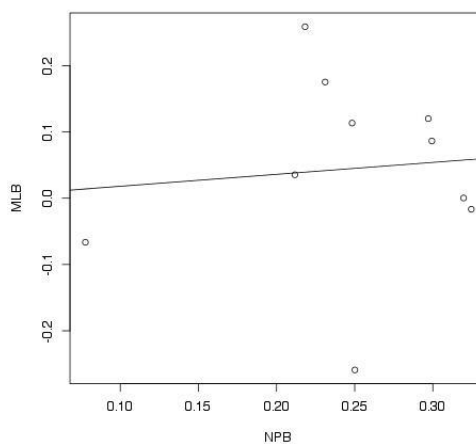
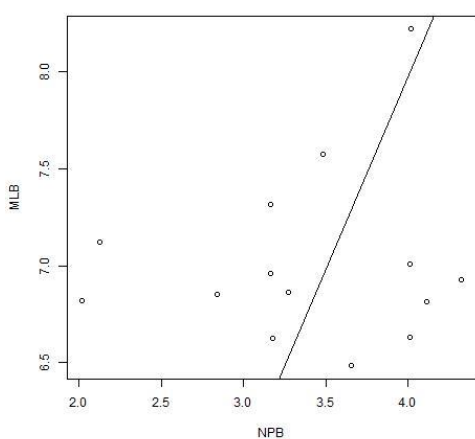


図 12 投手 BABIP の回帰直線

図 13 投手 LOB% の回帰直線

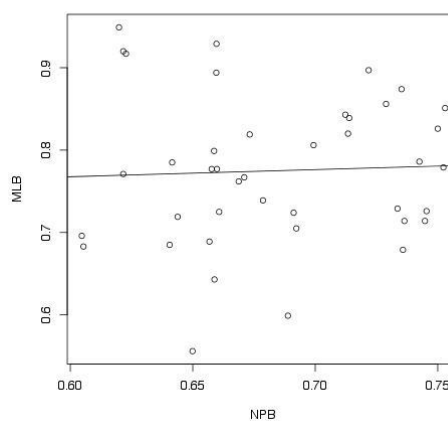
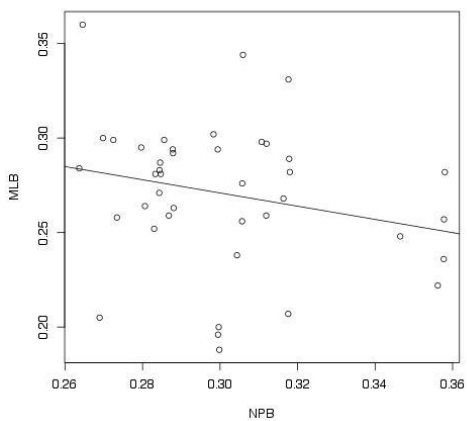


図-14 投手 HR/9 の回帰直線

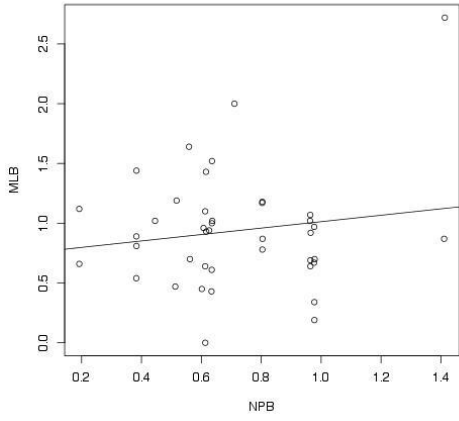


図-15 投手 H/9 の回帰直線

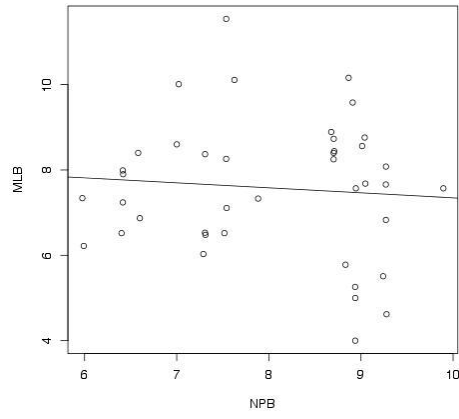


図-16 打者 BABIP の回帰直線

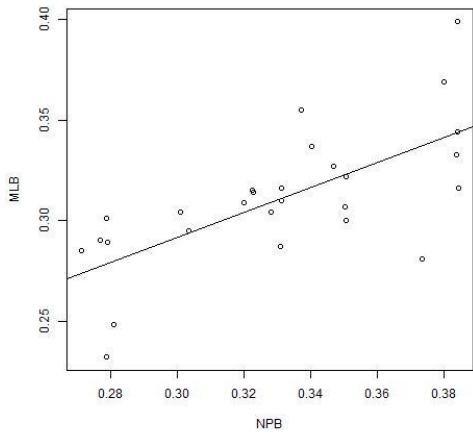


図-17 打者 ISO の回帰直線

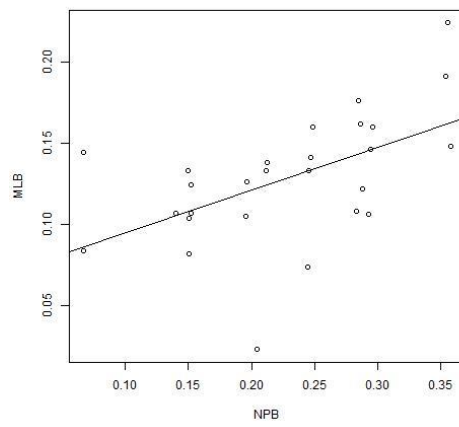
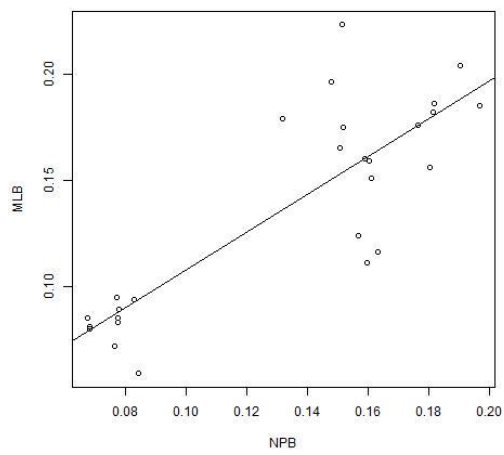


図-18 打者 K% の回帰直線





以下の式は、各説明変数の変化%の AR (1) モデルの推定式である。これら式からは、K% の変化%以外では符号変化を伴うアップダウンを描いていることを示している。このモデルは、以降で CNS から NPB へのトランザクションに伴う将来推定を行う際に利用する形である。式の下に示す表は推定結果である。

(投手)

$$BABIPRate_t = -0.1009 + 0.0006 \cdot BABIPRate_{t-1} + \varepsilon_t$$

$$LOB\%Rate_t = -0.1894 + 0.0023 \cdot LOB\%Rate_{t-1} + \varepsilon_t$$

$$H/9Rate_t = -0.2607 - 0.0007 \cdot H/9Rate_{t-1} + \varepsilon_t$$

$$HR/9Rate_t = -0.0392 - 0.0007 \cdot HR/9Rate_{t-1} + \varepsilon_t$$

(打者)

$$BABIPRate_t = -0.1609 + 0.0007 \cdot BABIPRate_{t-1} + \varepsilon_t$$

$$ISORate_t = 0.0117 - 0.0056 \cdot ISORate_{t-1} + \varepsilon_t$$

$$K\%Rate_t = -0.3937 + 0.0035 K\%Rate_{t-1} + \varepsilon_t$$

表 20 NPB 投手モデルにおける説明変数の変化%の AR(1)モデル推定結果

説明変数	係数	切片
BABIP	-0.1009	0.0006
LOB%	-0.1894	0.0023
HR/9	-0.0392	-0.0007
H/9	-0.2607	-0.0007

表 21 MLB 打者モデルにおける説明変数の変化%の AR(1)モデル推定結果

説明変数	係数	切片
log (BABIP)	-0.1609	0.0007
log (ISO)	0.0117	-0.0056
log (K%)	-0.3937	0.0035

### 第 3 節 CNS

ここでは、MLB や NPB と同様に、CNS のデータを用い、説明変数の有意度合、モデル式全体が有する説明力の把握を行う。

まず、各変数の係数を整理する。表 22 と 23 は、CNS パネルデータにおける投打モデルの係数と、その決定係数を示している。投手における BABIP 項の有意性が非常に低く、符号も理論とは合致していない。しかし、それ以外の説明変数と決定係数は共に問題ない値となった。打者においては、符号関係、決定係数共に有意であるといえる。

表 22 CNS データにおける投手モデルの係数

説明変数	係数	p 値
BABIP	-4.461592	0.0042
LOB%	-11.378549	2e-16
HR/9	0.876381	2e-16
H/9	0.463571	2e-16
R-Squared	0.87042	

表 23 CNS データにおける打者モデルの係数

説明変数	係数	p 値
log (BABIP)	0.786687	2.2e-16
log (ISO)	0.258673	2.2e-16
log (K%)	-0.113138	2.2e-16
R-Squared	0.79124	

図 19 と 20 には、設定モデルにより導出された主体別 Ability 項のヒストグラム、表 24、25 では各ポジション別で Ability 項のトップ/ワースト 5 人を示す。

図 19 投手 Ability のヒストグラム

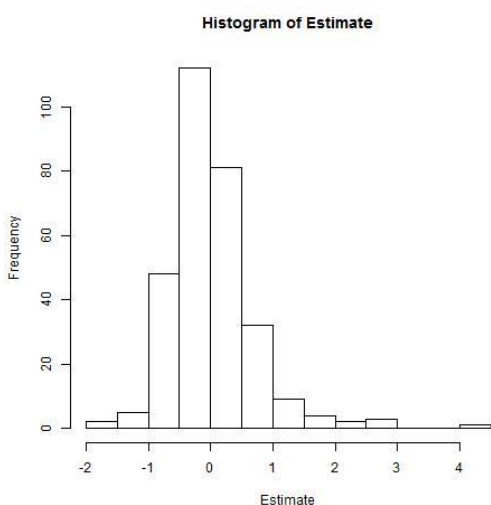


図 20 打者 Ability のヒストグラム

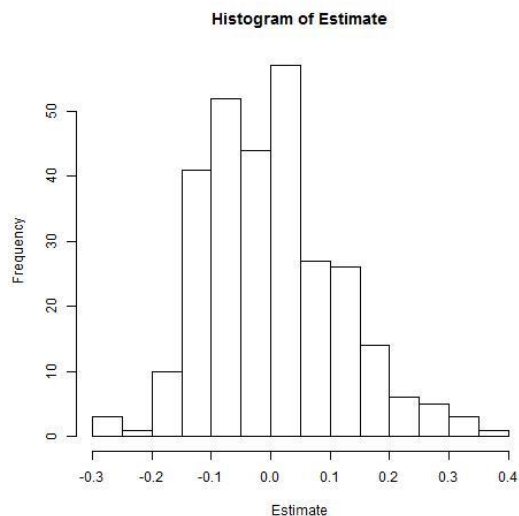


表 24 CNS 投手 Ability

順位	選手名	主体別ダミー	p 値
1	PERERA FROMETA,	-1.650660936	0.000165062
2	JOSE BETANCOURT	-1.647600928	0.000110663
3	ISRAEL SANCHEZ	-1.451201399	0.000692126
4	ALEXEI MILAN	-1.444095665	0.002781054
5	PABLO FERNANDEZ	-1.381293343	0.002230453
...			
290	YUSNIEL FLORES	2.440337879	3.88e-05
291	ANGEL MARQUEZ	2.714747061	1.62e-10
292	RAIMUNDO VAZQUEZ	2.809808117	6.33e-09
293	ORLEIDIS SALAZAR	2.88872202	1.39e-06
294	HANCEL DELGADO	4.087718936	1.84e-11

表 25 CNS 打者 Ability

順位	選手名	主体別ダミー	p 値
1	Jose Abreu	0.371047684	7.55e-15
2	JOSE VELAZQUEZ	0.337497266	1.14e-09
3	JOAN PEDROSO	0.333787276	6.83e-12
4	ROLANDO MENDEZ	0.308141902	1.86e-08
5	ENRIQUE DIAZ	0.298427924	1.50e-07
...			
286	LEANDRO LAMADRID	-0.18851	9.81E-05
287	ADRIANO GARCIA	-0.189773242	0.001128481
288	YAISMEL ALBERRO,	-0.249382323	2.56e-05
289	FRANKLIN ABALLE,	-0.266716985	5.69e-07
290	DANIEL AGUILERA	-0.286465431	6.64e-06

P 値を見る限り、投打共に非常に有意な係数が算出された。NPB では個体差に大きな格差が生まれたが、CNS は MLB と同等のレベルの個体差となった。投手ダミーの上位は MLB, NPB と違い、負の数となった。

次に、CNS のデータを MLB に変換する際に用いるダミー変数について表 26 に整理する。また、図 21、22 は投手、打者変換式の回帰直線を示している。尚、この回帰直線は原点を通過するように定めた。

表 26 CNS データにおける投手・打者 Ability の係数

	係数	p 値	R-Squared
投手	9.926	0.336	0.7469
打者	-0.09223	0.866	0.01105

投手ダミーは有意な値を取ったのに対し、打者ダミーの決定係数が非常に脆弱である。

図 21 投手 Ability の回帰直線

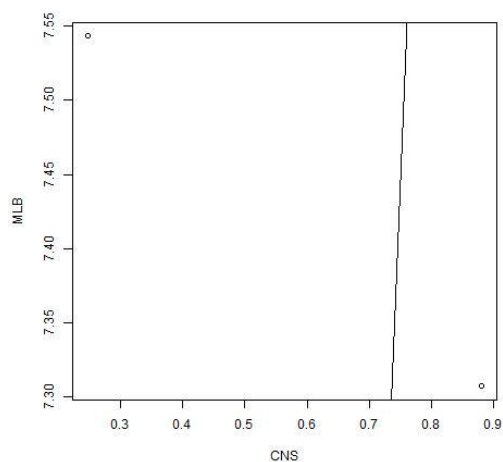
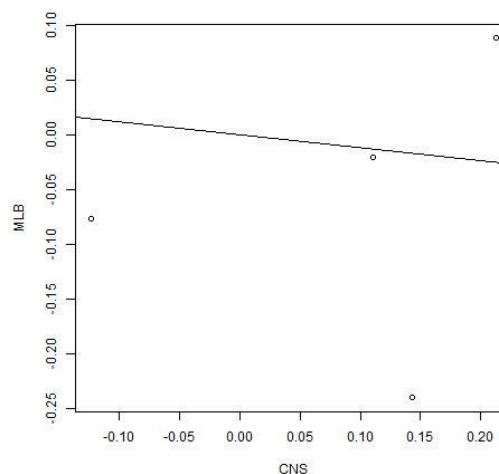


図 22 打者 Ability の回帰直線



次に、CNS のデータを MLB に変換する際の時系列分析で用いる、変換式の説明変数について表 27、28 に整理する。尚、図 23-29 は各説明変数の回帰直線を示している。

投手変換式においては、LOB%項と H/9 項の符号関係が設定した理論とは逆になり、BABIP 項と HR/9 項の決定係数は非常に脆弱である。しかし、H/9 項に関しては非常に高い有意性を示している。

表 27 CNS データにおける投手変換式の係数 0

説明変数	係数	p 値	切片	p 値	R-Squared
BABIP	0.07144	0.888	0.27062	0.101	0.005563
LOB%	0.17726	0.0705	0.46873	0.0164	0.5998
HR/9	0.1011	0.885	0.5269	0.452	0.00588
H/9	-0.3894	0.000549	0.3800660	0.000259	0.962

表 28 CNS データにおける打者変換式の係数

説明変数	係数	p 値	切片	p 値	R-Squared
log (BABIP)	0.08766	0.833	0.26259	0.163	0.01251
log (ISO)	0.30434	0.0263	0.06939	0.0760	0.7475.
log (K%)	-0.64752	0.1689	0.27959	0.0036	0.4128

打者変換式においては、設定した理論と符号関係が一致したが、BABIP 項の決定係数が非常に脆弱である。ISO 項の係数と決定係数は有意な値を算出した。

図 23 投手 BABIP の回帰直線

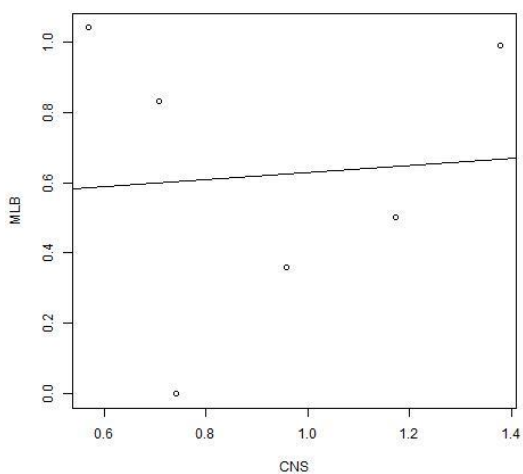


図-24 投手 H/9 の回帰直線

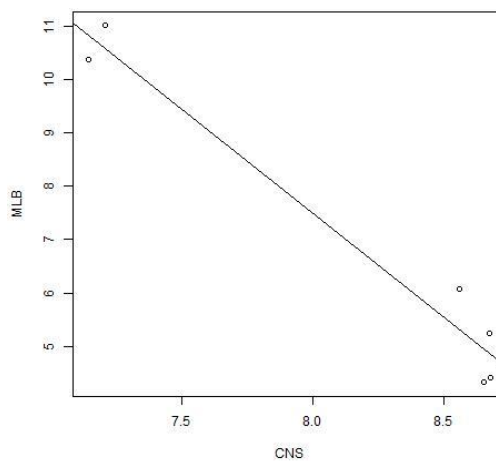


図 25 投手 HR/9 の回帰直線

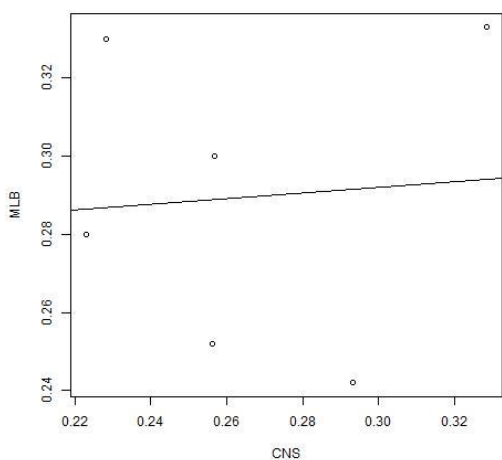


図-26 投手 LOB% の回帰直線

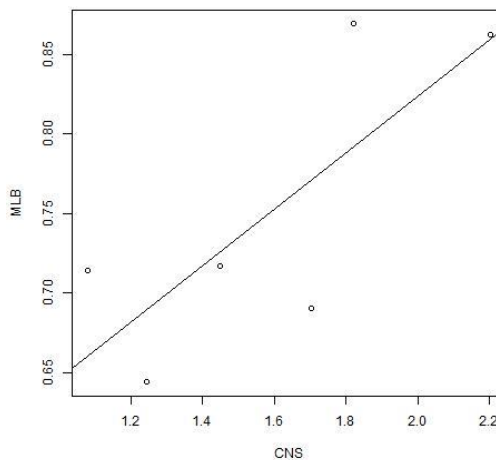


図-27 打者 BABIP の回帰直線

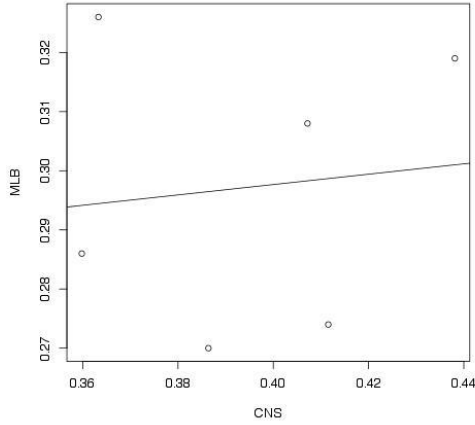


図-28 打者 ISO の回帰直線

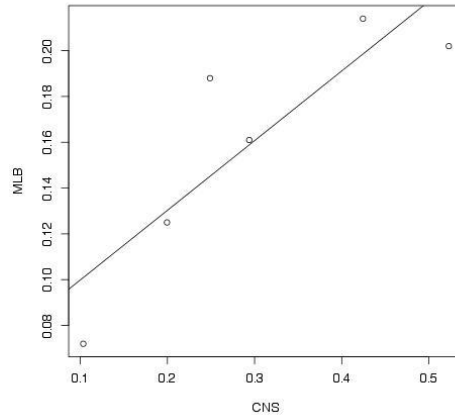
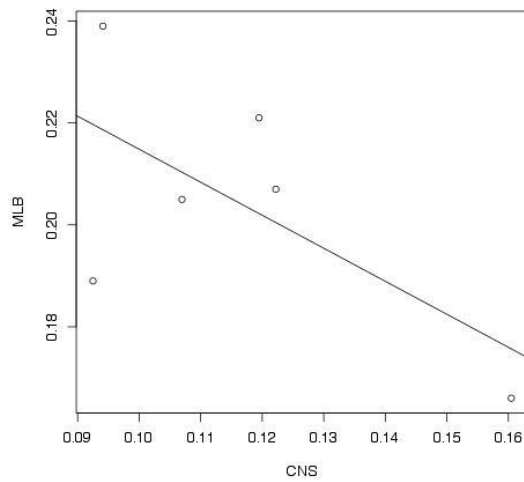


図-29 打者 K% の回帰直線



以下の式は、各説明変数の変化%の AR (1) モデルの推定式である。これら式からは、LOB%、HR/9、ISO、K%の変化%以外では符号変化を伴うアップダウンを描いていることを示している。このモデルは、以降で CNS から NPB へのトランザクションに伴う将来推定を行う際に利用する形である。式の下に示す表 29 は推定結果である。

(投手)

$$BABIPRate_t = 0.2257 - 10.1567 \cdot BABIPRate_{t-1} + \varepsilon_t$$

$$LOB\%Rate_t = 0.5802 + 6.003 \cdot LOB\%Rate_{t-1} + \varepsilon_t$$

$$H/9Rate_t = 0.3871 - 12.9349 \cdot H/9Rate_{t-1} + \varepsilon_t$$

$$HR/9Rate_t = 0.6928 + 2.102 \cdot HR/9Rate_{t-1} + \varepsilon_t$$

(打者)

$$BABIPRate_t = 0.6722 - 0.2391 \cdot BABIPRate_{t-1} + \varepsilon_t$$

$$ISORate_t = 0.6408 + 5.7944 \cdot ISORate_{t-1} + \varepsilon_t$$

$$K\%Rate_t = -0.423 + 0.0417 K\%Rate_{t-1} + \varepsilon_t$$

表 29 CNS 投手モデルにおける説明変数の変化%の AR(1)モデル推定結果

説明変数	係数	切片
BABIP	0.2257	-10.1567
LOB%	0.5802	6.003
H/9	0.3871	-12.9349
HR/9	0.6928	2.102

表 30 CNS 投手モデルにおける説明変数の変化%の AR(1)モデル推定結果

説明変数	係数	切片
log (BABIP)	0.6722	-0.2391
log (ISO)	0.6408	5.7944
log (K%)	-0.423	0.0417

## 第 4 節 検証

ここでは、今まで算出したデータと回帰式を用い、実際に CNS→NPB というトランザクションを経てプレーを行っている、Gourriel と Despaigne の 2 選手の成績予測を行う。そして、設定したモデルから将来推定した TA の値と、2014 年度の実際の TA の値を比較してモデルの完成度を推し量る。

表 31 では Yulieski Gourriel と Alfredo Despaigne を用いて行った検証結果を示す。

表-31 Yulieski Gourriel と Alfredo Despaigne の検証結果

	TA 推測値	TA 実測値
Yulieski Gourriel	0.819630348	1.017094017
Alfredo Despaigne	0.862222649	0.870588235

両選手とも推測値と実測値に大きな差が見られなかったため、今回設定したモデルはある程度機能していると言える。特に Alfredo Despaigne の値に関しては非常に近似している。

## 第4章 インプリケーション・課題

ここでは、これまでの検証結果を改めて整理し、本論の評価点、問題点をあげて今後に向けての提言を行い、本論文の結びとする。

2013年9月に施行されたキューバのスポーツ政策を受けて、今後日本においてキューバ人アスリートの雇用が活性化することが予想される。NPBでは2014年9月現在、既に4人のキューバ人が雇用契約を行い日本にてプレーをしている。本論文は、ここで発生する雇用リスクをパフォーマンスの将来予測を行うことで可視化し、更に野球における将来予測分野の先駆けとなるモデルを構築することを目指すものであった。

本論文では、全てのプレーは Ability と Luck から構成されることを前提にモデル立式を行い、Ability 部分を固定効果モデル、Luck 部分には自己回帰モデルを用いて分析を進めた。先行研究で触れた EqA は、高い精度での予測を可能にするために計 10 式を要したが、本論文においては打者の場合、個別効果モデルにおけるダミー変数、時系列分析における TA, BABIP, ISO, K% を求めるにとどまっている。より精度の高い将来推定を行うためには、今一度、説明変数、被説明変数の見直しを行い、変数を増やす必要があると考えられる。また、時系列分析においては AR(1) 過程を用いて推定を行ったが、AR(p) 過程を用いた方が初期値の影響を受け辛く、精度も増すと考えられる。しかし、AR 過程が自身の過去に回帰される以上、項を増やすためにはその増やす分の年度データが必要になる。ここで困難となったのがデータ入手の問題であり、これは特に CNS において顕著な課題点である。

今回は MLB、NPB、CNS 各リーグにおいてパネルデータを作成し、日米間、米玖間の相関関係を推定することで、NPB、CNS における成績を MLB に変換する関係式を立式し、同環境下での比較を可能にした。データを入手する際、NPB や MLB では公式機関によって見やすいデータが整備されているため、複数年分の確かなデータを入手するのが容易であったが、CNS に関しては 2006~2007 年度以前のデータを入手するのが不可能であり、近年になってデータバンクが整備されたため、パネルデータを作成する上で 2~4 年という短い連続データを用いる他なかったのは非常に残念な点であった。しかし、EqA とは違い、誰にでも入手可能なデータを用いて高いレベルの分析を可能にしたことは、評価すべき点であると考えられる。

最終的に、集めたデータから算出した固定効果モデルと AR(1) モデルを用いた関係式により MLB 下での比較が可能になったことで、CNS→NPB というトランザクションに伴う次



年度パフォーマンス推定を行うモデルが完成した。作成したモデルが正しく動くかを、検証パートにおいて、Yulieski Gourriel と Alfredo Despaigne の 2 選手を用い実証分析を行ったところ、TA 推定値が TA 実測値と近似したことから、今回設定したモデルを用いればある程度精度の高い推測が可能であると示すことができる。検証では良い値を算出することができたものの、現状 CNS→NPB のトランザクションを経た選手が非常にごく少数に留まっており、データのサンプル数が少ないことから、今回のモデルの有意性が必ずしも高いとは断言できないとも考えられる。

今一度モデルの係数を確認すると、NPB における、投手モデルの係数、打者ダミーの係数、投手変換式の係数は決定係数が非常に低いことが問題視できる。CNS では、打者ダミーの係数、投手変換式の BABIP 項と HR/9 項、打者変換式の BABIP 項の決定係数が非常に脆弱である。それにもかかわらず検証結果で近似した値を算出できたのは、各リーグに特徴があり、それが顕著になることで符号関係も変化してしまうことが考えられるためであると推測できる。これはリーグを隔てると成績予測が困難になることの表れでもあり、各リーグにおける特徴を細かく分析し、補正することで推定精度を向上させることが可能であると考えられる。リーグの特徴となる値を明確にすることが今後の課題である。

野球における将来予測分野はまだ未開の領域であり、研究者も少ない。その中である程度の有意性を示すことのできた本論文は、この分野において大きな貢献をしたと言っても過言ではないだろう。今後もこの分野が発展していくことを願い、本論文の結びとしたいと思う。

## 参考文献

- 戸部良也, 「強くて楽しいキューバ カストロの国の体験記」(2005), 風人社
- 福山博文, 「プロスポーツ選手の雇用契約における利用可能性: ヒューリスティックスの検証」, 2014年, 地域政策科学研究, 11, pp.27-40. (リンク)
- <http://ir.kagoshima-u.ac.jp/bitstream/10232/20718/1/27-40.pdf>

## Web 資料、データ出典

- Baseball Prospectus, “Glossary EqA”,  
<http://www.baseballprospectus.com/glossary/index.php?mode=viewstat&stat=61>
- Davenport, “Davenport Translation”,  
[https://groups.google.com/forum/embed#!topic=searchin/rec.sport.baseball.analysis/after\\$3A1995\\$2F01\\$2F01\\$20before\\$3A1995\\$2F01\\$2F31](https://groups.google.com/forum/embed#!topic=searchin/rec.sport.baseball.analysis/after$3A1995$2F01$2F01$20before$3A1995$2F01$2F31), 1995
- Serie Nacional de Beisbol  
([http://www.inder.cu/beta/competitions/baseball/Campeonato\\_Nacional/Serie\\_Nacional/](http://www.inder.cu/beta/competitions/baseball/Campeonato_Nacional/Serie_Nacional/), 8/19)から 2006～2013年
- NPB 公式ホームページ (<http://www.npb.or.jp>, 8/20)より 2009～2013年
- FanGraphs.com” (<http://www.fangraphs.com>, 8/18) より 2009～2013年
- 鷲田康, 「プロ野球亭日乗 日本球界に乗り込むキューバの英雄。門戸開放の陰に『亡命』と『裏開催』」(2014年5月24日), <http://number.bunshun.jp/articles/-/820844>, (検索日: 2014年8月25日)

## 補論 キューバ野球の概況

本論では CNS と NPB2 リーグを結ぶ人材ルートをテーマとして取り扱うが、ここでは補論として、キューバ政府がプログラム施行に至った要因について、同国の沿革を整理した後、その概況を述べる。

### 補論 A キューバ革命に伴う対外交流の変化

1864 年に Nemesio, Ernesto の Guilo 兄弟により野球のルール、用具が初めて伝来された後、キューバの野球はアメリカ合衆国との相互的な歩みの中で成熟を辿った。1869 年に Steven Bellan がアメリカ国内のセミプロチーム『トロイ・ヘイメイカーズ』でプレーをしたことを皮切りに、アフリカ系アメリカ人を中心としたリーグ『ニグロ・リーグ』(Negro League)、そして『メジャーリーグ・ベースボール』(以下、MLB) との交流が 1950 年代まで行われた。この期間にニグロ・リーグでプレーを行った Christobal Torriente, Jose Mendez 等のキューバ人選手は、後年にアメリカ野球殿堂 (National Baseball Hall of Fame 以下、HOF) 入りを果たしている。

一方 1950 年代は、アメリカ合衆国の支援を受け、アメリカ帝国主義的かつ反民族的な独裁政権を確立した Fulgencio Batista の下で、キューバの社会悪の深刻化が顕著となった時期でもある。Batista は 1952 年に国内砂糖生産の制限、1955 年には経済社会発展銀行 (BANDES) を設置し、公的資金を独裁に奉仕している企業、労働者連合に回す等、社会全体の向上たる視点で評価するといずれも非合理的な政策を行った。これに伴い、1958 年にキューバ政府は公的負債 12 億ペソを抱え破産に追い込まれ、また失業者人口も経済活動人口の 1/3 を占める 738000 人に上った。

Batista 政権下での社会悪の深刻化に対し、キューバでは Fidel Castro, Che Guevara を中心とした「キューバ革命」が 1959 年に成立した。1960 年に「キューバの大衆と革命政府の密接な結びつきを示し、革命的願望を維持する人民の意志を明確に表わした」(キューバ教育省, P464) 機会とされている第 1 次ハバナ宣言では、「アメリカ大陸と世界の人民に向けて、人間による人間の搾取と、帝国主義的支配のためにアメリカ大陸人民が被っている経済的・政治的・社会的悪を非難し、これらの人民は経済的・政治的・社会的復権と抑圧と搾取からの解放のために闘う義務があること」(キューバ教育庁, P464) を宣言し、これに対しアメリカ合衆国も 1961 年に食料、医薬品を入れた全面的な禁輸措置を実行<sup>12</sup>、今日まで経済、物流いずれの側面でも断交している。

---

<sup>12</sup> 1962 年に禁輸措置は「人道的措置を称する食料、医薬品のみ」は例外として除外されている

これに伴い、野球でも例外無く、キューバと MLB との結びつきは急速な冷え込みを示した。キューバ革命が成立した後、キューバと MLB が交流を行った事例は 1999 年に開催されたボルティモア・オリオールズとのエキシビション・マッチが唯一である。2006 年から行われている MLB 主催の国際大会『ワールド・ベースボール・クラシック』でも、全 3 回の大会でキューバとアメリカ合衆国と同じグループに属したことは 1 度も無く、また両国の対戦も 2014 年現在実現していない。

また、キューバ政府はアメリカ合衆国以外に対しても、自国のアスリートが国外にてプロ契約を結ぶことを 1961 年から禁止している。ただし、これは対外交流からの絶対的な離脱を意味するものではなく、キューバ革命後首相に就任した Castro は技術研修を名目とした海外への選手、コーチの派遣を推奨している。戸部 (2005, pp.58~68) では、以下の記述を見ることが出来る。

“すでにこの頃 (注 文中からは 1994~1996 年の期間と読み取ることが出来る)、キューバの野球関係者は、中国、ロシア、東欧、メキシコ、ニカラグアなどへ指導者を相当数派遣していた。

特に中国の場合は、選手の交流も行われていて、キューバへ留学する選手やコーチもいるし、逆にキューバからコーチと選手団がやってきて、試合を行ったり、指導に当たっていた。

メキシコにも、キューバからコーチは 10 名以上も出かけている。

(中略) マルチネス・コロナに言わせると、「スポーツ技術者の海外派遣」である。

1991 年ですでに 25 ヶ国に出かけており、92 年には 20 ヶ国に 259 名を派遣して 140 万ドルを稼いでいる。これが年を追うごとに大きくなっている。まさにキューバは、スポーツが外交の重要な一端を背負うと同時に、スポーツ活動のための資金不足をも、技術者の派遣によって補うようになった。(P188~189)”

具体例としては 2014 年 8 月に政府プログラムにより読売ジャイアンツに研修コーチとして入団した Carlos Soto が、日本以外にもニカラグア、イタリア、ベネズエラでもコーチ研修を行っている。またキューバは 1990 年代末~2004 年にかけて日本の社会人野球チーム『シダックス』に選手の派遣も行っており、Antonio Pacheco, Orestes Kindelan, Victor Mesa 等、代表経験を持つ選手が来日したケースも見受けられる。

## **補論 B キューバ野球のシステム**

キューバ野球は絶対的な鎖国下で醸成されたものではないが、今日に至るまでの競技レベルに至った要因は、国内で形成された「スポーツ学校」を土台とした育成システムにあ

る。以下図-30は、戸部（2005, pp.58~68）の記述に基づき、キューバにおけるアスリート育成のシステムである『ステート・アマ・システム』を整理したモデル図である。

図 30 キューバのアスリート育成システム

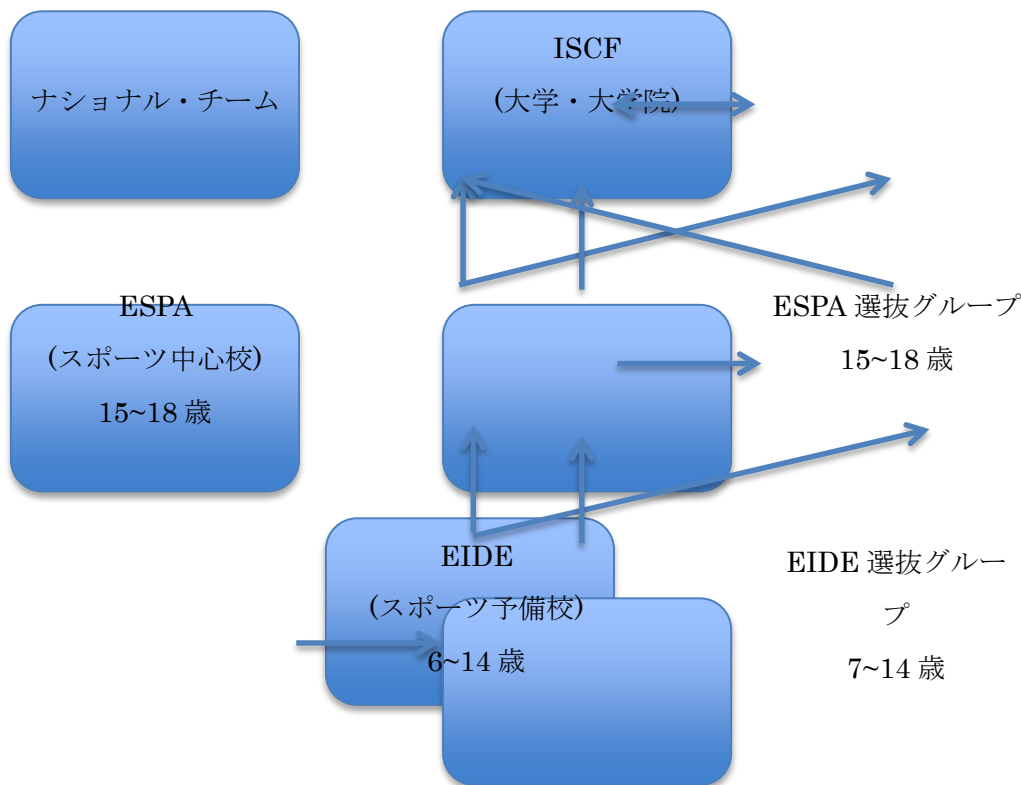


図 30 において、EIDE は小・中学校、ESPA は高校、ISCF は大学・大学院との区分である。キューバにおいてスポーツ学校のシステムが完成したのは 1960 年であり、以降は革命後に国会評議会議長となった Castro の下、スポーツ振興政策により規模の拡大が押し進められた。キューバでアスリートになる上では、テスト、もしくは一般小学校の体育教師からの推薦を経て、EIDE に入学することが起点となる。

キューバのスポーツ学校システムの特徴は、以下 2 点を示すことができる。

1. キューバのアスリートが「ステート・アマ・システム」下であり、アスリートは雇用契約上は国家公務員と同じ扱いを受ける。アスリートには一定の給与、住宅、自動車の貸与等、国内ヒエラルキーにおいて上位の環境での生活が保障されているが、給与レベルは資本主義国と比較し極端に低い。
2. キューバのスポーツ学校システムは垂直統合的であり、「合格点を取っていればトコロテン式に上がってゆけるエリート学校」（戸部, P58）である。システム上、小・中学校とナショナル・チームは結び付きを持っている。