Korean Journal of Sport Science
2018, Vol. 29. No. 3, 520-533.
https://doi.org/10.24985/kjss.2018.29.3.520

체육과학연구
ⓒ Korea Institute of Sport Science

The development of a valuation model to determine the true market value of professional baseball players

Sung-Bae Roger Park*, Tae-Geun Kwon & Jong-Hwan Jeon
Hanyang University

[Purpose] The main purpose of this current study is two-fold. Firstly, it attempts to develop a model to determine the true market value of Korean professional baseball players (hitters only) solely based on their athletic performances on the field. Secondly, it is to provide the evidential data for the market value of baseball players in Korea. **[Methods]** The statistical data and performance information were obtained from baseball almanac from KBO from 1997 to 2016. Seven hundred and ninety three players were included for data analysis. Principal component factor analysis was utilized to eliminate multicollinearity among 12 sabermetrics indices (OPS, GPA, SECA, TA, RC, RC/27, XR, ISO, PSN, sOBA, %OW, BABIP) and increase power of explanation of the proposed model with KMO(=0.77), p<0.001. **[Results]** The proposed model was successfully developed with Y_{Salary} = Years of Experience*921.5 + FA (free agent)*53528.9 + PHI(Power Hitter Index)*7313 + CHI(Contact Hitter Index)*5893.6. Furthermore, the proposed model explained 64.5% of variances of the market value for the Korean professional baseball players and proved to be statistically valid. **[Conclusions]** The newly developed model in this study was very helpful for us to identify the variables that affect the true market value of baseball players. It is expected that this model could make an important contribution in determining true market value of the baseball players in Korea.

Key words: Salary, Valuation, Sabermetrics, Component analysis

서 론

연구배경

기존 선수 연봉 책정 방식의 문제점

한국프로야구는 2000년대 중반 이후 대한민국 야구 대표팀의 연이은 국제대회에서의 호성적과 선수들의 인지도 상승에 힘입어 인기가 급상승했으며 이에 따라 한국프로야구 리그는 외형적인 측면에서 비약적인 성장을 이루며 대한민국 최고의 국민 스포츠로 자리 잡게 되었다. 1991년 8개 구단 체제가 구축된 한국프로야구 리그는 2012년 NC 다이노스, 2013년 KT 위즈의 리그 참가와 함께 10개 구단 체제로 외연을 확장했으며 현 체제 구축 이후 해마다 지속적인 관중 증가 추세를 보여 2017년 역대 최다인 840만 명을 돌파한 상황이다 (www.koreabaseball.com). 지속적인 리그의 성장과 구단 간 경쟁의 심화 속에서 선수 연봉 또한 가파른 상승 추세를 보이며 2018년 기준 선수 평균연봉 1억 5026

논문 투고일 : 2018. 04. 26.
논문 수정일 : 2018. 07. 30.
게재 확정일 : 2018. 08. 23.
* 교신저자 : 박성배(srogerpark@hanyang.ac.kr).
* 이 논문은 2017년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임(2017S1A5A2A01025770).

만원을 기록했다(KBO Press Release, 2018). 특히 최근 자유계약(FA) 선수시장에서 100억이 넘는 계약이 연이어 체결되면서 객관적인 선수 가치 평가와 합리적인 연봉 기준 책정에 대한 사회적 요구가 증가하고 있는 상황이다.

한국프로야구 구단들의 기존 선수 연봉 책정 방식은 크게 연봉 산정 기준들이 프로스포츠 선수의 특수성을 반영하기에 부족하다는 점과 객관적인 선수 가치 평가에 대한 인식이 부족하다는 두 가지 문제점을 내포한다. 프로야구선수의 연봉을 책정하는 방식에 대한 논의는 국내에서 오랫동안 지속되어 왔으며 30여 년간 한국프로야구가 지속되면서 구단들은 선수 연봉 책정에 대한 방식을 고안해내는 과정에서 적지 않은 노하우를 축적해왔다. 구단들은 각자 나름의 특성을 반영한 연봉 산정 기준을 가지고 있으며 이를 점수화하여 선수와의 연봉협상에 자료로 활용한다고 알려져 있다. 하지만 아직도 특정 구단에서는 구단의 모기업에서 사용하고 있는 인사 고과 방식을 프로야구단 소속 선수들에게 그대로 적용하고 있는 것이 현실이다. 선수의 가치를 공정하게 평가하기 위해 구단은 운영팀, 육성팀 등에 연봉 산정 담당자를 두고 있으며, 연봉을 입증하기 위한 다양한 기준을 두고 있다(Park, 2008). 구단은 데이터로 산출한 고과를 근거로 책정한 연봉을 선수들에게 제시한다. 한국프로야구 모구단의 연봉 고과 산출 기준을 보면 구단 고과(50%), 정규 시즌 성적(20%), 타석 수 또는 투구 이닝(10%), 1군 등록 일수(10%), 코치 고과 (10%) 등의 항목으로 구성된다(Sung, 2014). 이와 같이 선수들의 연봉은 성적에 비례하여 객관적으로 평가되기 보다는 구단과 선수 간의 인간적인 관계나 구단의 성과와 같은 외부적 요소에 의해 큰 영향을 받고 있다. 또한 선수는 성적과 관련된 직접 정보뿐만 아니라 통계적 가공을 거친 다양한 정보의 획득이 쉽지 않을 뿐만 아니라 보류제도 (Reserve Clause)로 인해 입단 이후 수년간 구단에 종속되는 특수한 관계라는 요인이 발생한다. 결국 구단에 비해 현저히 낮은 협상력으로 인해 연봉협상에서 불리한 상황에 처할 가능성이 높다.

그동안 한국야구위원회(KBO)는 2001년 공정거래위원회의 에이전트 제도 시행 명령 조치에도 불구하고 미루어오던 국내 선수에 대한 에이전트 제도를 2018시즌부터 도입했으며 자격을 획득한 에이전트는 이미 활동을 개시하였다. 에이전트를 배제한 구단과 선수 개인 간의 선수 연봉 협상은 객관적 기준의 부재와 구단과 선수 간 정보의 비대칭 문제로 인해 공정한 협상이 이루어지기 어려운 구조이며 시장논리가 배제된 협상이 이루어질 수밖에 없기 때문에 선수 가치가 시장가치에 비해 왜곡될 가능성이 높았다. 에이전트 제도가 실행됨에 따라 시장논리에 입각한 공정한 연봉 협상의 요건은 충족이 되었으나 여전히 국내 프로야구 선수에 대한 객관적이고 합리적인 연봉 산정 기준은 매우 불투명한 상황이다. 따라서 향후 프로야구 에이전트 제도가 순기능을 발휘하기 위해서는 선수의 시장가치를 객관적으로 반영할 수 있는 연봉 산정 플랫폼의 도입이 절실한 상황이다.

정량적 선수가치평가 연구의 부족

프로스포츠 선수의 연봉은 전통적으로 과거의 기록과 성적 데이터를 바탕으로 미래의 성적을 예측하고 이를 보상하는 개념으로 지급되고 있다. 이로 인해 선수의 부진, 부상 등과 같은 미래의 불확실한 변수들을 제대로 반영하지 못하고 선제적으로 연봉이 책정되어 선수의 가치를 제대로 반영하지 못하는 문제점을 안고 있는 것이다. 미국 메이저리그(MLB)에서는 이러한 문제점을 보완하기 위해 선수 연봉에 다양한 옵션과 인센티브를 제공하고 이를 통해 선수 계약 실패에 대한 위험부담을 줄이려는 노력을 하고 있다. 이에 비해 한국 프로야구는 선수 계약과 관련된 옵션과 인센티브계약과 같은 위험부담을 줄이기 위한 안전장치가 부족한 현실이며 계약 보장금액이 지나치게 높아 결과적으로 선수들의 계약금액과 총 연봉규모가 지속적으로 상승하고 있다. 또한 자체적인 선수 가치 평가에 대한 기준이 마련되지 않아 이에 대한 개선이 요구되고 있는 상황이다.

선수 가치 책정에 대한 학계의 연구는 크게 두 가지로 나뉜다. 첫째, Scully(1974)가 고안해 낸 구단 수익방정식을 이용한 개별 선수의 가치 책정에 대한 연구로 노동경제학의 균형임금이론을 반영하여 선수의 가치를 한계수익생산(MRP; Marginal Revenue Product)의 형태로 책정하는 방식이다. 구단의 생산물인 수익과 승리에 관련하여 선수들의 기여도를 통계적으로 검증하기 위한 방법으로 선수의 객관적인 가치를 추정하기 위하여

구단 수익방정식과 승리방정식을 추정하고 이를 통해 개별 선수의 한계수익생산을 도출하는 방식을 고안해냈다. Krautmann(1999)은 프로스포츠 노동시장의 불완전 경쟁 상태를 고려하지 못한 Scully(1974)의 방식에 문제점을 지적하고 균형 임금을 한계수익생산으로 측정하기 위해 프로스포츠 노동시장이 완전경쟁상태에 도달하는 자유계약(FA) 선수시장으로 대상을 한정하여 연구를 수행했다. FA선수들의 연봉 자료와 개인성적을 바탕으로 세부 성적이 연봉 결정에 미치는 영향을 알아보고 이를 통해 연봉 방정식을 추정하여 FA 대상자가 아닌 선수들까지 포함된 일반화된 선수 가치 평가 모형을 만들어냈다. 또한 Depken & Wilson(2004)은 기존의 수익방정식 및 승리방정식을 통해 추정된 개별 선수의 한계수익생산(MRP)을 선수의 실제 연봉과 비교하여 노동수요자인 구단주의 잉여를 알아보고 이를 통해 선수들의 연봉 적정성을 알아보려는 연구를 수행했다. 두 번째 방식은 선수들의 개인 기록을 팀 승리와 관련된 기여도로 환산하고 이를 선수 연봉과 비교하여 선수 연봉의 적정도를 알아보려는 시도이다. 이를 위해 가장 많이 사용되는 지표로는 Tango, Lichtman, & Dolphin (2007)이 고안해낸 WAR(Wins Above Replacement)이 대표적인데, 이는 미국프로야구(MLB) 소속의 특정 선수를 동일포지션의 평균적인 선수로 대체 기용했을 때를 가정하여 승리기여도를 승수의 개념으로 알아내는 방식이다. 주로 WAR로 환산된 선수의 기여도를 선수 연봉과 비교하여 선수의 연봉 적절성을 알아보는 연구를 수행하고 있는 것이다.

선수가치를 평가하는 모델로는 구단 수익에 대한 선수 개개인의 기여도를 평가하여 금액적 가치를 환산하는 Scully(1974)의 한계수익생산 함수 모형을 통한 선수 가치 정량화 방식과 개인 기록에 대한 분석을 통해 지급된 연봉과의 상관관계를 밝히고 그 적절성을 평가하는 Tango et al.(2007)의 WAR 계산 방식이 존재한다. 국내 연구의 경우 후자의 상관관계 관련 연구가 주류를 이루고 있으며 전자의 선수 가치 평가 및 적정 가치 환산과 같은 정량적인 연구는 많이 이루어지지 않은 상황이다. 기존의 해외 연구들을 살펴보면 Scully(1974)가 사용한 수익방정식 및 승리방정식을 기초로 하여 선수들의 가치를 한계수익생산(MRP)으로 환산하고 적정 연봉을 찾아내는 정량적인 연구들이 오랫동안 이루어져 왔다(Depken & Wilson, 2004; Fields, 2001; Krautmann, 1999; Sommers & Quinton, 1982). 기존의 연구들을 활용하여 충분히 한국프로야구 실정에 맞는 모형을 구축할 수 있으며, 프로야구선수들의 적정 연봉에 대한 관심과 우려가 높아지고 있는 현 상황에서 선수들의 적정 가치를 평가하는 정량적인 연구를 통해 객관적인 정보를 제공할 필요가 있다.

야구는 타격, 투구, 수비, 주루와 같이 다양한 상황 속에서 구분된 기록들이 존재하며 다른 종목에 비해 확률과 데이터로 상황을 설명하기 쉬운 장점으로 인해 20세기 초부터 일찍이 다양한 형태의 통계적인 분석 및 연구가 이루어졌다. 현재의 야구 통계 기록이 각광 받기 시작한 것은 미국의 빌 제임스의 연구와 밀접한 관련이 있으며 그는 선수들의 1차적인 성적 데이터를 가공하여 선수 기록과 가치를 객관화하는데 도움을 줄 수 있는 지수인 세이버메트릭스(Sabermetrics)를 고안해냈다(Choi, 2009). 안타, 타점, 도루와 같은 단순 성적 데이터와 달리 세이버메트릭스 지수들은 득점과 승리에 대한 개별 선수의 기여도 및 파급력을 분석할 수 있어 선수들의 실질적인 가치를 객관적으로 평가를 하는데 유용하다고 알려져 있다. 이에 단순 성적 지표보다는 환산된 세이버메트릭스 지표를 통해 선수의 가치를 다각도로 평가하는 것이 객관적인 선수가치평가에 적합하다고 할 수 있다(Choi, 2009; Lee, 2014; Lee & Kim, 2016).

프로야구선수의 가치평가와 관련된 국내연구로는 선수들의 경기력과 연봉 간의 상관관계 및 연봉 결정요인(Kim, 2001; Lee, 2001; Myung et al., 2016; Park, 2008; Park et al., 2016)을 밝혀낸 연구와 선수들의 경기력 분석을 통해 연봉의 적절성을 평가(Oh & Lee, 2013; Oh & Lee, 2016; Seung & Kang, 2012)하는 연구들이 주류를 이루어 왔다. 이러한 연구들을 통해 선수 연봉과 경기력 간의 상관관계를 알아보고 연봉결정요인을 찾아내는 작업을 통해 선수 연봉의 적절성을 알아볼 수 있는 근거를 제시하였다.

효율적인 선수 노동시장을 위한 가치증대 노력

한국프로야구 선수 시장은 현재 제한된 공급 속에서 구단들의 과열된 수요 경쟁으로 인해 급격한 임금증가

현상을 보이고 있다. 1997년부터 2017년까지 한국 프로야구 선수 연봉의 변화를 살펴보면 하위 30%에 해당하는 선수들의 연봉은 1997년 2,000만원에서 2017년 3,000만원으로 50% 상승에 그친 반면, 상위 1%에 속하는 선수들의 연봉은 1997년 1억 2,000만원에서 2017년 12억으로 20년 새 약 10배가 상승했다(Park, 2017). 노동경제학의 균형임금이론은 노동의 수요와 공급이 균형 상태에 도달할수록 전체 시장의 사회적 후생이 증가한다는 측면에서 긍정적인 평가를 받고 있다. 스포츠 선수시장에서 선수들의 객관적인 가치를 책정해야 하는 이유는 결국 스포츠 리그 전체 시장의 사회적 후생을 증가시키기 위함이다. 선수들의 가치를 환산해내고 선수의 적정 연봉을 제시함으로써 선수 가치에 대한 기준이 부재한 상황에서 소모적인 논쟁을 하기 보다는 객관적인 기준을 가지고 합리적인 계약이 이루어질 수 있도록 학계에서 현실에 맞는 합당한 기준을 제시해 줄 필요가 있다. 효율적인 선수 노동시장의 형성과 동시에 프로야구시장의 지속 가능한 발전을 위해 현시점에서 한국프로야구 시장에 대한 경제적인 분석과 적정 임금에 대한 연구가 시급한 상황이다.

연구목적

본 연구는 한국프로야구의 선수 노동시장에 맞는 객관적이고도 합리적인 '한국형 프로야구선수 적정 가치 평가 모델'을 개발하고 이를 통해 적정 연봉에 대한 근거 자료를 제시하고자 한다. 또한, 기존 세이버메트릭스 지수들에 비해 보다 단순하면서도 설명력이 높은 한국형 모델을 제시함으로써 프로야구 관계자들로 하여금 선수 가치 평가에 대한 이해도를 높이는 동시에 그 활용도를 최대화하는데 이 연구의 목적이 있다.

연구방법

연구대상

본 연구는 한국프로야구(KBO 리그)에서 1997년부터 2016년까지 활약한 타자 중 연구대상의 통일성 및 성적 지표에 대한 균일성을 보장하기 위해 규정타석을 충족하고 차기 년도 계약을 마친 793건 만을 연구대상에 포함시켰으며, 본 연구에 필요한 자료는 한국프로야구위원회에서 발간한 연감과 KBO에서 제공된 데이터를 바탕으로 재구성 및 편집을 하였다. 한국야구위원회의 공식야구규칙에 따르면 규정 타석은 해당 시즌 경기수의 3.1을 곱한 값에서 소수점을 버린 값으로 규정하고 있다. 위의 조건을 충족한 본 연구의 연구대상은 〈Table 1〉과 같다.

Table 1. KBO League At Bats requirement and subjects by season (1997 ~ 2016)

Season	Game	At Bats	Subjects
1997	126	390	36
1998	126	390	41
1999	132	409	37
2000	133	412	35
2001	133	412	39
2002	133	412	34
2003	133	412	37
2004	133	412	42
2005	126	390	38
2006	126	390	35
2007	126	390	40
2008	126	390	36
2009	133	412	37
2010	133	412	43
2011	133	412	36
2012	133	412	40
2013	128	396	46
2014	128	396	50
2015	144	446	44
2016	144	446	47
Total			793

자료수집 및 방법

인구통계학적 정보 및 계약 정보

연봉 책정 년도를 기준으로 한 연구대상들의 '연령', '연차', 그리고 '연봉'을 인구통계학적 변수로 분석에 사

용하였다(〈Table 2〉 참조). 개별선수들의 성적 이외에도 인구통계학적 데이터를 연봉 산정 모델에 활용하였다. 종속변수로 사용한 연봉의 경우 단일 년도 명목 연봉에 2016년을 기준으로 소비자물가지수를 반영하여 물가 변화를 배제한 실질연봉으로 변환하였다. 통계청에서는 소비자물가지수에 대해 소비자가 구입하는 제품이나 서비스의 가격변동을 나타내는 지수로 도매물가지수와 함께 일상생활에 직접 영향을 주는 물가의 변동을 추적하는 중요한 경제지표의 하나이며 도시소비자 가구가 구입하는 제품 및 서비스 가격의 변동을 측정할 수 있도록 편제하고 있다. 소비자물가지수는 일정량의 제품 및 서비스를 경상가격으로 구입할 수 있는 비용의 변동을 나타내며 서울을 포함한 전국 주요도시에서 매월 489개 상품 및 서비스 품목의 가격을 조사하여 '라스파이레스식(Laspeyres formula)' 에 의하여 식료품·비주류음료, 의복·신발, 보건 의료 등 12가지로 세분화하여 산출하고 있다(www.kostat.go.kr). 연봉 데이터의 경우 원데이터를 그대로 사용했을 경우 왜도와 첨도가 각각 2.98과 10.49로 정규성(data normality)을 벗어났기 때문에 이를 해결하기 위해 로그함수를 취한 결과, 〈Table 2〉에서 볼 수 있듯이 연봉 데이터의 정규성 가정을 만족하였다.

Table 2. Player demographics and contract details

	Age	Year	Salary (Unit: 10,000 won)	FA
Max.	19	0	1188	0
Min.	40	22	240000	1
Average	28.63	8.04	23078.32	.18
Median	28	8	10275	0
SD	4.12	3.89	32955.94	.39
Skewness	.21	.53	.46	1.64
Kurtosis	-.47	.05	-.362	.74
N	793	793	793	793

소비자물가지수는 노사 간 임금협상을 하거나 정부가 최저생계비, 공적연금 급여 등을 산정할 때 중요한 지표로 활용된다 (〈Table 3〉 참조). 한국은행의 통화정책인 물가안정목표제도 소비자물가지수를 대상 지표로 사용하고 있다(www.bok.or.kr). 아울러 FA 선수의 경우, 구단에 따라 계약금을 일시불로 지급하거나 계약기간에 따라 연봉과 같이 지급하는 두 가지 경우가 있어 통일성을 유지하기 위해 계약 총액에 계약기간을 나누어 연봉을 추산하였다. 예를 들어, 계약기간 4년 총액 80억 원인 A 선수가 있다고 가정하면, A 선수의 단일년도 연봉을 20억 원으로 계산하였다. 연구대상 중 FA 계약자에 대해 1, 미계약자에 대해 0의 결과값을 부여하여 'FA유무'를 구분하였으며 이를 통해 선수들의 계약 정보를 세분화하였다.

Table 3. Yearly Consumer Inflation Rate (1997 ~ 2016)

Year	Inflation rate	Present Value	Year	Inflation rate	Present Value
1997	4.4	.590	2007	2.5	.813
1998	7.5	.638	2008	4.7	.851
1999	.8	.643	2009	2.8	.875
2000	2.3	.658	2010	3	.901
2001	4.1	.685	2011	4	.938
2002	2.8	.704	2012	2.2	.958
2003	3.5	.729	2013	1.3	.971
2004	3.6	.755	2014	1.3	.983
2005	2.8	.776	2015	0.7	.990
2006	2.2	.793	2016		1

타자 성적 지표

타자 성적 지표는 타율(AVG), 출루율(OBP), 장타율(SLG), 출장경기(G), 타석(PA), 타수(AB), 득점(R), 안타(H), 1루타(1B), 2루타(2B), 3루타(3B), 홈런(HR), 총루타수(TB), 타점(RBI), 도루(SB), 도루사(CS), 볼넷(BB), 사구(HBP), 고의사구(IBB), 삼진(SO), 병살타(GDP), 희생타(SH), 희생플라이(SF)로 나누어 데이터를 〈Table 4〉와 같이 반영하였다.

세이버메트릭스 관련 지표

본 연구에서는 한국프로야구 선수 중 타자를 중심으로 조사를 실시하였으며 제공된 자료를 기반으로 타자의 경기력을 설명할 수 있는 세이버메트릭스 지수로 분석을 시도하였다. 수많은 세이버메트릭스 지표들이 있으나 Yang, Cho, Bae, & Jung(2015)의 연구에서 전문가의 검증을 거쳐 선별해낸 13개의 세이버메트릭스 지표 중 제공된 원시데이터로 환산이 가능한 12개 지표를 연구에 활용하였으며 지표의 통계적 특징은 아

Table 4. Hitter Statistics

Statistic	Max.	Min.	Average	Median	SD	Skewness	Kurtosis	N
AVG	.185	.376	.291	.290	.029	.078	.074	793
OBP	.245	.478	.370	.368	.037	.198	.080	793
SLG	.259	.739	.444	.439	.080	.589	.341	793
G	93	144	122.80	124.00	8.913	-.410	.350	793
PA	390	672	495.68	493.00	57.696	.267	-.574	793
AB	327	600	428.48	427.00	50.786	.321	-.400	793
R	26	135	66.36	64.00	18.617	.522	.175	793
H	65	201	125.01	123	22.676	.426	-.030	793
1B	36	174	87.47	86.00	18.070	.543	.527	793
2B	5	46	22.07	22.00	6.638	.415	.054	793
3B	0	17	2.13	2.00	2.219	1.793	4.933	793
HR	0	56	13.34	12.00	9.548	1.067	1.528	793
TB	91	377	191.36	186.00	46.626	.736	0.569	793
RBI	13	146	63.62	62.00	23.123	.668	0.388	793
SB	0	66	12.02	7.00	12.491	1.667	2.670	793
CS	0	21	5.19	4.00	3.977	1.079	1.094	793
BB	12	124	49.88	47.00	17.337	.789	.681	793
HBP	0	31	7.06	6.00	4.667	1.243	2.209	793
IBB	0	27	2.41	2.00	2.989	3.103	16.441	793
SO	27	161	70.48	68.00	21.359	.739	.471	793
GDP	1	23	9.75	9.00	4.118	.447	-.122	793
SH	0	36	6.21	5.00	6.283	1.325	1.916	793
SF	0	16	4.01	4.00	2.380	1.015	1.697	793

Table 5. Sabermertics statistics

Statistic	Max.	Min.	Average	Median	SD	Skewness	Kurtosis	N
OPS	1.20	.50	.81	.81	.11	.49	.25	793
GPA	.40	.178	.28	.28	.03	.43	.19	793
SECA	.64	.10	.28	.27	.08	.81	1.00	793
TA	1.46	.42	.82	.80	.16	.72	.68	793
RC	174.52	22.98	74.02	70.1	23.42	.92	1.16	793
RC/27	13.90	1.96	6.10	5.77	1.88	.93	1.10	793
XR	148.75	21.34	70.94	69.00	19.91	.67	.58	793
ISO	.41	.02	.15	.15	.07	.71	.52	793
PSN	35.34	.00	7.80	6.72	5.78	1.20	1.89	793
wOBA	.66	.26	.42	.42	.06	.57	.47	793
OW%	.89	.16	.59	.59	.13	-.18	-.28	793
BABIP	.41	.21	.32	.32	.03	-.01	.13	793

래 〈Table 5〉와 같다. WAR의 경우 선수의 가치를 승수로 환산하여 확인할 수 있다는 점에서 평가지표로 활용되고 있으며 이와 유사한 지수를 개발하는 방향으로 연구(Lee & Kim, 2016; Oh & Lee, 2016)들이 진행되고 있다.

자료처리방법

독립변수 설정

본 연구는 타자의 객관적인 현재가치를 파악하기 위

해 성적데이터를 포함하여 가장 설명력 높은 타자가치평가 모델 개발을 목적으로 하였으며 이를 위해 성적 이외에 통상적인 선수 가치에 가장 큰 영향을 준다고 판단되는 FA 유무와 연차를 독립변수로 설정하였다.

FA 유무

한국야구위원회 규약에서는 FA(Free Agent) 에 대해 'FA 자격요건'을 갖추어 모든 구단과 선수계약을 체결할 수 있는 권리를 획득한 선수로 정의하고 있으며 권리를 획득한 선수는 권리 행사 여부를 스스로 결정할 수 있다고 밝히고 있다. 한국프로야구에서는 1999년 한국프로야구 선수협회의 창설과 사회적 요구에 따라 FA 제도가 도입되었으며 한국프로야구 규약에서는 구단에서 방출 당한 자유계약선수와의 구별을 위해 '자유계약선수'라고 정의하고 있다. 한국야구위원회 규약 162조에 따르면 선수 개인이 FA 자격을 취득하기 위해서는 고졸 선수는 9시즌, 대졸 선수는 소속구단 혹은 국내 다른 구단과 FA 계약을 할 시에는 8시즌이면 되지만 FA로 해외 팀과 계약을 할 경우는 9시즌을 채워야 한다고 규정하고 있다. 또한 등록일수는 타자의 경우 정규시즌의 2/3 이상 출전, 투수의 경우 규정 이닝의 2/3 이상을 투구하거나 1군 등록 기간이 145일을 넘긴 시즌을 시즌 충족 요건으로 제한하고 있다(KBO Regulation, 2017).

신인급 선수들의 경우 보류 조항에 묶여 낮은 연봉과 제한된 연봉 인상률로 충분히 보상 받지 못한 대가를 FA제도를 통해 보상받는다는 인식이 강하며 FA 자격 취득을 기점으로 선수들의 협상력 및 계약 규모가 큰 차이를 보인다고 판단된다. 일부 정상급 FA 선수의 경우, 제한된 공급 속에 구단들 간의 치열한 경쟁으로 인해 최근 선수들의 계약금액이 폭등하고 있는 상황이다. Kim (2007)은 과도한 선수 영입 경쟁이 FA 선수와 비 FA 선수 간의 연봉 산정에 있어 객관성을 저해시키고 몸값을 폭등 시킨 요인이라 비판하고 있으나 본 연구에서는 선수계약 제도의 본질적 문제 지적의 차원을 넘어 실질적인 선수가치평가모델 개발에 주안점을 두었기에 FA 유무 또한 선수가치평가에 있어 중요한 잣대가 된다고 판단하였다. 실제로 FA유무가 선수가치평가에 미치는 영향을 판단하기 위해 FA 자격 취득 선수는 '1', FA 자격 미취득 선수는 '0'으로 분류하여 통계적 검증을 시도한 결과 아래 FA 자격 취득 유무에 따라 그룹 간 유의미한 차이를 보인다는 것이 입증되었으며 FA 유무가 선수가치 평가에 큰 영향을 미치는 것으로 밝혀져 선수가치평가모델의 독립변수로 사용할 수 있음이 분석결과로 나타났다〔$t(793) = -14.94$ ($p<.05$)〕.

연차

앞서 FA 유무에서 밝힌 바와 같이 선수의 가치는 FA 취득 여부에 따라 크게 좌우되며 FA 자격은 연령이 아닌 데뷔 연차로 판별되기에 연차를 가치평가모델의 변수로 선정하는 것이 논리적으로 타당하다. 또한 연차와 연령은 .82의 상관관계를 보여 각각 설명 변수로 사용할 경우 왜곡된 결과가 나올 가능성이 높아 연령이 아닌 연차를 변수로 사용하였다. 연차는 일정 기준을 충족해야만 FA로 인정받을 수 있기에 본 연구에서 규정한 규정타석을 충족한 연구대상에 대해 연차에 따른 연차 별 선수 연봉의 차이를 살펴본 결과 연차에 따라 그룹 간 유의미한 차이를 보이고 있으며 연차가 선수 가치 평가 모델의 독립변수로 사용될 수 있음을 증명하였다, $F(22,770)=17.93$, $p<.001$.

세이버메트릭스 지표

소개된 데이터에서 볼 수 있듯이 선수의 가치를 평가하기 위한 근거자료로 활용되는 성적데이터는 통계적 가공을 거치지 않은 성적 지표와 통계적 가공을 거친 세이버메트릭스 지표가 존재한다. 성적 지표는 가공을 거치지 않은 데이터로 지표 간의 연관성이 적어 개별 지표를 통계적으로 활용하기 용이한 장점을 가지고 있으나 팀 득점 기여도, 주자 진루 능력과 같이 승리에 대한 기여도를 종합적으로 판단하기 어려운 단점을 가지고 있기에 선수평가의 지표로 사용하기에는 부족하다는 평가를 받고 있다. 이에 반해 세이버메트릭스 지표는 다양한 요소를 종합적으로 반영하여 선수의 공격력을 좀 더 객관적으로 판단할 수 있다는 장점이 있지만 세이버메트릭스 지표 환산에 중복적으로 활용되는 변수들로 인해 지표들 간에 다중공선성(Multicollinearity) 문제가 발생할 수 있다는 단점이 존재한다. 회귀분석에서 사용된 입력변수들 간의 상관관계가 존재할 경우 예측 변수들 간의 상관도로 인해 분산을 증가시켜 회귀분석 계수의 신뢰도를 하락시키는 문제점을 야기할 수 있다.

Table 6. Correlation between Sabermetrics Statistics

	OPS	GPA	SECA	TA	RC	RC/27	XR	ISO	PSN	wOBA	%OW	BABIP
OPS	1	.99**	.84**	.97*	.91*	.97**	.90**	.86**	.29**	.94**	.91**	.502**
GPA		1	.83*	.97*	.91*	.98**	.90**	.81**	.28**	.96**	.92**	.55**
SECA			1	.90*	.78*	.83**	.78**	.87**	.37**	.89**	.76**	.18**
TA				1	.92*	.98**	.91**	.80**	.36**	.98**	.91**	.51**
RC					1	.93**	.99**	.74**	.36**	.88**	.84**	.52**
RC/27						1	.91**	.78**	.27**	.95**	.91**	.56**
XR							1	.74**	.38**	.88**	.83**	.49**
ISO								1	.30**	.75**	.70**	.10**
PSN									1	.38**	.27**	.14**
wOBA										1	.90**	.49**
%OW											1	.52**
BABIP												1

*p<.05, **p<.01

〈Table 6〉에서도 볼 수 있듯이 세이버메트릭스 지표들 중에는 매우 높은 상관관계를 보인 지표들이 다수 존재하고 있으며 이러한 지표들을 개발 모델의 독립변수로 그대로 사용할 경우 설명력이 저하된다. 이와 같은 문제점은 세이버메트릭스 지표를 활용한 유사연구에서 동일하게 지적되어 온 것으로 Lee(2014)는 타자능력 측정 지수를 개발하기 위해 상관계수가 큰 세이버메트릭스 지표들을 이용하여 변수의 개수를 축약하고 설명력 높은 모델을 만들기 위해 주성분회귀분석을 활용하였으며 이를 통해 우수 타자 지표를 개발하였다. 또한 Lee & Kim(2016)은 세이버메트릭스 지표인 WAR(Wins Above Replacement)을 중심으로 유사 타격 능력 지수(index)를 제안하고자 연구를 하였고 세이버메트릭스 지표인 OPS, GPA, wOBA 간에 매우 높은 상관관계를 보여 본 연구의 데이터와 유사한 경향을 보였다. 이러한 문제점을 해결하기 위해 이들은 세이버메트릭스 지표들의 가중평균을 계산하여 동일 선수의 WAR 지표와 비교하였으며 데이터 개수에 비해 변수의 개수가 많아 변수들 간의 다중공선성이 발생할 가능성이 높다는 문제점을 해결하기 위해 주성분분석을 활용하여 변수의 차원을 축소하고 설명력이 높은 성분으로 변수를 축약하는 방법을 시도하였다. 또한 Seung & Kang(2012)은 표준화계수가 크다고 판단되는 세이버메트릭스 지표 6개를 선정하여 주성분회귀분석을 통해 연봉과 성적 간의 관계를 설명하고 구단별 분석을 실시하였다.

주성분 회귀분석

12개의 세이버메트릭스 변수를 모두 사용하여 성적 지표를 만들 경우 데이터 간의 측정 단위 및 크기가 차이를 보이기 때문에 변수를 표준화하는 작업을 선행하였다. 세이버메트릭스 변수는 설명 변수들 간의 높은 상관관계로 인해 다중공선성(Multicollinearity) 문제를 야기할 수 있어 변수의 차원을 축소하고 설명력을 높일 수 있는 방법이 필요하다. 이에 주성분 회귀분석을 통해 주성분 변수를 얻어 이를 설명 변수로 이용함으로써 다중공선성 문제를 해결하였다(Bae et al., 2012; Lee, 2014; Lee & Kim, 2016).

주성분 개수를 선택할 때, 상관계수행렬을 이용하여 일반적으로 고유치 값이 1 이상인 주성분과 총 변동의 설명력이 80% 이상인 주성분 변수를 선택할 수 있다(Jolliffe, 1986). 성분 부하 값이 크다는 것은 그에 대응하는 원 변수의 영향이 크다는 것을 의미하므로 성분 부하 값이 큰 변수를 파악하여 주성분의 이름을 부여하고 아래의 〈Equation #1〉을 통해 주성분 점수를 구하게 된다.

$$y_{rj} = \underline{e_j}' \frac{(\underline{x_r} - \underline{\mu})}{\underline{\delta}} \quad \text{(Equation \#1)}$$

y_{rj}는 r번째 개체의 j번째 주성분 점수를 뜻하며 $\underline{e_j}'$는 j번째 주성분의 고유벡터를 뜻하며, $\underline{x_r}$는 r번째 개체의 측정치 벡터를 나타낸다. $\underline{\mu}$는 모평균 벡터이고, $\underline{\delta}$는 모표준편차 벡터를 나타낸다. 주성분 변수가 회귀분석의 설명변수 측정치가 되는 회귀모형은 아래와 같다(Sung, 1998).

$$y_i = \beta_0 + \beta_1 PRIN_1 + \beta_2 PRIN_2 + \\ \ldots\ldots + \beta_p PRIN_p + \epsilon_p \quad i = 1, \ldots, k$$

연구결과 및 논의

연구절차 및 자료처리 결과

본 연구의 연구대상은 1997년부터 2016년까지 20년간 한국프로야구(KBO리그)에서 활동한 대한민국 국적 타자 중 규정타석을 충족한 793건이며 연봉, 성적데이터, 연차, FA 유무를 분석 대상으로 삼았다. 개별 대상의 성적데이터는 EXCEL 2016을 통해 데이터 정제작업을 거쳤으며 2016년을 기준으로 물가상승률을 적용하여 연봉을 균등화하는 작업을 거쳤다. 다음으로는 IBM SPSS Statistics 21을 활용하여 기술통계, 상관분석, 주성분분석 등을 실시하였다. SPSS의 주성분 분석을 통해 도출된 주성분 변수를 포함하여 IBM Modeler 17의 분석에 필요한 자료를 만들었으며 이를 활용하여 선형회귀모형을 도출하였다. 마지막으로 본 회귀분석에 사용된 관측치들간의 독립성 가정을 검증하였다. 본 연구분석에 사용된 Durbin-Watson값은 2.036이 나왔으며 dL값은 1.621이고, dU값은 1.919로 나타났다 (Savin & White, 1977). Brook (2002)의 저서에서 제시한 Durbin-Watson 통계치를 이해하는 방법에 따르면, 본 연구에서 나타난 Durbin-Watson값(=2.036)은 dL값이 1.919와 4-dU값인 2.018 사이에 있고 이러한 경우에 자기 상관이 없다는 가정을 기각할 수 없기 때문에 본 OLS 분석에 대한 정당성을 제공하였다.

주성분 회귀분석 결과

주성분 회귀분석을 활용하기 위하여 사용될 데이터가 요인분석의 기본 가정을 충족하는지 검증하기 위해 KMO(Kaiser-Meyer-Olkin)와 Bartlett 검정을 사용하였다. 표본적합도 KMO 값은 .77로 권장치인 .5를 상회했으며, Bartlett의 구형성 검정은 유의확률 .001로 나타나 요인분석에 적합함을 확인할 수 있었다, KMO(=.77)〉.5, p(Bartlett)=.001.

다음으로 12개의 세이버메트릭스 지표에 대해 주성분 분석을 실시하였으며 고유값 중 1보다 큰 성분 1과 2에 대응하는 주성분이 전체의 85.57%를 설명하고 있어 두개의 주성분이 존재한다고 판단할 수 있으며 이러한 추론은 해당 스크리 플롯(Scree Plot)인 〈Fig. 1〉에서도 동일하게 나타났다. 이를 통해 나타난 주성분 행렬 및 성분점수는 〈Table 7〉과 같이 나타났다.

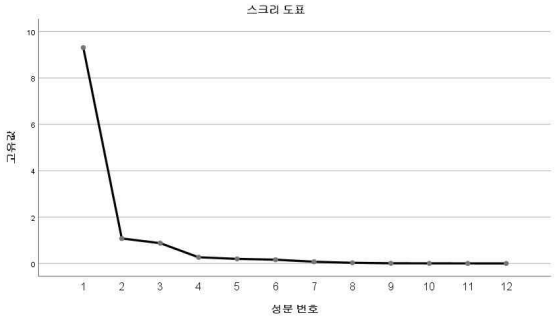


Fig 1. Scree Plot

SPSS를 이용하여 요인을 생성한 이후 각각의 요인점수를 산출하였다. 본 연구에 필요한 2개 주성분의 표준화된 변수식은 〈Equation #2〉과 〈Equation #3〉과 같다.

Prin1 = .106 OPS + .106 GPA + .095 SECA + … + .099 %OW + .055 BABIP (Equation #2)

Prin2 = .012 OPS + .073 GPA - .334 SECA …+ .102 %OW + .727 BABIP (Equation #3)

Table 7. Principal component analysis component matrix and component score coefficient matrix

Component matrix			Component score coefficient matrix		
	Component			Component	
	1	2		1	2
OPS	.98	.01	OPS	.11	.01
GPA	.98	.08	GPA	.11	.07
SECA	.88	-.36	SECA	.10	-.33
TA	.99	-.00	TA	.11	-.00
RC	.95	.06	RC	.10	.06
RC/27	.98	.10	RC/27	.11	.10
XR	.94	.03	XR	.10	.03
ISO	.83	-.41	ISO	.10	-.38
PSN	.37	-.36	PSN	.04	-.34
wOBA	.97	.02	wOBA	.10	.01
%OW	.92	.11	%OW	.10	.10
BABIP	.52	.78	BABIP	.06	.73

모델평가

모델 검증

선형회귀분석 모델링을 통해 도출된 선수가치평가 모델의 결과를 확인해본 종속변수인 선수 연봉과 선수 연차, FA 유무, Prin1, Prin2 간에 R은 .803의 값을 보여 강한 양의 상관관계를 보이는 것을 확인할 수 있었으며, R^2은 .645, 수정된 R^2은 .643의 값을 보여 양호한 설명력을 보인다는 것을 확인할 수 있었다〔F(4, 786)=357.433, p<.001.〕. 통계적 검증을 통해 도출해낸 선수가치평가 모델의 변수 및 계수의 결과는 〈Table 8〉과 같다. 모형의 계수들은 모두 유의수준 이하의 유의확률을 보여 통계적으로 유의미한 계수인 것을 확인할 수 있었다.

Table 8. model coefficients and statistical results

Variable	B	β	t
(constant)	-64.104		-16.549***
Year	.034	.404	17.562***
FA	.716	.563	25.728***
Prin1	.102	.208	9.745***
Prin2	-.023	-.046	-2.062*
R^2	.645		
F	357.433***		

***p<.001, *p<.05

모델 분석

생성 계수 분석

주성분분석을 통해 생성된 제1주성분(Prin1)과 제2주성분(Prin2)에 대해 살펴본 결과, 〈Equation #4〉와 같이 구성된 Prin1의 경우 장타력, 출루율, 득점생산력 등 다양한 종합적인 타격능력이 고르게 반영된 강타자의 면모를 반영하고 있는 지표라 판단된다. 또한 상대적으로 호타준족의 정도를 나타내는 PSN과 인플레이 타구 비율을 나타내는 BABIP가 상대적으로 적게 관련된 것을 볼 수 있는데 통상적으로 BABIP는 타자의 타격스타일에 따라 다른 값을 가지게 되며 라인드라이브 〉 그라운드볼 〉 플라이볼 순으로 BABIP가 높게 형성된다는 점을 고려할 때 장타 위주의 선수들이 Prin1에 높은 수치를 보일 가능성이 높다. 이에 Prin1을 강타자 지수(Power Hitter Index, PHI)로 명하였다.

〈Equation #5〉와 같이 구성된 Prin2의 경우 SECA, ISO, PSN이 음의 관계를 가지며 BABIP가 상대적으로 높은 기여도를 가진다는 점이 특징적이다. SECA, ISO, PSN은 모두 장타율이 반영된 세이버메트릭스 지표로서 Prin2가 타자의 장타력과 반하는 결과를 보인다는 것을 예측할 수 있다. 또한 BABIP가 높은 기여도를 보인다는 것은 라인드라이브나 그라운드볼 위주의 교타자들이 높은 수치를 보일 가능성이 크다. 이에 Prin2를 교타자 지수(Contact Hitter Index, CHI)로 명하였다. 이를 종합하여 〈Equation #6〉와 같은 모델을 도출할 수 있었다.

강타자 지수 (PHI) = .106 OPS + .106 GPA + .095 SECA + … + .099 %OW + .055 BABIP 〈Equation #4〉

교타자 지수 (CHI) = .012 OPS + .073 GPA - .334 SECA + … + .102 %OW + .727 BABIP 〈Equation #5〉

선수연봉 = 연차 * 921.5 + FA 유무 * 53528.9 + PHI * 7313.0 + CHI * 2197.3 + 5893.6 (단위: 만원) 〈Equation #6〉

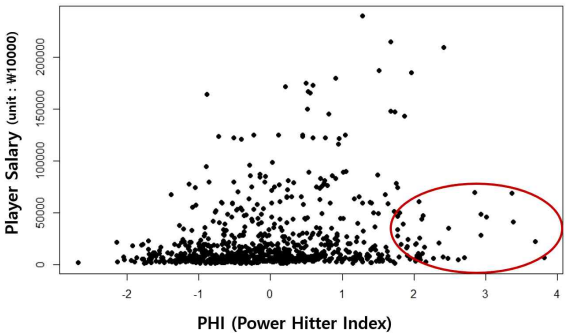


Fig 2. PHI and Player Salary Graph

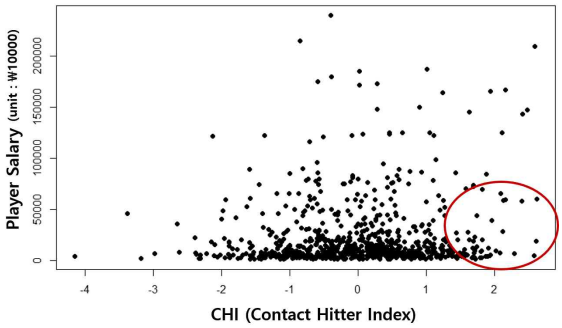


Fig 3. CHI and Player Salary Graph

Table 9. POS rank in PHI

Rank	POS_PHI	
1	Yang, Jun-Hyuk (1999)	1.22
2	Park, Jae-Hong (1997)	.73
3	Jang, Seong-Ho (2001)	.68
4	Kim, Tae-gyun (2003)	.67
5	Choi, Ik-Seong (1997)	.64
6	Shin, Dong-Joo (1997)	.61
7	Lee, Seung-Yeop (1999)	.54
8	Ma, Hae-Young (1999)	.54
9	Lee, Byung-Gyu (1999)	.52
10	Koo, Ja-Wook (2015)	.48
11	Jang, Seong-Ho (1999)	.47
12	Kim, Jae-Hwan (2016)	.42
13	Shim, Jeong-Soo (1999)	.42
14	Kim, Hyun-Soo (2008)	.41
15	Kim, Ki-Tae (1997)	.41

Table 10. POS rank in CHI

Rank	POS_CHI	
1	Kim, Soo-Yeon (2001)	1.25
2	Koo, Ja-Wook (2015)	.68
3	Park, Min-Woo (2014)	.65
4	Lee, Jong-Wook (2006)	.62
5	Lee, Byung-Kyu (2011)	.59
6	Lee, Dae-Hyung (2007)	.58
7	Kim, Jong-Ho (2013)	.57
8	Jang, Won-Jin (1999)	.56
9	Park, Yng-Taek (2002)	.55
10	Lee, Hyun-Gon (2007)	.53
11	Jeong, Bo-Myung (2007)	.50
12	Kim, Won-Seop (2008)	.47
13	Jeong, Su-Seong (2005)	.46
14	Shin, Jong-Gil (2013)	.43
15	Kim, Hyun-Soo (2008)	.43

결과 분석

강타자 지수(PHI), 교타자 지수(CHI)와 선수연봉 간의 관계를 각각 살펴본 결과 선수연봉과 강타자 지수 간에 보이는 결과는 〈Fig. 2〉와 같았으며 분포 추세선을 따라 우상향 위치에 속한 그룹의 경우 리그 내 우수한 성적을 남긴 강타자들이 주로 위치하고 있는 것을 확인할 수 있었다. 또한 〈Fig. 3〉을 통해서도 변수 정의 단계에서 규정한 교타자의 개념이 분포 추세선을 따라 우상향한 분포에서 강하게 나타나 교타자 지수로서 적합한 것을 확인할 수 있었다.

추세선을 따라 높은 PHI, CHI 지표를 보인 선수들은 우수한 성적 이외에도 저연봉으로 우수한 성적으로 거두었다는 측면에서 논의할 가치가 있다. 이들의 성과를 측정하기 위해 연봉을 PHI, CHI 값으로 나누어 연봉대비 성적(Performance on Salary, POS)의 개념으로 지수화 하여 이들의 성과를 알아본 결과 다음 〈Table 9, 10〉과 같은 결과를 얻을 수 있었다. POS 지표로 환산하여 높은 결과값이 나온 선수들은 PHI의 경우 대부분 위의 〈Fig. 2〉에서 표시된 영역에 위치하는 경우들로 비 FA 자격에 신인급 연차를 가진 선수들이 주류를 이루고 있었다. CHI의 경우에도 마찬가지로 비 FA 자격의 선수들이 상위권 대부분을 차지하고 있었으며 위의 〈Fig. 3〉

에 표시된 영역에 위치하고 있는 것을 확인할 수 있었다.

결론 및 제언

본 연구에서는 한국프로야구의 선수 노동시장 조사를 통해 선수의 가치평가를 위한 정량적 연구를 실시하였다. 선수 가치의 금전적 환산을 통한 직관적인 모델 개발을 목표로 함과 동시에 단순화된 변수를 활용한 선수가치평가 모델 개발을 목표로 하였다. 이를 위해 선수성적 변수 이외에 연봉에 결정적인 영향을 미친다고 판단되는 연차와 FA 유무를 독립변수로 추가하여 모델링을 실시하였다. 1997년부터 2016년까지 한국프로야구(KBO 리그) 내에서 규정타석에 진입하고 차기연도까지 계약이 유지된 타자들을 대상으로 가치평가 모델링을 실시하였다. 선수성적을 세이버메트릭스 지표로 환산하여 선수 성적 기록의 객관성을 향상 시켰으며 선수연봉과의 상관관계를 알아보았다. 주성분 회귀분석 결과 도출된 두개의 주성분을 검토한 결과 제1주성분(Prin1)은 타자의 전반적인 공격력을 종합적으로 반영함과 동시에 장타 위주의 타격을 반영하는바 강타자 지수(Power Hitter Index, PHI)로, 제2주성분(Prin2)은 장타와 음의 상관관계를 가짐과 동시에 단타 위주의 타격 성향을 잘 반영하는바 교타자 지수(Contact Hitter Index)로 명하였다.

본 연구를 통해 개발된 선수가치평가 모델의 수정된 R^2은 64.5%로 높은 설명력을 가진 것으로 드러났으며 지난 20년간의 연구대상들의 실제 연봉과 추정 연봉 간의 비교를 실시한 결과 최근 우려되고 있는 선수 연봉 증가 추세와 과다 책정의 문제점을 발견할 수 있었다. 성적지표로 사용된 강타자 지수(PHI)와 교타자 지수(CHI)를 통해서도 선수의 가치를 간접적으로 확인할 수 있었는데 PHI가 2를 상회하는 경우 대부분의 선수들이 해외진출 및 소위 FA 대박을 터트리며 금전적인 보상으로 이어지는 경우들이 대다수인 것이 발견되었다. CHI의 경우에도 2에 가까운 수치를 보이는 선수들의 경우 PHI와도 차별화된 방식으로 인정받으며 FA 계약에서 좋은 결과를 얻는 경우들이 다수 보였다. 선수들의 성적을 연봉을 기준으로 비교해보기 위해 POS (Performance On Salary)라는 지표를 생성하여 비교해본 결과 PHI의 경우 FA 자격을 갖추기 이전의 신인급 선수들이 우수한 효율(연봉대비 성적)을 지녔던 것으로 나타났으며, CHI의 경우에도 FA 자격 이전의 선수들이 효율성 면에서 상위권의 대부분을 차지하고 있는 것을 확인할 수 있었다. 이는 한국야구위원회(KBO 리그) 내에서 FA 자격 취득 이전의 선수들이 저평가 받고 있는 현상이 반영된 것으로 연봉 협상만으로는 선수들의 가치가 제대로 반영되지 못하는 경우가 많다는 것을 알 수 있다. 이는 FA 자격 유무가 선수의 가치평가에 큰 영향을 주는 것으로 본 연구의 결과물인 가치평가모델에서도 동일한 현상을 발견할 수 있었다. 국내의 경우 선수계약에 있어 인센티브 조항과 옵션 계약과 같이 성과 반영 계약이 활발하지 않은 상황이다. FA 계약 대상자들에 대한 합리적인 계약과 더불어 FA 자격 취득 이전 선수들에 대한 다양한 보상 장치 개발을 통해 보다 현실적인 선수 계약이 필요하다고 판단된다.

실제 프로야구 선수들의 계약은 당사자들 간의 협상력, 정보력, 시장 상황 등 다양하고 복잡한 상황 속에서 이루어지기에 단순한 모델로 선수의 가치를 예측하는 데는 무리가 따를 수 있다. 합리적인 구단 경영을 추구하는 목소리가 커지고 정량적 선수가치평가에 대한 사회의 요구가 점점 증가하고 있다. 이러한 상황에서 본 연구에서 제시된 선수가치평가 모델은 현재 반복되고 있는 비FA 선수들의 저연봉 및 일부 FA 선수들의 몸값 거품 논란에 대해 학계의 관점에서 새로운 근거자료를 제공한다는 측면에서 큰 의미가 있다. 또한 개발된 두 개의 성적 지표를 통해 선수의 가치를 사전에 재평가 하고 직관적인 예측을 통해 선수 영입 및 재계약에 도움을 줄 수 있다는 측면에서 의의가 있을 것으로 기대된다.

참고문헌

Bae, J. Y., Lee, J., M., & Lee, J. Y. (2012). Predicting Korea Pro-Baseball Rankings by Principal Component Regression Analysis, *Communications of the Korean Statistical Society*, 19(3), 367-379.

Bank of Korea. (1 January, 2018). Retrived from http://bok.or.kr.

Brooks, C. (2002). *Introductory Econometrics for Finance*. Cambridge University Press.

Choi, K. H. (2009). The introduction of Sabermetrics and Record Analysis in Korea, *Journal of Social Science, 25*(1), 129-139.

Depken, C. A., & Wilson, D. P. (2004). Labor markets in the classroom : Marginal product in Major League Baseball, *Journal of Economics and Finance Education, 3*, 12-24.

Fields, B. (2001). *Estimating the value of Major League Baseball players.* Unpublished master's thesis, East Carolina University, Greenville, NC, USA.

Jolliffe, I. T. (1986). Principal component analysis and factor analysis. *In Principal component analysis* (pp. 115-128), New York: Springer.

KBO Press Release. (2018). Announcement of registration of KBO league players in 2018, February 19, KBO, https://www.koreabaseball.com.

Kim, E. S. (2001). The Relation of Game Performance and Annual Salary for Korean Professional Baseball Players, *Journal of Korean Sociology of Sport, 14*(1), 15-24.

Kim, S. H. (2007). *The current programs and improvement measures on free agent in Korea professional baseball league.* Unpublished master's thesis, Dong-Eui University, Busan.

Krautmann, A. C. (1999). What's wrong with Scully estimates of a player's marginal revenue product, *Economic Inquiry, 37*(2), 369-381.

Lee, M. G. (2001). *Relationship between Performance Ability of Professional Baseball Batters and Annual Salary based on Sabermetrics.* Unpublished master's thesis, Kookmin University, Seoul.

Lee, J. T. (2014). Measurements for hitting ability in the Korean pro-baseball, *Journal of the Korean Data & Information Science Society, 25*(2), 349-356.

Lee, J. Y. (2001). The Relationship between Annual Salary and Ability of Korean Professional Baseball Batters. *Journal of Korean Sociology of Sport, 16*, 973-981.

Lee, J. Y., & Kim, H. G. (2016). Suggestion of batter ability index in Korea baseball - focusing on the sabermetrics statistics WAR, *Korean Journal of Applied Statistics, 29*(7), 1271-1281.

Myung, W. S., Won, Y. S., & Lee, M. G. (2016). The Study on the Determinants of Korean Professional Baseball Players' Salaries Using Decision Tree Analysis. *Journal of Sport and Leisure Studies, 65,* 63-77.

Oh, T. Y., Lee, & Y. H., (2013). Measurement of Monopsony Exploitation in Salary Determination; Case of Major League Baseball. *Korean Journal of Sport Management, 18*(3), 1-15.

Oh, T. Y., Lee, & Y. H., (2016). Value Evaluation Model for Korean Professional Baseball Players. *Korean Labor economic Association, 39*(2), 113-139.

Park, S. R. (2015). Suggestion for evaluation of market value of professional baseball players. *Sports Industry Issue Paper.* Seoul: Korea Institute of Sports Science.

Park, S. R. (2017). *Sports Business Insight* (1st ed.). Seoul: People and Thought.

Park, S. R., Lee, W. Y., & Jeon, H. K. (2016). Analysis of Critical Factors affecting Market Value of Baseball Players (Hitters) in Korea. *Journal of Sport and Leisure Studies, 66*, 55-65.

Park. S. H. (2008). Performance Factors Affecting the High Annual Salary of Korean Professional Baseball Batters. *Korean Journal of Sports Science, 17*(2), 485-494.

Savin, E., & White, K. (1997). The Durbin-Watson Test for Seria Correlation with Extreme Sample Sizes or Many Regression. *Econometrics, 45*(8), 1989-1996.

Scully, G. W. (1974). Pay and performance in major league baseball, *The American Economic Review, 64*(6), 915-930.

Seung, H. B., & Kang, K. H. (2012). A study on relationship between the performance of professional baseball players and annual salary. *Journal of the Korean Data & Information Science Society, 23*(2), 285-298.

Sommers, P. M., & Quinton, N. (1982). Pay and performance in major league baseball: Thecase of the first family of free agents. *Journal of Human Resources, 17*(3), 426-436.

Statistics Korea, (2018, January 1). Retrieved from http://kostat.go.kr.

Sung, H. H. (2014), *Stove League final 'Money Wars' negotiations* on December 7, Hankook News from http://www.hankookilbo.com/v_print.aspx?id=2c71215af1bb47bf9b45cd85153952ed.

Sung, W. H. (1998). Applied Multivariate Analysis, Seoul: Tamjin.

Tango, T. M., Lichtman, M. G., & Dolphin, A. E. (2007). *The book: Playing the percentages in baseball*. Potomac Books, Inc..

Yang, D. E., Cho, E. H., Bae, S. W., & Jung. S. W. (2015). Analysis of Professional Korean Baseball Batter's Performances Factors. *Journal of Sport and Leisure Studies, 60*, 305-313.

한국프로야구 선수들의 연봉 산정 모델 개발

박성배 · 권태근 · 전종환(한양대학교)

〔목적〕 이 연구의 목적은 크게 다음과 같은 두 가지로 나눌 수 있다. 첫째, 한국프로야구 선수들(타자들)의 성적에 따른 시장가치를 결정할 수 있는 모델을 개발하는 것이며 둘째, 이 모델을 기반으로 하여 선수들의 시장 가치에 대한 객관적인 근거 자료를 제공하는 것이다. **〔방법〕** 1997년부터 2016년까지 활약한 선수 중 각 해당연도 규정에 따라 최소 타석 기준을 충족한 793명의 자료를 한국야구위원회위원회가 제공하는 연감의 데이터를 바탕으로 재구성 및 편집 과정을 거친 후 통계 분석을 하였다. 12개의 세이버메트릭스 지수들의 다중공선성을 제거하고 동시에 모델의 설명력을 높이기 위해 주성분 분석을 실행하여 강타자 지수 (PHI)와 교타자 지수 (CHI)를 얻었다. **〔결과〕** 선수들의 연봉에 영향을 미치는 경력(연수)과 FA의 유무를 독립변수로 하여 선형회귀분석을 하여 다음과 같은 모델을 얻었다. 선수연봉 = 연차*921.5 + FA유무*53528.9 + 강타자 지수*7313 + 교타자 지수*2197.3 + 5893 (단위: 만원). 제안된 모델은 한국프로야구 내 선수가치의 64.5%를 설명하는 것으로 나타났으며 통계적 유효성이 검증되었다. **〔결론〕** 본 연구를 통해 새롭게 개발된 모델을 통해 선수의 가치평가에 영향을 주는 변수들을 밝혀낼 수 있었으며 이러한 결과가 시장에 적용된다면 프로야구선수 가치평가 시장에 큰 공헌을 할 수 있을 것으로 기대한다.

주요어: 연봉, 가치평가, 세이버메트릭스, 주성분 분석