

勞 動 經 濟 論 集
 第39卷 第2號, 2016. 6. pp.113~139
 © 韓 國 勞 動 經 濟 學 會

한국프로야구선수의 가치평가모형

오 태 연* · 이 영 훈**

본 연구는 한국프로야구리그(KBO)에서 활동하는 개별선수의 승리에 대한 기여도, 즉 생산성을 추정하는 모형의 구축을 목적으로 한다. 기존의 지표인 대체선수대비승수(WAR)의 단점인 계산방법의 복잡성을 해결하기 위하여 KBO 공식 홈페이지에서 제공하는 기록만을 이용한 계량모형을 구축함으로써 새로운 시즌의 기록만 대입하면 생산성을 생성할 수 있게 하였다. 또한, Scully(1974)와 Krautman(1999) 등 미국 메이저리그 선수의 가치추정모형이 소수의 기록에만 의존하였던 점을 개선하여 KBO가 발표하는 61개 기록을 모두 사용하여 현실에 적용할 수 있는 모형을 구축하였다. 2002-14년 팀기록의 패널 자료를 이용하여 득점 및 실점방정식을 추정하여 회귀분석에 기초한 대체선수대비승수를 추산하였다. 추정 결과, WAR과 상관계수가 타자는 0.869, 투수는 0.882로 나타나 두 지표가 비교적 유사한 것으로 나타났다. 추정 결과를 근거로 2014년 시즌 후 FA선수들의 계약 내용을 평가하면 비교적 생산성 추정치와 계약액수의 상관관계가 높아서 구단이 FA계약에서 합리적인 판단을 하고 있는 것으로 나타났다.

주제어: 한국프로야구, 선수노동시장, 한계생산, 대체선수대비승수, 패널 자료

논문 접수일: 2016년 1월 8일, 논문 수정일: 2016년 6월 13일, 논문 게재확정일: 2016년 6월 17일

* 서울대학교 체육교육과 박사과정(windfrog@snu.ac.kr)

** (교신저자) 서강대학교 경제학과 교수(yhnlee@sogang.ac.kr)

I. 서론

프로야구선수의 연봉은 학문적으로뿐만 아니라 현실적으로도 많은 관심을 받는 주제이다. 노동경제학적 관점에서는 선수의 가치를 추정하여 실제 연봉과 비교함으로써 선수노동시장의 차이에 따라 연봉이 어떻게 결정되는지를 비교할 수 있다. 실무적으로는 각 프로야구단이 매년 선수들과 연봉 협상, 특히 자유계약선수(free agency: FA)들과 연봉 협상 시 선수의 가치를 정확히 아는 것이 중요하다. 이에 프로야구 선수들의 가치 평가 및 연봉에 관한 다양한 연구가 진행되어 왔다(Scully, 1974; Zimbalist, 1992; Krautmann, 1999; Rockerbie, 2010; 오태연·이영훈, 2013).

이들 연구는 방법론 측면에서 두 가지로 구분할 수 있다. 첫째는 Scully(1974)가 제안한 방법으로 노동시장이 완전경쟁이면 균형임금은 노동자의 한계수익생산(Marginal Revenue Product)과 같다는 점에 기초를 두고 있다. 따라서 선수의 가치인 한계수익생산을 추정하기 위하여 수익방정식과 승리생산함수식을 추정하여 개별선수의 한계수익생산을 도출하는 방법을 제시하였다. 이 방법론은 총 3단계로 구성되어 있는데, 1단계에서 수익방정식을 추정함으로써 팀성적(승률)의 한계수익을 계산하고, 2단계에서 승리생산함수식을 추정하여 각 팀기록의 한계생산을 추정한다. 마지막으로 3단계에서 개별선수의 기록이 팀기록에 미친 기여도를 산출하여 1, 2단계에서 도출한 한계생산 및 한계수익의 추정치에 적용하여 개별선수의 한계수익생산을 산출한다. 즉 개별선수들이 팀기록에 미친 기여도를 계산하고, 팀기록이 팀승률에 미치는 한계생산을 추정하며, 팀승률이 팀수익에 미치는 영향을 추정하여 1승의 가치를 추계함으로써 개별선수가 팀수익에 미친 영향을 산출하는 방법이다.

둘째는 Krautmann(1999)의 자유시장접근법(Free Market Return Approach)이다. 이는 FA시장은 완전경쟁시장이라는 가정에서 출발한다. 완전경쟁시장에서는 연봉이 선수의 가치와 동일하다는 이론을 이용하여 FA선수들의 연봉 자료를 이용하여 연봉 방정식을 추정하고 추정 결과를 바탕으로 다른 선수들의 적정 연봉을 산출한다. 즉 FA선수의 연봉을 종속변수로, 공격 및 수비 기록을 독립변수로 설정하여 연봉 방정식을 추정하고 FA자격이 없는 다른 선수들의 기록을 대입하여 적정 연봉을 추계한다. 그러나 FA노동

시장이 완전경쟁이라는 가정이 위배된다면 자유시장접근법은 적정 연봉을 산출하는 데 적합하지 않다. KBO(Korea Baseball Organization)의 경우 매년 FA선수가 다수 있으나) 수비 위치별로 분류하면 소수에 불과하고, 각 구단이 FA선수 영입을 통해 보강하려는 선수의 포지션도 서로 다르기 때문에 다수의 수요자와 공급자가 존재하는 완전경쟁시장으로 가정하기 어렵다. 따라서 이 방법은 KBO 선수의 가치를 평가하는 데 적합하지 않다.

후속 연구들은 위에서 언급한 Scully(1974), 또는 Krautmann(1999)의 방법을 보완한 연구들이다. 오태연·이영훈(2013)은 운회귀모형(Spline Regression Model)을 설정하여 수익방정식을 추정함으로써 1승의 가치를 동일하게 가정하는 선형회귀모형의 단점을 보완하였다. 예를 들어, 포스트시즌 진출을 다투는 팀에 1승과 하위권에 머물러 포스트시즌 진출이 어려운 팀에 1승의 가치가 다름을 모형에 적용하였다. 구체적으로 메이저리그(MLB) 팀의 패널 자료를 이용하여 승률구간별로 한계수익을 차등하게 추정하는 모형을 제시하였다. Rokerbie(2010)는 승률생산함수의 종속변수인 승률이 0에서 1 사이의 값을 갖는 제약된 종속변수(limited dependent variable)라는 점을 고려한 계량모형을 설정하였다. 즉 로지스틱 승률함수를 추정함으로써 선형승률방정식을 보완하였다.

그러나 위에서 언급한 두 가지 방법을 적용하여 한국프로야구(KBO) 선수들의 가치 평가 및 적정 연봉을 추정하기에는 한계가 있다. 첫째, 국내 현실상 Scully(1974) 모형의 1단계인 수익방정식의 추정이 어렵다. 이는 KBO에 속한 구단의 수익 자료가 충분히 축적되지 않았기 때문이다. 둘째, 대기업의 지원을 받는 KBO 구단은 이윤추구를 제1 목적으로 하는 MLB 구단과는 경영 목적이 다르기 때문에 Scully 모형을 그대로 적용하는 것이 적절치 못하다.²⁾ 셋째, 이미 언급한 대로 Krautmann(1999)의 가정인 완전경쟁 FA시장을 충족한다고 보기 어렵기 때문이다.

이에 본 연구에서는 개별선수의 승리에 대한 기여도를 추정하는 데 중점을 둔다. 즉 개별선수의 한계생산을 추정함으로써 궁극적으로는 선수의 가치 및 적정 연봉에 대한 연구의 기초를 다지고자 한다. 독립적인 경영을 하는 MLB 구단과 달리 KBO 구단은 넥센 히어로스를 제외하고는 대기업을 모기업으로 두고 있다. 따라서 경영 목적에서도 차이가 난다. MLB 구단이 이윤추구를 목적으로 한다면 KBO 구단은 모기업의 효용을 극대화한다고 볼 수 있다(Fort, Kang & Lee, 2015; Jang & Lee, 2016). MLB에서는 1승이 가져오는 한계수익이 연고 도시에 따라 큰 차이를 보이지만 KBO 구단의 모기업은

1) 2015년 시즌 후에는 FA가 23명이었다.

2) KBO 구단의 경영 목적에 관하여는 Jang and Lee(2016)를 참조.

모두 전국 소비자를 대상으로 하기 때문에 팀 성적이 가져오는 한계효용 또는 한계수입이 비교적 유사하다고 가정할 수 있다. 따라서 KBO에서 선수의 한계수익생산에 대한 정보는 한계생산에 모두 내포되어 있다. 선수의 한계생산 추정을 통한 가치평가 모형은 다음과 같은 측면에서 노동경제학 문헌에 기여할 것이다. 첫째, 본 모형을 통하여 도출한 선수의 생산성과 자유계약선수들의 연봉을 비교함으로써 자유계약선수들의 ‘몸값 과열’ 논쟁을 평가할 수 있다. 둘째, KBO 선수들은 경력에 따라 직면하는 선수노동시장이 다르다. 1-3년차 선수들은 보류 조항으로 인하여 수요독점시장에 있으며, 3-7년차 선수들은 보류 조항과 연봉조정제도로 인하여 완화된 수요독점시장에 있으며, 8년차 이후는 FA자격을 취득하면서 경쟁시장에 직면한다고 볼 수 있다. 따라서 본 모형을 통해서 추정한 선수 가치와 선수 연봉을 노동시장 구조와 연결하여 분석함으로써 연봉 결정에 노동시장 구조가 미치는 영향을 분석할 수 있다.

선수 성과에 대한 연구는 다양하게 진행되어 왔다(James, 2000; 오광모·이장택, 2003; Kang·Lee·Kwok, 2007; 남영식·박호정·장재열, 2012; 이장택, 2014). 이장택(2014)은 주성분분석(Principal Component Analysis)을 통해 KBO 타자 능력을 평가하는 지표를 제시했다. 이 연구에서는 타점, 출루율, 장타율 등 특정 기록에만 의존하여 타자를 평가했던 기존방법과 달리 모든 기록을 사용하여 타자 능력을 평가하였다. 또한 오광모·이장택(2003)은 요인분석, 신경망분석 및 의사결정나무분석을 이용하여 타자의 연봉 예측 모형을 제안하였다. 이외에도 남영식·박호정·장재열(2012)은 확률변경모형(Stochastic Frontier Model), Kang, Lee and Kwok(2007)은 자료포락방법(Data Envelopment Analysis)을 이용하여 선수의 생산성을 추정하였다. 남영식 외(2012)의 연구는 타자의 성과지표로서 OPS³⁾를, 투수의 성과지표로서 WHIP⁴⁾을 사용하여 이에 대한 선수의 연봉, 지역 연고 일치 여부, 연령, 해외리그 경험 여부, 자유계약 여부 및 경기 수 등을 독립변인으로 하여 효율성을 추정하였으며, Kang et al.(2007)의 연구는 각 구단의 총 연봉 효율성을 추정하였다.

James(2000)는 제시한 선수의 생산성지표로 WinShare(WS)를 제시하였다. WS는 승리에 대한 개별선수의 기여도를 계산한 것으로, WS=3은 이 선수가 기여한 팀승리가 3경기임을 의미한다. 그러나 WS지표에서는 선수가 패배에 기여한 부분을 간과한 단점이 있다. 따라서 소속 선수의 WS를 모두 합친 경기수가 팀승수와 같지 않은 문제점이 있

3) OPS=출루율(OBP)+장타율(SLG).

4) 투수의 이닝당 출루허용 수.

다. 최근에 가장 많이 사용되는 지표는 대체선수대비승수(Wins Above Replacement: WAR)이다. WAR은 미국의 Baseball-Reference와 Fangraphs 등 야구전문 통계기관에서 독립적으로 개발하였는데, 각각 bWAR, fWAR이라 한다. WAR은 미국 MLB에서 특정 선수가 해당 포지션의 AAA⁵⁾ 선수를 대체했을 때 기여하는 승수를 나타내는 지표이다. 즉 WAR=2인 선수를 기용하면 AAA 선수를 기용할 때보다 2승을 더 거둘 수 있음을 의미한다. WAR은 +/- 모든 숫자로 도출되기 때문에 WS에서 간과했던 패배 기여도를 고려하고 있으며, 시계열 변화에도 일관된 기록을 도출하여 최근에 가장 많이 이용되고 있다. 한국프로야구에서도 WAR을 대표적인 선수생산성 지표로 이용하고 있다. KBO는 미국과 달리 AAA가 존재하지 않으며, 2군에 속한 대체선수의 기량에 대한 명확한 통계자료가 부족하여 대체선수에 대한 가정이 필요하다. 일반적으로 국내 프로야구에서 야수는 600타석에서 3승을, 투수의 경우 9이닝당 0.175승을 기여하는 선수를 기준으로 가정하고 추가적인 기여도를 WAR로 계산한다(www.statiz.co.kr). 또한, WAR을 계산하기 위해서 연도별 리그 평균치, 구장 요인(park factor), 연도별 가중치 등 다양한 지표가 필요하다. 따라서 WAR을 공표되는 기록들의 조합만으로 계산할 수 없기 때문에 지표를 제공하는 기관 업체에 따라 차이가 나타난다.

본 연구에서는 WAR의 단점을 고려한 선수 가치를 추정하는 모형을 제시하고자 한다. 첫째, KBO가 발표하는 팀 및 선수 기록에 근거한 평가모형을 구축하여 지표 산출의 용이성을 높인다. 즉 KBO 공식 홈페이지에서 제공하는 선수 기록인 72개 변수 중 수비 기록을 제외한 61개 변수를 이용하는 통계모형을 구축함으로써 새로운 시즌의 기록만 대입하면 손쉽게 지표를 생성할 수 있다.⁶⁾ 둘째, 각 요인이 팀승리에 미치는 기여도를 회귀분석에 기초하여 도출한다. 즉 회귀분석에 기초한 대체선수대비승수인 RBWAR(Regression-Based Wins Above Replacement)을 제시한다. 본 연구에서 제시하는 RBWAR 산출모형의 구성은 다음과 같다. 첫째, 2002-2014년 61개 팀기록을 이용한 득점 및 실점방정식을 각각 패널자료 모형(panel data model)을 이용하여 추정한다. 이 과정에서 61개 기록을 모두 이용할 경우에 발생하는 자유도(degrees of freedom) 감소의 문제를 해결하기 위하여 중복된 정보를 포함하는 기록을 제거하고 요인분석(factor analysis)을 이용하여 각 기록을 유형화하여 회귀분석모형에 적용한다. 둘째, 대체선수

5) AAA란 미국 프로야구리그의 하나로 메이저리그의 바로 밑 단계의 마이너리그를 의미한다.

6) 본 연구에서는 WAR에서 사용하는 수비포지션별 가중치를 이용하여 야수들의 수비를 통한 기여도를 산출하였다. 개별선수의 수비 기록을 이용하는 연구는 추후의 연구로 미룬다.

대비승수를 계산하기 위한 기준으로 각 기록의 표준편차를 산출하여 각 기록의 평균에서 $(0.5 \cdot \text{표준편차})$ 가 부족한 가상의 선수를 설정하고 이 선수 대비 승수기여도를 산출하였다. 즉 타자의 경우 출루율과 장타율을 고려한다면 1군 선수 평균 출루율보다 표준편차의 1/2만큼 출루율이 낮으며, 장타율도 동일하게 평균보다 낮은 선수를 설정하여 이 선수대비 승리기여도를 산출한다. 한 기록이 정규분포를 따른다면 100명의 선수 중 84번째 선수를 가상의 비교대상으로 설정한 것이다.

본 연구에서 제시하는 KBO 선수의 승리기여도 추정모형은 기본적으로 세 단계로 구성된 Scully 모형의 2, 3 단계와 동일하다. 그러나 Scully 모형에서는 공격 기록으로 출루율(on-base percentage), 장타율(slugging percentage), 투수 기록으로 삼진(strikeouts), 워크(walks) 등 임의로 선정한 소수의 기록에만 의존한 반면, 본 연구에서는 61개의 다양한 기록을 모형에서 고려하였다는 장점이 있다. 따라서 현실적으로 구단에서 선수기여도를 산출하는 시스템을 구축하는 데 도움이 될 수 있다. 또한, 기존 WAR과 달리 본 모형은 회귀분석을 통하여 각 기록의 가중치를 추정하는 장점이 있으며, 공식적으로 발표하는 기록에만 의존하여 기여도를 산출하는 수월성이 있다.

본 연구의 구성은 다음과 같다. 제2장에서는 표본자료 및 계량모형을 자세히 설명하고, 제3장에서는 회귀방정식 설정 및 실증분석 결과를 논의한다. 제4장에서는 추정된 RBWAR을 WAR과 비교 분석하며, 제5장에서 결론으로 논문을 마감한다.

〈표 1〉 전체 변수

범주	수집 자료
타격 (24개)	경기 수, 타석, 타수, 득점, 1루타, 2루타, 3루타, 홈런, 총루타, 타점, 희생타, 희생플라이, 볼넷, 고의사구, 사사구, 삼진, 더블플레이, 장타율, 출루율, OPS, 멀티히트수, 득점권타율, 대타타율
투수 (30개)	경기 수, 승수, 패수, 세이브 수, 홀드 수, 승률, 투구이닝, 피안타, 피홈런, 볼넷, 사사구, 삼진, 실점, 자책점, WHIP, 완투 수, 완봉 수, 퀄리티스타트, 불론세이브, 상대타자수, 총투구 수, 이닝당투구 수, 피안타율, 피2루타, 피3루타, 희생타, 희생플라이, 고의사구, 와일드피치, 보크
주루 (7개)	경기 수, 도루시도 수, 도루, 도루자, 도루성공률, 주루사, 견제사
기타	선수연봉, WAR(2014년 시즌)

II. 연구방법

1. 자료수집 및 변인정리

본 연구에서 사용한 자료는 KBO 팀 성적과 개인성적의 패널 자료이다. KBO 공식 홈페이지 기록실에 있는 자료만을 이용하였으며, 2015년 시즌 팀 기록과 동일한 형태의 기록이 존재하는 2002년 시즌부터 자료를 수집하였다. 개별선수의 기록은 회귀분석에 사용하는 것이 아니므로 본 연구에서 표본 기간도 팀 기록과 동일한 기록을 수집하였다. 야구 통계전문 업체인 (주)스포츠투아이의 도움을 받아 2014년 시즌 1군 경기에 한 번이라도 출전한 경험이 있는 모든 선수들의 자료를 수집하였다. 팀 단위의 타격 기록과 투수력 기록 및 주루 관련 기록과 선수 연봉과 WAR(2014년 시즌)을 수집하였다. 표본 자료의 종류에 대해서는 위의 <표 1>에 자세히 서술하였다.

<표 1>에 포함된 KBO에서 제공하는 61개 기록 중 중복된 정보를 포함한 기록을 사전적으로 제거하였다. 예를 들어 총 루타는 1루타, 2루타, 3루타, 홈런의 합이다. 따라서 총 루타는 추가적인 정보를 제공하지 않으므로 표본에서 제외하였다. 또한 KBO에서 제공하는 61개 기록은 누적자료와 비율자료로 구분된다. 누적자료는 팀 기록이 각 선수 기록의 합으로 계산된 것으로 타점, 홈런 수 등이다. 비율자료는 타석 혹은 투구이닝 대비 기록으로 팀 기록은 개별선수의 가중평균이다. 타율, 평균자책점 등이 이에 속한다. 따라서 누적자료와 비율자료 간에는 서로 유사한 기록을 다른 방식으로 표현하기 때문에 중복된 정보를 포함하는 경우가 있다. 이와 같이 중복된 정보를 포함한 변수가 여러 개 있으면 정보의 손실이 없는 범위에서 표본에서 변수를 제외하였다. 예를 들어, 각 루타, 타석 수 등 개별 기록이 있으면 타율을 계산할 수 있다. $(1루타 수 + 2루타 수 + 3루타 수 + 홈런 수) / (타석 수 - 볼넷 수 - 고의사구 수 - 사사구 수 - 희생타수)$. 따라서 표본에서 타율을 제외하여도 정보의 손실이 없다. 또한 고의사구와 같이 선수의 순수한 능력을 나타내기보다는 경기상황에 따른 감독의 작전의 결과물인 기록을 표본에서 제외하였다. 이를 바탕으로 본 연구에서 사용한 기록은 <표 2>에 정리하였다.

2. 실증분석모형

본 연구는 3단계의 연구방법을 통해 최종적으로 선수의 경기력을 정확하게 추정하는 지표를 개발하는 것을 목적으로 한다. 첫 번째 단계는 <표 2>에 있는 다양한 기록을 요인분석(factor analysis)을 통하여 요인별로 분류한다. 요인분석이란 다양한 변수를 유사한 잠재변인으로 결합하는 연구방법이다. 본 연구에서는 타격, 투수 및 주루 요인이 각각 어떠한 기록으로 구성되었는지, 해당 요인은 어떤 기록들의 선형결합으로 이루어져 있는지에 대해 요인분석을 통하여 연구하였다. 요인을 도출하는 과정에서 요인점수는 일반적으로 활용하는 회귀분석법을 이용하여 추정하였다.

두 번째 단계로 요인분석에서 도출한 잠재변인들을 득점 및 실점방정식에 설명변수로 포함하여 회귀분석을 하였다. 득점방정식에서는 타격 및 주루 요인을 독립변수로 포함하였으며, 실점방정식에서는 투수 요인을 독립변수로 설정하여 회귀분석을 하였다. 득점 및 실점방정식은 각각 아래와 같은 패널자료 모형을 설정하여 추정하였다. 즉 각 구단의 개별효과와 시간효과를 고려한 모형을 설정하고, 가설검정을 통하여 각 효과의 통계적 유무에 판단을 하여 식 (1)과 같은 형태의 최종 회귀모형을 설정하였다.

$$y_{it} = x_{it}\beta + \alpha_i + \lambda_t + \epsilon_{it} \quad (1)$$

y_{it} 는 종속변수로 팀득점 또는 팀실점을 사용하였다. x_{it} 는 독립변수로 득점회귀식에서는 타격 및 주루 요인, 실점회귀식에서는 투수 요인을 사용하였다. α_i 는 개별효과(individual effect)로 각 팀의 특징적인 효과를, λ_t 는 시간효과(time effect)로 각 시즌의

<표 2> 최종 사용 변수

범주	지표
타격	타점, 1루타, 2루타, 3루타, 홈런, 득점권타율, 사사구, 사구, 희생플라이, 희생타, 병살타, 대타타율
투수	피1루타, 피2루타, 피3루타, 피홈런, 사구, 사사구, 삼진, 희생플라이, 희생타, 이닝당투구 수, 와일드피치
주루	피견제사, 주루사, 도루성공률

〈표 3〉 기록별 요인 분석

변인		요인 1	요인 2	요인 3	요인 4
타격기록	타점	.948	.084	.250	.055
	1루타	.842	.237	.236	.154
	홈런	.838	-.257	-.085	-.015
	2루타	.725	.210	-.058	.166
	득점권타율	.611	.276	.377	.000
	사사구	.472	-.217	-.220	-.429
	3루타	.098	.862	.096	-.099
	대타타율	-.002	.046	.714	-.051
	희생타	.247	.452	.557	.242
	사구	.428	-.373	.539	.237
	병살	.141	-.222	.102	.863
	희생번트	-.193	-.487	.382	-.599
	고유효값	4.172	1.663	1.345	1.297
	분산(%)	34.769	13.860	11.212	10.806
누적분산(%)	34.769	48.628	59.841	70.647	
Bartlett의 구형성 검정=711.086 df=66 , sig=.000					
변인		요인1	요인2	요인3	요인4
투구기록	피희생타	.737	-.004	-.077	.072
	폭투	.686	.017	-.262	.217
	피2루타	.677	.134	.014	-.335
	피3루타	.677	.138	.179	-.275
	피1루타	.169	.816	-.019	.008
	피홈런	.109	.813	.098	-.181
	사구	-.087	.562	.389	.321
	사사구	-.386	.477	-.352	-.202
	삼진	-.012	.066	.802	-.112
	이닝당투구수	.572	-.066	-.591	-.041
	피희생번트	-.048	-.061	-.058	.853
	고유효값	2.609	2.044	1.209	1.062
	분산(%)	23.719	18.577	10.988	9.655
	누적분산(%)	23.719	42.297	53.284	62.939
Bartlett의 구형성 검정=245.524 df=55 , sig=.000					
변인		요인1	요인2		
주루기록	견제사	.799	.339		
	주루사	.702	-.504		
	도루성공률	.089	.870		
	고유효값	1.141	1.124		
	분산(%)	38.044	37.477		
	누적분산(%)	38.044	75.521		
Bartlett의 구형성 검정=6.126 df=3 , sig=.106					

〈표 4〉 타격 기록 요인 점수 구성

기록범주	요인변수	요인점수 구성
타격기록	BAT1	0.251·타점 + 0.206·루타+ 0.283·홈런 + 0.210·2루타+ 0.126·득점권타율 + 0.217·사사구
	BAT2	0.534·3루타
	BAT3	0.503·대타타율 + 0.326·희생타 + 0.327·사구
	BAT4	0.622·병살 - 0.405·희생번트
투구기록	PIT1	0.318·피희생타 + 0.286·폭투 + 0.261·피2루타 + 0.285·피3루타
	PIT2	0.439·피1루타 + 0.414·피홈런 + 0.296·사구 + 0.293·사사구
	PIT3	0.602·삼진 - 0.382·이닝당투구 수
	PIT4	0.753·피희생번트
주루기록	RUN1	0.703·견제사 + 0.614·주루사
	RUN2	0.773·도루성공률

특징적인 효과를 의미한다. 마지막으로 ϵ_{it} 는 오차항으로 특유성(idiosyncratic error)를 가정한다. 또한, 강외생성(strictly exogeneity)을 가정한다: $E[\epsilon_{it}/x_{it}, \alpha_i] = 0$. 득점 및 실점방정식 추정 결과에서 오차항의 자기상관 또는 이분산 가능성을 고려하여 추정량의 분산에 대한 일치추정량인 Newey-West의 HAC(heterogeneity-autocorrelation-consistent) 분산추정량을 이용하였다.

마지막으로 회귀분석을 통해 도출한 요인들의 득점 혹은 실점에 대한 한계생산 추정치를 이용하여 개별선수의 득점 혹은 실점에 대한 기여도를 계산하였다. 이 과정에서 선수들의 경기 출전 타석 수 혹은 이닝 수를 고려하였는데 이를 가중치로 이용하여 선수가 기여한 득점 및 실점을 계산하였다. 그리고 개별선수가 기여한 득실차로 인하여 발생한 승리 혹은 패배 기여를 계산하여 최종적인 RBWAR을 산출하였다.

Ⅲ. 추정 결과

1. 요인분석 결과

요인분석은 Barlett의 구형성검정을 통하여 요인분석의 타당성을 검증하였으며 주성분분석을 통하여 추출한 초기 고윳값(eigen value)의 합계가 1보다 큰 요인들만을 추출하였다. 각 요인에 속한 변수는 Kaiser의 정규화된 베리맥스를 통하여 성분행렬을 도출하였으며, 이를 가장 일반적으로 사용하는 회귀분석 방법을 사용하여 성분점수의 계수행렬을 도출하였다. 구체적인 분석결과는 <표 3>과 <표 4>에 정리하였다. 연구 결과, 타격의 경우 4개 요인으로 정리되었으며, 요인 1은 타점, 1루타, 홈런, 2루타, 득점권타율, 사사구 등 6개 타격 기록의 결합, 요인 2는 3루타 단독, 요인 3은 대타타율, 희생타 및 사구의 결합, 요인 4는 병살 및 희생번트의 결합이다. 투수 기록을 보면 총 11개 투수기록이 4개 요인으로 정리되었다. 마지막으로 주루기록은 2개 요인으로 정리되었다.

2. 득점 및 실점방정식 추정 결과

세 번째 단계로 패널자료 모형을 구성하여 득점방정식 및 실점방정식을 추정하였다. 득점방정식은 각 팀의 총 득점을 종속변수로 설정하고 요인분석에서 도출한 4개 타격 요인, 2개 주루 요인을 독립변수로 설정하였다. 실점방정식은 득점방정식과 유사하게 각 팀의 총 실점을 종속변수, 4개 투수 요인을 독립변수로 설정하였다. 본 연구에서 사용한 자료는 2002년부터 2014년 동안의 프로야구 팀들의 성적에 대한 불균형패널 자료이다. 득점 또는 실점의 회귀방정식에 개별효과(individual effect) 및 시간효과(time effect)의 포함 여부를 결정하기 위하여 두 효과의 존재 유무를 각각 F검정을 통하여 실시하였다. 이 검정은 고정효과 또는 임의효과에 무관하게 일치성이 있는 Within 추정치를 이용하여 실행하였다. 검정 결과를 토대로 최종 회귀식을 설정한 후에 고정효과(fixed effects)와 임의효과(random effects) 사이의 적절한 추정방법을 정하기 위하여 하우스만 검정을 실시하였다.

〈표 5〉 효과 검정 결과

방정식	효과	구분	통계값
득점방정식	시간효과 (time effect)	F	2.3918
		df1/df2	12/87
		P	0.010
	개별효과 (individual effect)	F	0.8976
df1/df2		9/90	
P		0.531	
실점방정식	시간효과 (time effect)	F	5.7418
		df1/df2	12/89
		P	.000
	개별효과 (individual effect)	F	.6058
		df1/df2	9/92
		P	.7891

〈표 6〉 회귀 분석 결과

모형 이름	변수	within		GLS		
		EST.	S.E	EST.	S.E	
득점방정식	(계수)	-	-	543.978	16.446**	
	BAT1	77.081	2.169**	76.690	39.659**	
	BAT2	6.882	2.236**	7.558	4.345**	
	BAT3	13.457	2.456**	13.862	5.724**	
	BAT4	0.209	1.895	0.427	.848	
	RUN1	0.102	0.211	0.134	.609	
	RUN2	81.160	26.350**	106.677	.001**	
	R-sq	0.923		0.933		
	Hausman test					
	Chi-sq(df)	2.401 (6)				
p	.879					
모형 이름	변수	within		GLS		
		EST.	S.E	EST.	S.E	
실점방정식	(계수)	-	-	604.441	4.439**	
	PIT1	14.492	3.512**	15.203	3.469**	
	PIT2	80.580	5.691**	79.670	4.546**	
	PIT3	-10.415	3.622**	-9.369	4.518*	
	PIT4	-7.980	2.849**	-5.948	2.130**	
	R-sq	0.844		0.843		
	Hausman test					
	Chi-sq(df)	28.721 (4)				
	p	.000				

득점방정식의 기본 모형은 식 (2)와 같다.

$$R_{it} = \beta_0 + \beta_1BAT1_{it} + \beta_2BAT2_{it} + \beta_3BAT3_{it} + \beta_4BAT4_{it} + \beta_5RUN1_{it} + \beta_6RUN2_{it} + \alpha_i + \lambda_t + \epsilon \quad (2)$$

<표 5>에 제시한 대로 개별효과 α_i 와 시간효과 λ_t 에 대한 검정을 실시하였다. 검정 결과에 의하면 시간효과가 불변이라는 가설은 5% 유의수준에서 기각된 반면, 개별효과가 불변하다는 가설은 기각할 수 없었다. 따라서 최종추정방정식에는 개별효과를 제외하고 시간효과만 포함하였다. <표 6>에 득점방정식을 within 추정법과 GLS 추정법으로 도출한 추정치를 정리하였다. 먼저 하우스만 검정통계량이 2.401이며 p-값이 0.879이어서 임의효과 설정이 옳다는 귀무가설, 즉 시간효과와 설명변수 간에 상관관계가 없다는 귀무가설을 기각할 수 없었다. 따라서 within 추정량과 GLS 추정량 모두 일치추정량이지만 후자가 더 효율적인 추정량이어서 득점방정식의 최종 모형은 시간효과가 포함된 임의효과 모형으로 결정하였다. GLS 추정치를 기준으로 추정 결과를 논의하면 타격 부문 요인 중에서 요인 1-3은 10% 수준에서 통계적 유의성을 가지며 주루요인 중에서는 요인 2만 유의적인 것으로 나타났다.

실점방정식을 추정할 때에도 득점방정식과 동일한 검정을 통해 최종 모형을 도출하였다. 실점방정식의 기본모형은 식 (3)과 같다.

$$ER_{it} = \beta_0 + \beta_1PIT1_{it} + \beta_2PIT2_{it} + \beta_3PIT3_{it} + \beta_4PIT4_{it} + \alpha_i + \lambda_t + \epsilon_{it} \quad (3)$$

실점방정식의 시간효과 및 개별효과 검정결과는 <표 5>에 나타나 있다. 실점방정식 역시 득점방정식과 동일하게 개별효과는 차이가 없으며, 시간효과가 동일하다는 귀무가설이 기각되어 최종 모형에는 시간효과만이 포함하였다. 하우스만 검정 결과에 의하면 통계량이 28.721이며 p-값이 0.00이어서 임의효과 설정이 옳다는 귀무가설이 기각되었다. 따라서 GLS 추정량은 불일치추정량이 되므로 고정효과 모형을 채택하고 이때 가장 효율적인 within 추정치를 이용하였다. 추정 결과에 의하면 투수요인 4개가 모두 5% 수준에서 통계적 유의성을 갖는 것으로 나타났다(표 6).

3. 기록별 득실점에 미치는 영향력

〈표 7〉 변인별 영향력

	변인	N	평균	표준편차	impact	impact(%)	유형
타격	타점	106	569.830	80.950	0.239	2.39	누적기록
	1루타	106	1170.443	79.786	0.199	1.99	누적기록
	홈런	106	108.783	35.013	0.623	6.23	누적기록
	2루타	106	203.245	23.342	0.693	6.93	누적기록
	득점권타율	106	0.268	0.017	964.716	9.65	비율기록
	사사구	106	72.651	16.185	1.033	10.33	누적기록
	3루타	106	19.047	7.124	0.516	5.16	누적기록
	대타타율	106	0.218	0.040	170.548	1.71	비율기록
	희생타	106	37.840	7.901	0.555	5.55	누적기록
	사구	106	468.170	58.828	0.075	0.75	누적기록
	병살	106	101.509	14.379	0.009	0.09	누적기록
	희생번트	106	84.085	25.803	-0.003	0	누적기록
	투수	피희생타	106	37.840	8.360	0.299	2.99
폭투		106	51.000	13.617	0.165	1.65	누적기록
피2루타		106	203.245	28.321	0.072	0.72	누적기록
피3루타		106	19.047	7.139	0.314	3.14	누적기록
피1루타		106	1170.443	85.623	0.405	4.05	누적기록
피홈런		106	108.783	28.846	1.134	11.34	누적기록
사구		106	468.170	57.724	0.405	4.05	누적기록
사사구		106	72.651	16.656	1.390	13.90	누적기록
삼진		106	817.462	94.119	-0.122	-1.22	누적기록
이닝당투구수		106	16.779	0.835	-8.705	-87.05	누적기록
피희생번트		106	84.085	16.532	-0.116	-1.16	누적기록
주루	견제사	106	8.330	3.580	0.020	0.20	누적기록
	주루사	106	52.868	9.771	0.006	0.06	누적기록
	도루성공률	106	0.676	0.061	1126.600	11.26	비율기록

본 연구의 회귀분석에 사용한 자료는 요인분석 과정에서 모든 기록을 표준화하여 사용하였으며, <표 4>에 나타난 대로 각 요인은 표준화된 기록들의 선형결합이다. 따라서 회귀계수 추정치로부터 각 변수의 변화가 득점 또는 실점에 미치는 영향력을 직관적으로 판단하기 어렵다. 이에 <표 7>에 각 변인의 한계효과를 계산하여 추정 결과의 타당성을 정성적으로 평가할 수 있게 하였다. 한계효과 계산방법은 회귀계수를 요인점수구성 계수에 곱한 후 변인의 표준편차로 나누어 표준화한 값이 아닌 실제 값의 변화량에 대한 한계효과를 계산하였다. 각 변수의 한계효과를 살펴보면 타점의 경우 0.239를 기록하였는데 이는 타점이 1개 증가할 경우 팀 시즌 득점이 0.239점 증가함을 의미한다. 비율 기록의 경우 득점권 타율은 964.716의 한계효과를 기록하였는데 이는 득점권 타율이 1%p 오르면 팀 시즌 득점이 0.965점 상승함을 의미한다.

IV. 개별선수의 승리기여도 산출 결과

1. 선수들의 기여도 (RBWAR) 산출방법

<표 7>에 나타난 각 기록별 득점 또는 실점에 대한 영향도와 개별선수의 각 기록에 대한 기여도를 접목함으로써 개별선수의 팀 득점 및 실점에 미치는 기여도를 추계하였다. 여기에 개별선수의 출장 타석수 및 투구이닝을 고려하여 가중치를 설정하였다. 본 연구에서 추계하는 RBWAR, 즉 대체선수대비승수를 계산하기 위한 대체선수를 다음과 같이 설정했다. 모든 기록의 리그 평균기록에서 (0.5·표준편차) 만큼 부족한 기록을 산출하고, 모든 기록에서 이렇게 설정한 성적을 기록한 선수를 대체선수로 가상하였다. 이는 현재 WAR을 계산하는데 사용하는 방식과 유사한 방법이지만 국내 야구통계전문기관인 STATIZ는 타자는 600타석에서 3승, 투수는 9이닝당 0.175승을 고정시키는데 비해, 본 연구의 방법의 시즌별로 변하는 기록의 변동이 반영된다는 장점이 있다.

타격 및 주루 기록은 해당 타자와 대체선수 간의 기록차이에 한계효과를 곱한 후 그 선수의 총 타수의 팀 내 비중으로 가중치를 두어 득점기여도를 계산하였다. 투수 기록도 타자와 마찬가지로 각 투수 기록을 이닝당 기록으로 계산한 후 대체선수 대비 성적의 차이에 한계효과를 곱하여 이닝당 기여도를 계산하였다. 여기에 투수의 투구이닝의

팀 내 투구 비중을 곱하여 개별선수의 실점을 줄이는 데 대한 기여도를 계산하였다. 투수 기록을 이닝당 기록으로 사용한 이유는 기준이 되는 대체선수와의 정확한 비교를 위해서이다. 즉 대체선수가 1이닝당 기록한 지표에 비해 얼마나 차이가 있는지를 먼저 살펴본 후 이를 해당 선수의 경기이닝을 곱하여 대체선수가 해당 선수만큼 플레이했을 경우 기록했을 득점 및 실점 기여도와의 차이를 계산한 것이다. 이를 종합하면 다음과 같은 식으로 표현할 수 있다. 타격 및 주루지표 계산법과 투구지표 계산법은 아래 각각의 식으로 표현할 수 있다.

$$\beta \cdot (STAT_i - STAT_R) \cdot \frac{PA_i}{TPA} \quad (4)$$

$$\beta \cdot \left(\frac{STAT_i}{IP_i} - \frac{STAT_R}{IP_R} \right) \cdot IP_i \cdot \frac{IP_i}{TIP} \quad (5)$$

위 식들에서 β 는 <표 13>에 나타난 각 기록의 한계효과를, $STAT_i$ 는 각 선수의 기록을, $STAT_R$ 은 대체선수의 기록을, PA 는 선수의 타석수를, TPA 는 팀의 총 타석수를, IP 는 투수가 투구한 이닝을, TIP 는 팀의 총 이닝을 의미한다. 각 계산방법을 통해서 각 지표들이 미치는 득점 혹은 실점 기여도를 계산한 후 이를 각각 더하여 득점 및 실점기여도를 계산하였다. 이때 득점기여도를 산출하는 데 회귀분석 결과에 따라 유의한 값을 기록한 BAT1, BAT2, BAT3 및 RUN2만을 사용하였으며, 이에 따라 각각 세부분인인 bat1, bat2, bat3, bat4, bat5, bat6, bat7, bat8, bat9, bat10 및 run3를 사용하였다. 투수 기록은 모든 변인이 유의하였기 때문에 모든 기록을 사용하였다.

다음 단계로 야수들의 타격 및 주루기록, 투수들의 투구기록을 바탕으로 도출한 득점 및 실점 기여도를 바탕으로 하여 팀의 승리 기여도를 계산하였다. 우선 팀별 득점 및 실점에 대한 승리의 비율은 다음과 같이 계산하였다.

$$W = \frac{1}{N} \sum_{i=1}^N W_i = \frac{1}{N} \sum_{i=1}^N \frac{|\text{팀총득점}_i - \text{팀총실점}_i|}{|\text{팀승수}_i - \text{팀패배수}_i|} \quad (6)$$

여기서 W_i 는 팀 i 의 득실차에 대한 승패차의 비율이다. 각 팀별 W_i 의 평균을 최종적인 득실차에 대한 승차의 비율 W 로 사용하였다. 이를 통하여 각 선수들의 득점 혹은

〈표 8〉 포지션별 조정 승수

포지션	기준 이닝	조정 승수
포수	1458	1.25
1루수	1458	-1.25
2루수	1458	0.25
3루수	1458	0.25
유격수	1458	0.75
좌익수	1458	-0.75
중견수	1458	0.25
우익수	1458	-0.75
지명타자	600	-1.75

실점 기여도에 대한 승리 기여를 계산하였다. 즉 선수의 득점 기여가 R 이라 하면 승리 기여는 R/W 로 계산하였다.

마지막으로 선수들의 수비에 대한 기여를 계산하여 최종적인 **RBWAR**을 도출하였다. **WAR**을 계산할 때에는 각 선수들의 수비 기록과 함께 수비 위치에 따라 가중치를 더해주는데 본 연구에서는 수비 기록은 추후 연구로 미루고 수비 위치에 따른 가중치만을 사용하였다. 즉 **STATIZ**에서 **WAR**을 추정할 때 사용하는 수비 위치 가중치를 본 연구에서도 그대로 사용하였다. 사용한 가중치는 <표 8>에 제시하였다.⁷⁾ 이는 각 수비 위치별 기준이닝당 승리 기여를 의미하며, 예를 들어 포수가 1458이닝 동안 수비를 하였으면 팀 전체 승수에 1.25승만큼을 추가적으로 기여한 것을 의미한다. 만약 한 선수가 3루수로 200이닝, 유격수로 300이닝을 출전하였다고 하면 이 선수의 수비 위치 조정은 다음과 같다.

7) 수비를 통한 야수의 기여도를 정확히 추정하기 위해서는 각 수비위치에서 상대적 수비능력을 고려하여야 한다. 본 연구의 표본에 대비되는 **KBO**가 발표하는 13개 수비기록은 수비위치별 세분화된 기록을 포함하지 않는다. 따라서 본 연구에서는 모든 야수가 수비위치로 동일한 능력을 갖는다고 가정하고 수비능력을 반영하는 보다 세분화된 연구는 추후연구로 미룬다.

$$\frac{200}{1458} \times 0.25 + \frac{300}{1458} \times 0.75 = 0.066 \quad (7)$$

즉 이 선수는 수비로 인하여 팀에 0.066승을 기여하였다고 계산하였다. 이때 야수의 수비위치 조정 기여승수를 PW라고 하면 선수가 기록한 승수기여도는 (RBWAR = R/W + PW)로 계산할 수 있다.

2. RBWAR 산출 결과

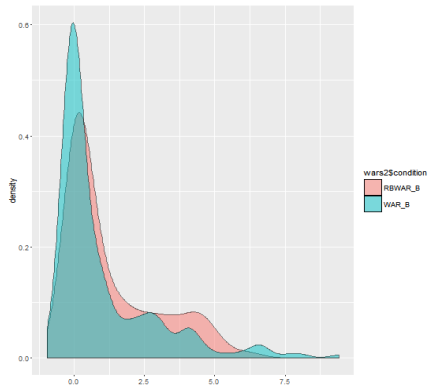
먼저 (주)스포츠투아이에서 발표한 2014년 시즌 개별선수의 WAR과 본 모형으로 산출한 RBWAR의 상관계수를 <표 9>에 정리하였다. 타자와 투수의 상관계수가 각각 0.869, 0.882로 매우 높게 나타났다. 투수의 경우 선발투수에 비해 구원투수의 상관계수가 0.670으로 더 낮다. 이는 구원투수를 평가하는 데 있어 RBWAR이 기존 WAR에 상대적으로 차이가 있음을 의미한다. 또한 [그림 1]과 [그림 2]에 각각 타자와 투수의 RBWAR과 WAR의 kernel 분포를 비교하였다. 두 그림 모두 양의 왜도(skewness)를 보여주어 소수의 슈퍼스타급 선수들과 다수의 보통선수들이 있음을 보여주고 있다. 기존 WAR과 비교할 때 타자의 경우 최상위 선수의 RBWAR이 기존 WAR에 비해서 낮은 값을 가지며, 투수의 경우 반대로 최상위 선수들의 RBWAR이 기존 WAR보다 더 높은 값을 갖는다.

<표 10>은 2014년 시즌 각 구단별 기여도 상위 3명의 타자를 정리한 것이다. 넥센 히어로스의 경우 야수 중 서건창, 강정호, 박병호 순으로 RBWAR이 높게 나타났다. 기존 WAR은 강정호, 서건창, 박병호 순으로 정리하여 비슷하지만 다소 차이가 있는 것

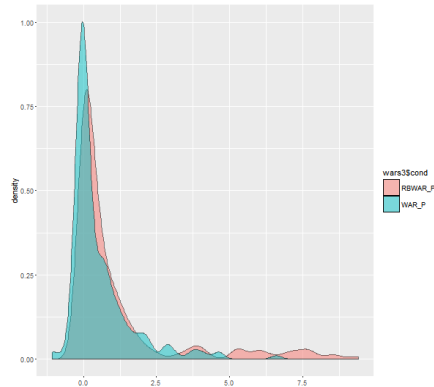
<표 9>. WAR과 RBWAR의 상관계수

항목		상관계수
타자		0.869
투수	투수 전체	0.882
	선발투수	0.847
	구원투수	0.670

(그림 1) 타자들의 WAR 커널분포



(그림 2) 투수들의 WAR 커널분포



으로 나타났다. <표 11>은 투수를 제외한 수비 위치별로 기여도 상위 3명의 타자를 정리하였다. 포수는 이재원이 5.18로 가장 높게 나타났으며, 2루수의 경우 서건창이 6.32로 나바로에 6.21에 비해 근소하게 기여도가 높은 것으로 나타났다. 그러나 기존 WAR에 의하면 서건창과 나바로의 격차는 1.5승 정도로 큰 차이가 난다. 좌익수의 경우 RBWAR이 김현수, 최형우, 김주찬 순으로 나타났으나 기존 WAR에 의하면 최형우가 김현수보다 기여도가 높게 나타났다.

<표 12>는 2014년 시즌 각 구단별 기여도가 높은 투수 3명을 정리하였다. KIA 타이거즈에서는 양현종의 기여도가 독보적으로 높으며, NC 다이노스는 3명의 선발투수가 근소한 차로 높은 RBWAR을 갖는 것으로 나타났다. <표 13>에는 선발투수와 구원투수를 분리하여 전체 상위 5명씩을 정리하였다. 선발투수 중 넥센 히어로즈의 벤헤젠의 RBWAR이 9.40으로 가장 높게 나왔으며, 국내 투수 중에서는 김광현이 7.76으로 유희관, 양현종의 7.70에 근소하게 기여도가 높은 것으로 나타났다. 이는 기존 WAR이 양현종을 4.62로 김광현의 3.97보다 근소하게 우세한 것으로 평가한 것과 차이가 난다.

마지막으로 <표 14>, <표 15>에서 2014년 시즌 후의 FA선수들의 계약 내용과 RBWAR을 비교하였다. RBWAR 기준으로 2014년 타자 FA 중 기여도 상위 3명은 김강민, 박용택, 최정인데 결과적으로 이 3명의 선수가 가장 많은 연봉 계약을 맺었다. 투수의 경우 윤성환의 RBWAR이 7.20으로 가장 높게 나타났고, 장원준이 5.99로 두 번째로 높은 기여도를 기록하였으며, 두 투수 모두 80억 원 이상의 계약을 체결하였다.

〈표 10〉 2014년 야수 팀별 RBWAR 상위 선수

팀명	선수명	포지션	OBP	SLG	WAR	RBWAR
KIA	안치홍	2루수	0.390	0.544	5.17	4.46
	이대형	중견수	0.372	0.401	2.3	3.66
	김주찬	좌익수	0.397	0.514	2.66	3.25
LG	박용택	중견수	0.43	0.461	4.23	4.99
	오지환	유격수	0.354	0.413	2.43	4.27
	이진영	우익수	0.384	0.425	2.29	3.14
NC	나성범	중견수	0.400	0.597	6.64	5.80
	테임즈	1루수	0.422	0.688	6.57	4.93
	박민우	2루수	0.392	0.399	3.89	4.77
SK	이재원	포수	0.413	0.507	4.05	5.18
	김강민	중견수	0.368	0.495	4.57	4.87
	최정	3루수	0.397	0.506	2.77	3.68
넥센	서건창	2루수	0.438	0.547	8.09	6.32
	강정호	유격수	0.459	0.739	9.42	5.05
	박병호	1루수	0.433	0.686	7.61	4.41
두산	정수빈	중견수	0.379	0.422	2.75	4.86
	김현수	좌익수	0.396	0.488	4.19	4.34
	민병현	우익수	0.395	0.500	3.7	4.29
롯데	황재균	3루수	0.388	0.475	2.78	4.62
	손아섭	우익수	0.456	0.538	6.74	4.37
	정훈	2루수	0.386	0.398	2.97	4.28
삼성	나바로	2루수	0.417	0.552	6.58	6.21
	채태인	1루수	0.368	0.484	2.84	4.60
	김상수	유격수	0.354	0.407	3.18	4.56
한화	정근우	2루수	0.391	0.416	4.6	4.30
	피에	중견수	0.373	0.524	4.06	3.79
	송광민	3루수	0.357	0.468	1.22	2.87

〈표 11〉 2014년 야수 포지션별 RBWAR 상위 선수

포지션	선수명	팀명	OBP	SLG	WAR	RBWAR
포수	이재원	SK	0.413	0.507	4.05	5.18
	양의지	두산	0.36	0.48	2.99	2.92
	최경철	LG	0.28	0.293	0.5	2.50
1루수	테임즈	NC	0.422	0.688	6.57	4.93
	채태인	삼성	0.368	0.484	2.84	4.60
	박병호	넥센	0.433	0.686	7.61	4.41
2루수	서건창	넥센	0.438	0.547	8.09	6.32
	나바로	삼성	0.417	0.552	6.58	6.21
	박민우	NC	0.392	0.399	3.89	4.77
3루수	황재균	롯데	0.388	0.475	2.78	4.62
	김민성	넥센	0.349	0.45	1.76	3.91
	최정	SK	0.397	0.506	2.77	3.68
유격수	강정호	넥센	0.459	0.739	9.42	5.05
	김상수	삼성	0.354	0.407	3.18	4.56
	오지환	LG	0.354	0.413	2.43	4.27
우익수	박환이	삼성	0.409	0.443	2.15	4.53
	손아섭	롯데	0.456	0.538	6.74	4.37
	민병현	두산	0.395	0.5	3.7	4.29
좌익수	김현수	두산	0.396	0.488	4.19	4.34
	최형우	삼성	0.426	0.649	6.17	3.73
	김주찬	KIA	0.397	0.514	2.66	3.25
중견수	나성범	NC	0.4	0.597	6.64	5.80
	박용택	LG	0.43	0.461	4.23	4.99
	김강민	SK	0.368	0.495	4.57	4.87

〈표 12〉 2014년 투수 팀별 RBWAR 상위 선수

팀명	선수명	ERA	WHIP	WAR	RBWAR
KIA	양현종	4.25	1.39	4.62	7.70
	임준섭	6.06	1.71	0.95	3.81
	홀튼	4.8	1.54	1.16	2.10
LG	리오단	3.96	1.27	2.24	7.08
	우규민	4.04	1.33	2.87	5.87
	류제국	5.12	1.38	2.27	5.20
NC	에릭	4.01	1.34	2.10	7.41
	찰리	3.81	1.45	2.04	6.81
	이재학	4.21	1.41	1.87	6.08
SK	김광현	3.42	1.49	3.97	7.76
	조지훈	27	4.5	-0.05	3.68
	울프	4.85	1.35	0.75	1.79
넥센	밴헤켄	3.51	1.32	6.60	9.40
	소사	4.61	1.48	1.70	3.83
	한현희	3.2	1.35	1.69	1.60
두산	니퍼트	3.81	1.3	4.15	8.39
	유희관	4.42	1.43	2.28	7.70
	노경은	9.03	1.97	0.37	2.52
롯데	옥스프링	4.2	1.34	3.77	8.69
	장원준	4.59	1.46	2.91	5.99
	유민	5.93	1.71	2.06	5.33
삼성	윤성환	4.39	1.37	3.68	7.20
	밴덴헐크	3.18	1.11	4.64	6.32
	배영수	5.45	1.57	2.83	4.15
한화	앨버스	5.89	1.59	3.18	5.33
	이태양	6.46	2.09	1.65	5.32
	안영명	4.52	1.54	1.41	2.14

〈표 13〉 2014년 투수 역할별 RBWAR 상위 선수

역할	선수명	팀명	ERA	WHIP	S2I WAR	2014평균
선발	밴헤켄	넥센	3.51	1.32	6.60	9.40
	옥스프링	롯데	4.20	1.34	3.77	8.69
	니퍼트	두산	3.81	1.30	4.15	8.39
	김광현	SK	3.42	1.49	3.97	7.76
	유희관	두산	4.42	1.43	2.28	7.70
구원	안영명	한화	4.52	1.54	1.41	2.14
	전유수	SK	5.00	1.61	0.35	1.68
	이민호	NC	5.01	1.49	-1.00	1.68
	한현희	넥센	3.20	1.35	1.69	1.60
	차우찬	삼성	5.60	1.56	0.69	1.58

〈표 14〉 2014년 타자 FA선수 기록

선수명	팀명	pos	OBP	SLG	RBWAR (2014)	WAR (2014)	연봉총액 (백만원)	계약 기간	RBWAR (2015)
김강민	SK	중견수	0.368	0.495	4.87	4.57	5,600	4	1.89
김경언	한화	우익수	0.397	0.467	1.78	1.87	850	3	2.07
나주환	SK	2루수	0.332	0.382	2.98	1.48	550	2	0.86
박경수	LG	2루수	0.344	0.315	1.16	0.04	1,820	4	3.89
박기혁	롯데	유격수	0.295	0.196	0.13	-0.28	1,140	4	2.76
박용택	LG	중견수	0.430	0.461	4.99	4.23	5,000	4	4.15
이성열	넥센	우익수	0.341	0.461	0.99	0.67	500	2	1.25
조동찬	삼성	3루수	0.360	0.365	0.23	0.14	2,800	4	-
조동화	SK	우익수	0.330	0.325	3.68	-0.29	2,200	4	1.20
차일목	KIA	포수	0.286	0.254	0.63	0.16	450	2	0.03
최정	SK	3루수	0.397	0.506	3.68	2.77	8,600	4	2.32

〈표 15〉 2014년 투수 FA선수 기록

선수명	팀명	역할	ERA	WHIP	RBWAR (2014)	WAR (2014)	연봉총액 (백만원)	계약 기간	RBWAR (2015)
권혁	삼성	구원	2.86	1.10	0.33	0.55	3,200	4	2.56
김사율	롯데	구원	5.79	1.58	1.42	0.84	1,450	4	0.11
배영수	삼성	선발	5.45	1.57	4.15	2.83	2,150	3	1.76
송은범	KIA	선발	7.32	2.00	1.20	0.58	3,400	4	0.88
안지만	삼성	구원	3.75	1.30	0.98	1.09	6,500	4	1.42
윤성환	삼성	선발	4.39	1.37	7.20	3.68	8,000	4	8.54
이재영	SK	구원	6.19	1.52	0.55	-0.13	450	2	0.19
장원준	롯데	선발	4.59	1.46	5.99	2.91	8,400	4	6.16

V. 결론 및 제언

본 연구는 한국프로야구선수들의 가치를 평가하는 계량모형 구축을 목적으로 시작하였다. 스포츠경제학 문헌에서 찾을 수 있는 미국 메이저리그의 선수 가치평가모형은 Scully(1974)와 Krautmann(1999)로 구분할 수 있다. 그러나 이들 계량모형의 접근법을 한국프로야구에 직접 적용할 수 없었다. 이는 첫째, KBO구단의 수익자료 부족으로 Scully 모형의 1단계에서 추정하는 승리의 한계수익을 추정할 수 없으며 둘째, Krautmann 모형의 가정(FA시장, 즉 선수노동시장이 완전경쟁이다)이 만족되지 않기 때문이다. KBO의 FA시장은 매년 각 수비 위치별로 소수의 선수만 배출되며, 한국프로야구에서는 FA선수의 보상조건이 메이저리그보다 강하기 때문에 완전경쟁시장으로 보기 어렵기 때문이다. 마지막으로 메이저리그 구단은 모두 이윤추구를 제1 목표로 하는 데 반해, 대기업을 모기업으로 두고 있는 KBO 구단은 이윤추구가 제1 목표가 아니기 때문에 Scully 모형의 이윤극대화 조건에서 도출한 개별선수의 한계수입 생산이 한국프로야구에서는 절대적인 가치기준이 아니다.

이에 본 연구는 연구 목적을 다소 축소하여 개별선수의 한계생산을 추정하는 계량모형을 구축하였다. 즉 개별선수의 승리기여도를 추계하는 모형을 구축하였다. 따라서 이는 WAR과 같은 기존 지표와 동일한 목적을 지니고 있으나 방법론상 많은 차이를 지닌다. 본 연구의 기본적인 방법론은 Scully 모형의 2, 3단계와 동일하다. 2단계인 승리생산함수 추정과 3단계인 개별선수의 팀 기록에 미친 기여도를 2단계 추정결과에 적용하여 팀 승리에 미친 기여도를 추계하는 방법이다. Scully 모형에서는 공격력을 대표하는 변수로 출루율, 장타율 등 소수의 기록을 임의로 설정한 단점이 있다. 본 연구에서는 KBO가 발표하는 공식 기록 72개 중 수비 기록을 제외한 61개 기록을 모두 이용하였고, 요인분석을 통하여 승리(득점 또는 실점)생 산함수 추정에 필요한 설명변수를 선정하였다. 따라서 본 계량모형은 Scully 모형의 단점을 보완하며 현실에 직접 적용할 수 있는 현실적인 모형으로 판단한다.

기존 WAR과 차이점은 다음과 같다. 팀 승리에 대한 기여도를 측정하는 모든 지표는 각 기록의 가중치로 구성되어 있다. WAR의 경우도 다양한 기록의 가중치이다. 그러나

그 가중치가 알려져 있지 않기 때문에 가중치의 비교를 통한 WAR지표의 타당성을 판단할 수가 없다. 또한, WAR의 구성요소 중에는 공식 발표 자료가 아닌 경기를 관찰하여 작성해야 하는 기록도 있다. 따라서 이 지표를 추산하기 위해서는 많은 노력을 요하며 주관성이 작용할 수 있는 여지도 있다. 이에 비하여 본 연구는 회귀분석을 통하여 각 기록의 가중치를 추정함으로써 각 기록의 팀 승리에 미치는 영향 정도를 통계적으로 검증할 수 있을 뿐 아니라 그 추정치를 비교함으로써 가중치의 타당성을 점검할 수 있는 장점이 있다. 또한, 모든 분석 절차가 순수하게 표본 자료에 의존하는 계량분석이어서 실제 구단의 선수가치 평가에 활용하는 데 간편한 장점이 있다.

개별선수의 기여도 추계치(RBWAR)를 비교해 보면, 기존 WAR과 상관계수가 높게 나타났지만 일부 선수는 다소 다른 결과가 나왔다. 2014년 FA선수들의 계약 내용을 비교해 보면 RBWAR값으로 판단할 때 구단의 선수계약이 어느 정도 합리성을 지닌 것으로 보인다. 즉 RBWAR이 높은 선수일수록 더 높은 연봉에 계약을 한 것으로 나타났다.

그러나 본 연구에서 구축한 선수가치 평가모형은 다음과 같은 향후 과제를 보완하여야 더 완성도를 높일 것으로 기대된다. 첫째, 팀 승리 기여도에서 더 나아가 구단 목적에 부합한 선수 가치를 평가하는 모형으로 발전해 나가야 한다. 이는 현재 구축한 Scully 모형의 2, 3단계에 Krautmann방법을 수정하여 적용하는 방법을 통해서 구축하는 방법을 고려할 수 있다. 둘째, RBWAR에서 나오는 특성 중 구원투수가 선발투수에 비해 현격하게 기여도가 낮게 추계되는 점,⁸⁾ 야수의 수비능력을 충분히 반영하지 못하는 점들을 개선해야 한다.

참고문헌

- 남영식·박호정·장재열. 「확률프론티어 기법을 활용한 한국 프로야구 선수들의 효율성 분석」. 『문화경제연구』 15권 2호 (2012, 8): 33-56.
- 오광모·이장택. 「데이터마이닝을 이용한 한국 프로야구 선수들의 연봉에 관한 모형연구」. 『한국스포츠사회학회지』 16권 2호 (2003, 11): 295-309.
- 오태연·이영훈. 「메이저리그야구 선수노동시장구조와 연봉 적절성」. 『한국스포츠산업경

8) 이는 기존 WAR에서도 나타나는 현상이다.

- 영학회지』 18권 3호 (2013, 5): 1-15.
- 이장택. 「한국프로야구에서 타자능력의 측정」. 『한국데이터정보과학회지』 25권 2호 (2014. 4): 349-356.
- Fort, Rodney, Kang, Joon-Ho and Young Hoon Lee. “KBO and international sports league comparisons.” In *The Sports Business in The Pacific Rim*. edited by Young H. Lee and Rodney Fort, pp. 175-193. Springer International Publishing, 2015.
- Jang, Hayley and Young Hoon Lee. “A Business Analysis of Asian Baseball Leagues.” *Asian Economic Policy Review* 11 (1) (January 2016): .95-112.
- James, Bill. *The New Bill James Historical Baseball Abstract*, Free Press; Revised Edition (June 2000).
- Kang, Joon Ho., Lee, Young Han, and Sihyeong Kwon. “Evaluating Management Efficiency of Korean Professional Baseball Teams using Data Envelopment Analysis.” *International Journal of Sport and Health Science* 5 (November 2007): 125-134.
- Krautmann, Anthony C. “What’s Wrong with Scully-Estimates of a Player’s Marginal Revenue Product.” *Economic Inquiry* 37 (2) (April 1999): 369-381.
- Rockerbie, Duane W. “Marginal Revenue Product and Salaries: Moneyball Redux.” *JEL Working Papers* (2010).
- Scully, Gerald W. “Pay and Performance in Major League Baseball.” *The American Economic Review* 64 (6) (December 1974): 915-930.
- Zimbalist, Andrew. *Baseball and Billions*, New York: Basic Books, (April 1992).

abstract

Value Evaluation Model for Korean Professional Baseball Players

Taeyeon Oh · Young Hoon Lee

The purpose of this study is to establish evaluation model that can explain marginal effects of baseball players of Korean Baseball Organization (KBO) on their team winnings in terms of productivity. We proposed econometric model with using variables that are provided from official homepage of KBO to overcome the complexity of mainly used productivity index: wins above replacement (WAR). Also, compare to the previous studies such as Scully(1974) or Krautmann(1999) that using limited indices of baseball stats, this study included 61 indices that are provided from official homepage of KBO. We estimated regression based WAR(RBWAR) by conducting panel regression with each team's statistics data of 2002 to 2014. As a results, RBWAR shows 0.869 correlation coefficient for batters and 0.882 for pitchers with WAR in 2014 that can be concluded that two indices shows similar results. From the results of estimation, we analyze the relationship between productivity and actual contract of free agent players in 2015 and it showed that teams have contracted reasonably.

Keywords: Korean Professional Baseball, athletes' labor market, marginal productivity, wins above replacement, panel data