Contract Year Phenomenon: Is that a true thing?

It's not uncommon to hear Fantasy Baseball enthusiasts say this: "Hey I just added this guy who's becoming a free agent at the end of the season, dude's gonna have a breakout year." Not just fantasy team managers, there is even a Wikipedia page for contract year phenomenon. The definition goes like this: "Contract year phenomenon is a term used in North American sports to describe the occurrence when athletes perform at a very high level in the season prior to their free agency eligibility. Most often, these athletes have seasons that are statistically better than previous years, but then once they sign their new contract, they return to their previous level of performance."

It's safe to say that 99% of the fantasy team managers have no experience working for an MLB team. However it still got me thinking, is it true that players tend to perform better in their contract years? Furthermore, can teams really take advantage of what fantasy managers have been harvesting, at least in their minds?

Fortunately, ESPN tracks the list of free agents every year. Well, it's a pretty long list. You'll be surprised to find there are many names that you don't even know. In order to pay more attention to players who were not at the substitute level, I only selected those players who eventually won major league contracts in that offseason. In terms of performance, I checked the stats that belong to the Dashboard tab on FanGraphs(G, PA, HR, R, RBI, BB%, K%, BABIP, AVG, OBP, SLG, wOBA, wRC+, WAR for batters; W, L, G, IP, K/9, BB/9, HR/9, LOB%, BABIP, HR/FB, vFA (pi), ERA, FIP, xFIP, WAR for pitchers). Data from 2010-2019 were included, to make it better representative of recent trends for the entire decade and to exclude the obviously strange 2020 season.

To make it easier to understand, let's say Player A played from 2004 to 2015 and was a free agent after the 2010 and 2014 seasons. Statistics for 2010 and 2014 were marked as contract years, while only statistics for 2011, 2012, 2013 and 2015 were marked as non-contract years. Then the statistics will be grouped by player. Cumulative statistics such as HR and W were expressed as averages, while non-cumulative statistics such as BB% and ERA are weighted by PA and IP, respectively. For convenience, player A will have only one row in my database, which contains variables such as G_contractyear, G_noncontractyear, HR_contract year, HR_noncontract year, etc.

My method has pros and cons here. By including data from 2010-2019, I get to pay attention to recent trends in the entire league, but at the same time I lost some data that may be important before 2010. In terms of statistics, the average of cumulative statistics takes into account of plate-appearances, to a certain extent, represents the player's health and role in the team, and the weighted average of non-cumulative statistics can represent their overall efficiency.

Figure below shows the whole process for this research:

```
ESPN List of Free Agents
            |
            v
        2010-2019

    Stats from FanGraphs

14 Batting Categories, 15 Pitching Categories
            |
            v
   Overall Contract Year Phenomenon

   498 Batters and 311 Pitchers included
          /                    \
         v                      v
Above Average Players Contract    First Time Contract Year
     Year Phenomenon                   Phenomenon

  66 Batters and 45 Pitchers      369 Batters and 230 Pitchers
        included                        included
```

Overall Contract Year Phenomenon

After a few lines of SQL and Python code, I came up with these four diagrams. The top two represent the batter's data, and the last two represent the pitcher's data.
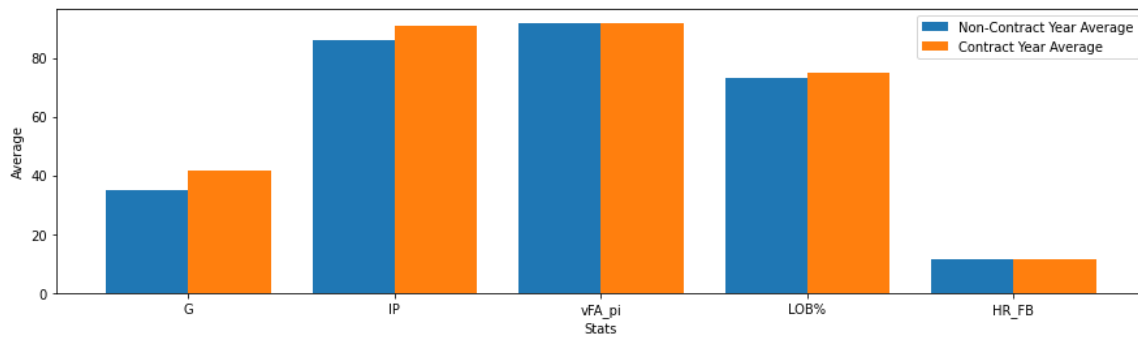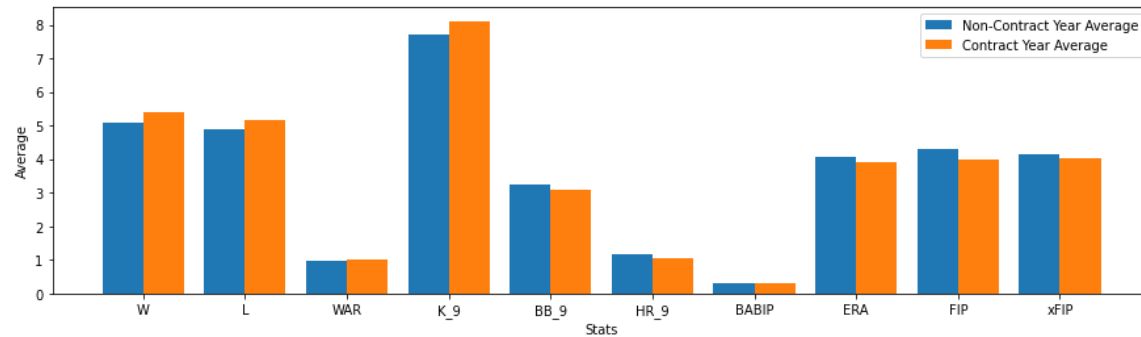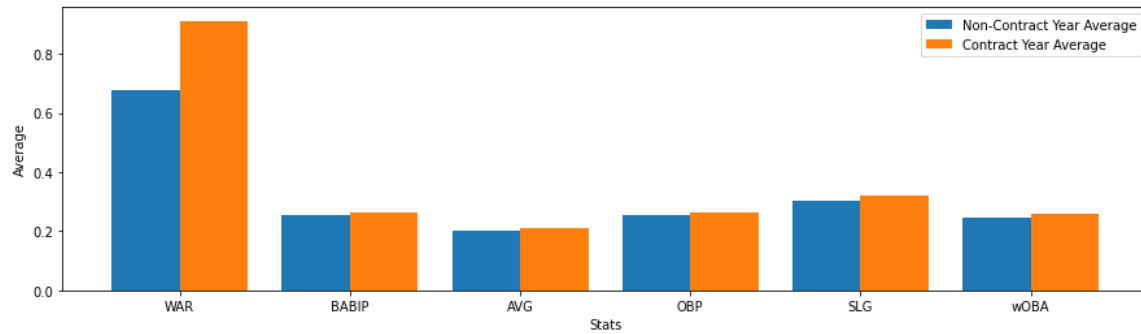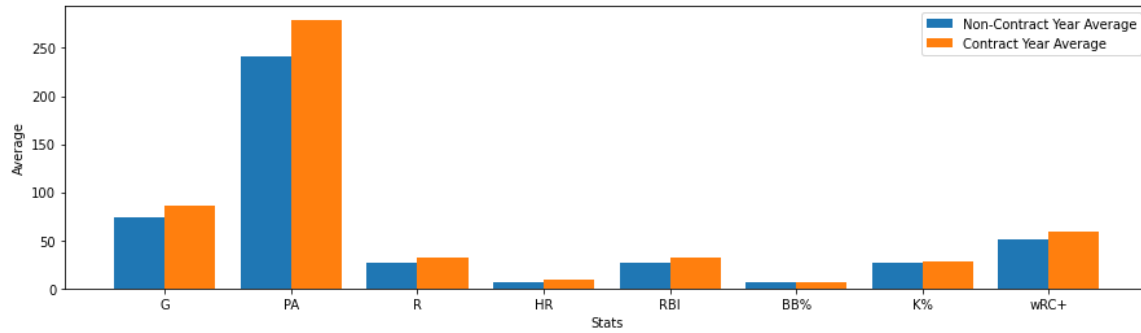
Well, the trend was obvious. Players in their contract years basically performed better on every indicator.

To further prove my findings, I performed paired T tests to see if the changes in the above statistics were considered statistically significant. The paired sample t-test is also called dependent sample t-test. It's an univariate test that tests for a significant difference between 2 related variables. P-value is the probability that your null-hypothesis happens. In this case, I come up with a null hypothesis that the players in non-contract years have the same performance as the players in contract years. I want to prove that it is wrong with the stats. I come up with a p-value of 0.05. It means that there is only 5% chances players in non-contract years is the same as those in contract years, which kind of proves that it is actually not, cause 95% chances indicate that. A lot of research use 0.05 as the threshold for significancy. Basically it's the smaller the better your research proves to be significant.

Stats	NonContract	Contract	Statistics	P_Value
G	74.165724	85.698963	9.741856	1.22E-20
PA	241.502473	279.29781	7.483446	3.31E-13
R	27.040565	32.493912	7.544062	2.19E-13
HR	6.429469	8.514193	8.260043	1.33E-15
RBI	26.30174	31.993613	7.647353	1.07E-13
wRC	51.278994	58.835882	2.674219	7.74E-03
WAR	0.677907	0.91241	4.250618	2.55E-05
BB%	6.453519	6.562512	0.359829	7.19E-01
K%	26.651119	27.842981	1.666383	9.63E-02
BABIP	0.253702	0.264058	1.511601	1.31E-01
AVG	0.200589	0.209072	1.826976	6.83E-02
OBP	0.255141	0.265359	2.053723	4.05E-02
SLG	0.303321	0.321467	3.124031	1.89E-03
wOBA	0.246536	0.257501	2.492008	1.30E-02

There were 498 batters in my database. The results showed that the performance was mostly better and statistically significant.

Stats	NonContract	Contract	Statistics	P_Value
G	35.038898	41.77074	7.661557	2.37E-13
IP	86.173957	91.10676	2.424646	1.59E-02
vFA_pi	91.882849	91.85714	-0.23341	8.16E-01
LOB	73.330325	75.10588	2.494935	1.31E-02
HR_FB	11.525566	11.39392	-0.18257	8.55E-01
W	5.095377	5.382208	1.664473	9.70E-02
L	4.903079	5.153912	1.601277	1.10E-01
WAR	0.98918	1.027594	0.662489	5.08E-01
K_9	7.714602	8.106902	3.686442	2.68E-04
BB_9	3.226853	3.077709	-1.30717	1.92E-01
HR_9	1.167332	1.051551	-1.0667	2.87E-01
BABIP	0.289773	0.291915	0.593476	5.53E-01
ERA	4.085802	3.900438	-1.4199	1.57E-01
FIP	4.285019	3.98219	-1.39399	1.64E-01
xFIP	4.13028	4.031046	-0.88389	3.77E-01

As for the 311 pitchers in my database, the performance was better as well, albeit only G, IP, LOB, W and K/9 were statistically significant with the p-value<0.05.

Above Average Players Contract Year Phenomenon

In addition to analyzing all the players who met my criteria, I wanted to see if players above average had different contract year phenomenon. I set the standard for players who averaged at least 2 WARs in non-contract years. Only 50 batters and 33 pitchers were selected in my database.

Stats	NonContract	Contract	Statistics	P_Value
G	126.013149	125.5202	-0.13997	0.88912
PA	519.457816	513.46338	-0.3728	0.71051
R	66.823749	65.282828	-0.58738	0.558986
HR	17.448882	18.299242	0.858309	0.393876
RBI	63.827122	64.017677	0.067472	0.946413
wRC	115.097683	110.66946	-1.85727	0.067804
WAR	2.943865	2.539773	-1.82263	0.072959
BB%	9.254618	9.654484	1.753163	0.084289
K%	18.327593	18.463219	0.381167	0.704323
BABIP	0.30409	0.292875	-3.2115	0.002054
AVG	0.269883	0.26212	-2.37043	0.020744
OBP	0.343066	0.339511	-1.06165	0.292323
SLG	0.448589	0.443281	-0.70395	0.48398
wOBA	0.342946	0.336985	-1.56891	0.121525

Stats	NonContract	Contract	Statistics	P_Value
G	28.574206	28.314815	-0.2264	0.82194
IP	165.742503	161.22889	-0.63961	0.52574
vFA_pi	91.892623	91.421667	-2.85048	0.00662
LOB	74.067143	73.518911	-0.47083	0.64009
HR_FB	10.271763	12.963656	3.892333	0.00033
W	10.873025	10.32037	-0.95078	0.34691
L	8.459462	8.357407	-0.2089	0.83549
WAR	2.898298	2.608889	-1.07726	0.28724
K_9	8.034448	8.23595	1.269023	0.21111
BB_9	2.530825	2.729561	1.445202	0.15549
HR_9	0.926038	1.169156	3.563357	0.0009
BABIP	0.289194	0.296197	1.771649	0.08338
ERA	3.604338	4.021925	1.837465	0.0729
FIP	3.598387	4.015119	3.038472	0.00399
xFIP	3.698088	3.925192	2.582694	0.01321

Due to the small number of cases, the results were not very significant, and the p-value of most statistics exceeds 0.05. The trend also showed that in this group of players, the contract year phenomenon almost disappeared.

First Time Contract Year Phenomenon

Finally, I wanted to exclude players who have contract years after the age of 32. Overall I felt this was more representative of the phenomenon of the first (at least first few) contract year of a player.

Stats	NonContract	Contract	Statistics	P_Value
G	74.216441	85.391399	7.997615	1.66E-14
PA	244.972215	282.69041	6.36714	5.74E-10
R	27.642945	33.100323	6.397261	4.81E-10
HR	6.699706	8.865563	6.911902	2.12E-11
RBI	26.847572	32.351278	6.1563	1.95E-09
wRC	52.795648	58.354844	1.656824	9.84E-02
WAR	0.721626	0.978512	3.981981	8.24E-05
BB%	6.626296	6.489186	-0.36714	7.14E-01
K%	26.975592	28.062554	1.308708	1.91E-01
BABIP	0.257278	0.264653	0.905713	3.66E-01
AVG	0.202127	0.207964	1.027238	3.05E-01
OBP	0.257778	0.263684	1.00408	3.16E-01
SLG	0.307179	0.322259	2.136142	3.33E-02
wOBA	0.249233	0.256803	1.45028	1.48E-01

Stats	NonContract	Contract	Statistics	P_Value
G	34.583337	40.36826	5.821387	1.96E-08
IP	89.443031	93.25227	1.494888	1.36E-01
vFA_pi	92.208971	92.17112	-0.27061	7.87E-01
LOB	73.520286	74.34918	0.988249	3.24E-01
HR_FB	10.872503	11.91901	2.465681	1.44E-02
W	5.251339	5.41	0.802082	4.23E-01
L	5.126106	5.269275	0.75324	4.52E-01
WAR	1.053209	1.074022	0.285901	7.75E-01
K_9	7.77366	8.107072	2.596366	1.00E-02
BB_9	3.206777	3.055222	-1.06742	2.87E-01
HR_9	1.134266	1.072296	-0.47862	6.33E-01
BABIP	0.290541	0.294941	1.121445	2.63E-01
ERA	3.961487	4.052734	0.728181	4.67E-01
FIP	4.226844	4.003018	-0.8165	4.15E-01
xFIP	4.13508	4.01449	-0.80991	4.19E-01

There were 369 batters and 230 pitchers in my database. The results were a bit like the database of all players, the results shown by the batters are more important, and the results shown by the pitchers are less.

Conclusions

All in all, no matter from the perspective of the entire league or from the first contract year, all players performed somewhat better in their contract years, and the performance of batters was more significant. As for the above-average players, the contract year phenomenon disappeared. From the perspective of the team or the fantasy manager, it might be wiser to bet on average/below average players who are in their contract years, because these players might be the ones to save the team.