

Expansion: does it add muscle or fat?

by
David W. Smith
June 26, 1999

Many things in baseball generate mixed feelings and opinions, but expansion creates especially strong reactions. On the one hand, communities have given consistent civic support in the competition for an expansion club and a successful effort is seen as a boon both financially and in terms of Major League status. On the other hand, expansion has always brought with it the criticism that the quality of play has been lowered by the addition of new players and teams. Perhaps the most frequent expression of this position is a phrase something like: "pitching has been diluted by the addition of so many rookies who would have still been in AAA without expansion". This conventional wisdom is often coupled to implicit criticisms of individual achievements in expansion seasons, such as the 61 home runs by Roger Maris in 1961 and the 70 by Mark McGwire last year. I am always suspicious of pronouncements like that, especially since they are rarely accompanied by any sort of rigorous evidence. I therefore decided to analyze the dilution question by using play by play data to examine the actual performance of rookies.

I decided it was important to cover all the expansion seasons, since there could be differences between them. I also wanted to make comparisons between expansion and non-expansion years. I settled on the last 40 consecutive seasons, 1959-1998, for my study. This period covers all the expansion years plus a two year lead in to the first one. The data for my work came from Retrosheet and Total Sports (see appendix for more details). Table 1 is a summary of the games examined:

	Games	Plate Appearances
AL	39,504	3,021,354
NL	36,373	2,765,335
Total	75,877	5,786,689

There were actually 76508 games played during these 40 seasons, which means we are missing accounts for 631 of them. Therefore, I had 99.2% of the possible games available for my analysis. Of the 40 seasons, we have all the games for 25 of them. The season with the poorest coverage is 1967, where we have 95.2%. There are more details on the missing games in the appendix, but I must offer one final note. Of the 631 missing games, 8 are from the AL and 623 are NL, a pretty amazing imbalance.

My fundamental approach begins by categorizing each batter and each pitcher as a rookie or as a veteran for each season. Of course, all players in their first years are rookies, but I

decided to use the definition of a rookie that determines eligibility for the rookie of the year award. For batters that means a player is still considered a rookie in a given season if he had no more than 130 at bats in previous years. For pitchers the standard is no more than 50 innings pitched. These values have been different over the years, but have been stable since 1971. There is also a criterion of fewer than 45 days on a Major League roster, but that information was not available to me and I used the appearance data to make the decisions. Therefore, some players were classified as rookies in two seasons; I decided not to extend it to a third, since very few would meet the service time limit. I can best show the consequences of this definition with two examples. First, Scott Rolen of the Phillies had exactly 130 at bats in 1996 and his season ended when he was hit by a pitch and suffered a broken wrist on September 7. Therefore, he was still a rookie in 1997 and in fact he won the 1997 NL rookie of the year award. I counted him as a rookie in both 1996 and 1997 for my analysis. The second example considers all players. In 1998, 208 players made their Major League debuts, but there were 321 players I categorized as rookies.

After players are categorized, then each plate appearance falls into one of four categories as shown in Table 2 with some summary data from 1959-1998:

	Plate Appearances	Percentage
Rookie Batters	597,514	10.3
Veteran Batters	5,189,175	89.7
Rookie Pitchers	825,965	14.3
Veteran Pitchers	4,960,724	85.7

Of course, it is not merely appearances that are important to my analysis, but performance. There are many ways to evaluate performance, but I chose the one that is pretty conventional in Sabermetric circles these days, namely on-base average plus slugging average, or OPS. Table 3 presents the OPS for each of group of player:

	OBA	SA	OPS
Rookie Batters	.302	.349	.651
Veteran Batters	.326	.391	.717
Rookie Pitchers	.337	.395	.732
Veteran Pitchers	.322	.386	.708

For pitchers, the values are what they allowed.

There are wide differences here, with veterans having the better of rookies in all cases. However, we will see in a few minutes that rookie batters and rookie pitchers not equally disadvantaged.

Before we get into the specific numbers in detail, it is important to have some sense of what a good OPS value is and what a poor one is. Since my topic is expansion and the possible effects of new players on expansion seasons, it is also necessary to see how this measurement may vary from year to year. The first three figures give this information.

Figure 1 presents the OPS for all players for the 40 seasons. The expansion seasons have been marked for convenient reference. There are wide swings here, but the changes in the six expansion years are mixed. Three show sharp increases (1969, 1977, and 1993), one a modest rise (1961), one no change (1998) and one actually saw a decrease (1962). Amidst the variation, there does appear a general upward trend from the low point of 1968, the famous "year of the pitcher".

Figure 2 gives the same information, but for the AL only. The patterns of changes are similar, but the increase from 1968-forward is much sharper.

Finally, figure 3 shows the NL results. Once again the expansion patterns are the same, but there is a surprising flat shape to the curve from 1971-forward. Although the DH accounts for the overall higher values in the AL, the reason for the different trends between the two leagues is not clear. There are two other points to make about these figures and figure 3 will do to show it. First, note that some big increases were followed by big decreases, especially 1970 to 1971, 1977 to 1978 and 1987 to 1988. This latter drop is especially puzzling, since 1987 was not an expansion year and the drop in 1998 was to levels lower than those that prevailed before the 1987 burst. The 1987 has been seen by many people as an example of a juiced ball, which the manufacturer and the leagues deny, but the changes are

striking. Second, one of the big increases, this one over a three year period from 1992 to 1993 to 1994, was not followed by a drop, but rather by a new plateau that has lasted for the last five years. Although the expansion year of 1993 was in the midst of the rise, the expansion of 1998 was invisible here.

With that general background in mind, let us turn now in detail to the rookie vs veteran comparisons. Figure 4 is a very messy graph that shows the differences between OPS allowed by rookie and veteran pitchers against all batters. One of the reasons the graph is messy is that the two lines overlap substantially. Or, said another way, it appears the rookie and veteran pitchers are not all that different, even though the rookies in general have higher OPS allowed.

Keep that picture in mind and compare it to Figure 5, which shows the differences between rookie and veteran batters against all pitchers. This graph is much less messy because the lines are so clearly different. Or, it appears that the rookie and veteran batters are very much different with the rookies having a much lower OPS.

Remember that the conventional wisdom I referred to at the start only considered talent dilution from having less qualified pitchers. I think the last two figures show that rookie batters are even more disadvantaged than pitchers are. However, I also think it is hard to look at these graphs and decide very concretely how much worse off the batters are. Therefore, I created Figure 6. The lines here are simply the composites from the two previous figures, formed by making ratios. That is, the top line is the pitcher data from figure 4 with the rookie OPS divided by the veteran OPS. The bottom line is the batter data with the rookie to veteran ratio. A value of 1.0 would mean that the rookies and veterans were equal to each other. Therefore, the greater the distance from 1.0, then the greater the difference between rookies and veterans. The data for the pitchers (top line) are very revealing. Although it is generally true that the rookies did worse, the expansion years have no relation to the peaks and valleys. In fact, for 3 of these 40 seasons, the rookie pitchers actually had a better OPS than the veterans! (ratio less than 1.0). For the batters the results are much more dramatic. The rookie batters never even approached the veterans. And, most importantly, once again there is no apparent relation of the relative performance of rookies to expansion.

The last feature I examined was the yearly difference in rookie usage. Figure 7 shows the percentage of plate appearances against rookie pitchers each year. Once again there is lots of noise, but for five of the six expansion seasons, the percentage has a high value, which makes sense. Only 1998 doesn't fit that pattern. The highest value on the graph, way out of line with the others, is for 1995. My interpretation is that the strike of

1994 limited the chance to bring up rookies in September and they appeared in early 1995 instead. You may also recall that the 1995 season missed some games at the start and there was concern for injuries to pitchers, which may also contribute to the increased rookie percentage as each individual pitcher was used less. I did not check the number of pitchers used per game during early 1995 as compared to the rest of the year, which would help to address this point.

Figure 8 presents the same percentage information for batters. There are peaks corresponding to the first four expansion, with 1993 and 1998 not fitting. Note that the 1995 burst we saw for pitchers did not occur for batters, lending strength to my argument about injury concern for pitchers.

Before I make my conclusions, I must mention four of the many other factors that I have not mentioned which can affect the expansion analysis. These are summarized in Table 4.

Table 4. Other Factors Affecting Expansion

1. Population Size
2. Rule Changes
3. Park Factors
4. Veterans Hanging on

1. Population Size

The dilution argument misses the changing demographics of the U.S. Not only is the population much higher than it was 40 years ago, there is now wider participation from groups of people who were previously excluded. Not only is there more complete inclusion of African-Americans, but there is also much greater presence of Latin-Americans and recently of Asians. However, a force in the other direction is the fact that the other professional sports offer more competition for talented young athletes than was true 40 years ago. On balance, it appears that the available population base is larger than it has ever been.

2. Rule Changes

Rule changes are usually subtle, but two of the biggest ones in the last half century were the changing of the strike zone in 1963 and 1969 and the mound in 1969, in perfect coordination with expansion. There have been other strike zone edicts, including the current year, but there is general agreement that the high strike was actively called in the mid-1960s and the changed height of the mound can't really be debated.

3. Park Factors

Expansion brings new parks into the Major Leagues. In the recent expansions, most of these new parks have been pretty neutral, with a few of them favoring the pitcher, but this was not always the case. Consider 1961 when the AL added Wrigley Field in Los Angeles and Metropolitan Stadium in Minnesota. The next year the NL added Colt Stadium in Houston and reactivated the Polo Grounds in New York. The net effect of these four parks was to favor the batter. In the other direction in the 1962 NL expansion, the Dodgers moved from the Coliseum to Dodger Stadium, a change which hugely favored the pitcher. The 1969 expansion saw the use of Sick's Stadium in Seattle, Jarry Park in Montreal, San Diego Stadium, and the reopening of Municipal Stadium in Kansas City. The first two of these were definite hitter's parks, but the latter two favored the pitchers. Of course, the granddaddy of all park effects in favor of the hitters came with the two parks in Denver.

4. Veterans Hanging on

It is reasonable to suggest that some veterans contemplating retirement will delay that decision in an expansion year. This raises the possibility of another kind of talent dilution. Unfortunately these players are not as easy to identify as rookies. For example, players may retire for injury reasons (Herb Score, Sandy Koufax, Tony Conigliaro, Kirby Puckett) and not just because of diminished performance. I considered a few ways to sort this out, but didn't like any of them, so I am not presenting information on such players. However, I thought it valuable to point out the question.

Conclusions

1. Overall offense can vary greatly from year to year, with moderate relation to expansion.
2. Rookie usage can vary greatly from year to year with moderate relation to expansion.
3. Rookies perform at lower levels than veterans, but rookie batters do worse in relative terms than rookie pitchers.
4. Rookie performance can vary greatly from year to year, with WEAK relation to expansion.
5. Conventional wisdom about diluted pitching due to expansion is not supported by the data.

Appendix

Play by play data come from Retrosheet (1959-1983) and Total Sports (1984-1998). More information about Retrosheet can be found at: www.retrosheet.org

Games from 1959-1975 needed by Retrosheet as of June 14, 1999
(all games from 1976-1998 are in hand)

Braves	393	(246 Atlanta, 147 Milwaukee)
Pirates	295	
Reds	236	
Astros	177	
Padres	81	
Cubs	39	
Expos	19	
KC A's	8	
Angels	7	
Dodgers	5	
Indians	1	
Mets	1	

Missing games by opponent combinations

Astros-Braves	89
Pirates-Reds	89
Braves-Pirates	68
Braves-Reds	64
Braves-Pirates	65
Padres-Braves	38
Braves-Reds	40
Pirates-Astros	45
Astros-Reds	20
Padres-Astros	20
Cubs-Braves	14
Cubs-Pirates	14
Padres-Reds	12
Padres-Pirates	9
Expos-Braves	9
Angels-KC A's	7
Expos-Reds	5
Reds-Cubs	6
Braves-Cubs	5
Padres-Expos	2
Dodgers-Pirates	3
Dodgers-Astros	2
Pirates-Expos	2
Indians-KC A's	1
Mets-Braves	1
Astros-Expos	1

OPS, All Players, 1959-1998

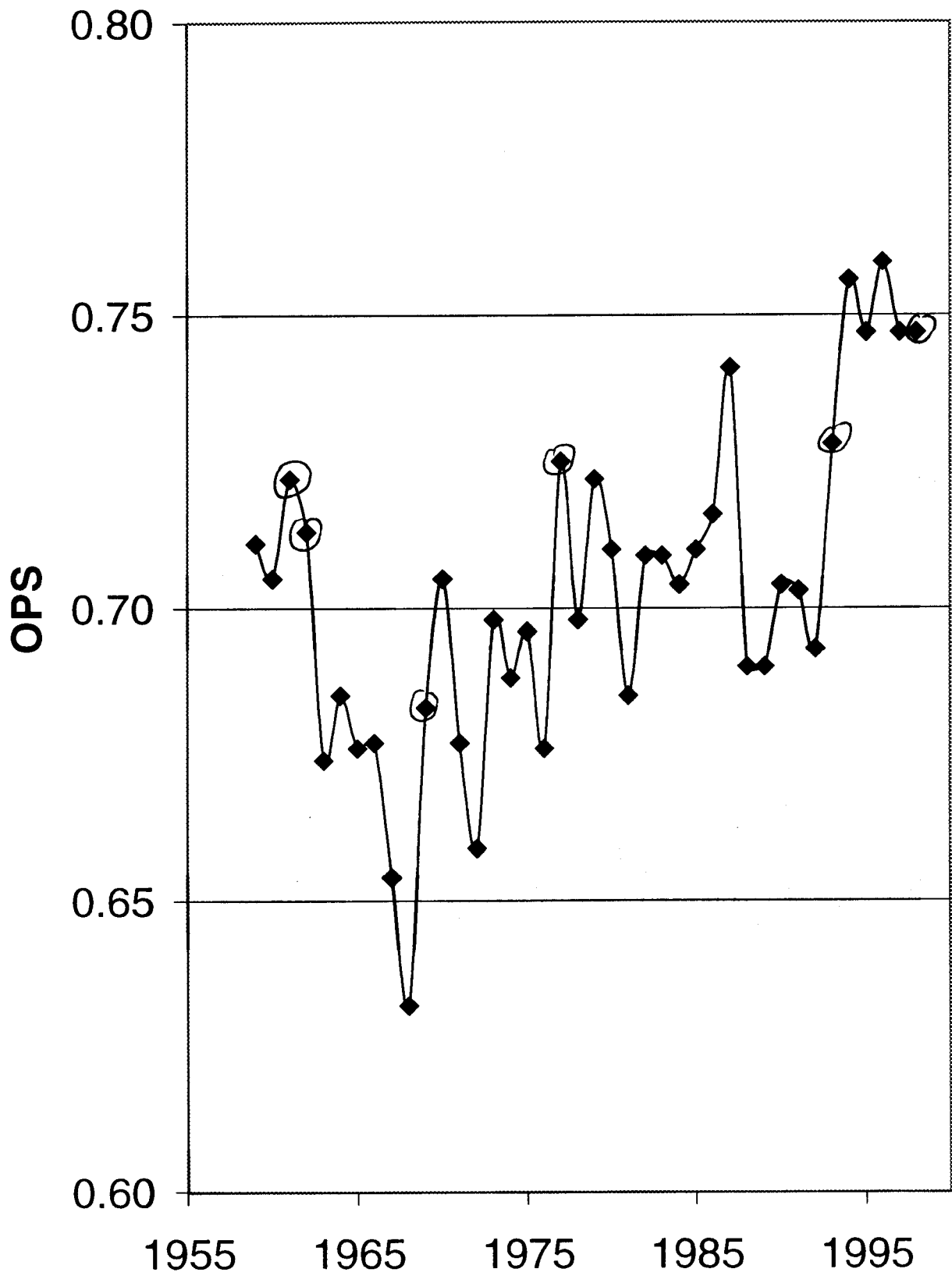


Figure 1

OPS, AL, 1959-1998

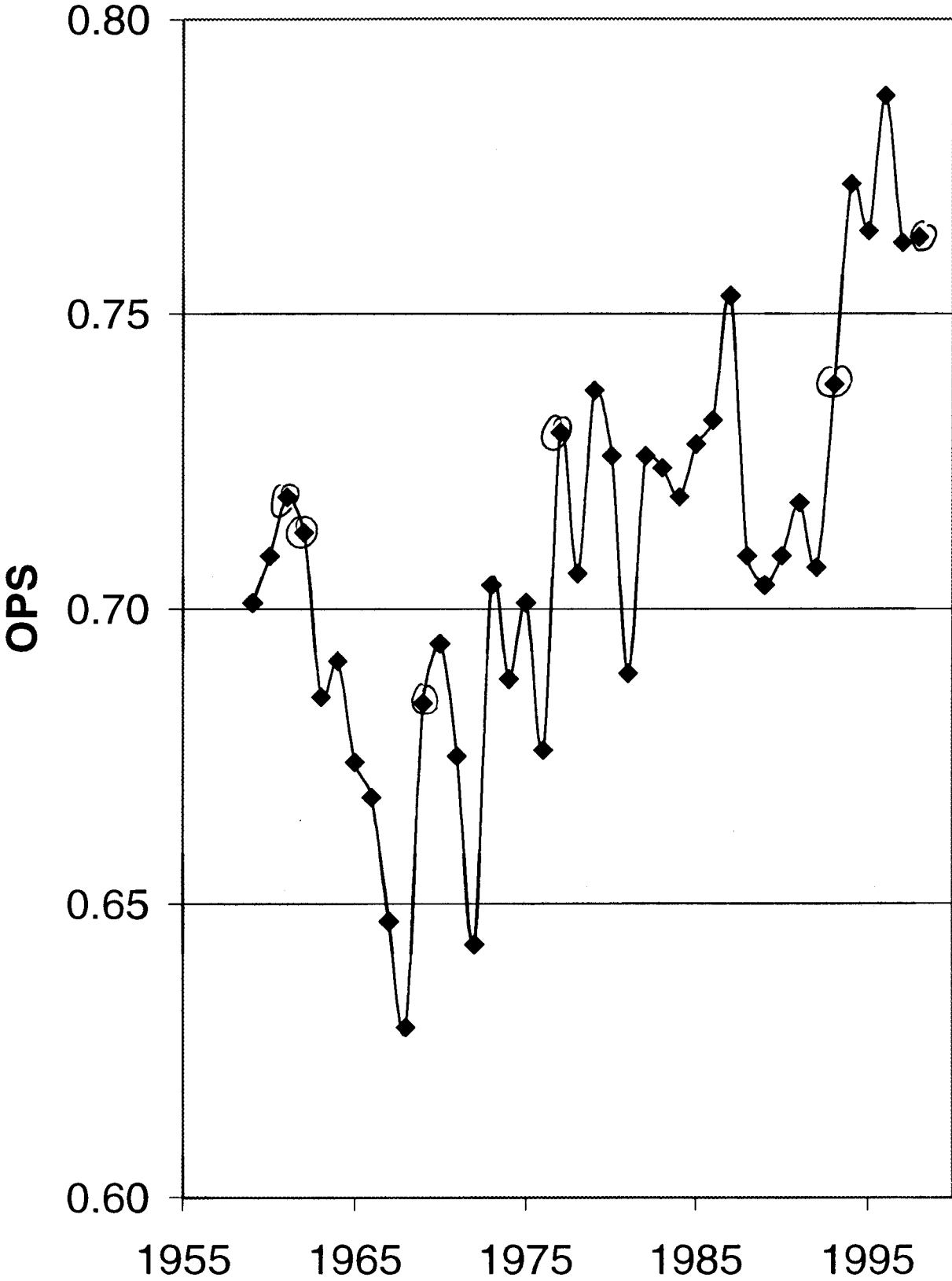


Figure 2

OPS, NL, 1959-1998

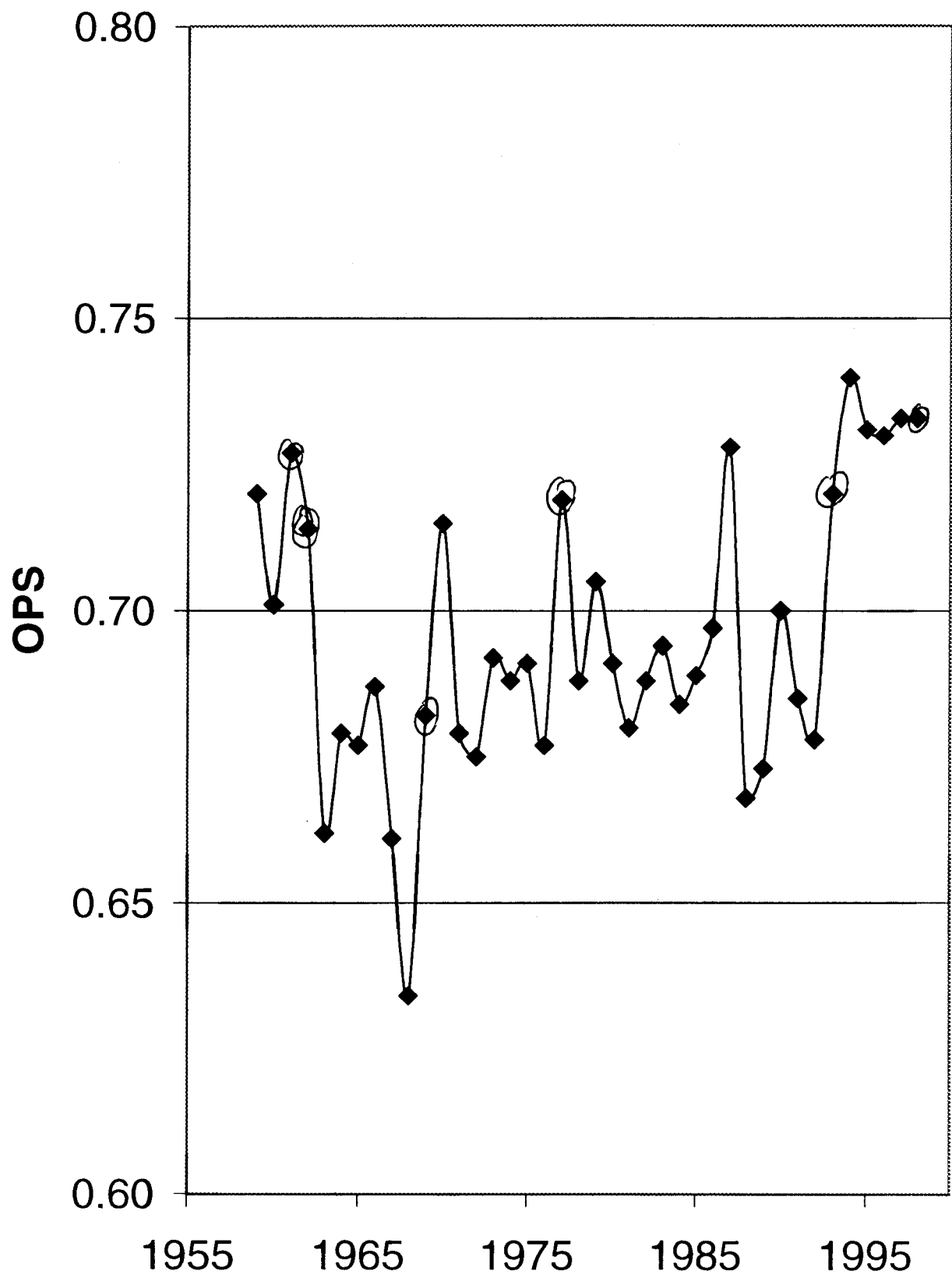


Figure 3

Rookie Pitchers and Veteran Pitchers Against All Batters

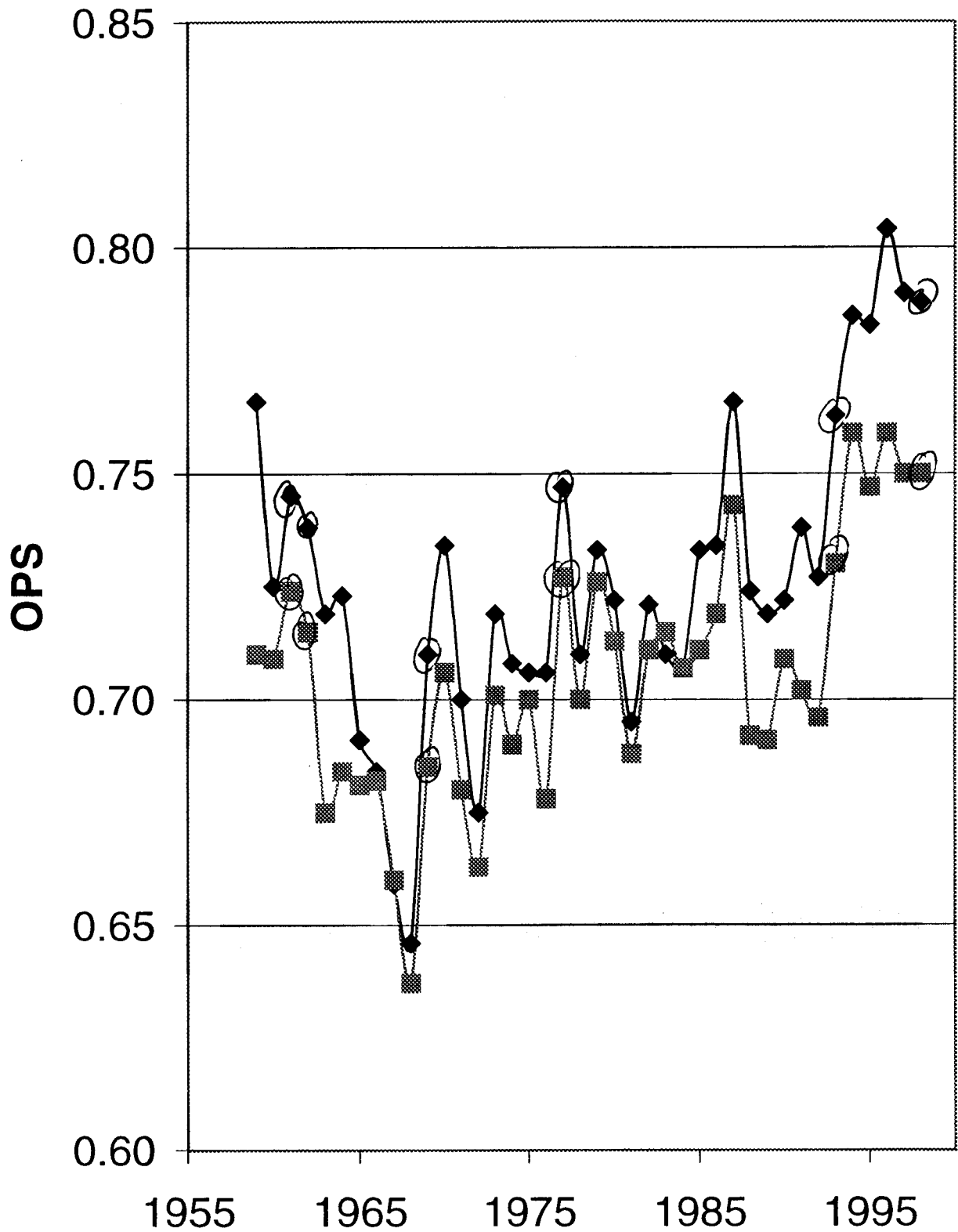


Figure 4

Rookie Batters and Veteran Batters Against All Pitchers

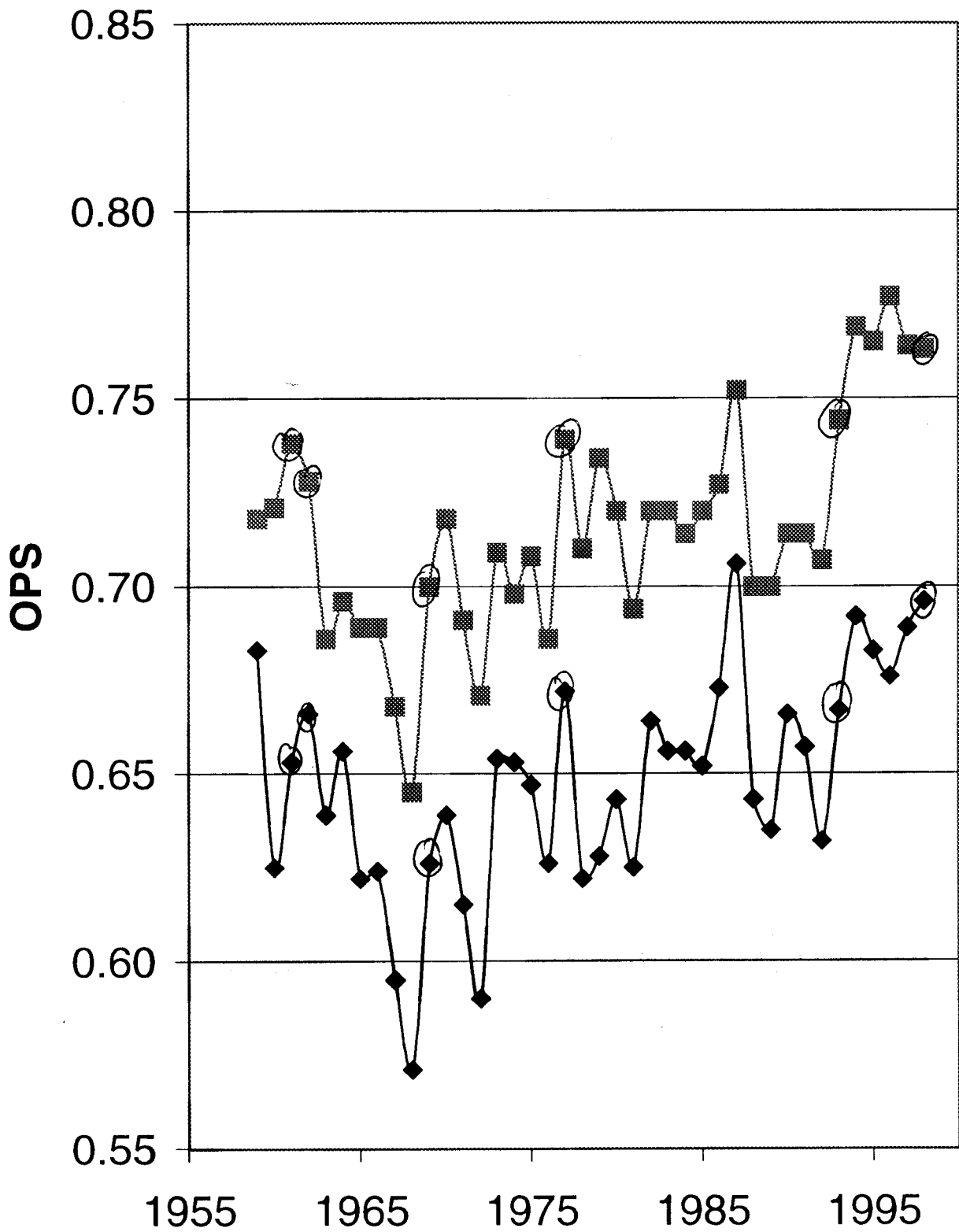


Figure 5

Ratio of Rookie to Veteran OPS for Pitchers and Batters

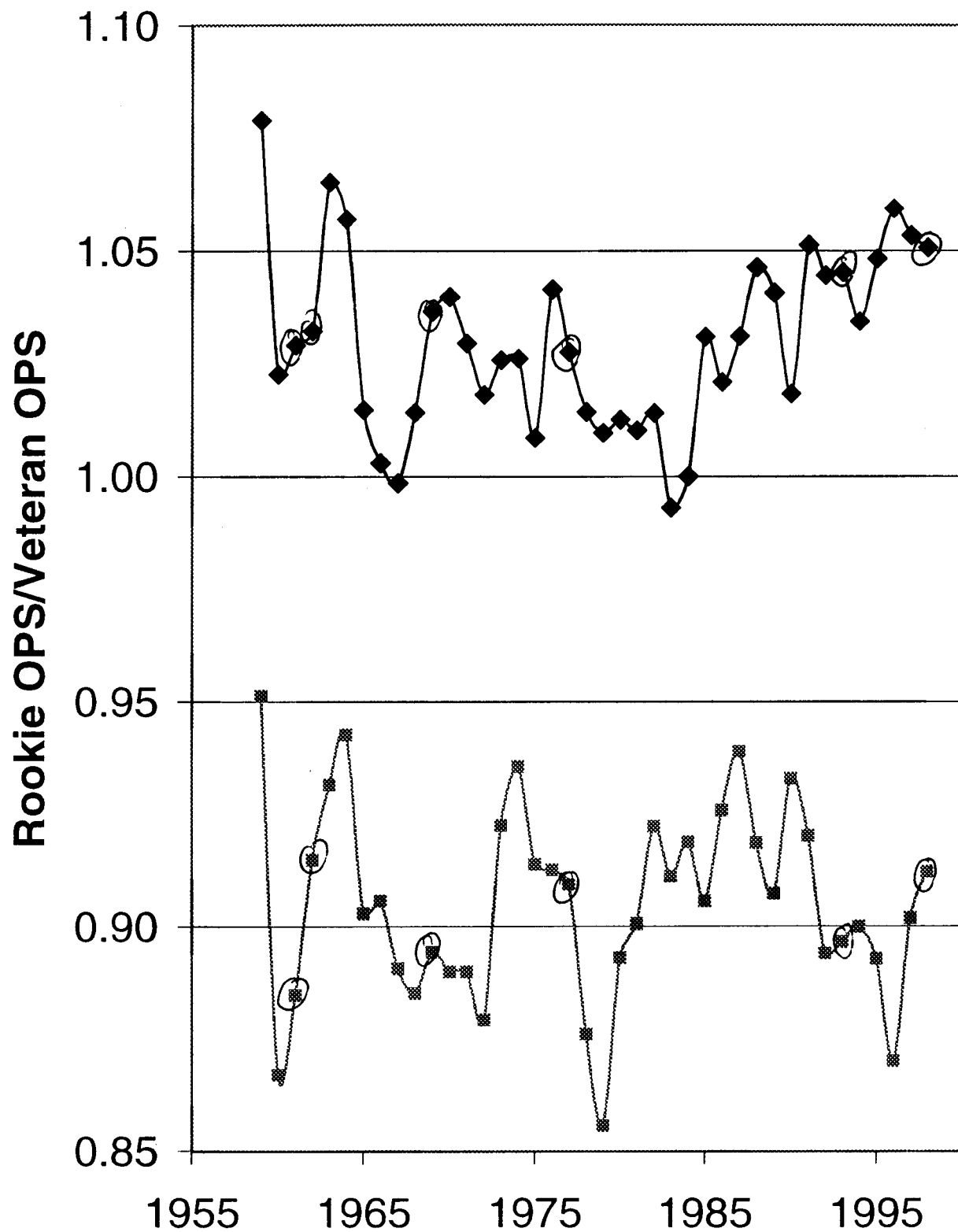


Figure 6

Percentage of Plate Appearances vs Rookie Pitchers

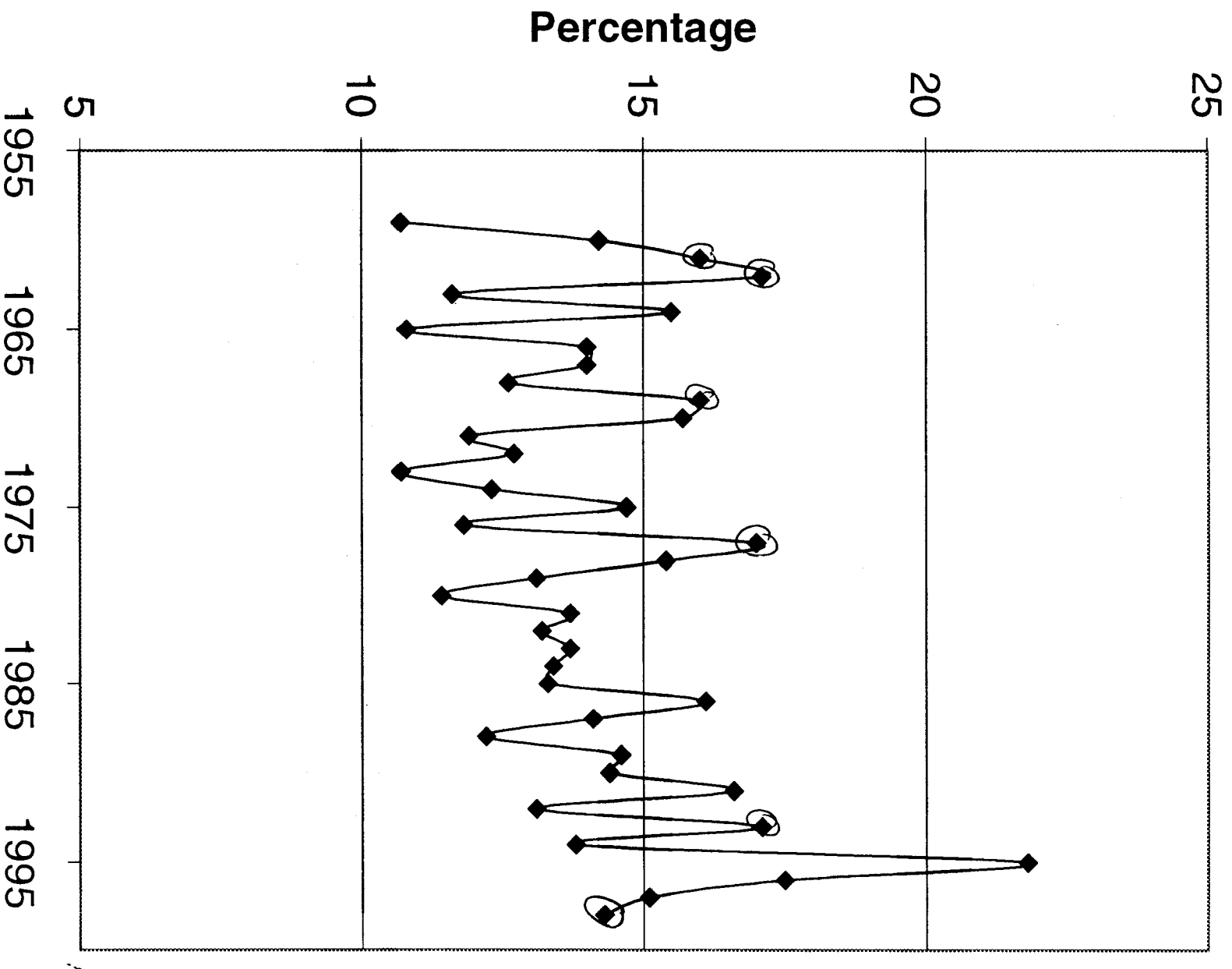


Figure 7

Percentage of Plate Appearances by Rookie Batters

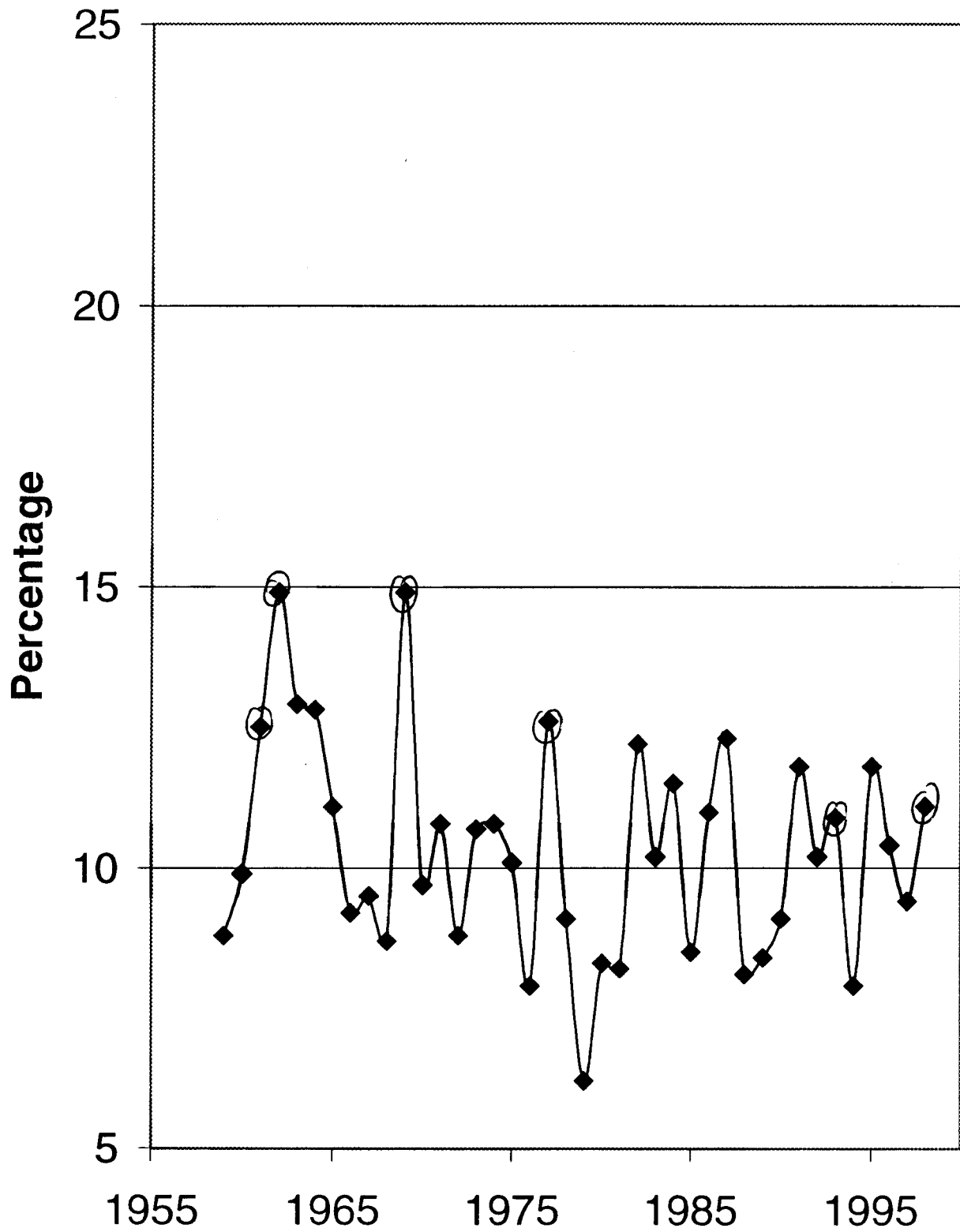


Figure 8

Average Number of Players and Rookies per Team

