Adjusted Home Run Frequencies

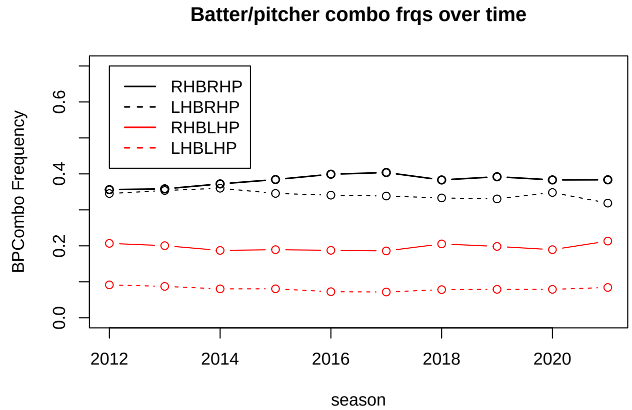
Jason Osborne and Rich Levine

Adjusted hr frequencies, 2012-2021.

Let us see how the batter-pitcher combination (`bpcombo`) frequencies have varied over time

```
##            PIT_HAND_CD
## BAT_HAND_CD          L          R
##           L 0.08054097 0.34107305
##           R 0.19685160 0.38153439

## [1] "by Batter Hand (row sums)"

##        L        R
## 0.421614 0.578386

## [1] "by Pitcher Hand (col sums)"

##         L         R
## 0.2773926 0.7226074

## [1] "Conditionally on Batter Hand"

##            PIT_HAND_CD
## BAT_HAND_CD         L         R
##           L 0.1910301 0.8089699
##           R 0.3403464 0.6596536

## [1] "Conditionally on Pitcher Hand"

##            PIT_HAND_CD
## BAT_HAND_CD         L         R
##           L 0.2903501 0.4720032
##           R 0.7096499 0.5279968

## [1] "All four relative freqs as a vector"

##   BAT_HAND_CD PIT_HAND_CD       Freq
## 1           L           L 0.08054097
## 2           R           L 0.19685160
## 3           L           R 0.34107305
## 4           R           R 0.38153439
```

These four bpcombo frequencies have changed little over time, though the preference for LHB when facing RHP may have decreased slightly.

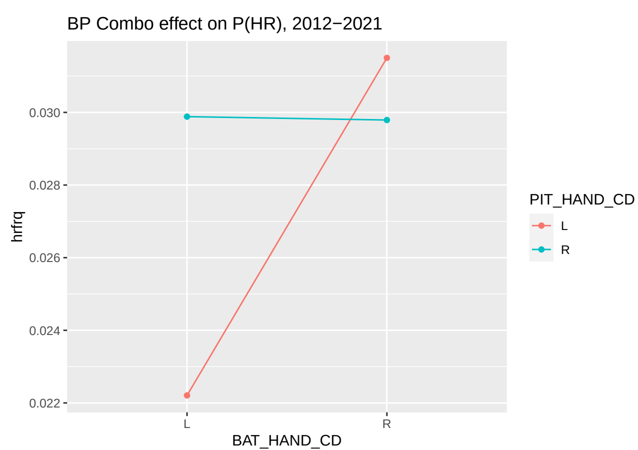


The effects of bpcombo on home run frequency can be investigated with an interaction plot, generated with data from 2012-2021:

```
allyrs.12vars %>% group_by(BAT_HAND_CD,PIT_HAND_CD) %>% summarize(hrfrq=mean(hr)) -> hrfrq.bp.era
hrfrq.bp.era
```

```
## # A tibble: 4 x 3
## # Groups:   BAT_HAND_CD [2]
##   BAT_HAND_CD PIT_HAND_CD  hrfrq
##   <chr>       <chr>        <dbl>
## 1 L           L           0.0222
## 2 L           R           0.0299
## 3 R           L           0.0315
## 4 R           R           0.0298
```

```
ggplot(hrfrq.bp.era,aes(y=hrfrq,x=BAT_HAND_CD,color=PIT_HAND_CD)) +
  geom_line(aes(group=PIT_HAND_CD)) + geom_point() +
  ggtitle("BP Combo effect on P(HR), 2012-2021")
```



Looking over this 10 year period, it can be seen that home runs are least likely when a LHB is facing a LHP. Remarkably, the effect of the batter hand only appears to matter when facing lefties. Wow!

For a given park, the frequencies of the four combinations can vary dramatically from one season to the next, depending upon the personnel of the home team and with the unbalanced schedules of years past, upon the personnel of other teams in the division. In light of bpcombo effects, home run frequencies for each park can be adjusted to league-wide bpcombo frequencies simply by reweighting the four conditional home run rates to the these frequencies.

```
allyrs.12vars %>% group_by(park,BAT_HAND_CD,PIT_HAND_CD) %>%
  summarize(hrfrq=mean(hr)) %>% pivot_wider(values_from=hrfrq,
                                names_from=c("BAT_HAND_CD","PIT_HAND_CD")) ->
  hrsummary.wide
hrsummary.wide %>% mutate(adjhr=bpfrqs.era[1,1]*L_L + bpfrqs.era[1,2]*L_R +
                                bpfrqs.era[2,1]*R_L + bpfrqs.era[2,2]*R_R) ->
  hrsummary.wide
# unadjusted
allyrs.12vars %>% group_by(park) %>% summarize(obshr=mean(hr)) -> hrfrqs.bypark
hrsummary.wide %>% inner_join(hrfrqs.bypark) -> hrsummary.wide
```

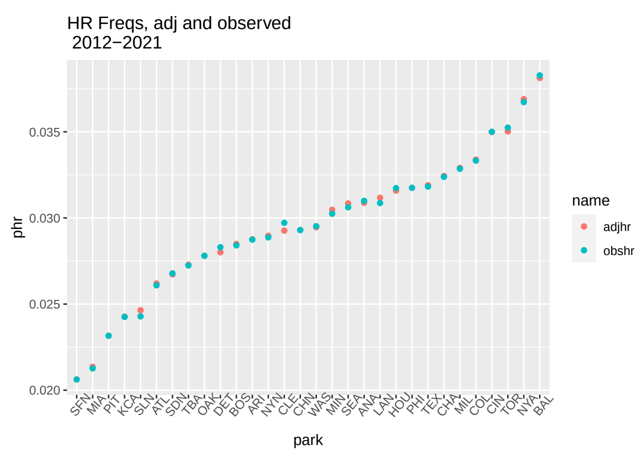
These adjusted frequencies can be plotted against park, along with the unadjusted frequencies. Further investigation of changes over time is warranted though, as a glance at 10-year averages still shows considerable variability in bpcombo frequencies across parks. It must be kept in mind that many players reside with the same team for long periods of time, so these 10 years are not at all independent. However, we average anyway ...

A technique worth mentioning in the construction of this plot is to achieve an ordering of parks on the horizontal axis according to either the observed or adjusted home run rate by so ordering the levels of park as a factor.

```
hrsummary.wide$park <- factor(hrsummary.wide$park,
                              levels=hrsummary.wide$park[order(hrsummary.wide$adjhr)])
hrsummary.wide %>% pivot_longer(cols=c("adjhr","obshr"),values_to="phr") ->
  hrsummary.tall
```

Now `ggplot` can be used ...

```
ggplot(hrsummary.tall) + geom_point(aes(y=phr,x=park,color=name)) +
  theme(axis.text.x=element_text(angle=50)) +
  ggtitle("HR Freqs, adj and observed \n 2012-2021")
```



Ok, let us consider those teams for which the observed and adjusted hr freqs were different.

```
hrsummary.wide %>% mutate(adjmnt=adjhr-obshr) %>%
  arrange(abs(adjmnt)) -> hrsummary.wide; hrsummary.wide %>% tail
```

```
## # A tibble: 6 x 8
## # Groups:   park [6]
##   park     L_L    L_R    R_L    R_R  adjhr  obshr    adjmnt
##   <fct>  <dbl>  <dbl>  <dbl>  <dbl>  <dbl>  <dbl>     <dbl>
## 1 TOR   0.0252 0.0330 0.0359 0.0384 0.0350 0.0352 -0.000223
## 2 MIN   0.0187 0.0275 0.0352 0.0332 0.0305 0.0302  0.000225
## 3 DET   0.0218 0.0261 0.0342 0.0278 0.0280 0.0283 -0.000294
## 4 LAN   0.0266 0.0340 0.0294 0.0305 0.0312 0.0309  0.000307
## 5 SLN   0.0184 0.0243 0.0312 0.0229 0.0246 0.0243  0.000357
## 6 CLE   0.0184 0.0322 0.0290 0.0291 0.0293 0.0297 -0.000453
```

These differences between observed relative frequencies are small, but the number of plate appearances is large:

```
allyrs.12vars %>% group_by(park) %>% summarize(pa=n(),hr=sum(hr)) %>%
  inner_join(hrsummary.wide) %>% mutate(hrdiff=adjmnt*pa) %>%
  arrange(abs(hrdiff)) -> hrsummary.wide ; hrsummary.wide %>% tail
```

```
## # A tibble: 6 x 11
##   park     pa    hr    L_L    L_R    R_L    R_R  adjhr  obshr    adjmnt hrdiff
##   <chr> <int> <int>  <dbl>  <dbl>  <dbl>  <dbl>  <dbl>  <dbl>     <dbl>  <dbl>
## 1 TOR   57373  2022 0.0252 0.0330 0.0359 0.0384 0.0350 0.0352 -0.000223  -12.8
## 2 MIN   58421  1767 0.0187 0.0275 0.0352 0.0332 0.0305 0.0302  0.000225   13.1
## 3 DET   57347  1623 0.0218 0.0261 0.0342 0.0278 0.0280 0.0283 -0.000294  -16.9
## 4 LAN   56243  1736 0.0266 0.0340 0.0294 0.0305 0.0312 0.0309  0.000307   17.2
## 5 SLN   56990  1384 0.0184 0.0243 0.0312 0.0229 0.0246 0.0243  0.000357   20.4
## 6 CLE   57104  1697 0.0184 0.0322 0.0290 0.0291 0.0293 0.0297 -0.000453  -25.9
```

Each of the six teams that have the largest absolute adjustment have an extreme value either for proportion of LHB or proportion of LHB. First the parks that host the *fewest* plate appearances by LHB:

```
allyrs.12vars %>% select(park,BAT_HAND_CD) %>% table %>% prop.table(margin="park") %>%
  as.data.frame() %>% pivot_wider(values_from=Freq,names_from=BAT_HAND_CD) %>%
  arrange(L) -> BHbyPark
BHbyPark %>% head
```

```
## # A tibble: 6 x 3
##   park      L     R
##   <fct> <dbl> <dbl>
## 1 ANA   0.380 0.620
## 2 DET   0.382 0.618
## 3 TOR   0.383 0.617
## 4 CHA   0.390 0.610
## 5 HOU   0.396 0.604
## 6 MIA   0.397 0.603
```

Now for the *most* plate appearances by LHB

```
BHbyPark %>% tail
```

```
## # A tibble: 6 x 3
##   park      L     R
##   <fct> <dbl> <dbl>
## 1 SFN   0.444 0.556
## 2 PHI   0.455 0.545
## 3 NYN   0.457 0.543
## 4 MIN   0.458 0.542
## 5 SEA   0.468 0.532
## 6 CLE   0.519 0.481
```

Note that DET and TOR both see large downward adjustment and host the 2^{nd} and 3^{rd} lowest LHB frequencies at 38%. CLE and MIN see large upward adjustment and host the most and third most LHB, respectively (52% and 46% LHB!).

The other two teams for which adjusments are largest have extremes for plate appearances involving LHP:

```
allyrs.12vars %>% select(park,PIT_HAND_CD) %>% table %>% prop.table(margin="park") %>%
  as.data.frame() %>% pivot_wider(values_from=Freq,names_from=PIT_HAND_CD) %>%
  arrange(L) -> PHbyPark
PHbyPark %>% head
```

```
## # A tibble: 6 x 3
##   park      L     R
##   <fct> <dbl> <dbl>
## 1 MIL   0.206 0.794
## 2 SLN   0.206 0.794
## 3 CIN   0.206 0.794
## 4 CLE   0.215 0.785
## 5 MIA   0.243 0.757
## 6 NYN   0.243 0.757
```

```
PHbyPark %>% tail
```

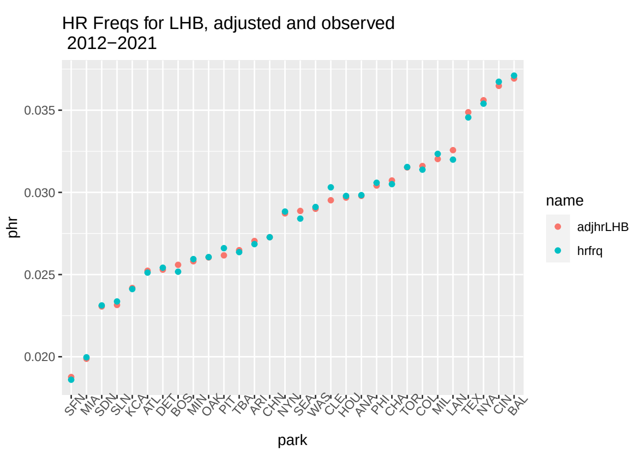
```
## # A tibble: 6 x 3
##   park      L     R
##   <fct> <dbl> <dbl>
## 1 SFN   0.320 0.680
## 2 CHA   0.323 0.677
## 3 TEX   0.326 0.674
## 4 BOS   0.327 0.673
## 5 SEA   0.339 0.661
## 6 LAN   0.362 0.638
```

Dodger Stadium (LAN) has seen the greatest number of plate appearances with a LHP (36%) while Busch (SLN) has seen the second fewest (21%.) The variation in frequency of LHB across parks (38% for ANA up to 52% for CLE) and LHP (21% for MIL up to 36% for LAN) is remarkable. Lineups and rotations are perhaps more stable than one might think given all the personnel changes by high-profile free agents.

HR v park plots separated by batterhand

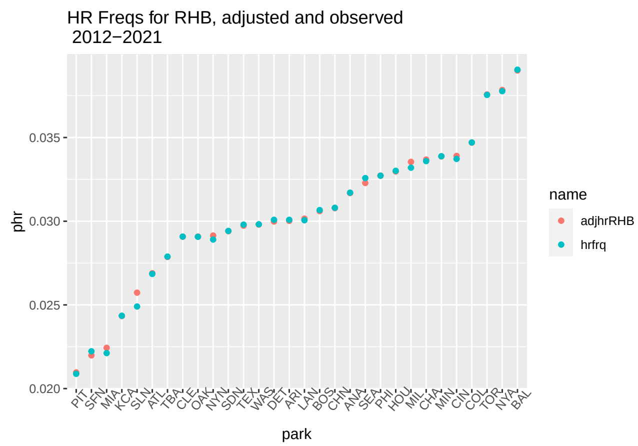
The adjusted HR Frequency for LHB is the weighted average of observed HR frequencies against LHP and RHP, with weights given by the conditionals from `bpcombos` on page 1

```
allyrs.12vars %>% select(BAT_HAND_CD,PIT_HAND_CD) %>% table %>%
  prop.table(margin="BAT_HAND_CD") -> bhtable

# compute observed hr frqs by park
allyrs.12vars %>% group_by(park,BAT_HAND_CD,PIT_HAND_CD) %>%
 summarize(hrfrq=mean(hr)) %>%
  pivot_wider(values_from=hrfrq,
              names_from=c("BAT_HAND_CD","PIT_HAND_CD")) ->
  hrsummary.wide
# compute weighted HR freq for LHB and for RHB
hrsummary.wide %>%
  mutate(adjhrLHB=bhtable[1,1]*L_L+bhtable[1,2]*L_R,
         adjhrRHB=bhtable[2,1]*R_L+bhtable[2,2]*R_R) ->
hrsummary.wide

# hrfrs by hand not adjusted for pitcher hand
allyrs.12vars %>% group_by(park,BAT_HAND_CD) %>%
  summarize(hrfrq=mean(hr)) %>%
  inner_join(hrsummary.wide) -> hrsummary.wide

# make LHB tall
hrsummary.wide %>% filter(BAT_HAND_CD=="L") -> hrsummary.wide.LHB
hrsummary.wide.LHB$park <- factor(hrsummary.wide.LHB$park,
 levels=hrsummary.wide.LHB$park[order(hrsummary.wide.LHB$adjhrLHB)])

hrsummary.wide.LHB %>%
  pivot_longer(cols=c("adjhrLHB","hrfrq"),values_to="phr") ->
  hrsummary.tall.LHB

# plot LHB
ggplot(hrsummary.tall.LHB) + geom_point(aes(y=phr,x=park,color=name))+
  theme(axis.text.x=element_text(angle=50)) +
  ggtitle("HR Freqs for LHB, adjusted and observed \n 2012-2021")
```



Similarly for RHB,

```
hrsummary.wide %>% filter(BAT_HAND_CD=="R") -> hrsummary.wide.RHB
hrsummary.wide.RHB$park <- factor(hrsummary.wide.RHB$park,
 levels=hrsummary.wide.RHB$park[order(hrsummary.wide.RHB$adjhrRHB)])

hrsummary.wide.RHB %>%
  pivot_longer(cols=c("adjhrRHB","hrfrq"),values_to="phr") ->
  hrsummary.tall.RHB

# plot RHB
ggplot(hrsummary.tall.RHB) + geom_point(aes(y=phr,x=park,color=name)) +
  theme(axis.text.x=element_text(angle=50)) +
  ggtitle("HR Freqs for RHB, adjusted and observed \n 2012-2021")
```



Leftover page from earlier draft

Were bpcombo freqs for `CLE` different from those of the rest of the league?

```
allyrs.12vars %>% select(park,BAT_HAND_CD,PIT_HAND_CD) %>% table %>%
  prop.table(margin=c("park")) %>% as.data.frame -> teambpfrqs.tall
teambpfrqs.tall %>% filter(park=="CLE") %>% print
```

```
##   park BAT_HAND_CD PIT_HAND_CD       Freq
## 1  CLE           L           L 0.06957481
## 2  CLE           R           L 0.14536635
## 3  CLE           L           R 0.44986341
## 4  CLE           R           R 0.33519543
```

```
# Glancing back at era frequencies:
bpfrqs.era.vec %>% print
```

```
##   BAT_HAND_CD PIT_HAND_CD       Freq
## 1           L           L 0.08054097
## 2           R           L 0.19685160
## 3           L           R 0.34107305
## 4           R           R 0.38153439
```

For whatever reason, there were considerably more PA involving Pitchers and Batters of the same hand (RHBRHP or LHBLHP) at Jacobs Field, resulting in upward adjustment to era frequencies (so long as CLE conditional HR rates not too different.)

For other teams, we compute team combo frequencies relative era combo frequencies

```
allyrs.12vars %>% group_by(BAT_HAND_CD,PIT_HAND_CD) %>% summarize(count=n()) %>%
  ungroup %>% mutate(relFreq=count/sum(count)) -> bptotals
teambpfrqs.tall %>% inner_join(bptotals) %>%
  mutate(team2era = Freq/relFreq) -> team2era
team2era %>%  arrange(park,BAT_HAND_CD,PIT_HAND_CD)  %>%
  filter(park %in% c("CLE","SLN","LAN","DET","MIN","TOR")) %>% print
```

```
##    park BAT_HAND_CD PIT_HAND_CD       Freq  count    relFreq  team2era
## 1   CLE           L           L 0.06957481 139051 0.08054097 0.8638437
## 2   CLE           L           R 0.44986341 588850 0.34107305 1.3189650
## 3   CLE           R           L 0.14536635 339857 0.19685160 0.7384565
## 4   CLE           R           R 0.33519543 658705 0.38153439 0.8785458
## 5   DET           L           L 0.06240954 139051 0.08054097 0.7748795
## 6   DET           L           R 0.31973774 588850 0.34107305 0.9374465
## 7   DET           R           L 0.22011614 339857 0.19685160 1.1181831
## 8   DET           R           R 0.39773659 658705 0.38153439 1.0424659
## 9   LAN           L           L 0.11238732 139051 0.08054097 1.3954056
## 10  LAN           L           R 0.30554914 588850 0.34107305 0.8958466
## 11  LAN           R           L 0.25007556 339857 0.19685160 1.2703761
## 12  LAN           R           R 0.33198798 658705 0.38153439 0.8701391
## 13  MIN           L           L 0.08060458 139051 0.08054097 1.0007898
## 14  MIN           L           R 0.37734719 588850 0.34107305 1.1063530
## 15  MIN           R           L 0.18827134 339857 0.19685160 0.9564126
## 16  MIN           R           R 0.35377690 658705 0.38153439 0.9272477
## 17  SLN           L           L 0.06211616 139051 0.08054097 0.7712368
## 18  SLN           L           R 0.33967363 588850 0.34107305 0.9958970
## 19  SLN           R           L 0.14395508 339857 0.19685160 0.7312873
## 20  SLN           R           R 0.45425513 658705 0.38153439 1.1906008
```

```
## 21   TOR           L           L 0.07205480 139051 0.08054097 0.8946354
## 22   TOR           L           R 0.31143569 588850 0.34107305 0.9131055
## 23   TOR           R           L 0.21663500 339857 0.19685160 1.1004991
## 24   TOR           R           R 0.39987451 658705 0.38153439 1.0480694
```

A reminder of which combos lead to the most HRs

```
allyrs.12vars %>% group_by(BAT_HAND_CD,PIT_HAND_CD) %>% summarize(hrfrq=mean(hr),hrsum=sum(hr))
```

```
## # A tibble: 4 x 4
## # Groups:   BAT_HAND_CD [2]
##   BAT_HAND_CD PIT_HAND_CD  hrfrq hrsum
##   <chr>       <chr>        <dbl> <int>
## 1 L           L           0.0222  3088
## 2 L           R           0.0299 17597
## 3 R           L           0.0315 10706
## 4 R           R           0.0298 19623
```