Geometric mean

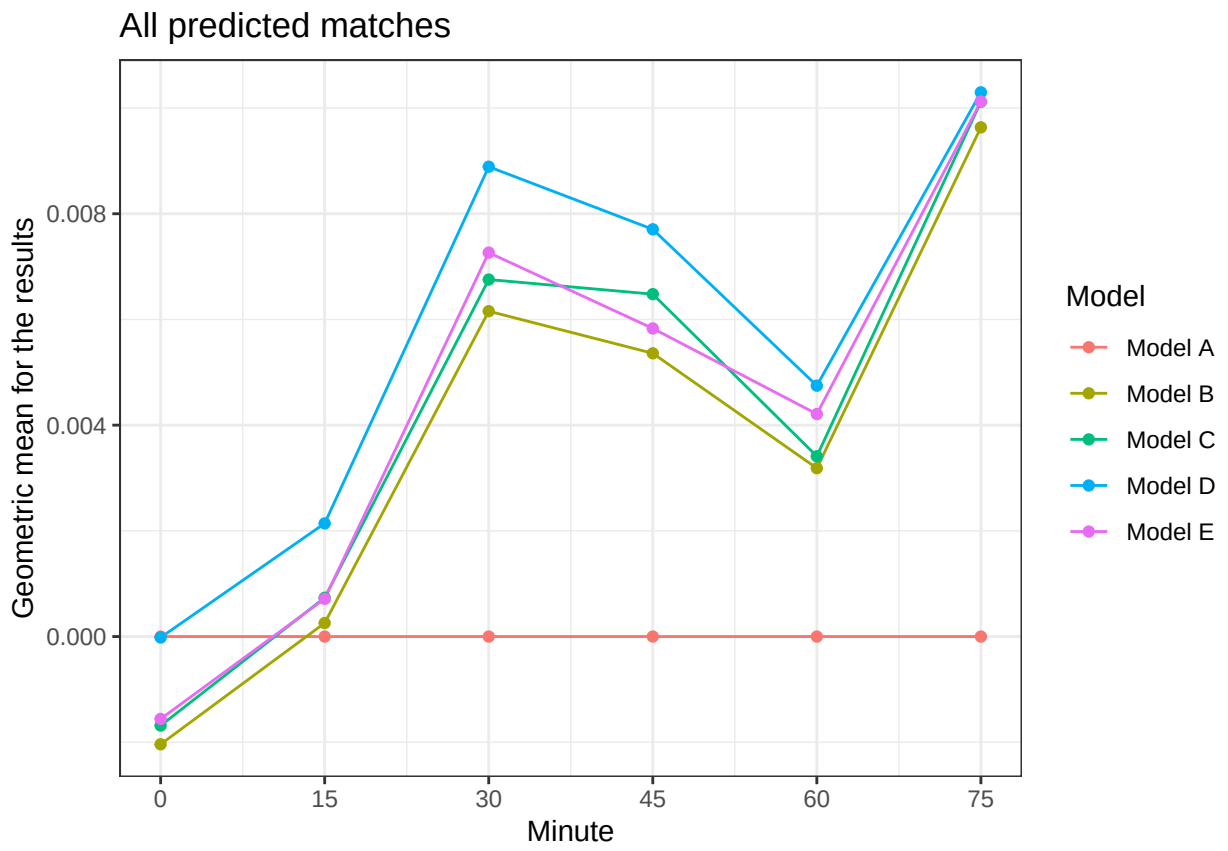
```
library(dplyr)
library(ggplot2)
library(tidyr)
library(knitr)

load("data/HDA_dc.RData")
load("data/first_matches.RData")

HDA = HDA_dc %>%
  anti_join(first_matches)

nrow(HDA)
```

```
## [1] 1858
```

```
HDA[,c(9:158)][which(HDA[,c(9:158)] == 0, arr.ind = TRUE)] = 10^-5
```

```
HDA = HDA %>%
  mutate()
```

```
results = tibble(GeoMean = apply(HDA[,c(99:128)], 2, EnvStats::geoMean),
                 Minute = as.integer(rep(c(0, 15, 30, 45, 60, 75), 5)),
                 Model = factor(c(rep("A", 6),
                                  rep("B", 6),
                                  rep("C", 6),
                                  rep("D", 6),
                                  rep("E", 6)))) %>%
  pivot_wider(names_from = "Model", values_from = "GeoMean", names_prefix = "Model ") %>%
  mutate(`Model B` = log(`Model B`) - log(`Model A`),
         `Model C` = log(`Model C`) - log(`Model A`),
         `Model D` = log(`Model D`) - log(`Model A`),
         `Model E` = log(`Model E`) - log(`Model A`),
         `Model A` = 0) %>%
  pivot_longer(cols = starts_with("Model"), names_to = "Model", values_to = "GeoMean")


results %>%
  ggplot(aes(x = Minute, y = GeoMean, col = Model)) +
  geom_line() +
  geom_point() +
  scale_x_continuous(breaks = c(0, 15, 30, 45, 60, 75)) +
  theme_bw() +
  ggtitle("All predicted matches") +
  ylab("Geometric mean for the results")
```

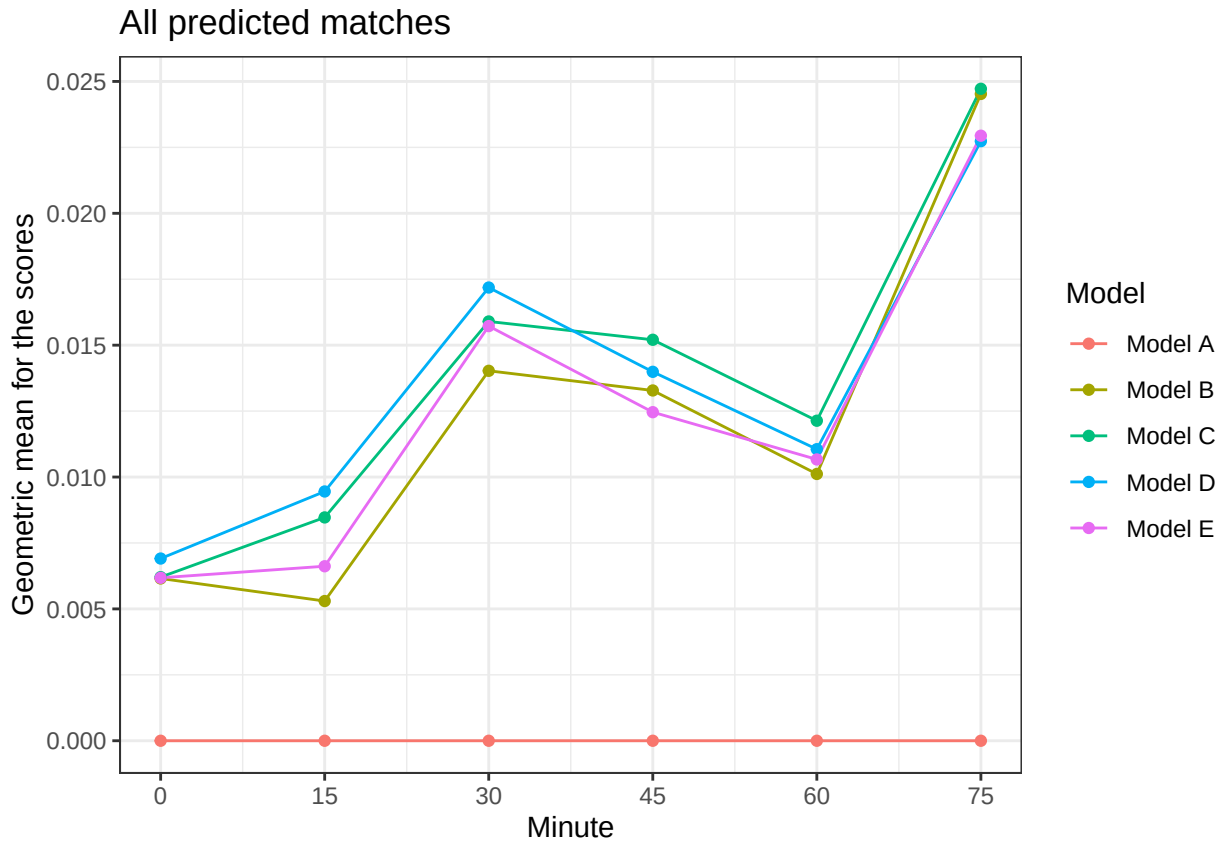



```
results %>%
  pivot_wider(id_cols = "Model", values_from = "GeoMean", names_from = "Minute",
              names_prefix = "Minute ") %>%
  kable()
```

Model	Minute 0	Minute 15	Minute 30	Minute 45	Minute 60	Minute 75
Model A	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
Model B	-0.0020390	0.0002557	0.0061560	0.0053594	0.0031854	0.0096346
Model C	-0.0016848	0.0007344	0.0067551	0.0064789	0.0034112	0.0101189
Model D	-0.0000181	0.0021412	0.0088903	0.0077061	0.0047470	0.0102976
Model E	-0.0015592	0.0007145	0.0072645	0.0058301	0.0042106	0.0101234

```
scores = tibble(GeoMean = apply(HDA[,c(129:158)], 2, EnvStats::geoMean),
                Minute = as.integer(rep(c(0, 15, 30, 45, 60, 75), 5)),
                Model = factor(c(rep("A", 6),
                                 rep("B", 6),
                                 rep("C", 6),
                                 rep("D", 6),
                                 rep("E", 6)))) %>%
  pivot_wider(names_from = "Model", values_from = "GeoMean", names_prefix = "Model ") %>%
  mutate(`Model B` = log(`Model B`) - log(`Model A`),
         `Model C` = log(`Model C`) - log(`Model A`),
         `Model D` = log(`Model D`) - log(`Model A`),
```

```
         `Model E` = log(`Model E`) - log(`Model A`),
         `Model A` = 0) %>%
  pivot_longer(cols = starts_with("Model"), names_to = "Model", values_to = "GeoMean")

scores %>%
  ggplot(aes(x = Minute, y = GeoMean, col = Model)) +
  geom_line() +
  geom_point() +
  scale_x_continuous(breaks = c(0, 15, 30, 45, 60, 75)) +
  theme_bw() +
  ggtitle("All predicted matches") +
  ylab("Geometric mean for the scores")
```




```
scores %>%
  pivot_wider(id_cols = "Model", values_from = "GeoMean", names_from = "Minute",
              names_prefix = "Minute ") %>%
  kable()
```

Model	Minute 0	Minute 15	Minute 30	Minute 45	Minute 60	Minute 75
Model A	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
Model B	0.0061534	0.0052974	0.0140277	0.0132818	0.0101162	0.0245261
Model C	0.0062048	0.0084680	0.0158982	0.0152038	0.0121350	0.0247208
Model D	0.0069092	0.0094519	0.0171848	0.0139893	0.0110556	0.0227346
Model E	0.0061788	0.0066170	0.0157233	0.0124595	0.0106665	0.0229416

```r
load("~/GitHub/soccer-live-predictions/soccer-live-predictions/scrape/data/reds.RData")

matches = reds %>%
  filter(Season > 2015, Half == 1) %>%
  select(Season, Match)

HDA_reds = HDA %>%
  inner_join(matches)
```

```
## Joining, by = c("Season", "Match")
```

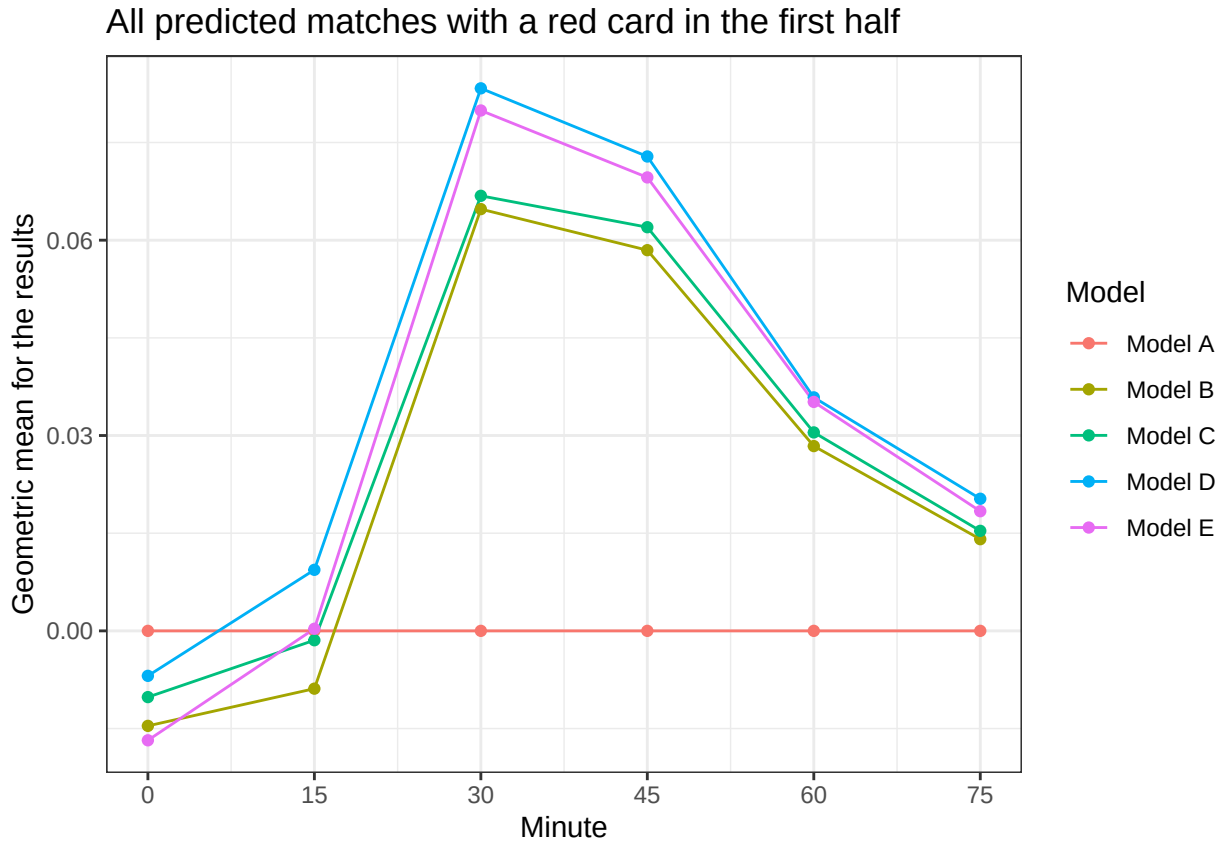
```r
HDA_no_reds = HDA %>%
  anti_join(matches)
```

```
## Joining, by = c("Season", "Match")
```

```r
nrow(HDA_reds)
```

```
## [1] 82
```

```r
results_reds = tibble(GeoMean = apply(HDA_reds[,c(99:128)], 2, EnvStats::geoMean),
                      Minute = as.integer(rep(c(0, 15, 30, 45, 60, 75), 5)),
                      Model = factor(c(rep("A", 6),
                                       rep("B", 6),
                                       rep("C", 6),
                                       rep("D", 6),
                                       rep("E", 6)))) %>%
  pivot_wider(names_from = "Model", values_from = "GeoMean", names_prefix = "Model ") %>%
  mutate(`Model B` = log(`Model B`) - log(`Model A`),
         `Model C` = log(`Model C`) - log(`Model A`),
         `Model D` = log(`Model D`) - log(`Model A`),
         `Model E` = log(`Model E`) - log(`Model A`),
         `Model A` = 0) %>%
  pivot_longer(cols = starts_with("Model"), names_to = "Model", values_to = "GeoMean")

results_reds %>%
  ggplot(aes(x = Minute, y = GeoMean, col = Model)) +
  geom_line() +
  geom_point() +
  scale_x_continuous(breaks = c(0, 15, 30, 45, 60, 75)) +
  theme_bw() +
  ggtitle("All predicted matches with a red card in the first half") +
  ylab("Geometric mean for the results")
```

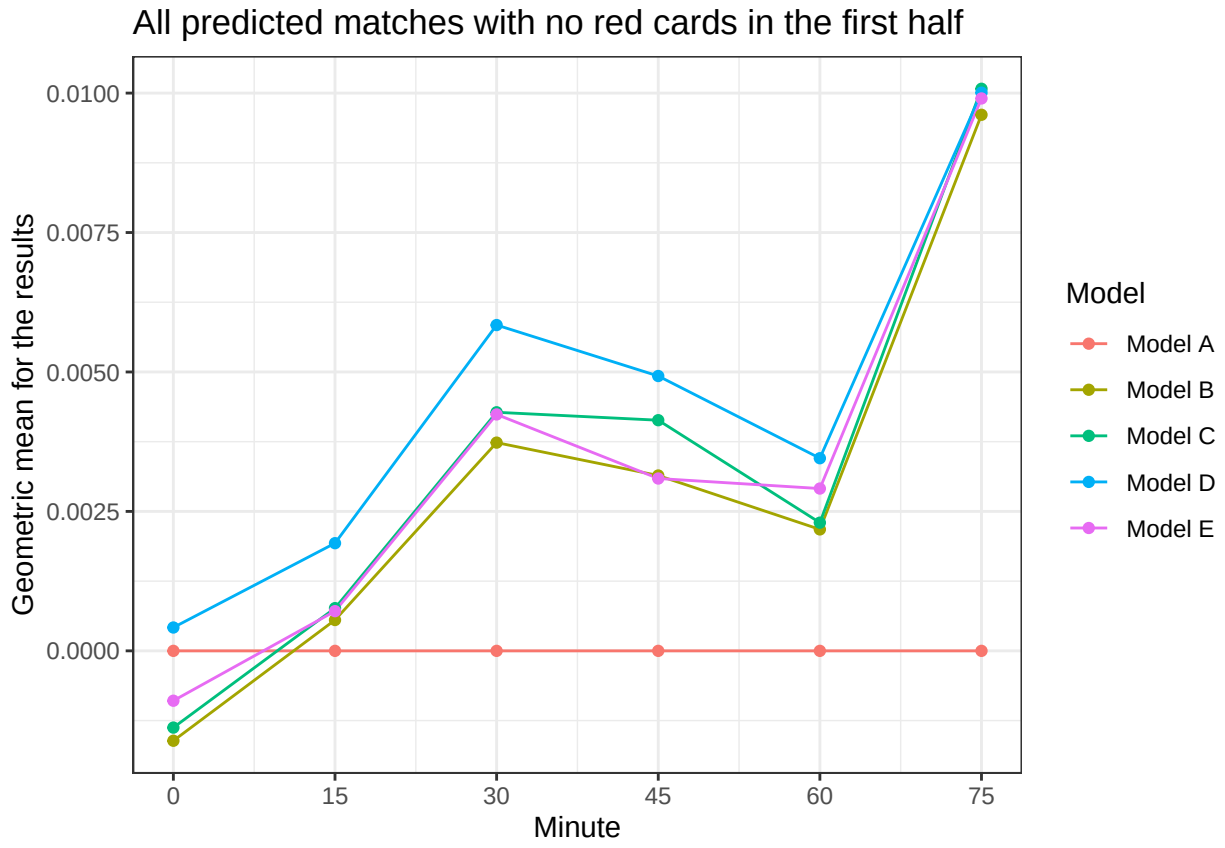



```
results_reds %>%
  pivot_wider(id_cols = "Model", values_from = "GeoMean", names_from = "Minute",
              names_prefix = "Minute ") %>%
  kable()
```

Model	Minute 0	Minute 15	Minute 30	Minute 45	Minute 60	Minute 75
Model A	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
Model B	-0.0145901	-0.0088729	0.0647713	0.0584654	0.0283797	0.0140808
Model C	-0.0101723	-0.0014382	0.0668147	0.0619944	0.0304802	0.0153576
Model D	-0.0069066	0.0093761	0.0833200	0.0728545	0.0358600	0.0202885
Model E	-0.0168054	0.0003236	0.0799168	0.0696394	0.0351626	0.0183776

```
results_no_reds = tibble(GeoMean = apply(HDA_no_reds[,c(99:128)], 2, EnvStats::geoMean),
                 Minute = as.integer(rep(c(0, 15, 30, 45, 60, 75), 5)),
                 Model = factor(c(rep("A", 6),
                                  rep("B", 6),
                                  rep("C", 6),
                                  rep("D", 6),
                                  rep("E", 6)))) %>%
  pivot_wider(names_from = "Model", values_from = "GeoMean", names_prefix = "Model ") %>%
  mutate(`Model B` = log(`Model B`) - log(`Model A`),
         `Model C` = log(`Model C`) - log(`Model A`),
         `Model D` = log(`Model D`) - log(`Model A`),
```

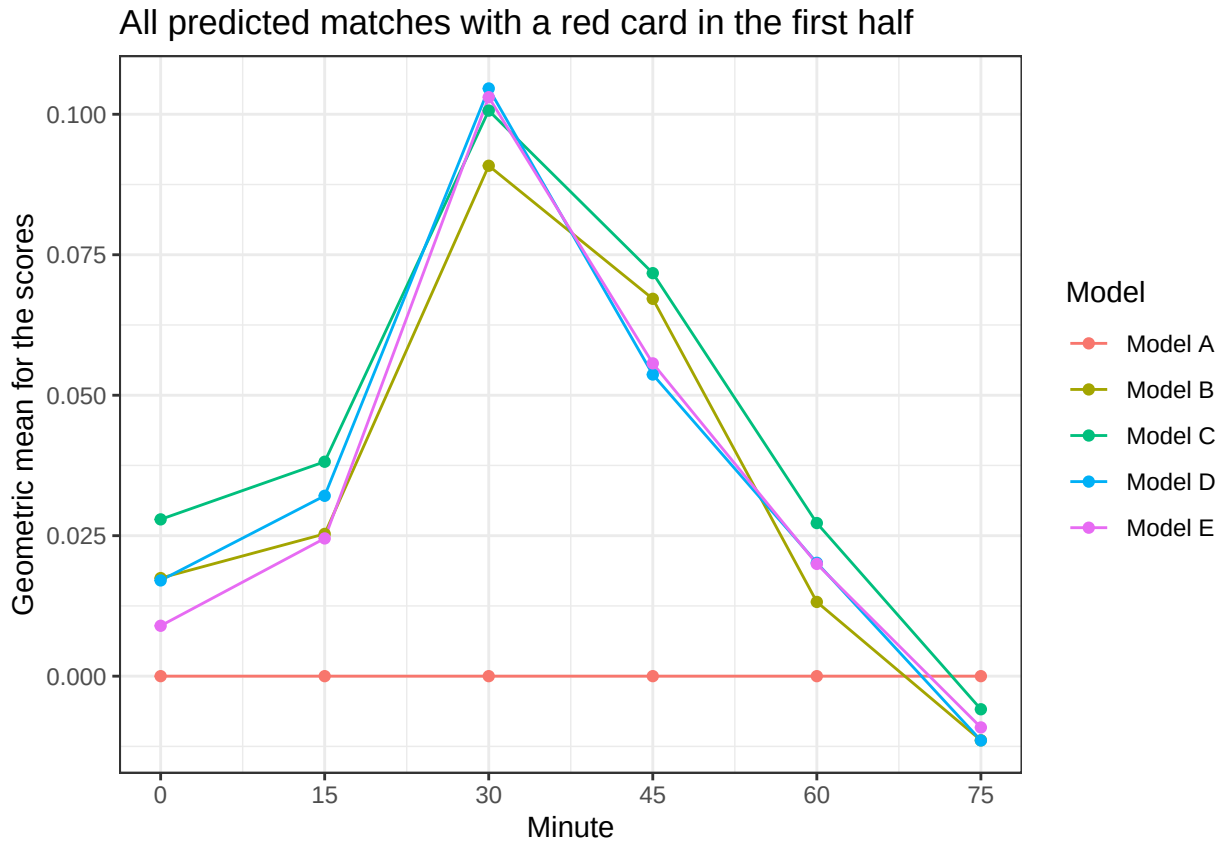
```
         `Model E` = log(`Model E`) - log(`Model A`),
         `Model A` = 0) %>%
  pivot_longer(cols = starts_with("Model"), names_to = "Model", values_to = "GeoMean")

results_no_reds %>%
  ggplot(aes(x = Minute, y = GeoMean, col = Model)) +
  geom_line() +
  geom_point() +
  scale_x_continuous(breaks = c(0, 15, 30, 45, 60, 75)) +
  theme_bw() +
  ggtitle("All predicted matches with no red cards in the first half") +
  ylab("Geometric mean for the results")
```




```
results_no_reds %>%
  pivot_wider(id_cols = "Model", values_from = "GeoMean", names_from = "Minute",
              names_prefix = "Minute ") %>%
  kable()
```

Model	Minute 0	Minute 15	Minute 30	Minute 45	Minute 60	Minute 75
Model A	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
Model B	-0.0016118	0.0005530	0.0037334	0.0031436	0.0021773	0.0096109
Model C	-0.0013757	0.0007657	0.0042772	0.0041352	0.0022982	0.0100769
Model D	0.0004188	0.0019302	0.0058413	0.0049274	0.0034536	0.0100071
Model E	-0.0008941	0.0007115	0.0042387	0.0030898	0.0029098	0.0099037

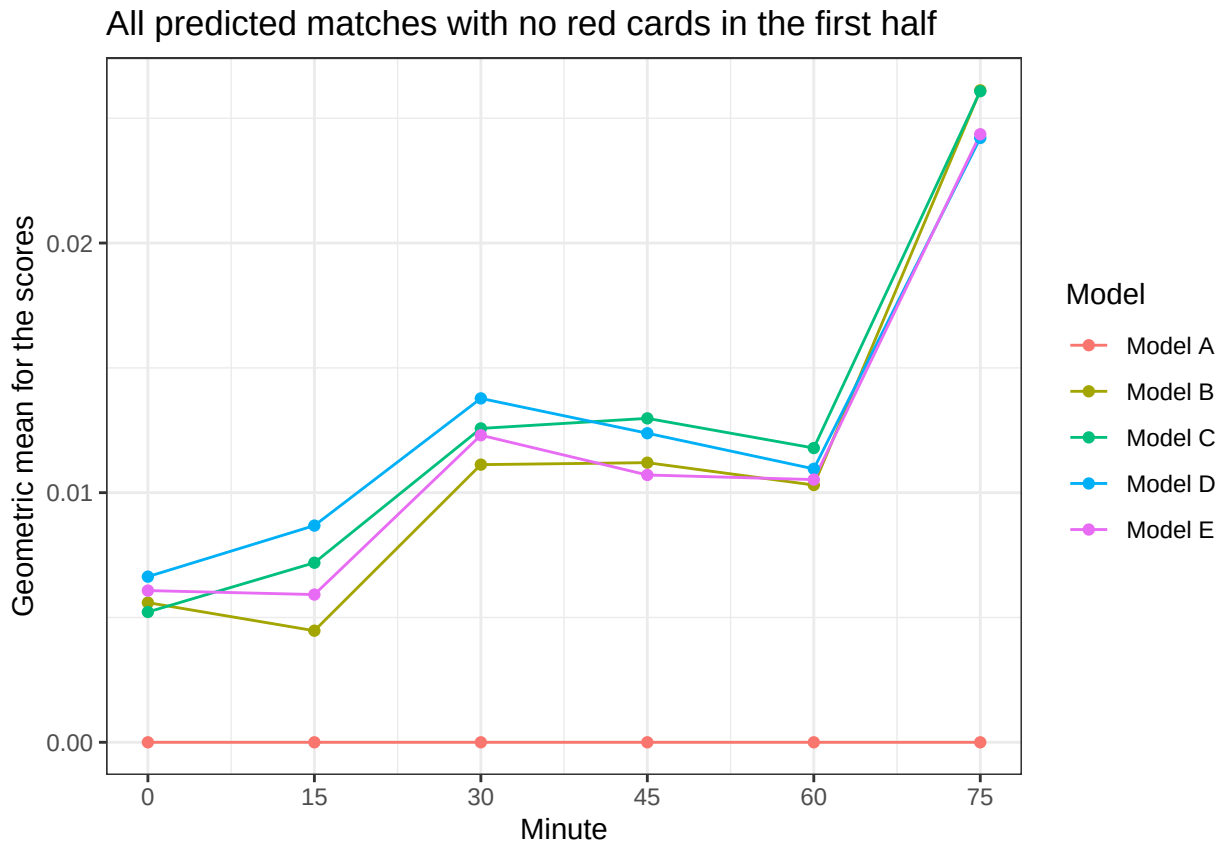
```r
scores_reds = tibble(GeoMean = apply(HDA_reds[,c(129:158)], 2, EnvStats::geoMean),
                     Minute = as.integer(rep(c(0, 15, 30, 45, 60, 75), 5)),
                     Model = factor(c(rep("A", 6),
                                      rep("B", 6),
                                      rep("C", 6),
                                      rep("D", 6),
                                      rep("E", 6)))) %>%
  pivot_wider(names_from = "Model", values_from = "GeoMean", names_prefix = "Model ") %>%
  mutate(`Model B` = log(`Model B`) - log(`Model A`),
         `Model C` = log(`Model C`) - log(`Model A`),
         `Model D` = log(`Model D`) - log(`Model A`),
         `Model E` = log(`Model E`) - log(`Model A`),
         `Model A` = 0) %>%
  pivot_longer(cols = starts_with("Model"), names_to = "Model", values_to = "GeoMean")

scores_reds %>%
  ggplot(aes(x = Minute, y = GeoMean, col = Model)) +
  geom_line() +
  geom_point() +
  scale_x_continuous(breaks = c(0, 15, 30, 45, 60, 75)) +
  theme_bw() +
  ggtitle("All predicted matches with a red card in the first half") +
  ylab("Geometric mean for the scores")
```



```
scores_reds %>%
  pivot_wider(id_cols = "Model", values_from = "GeoMean", names_from = "Minute",
              names_prefix = "Minute ") %>%
  kable()
```

Model	Minute 0	Minute 15	Minute 30	Minute 45	Minute 60	Minute 75
Model A	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
Model B	0.0174591	0.0253175	0.0908397	0.0671582	0.0132089	-0.0114157
Model C	0.0279034	0.0381595	0.1006545	0.0717241	0.0272477	-0.0058915
Model D	0.0170383	0.0321001	0.1045914	0.0536829	0.0201733	-0.0114408
Model E	0.0089658	0.0245147	0.1030408	0.0556664	0.0199778	-0.0091238

```
scores_no_reds = tibble(GeoMean = apply(HDA_no_reds[,c(129:158)], 2, EnvStats::geoMean),
                 Minute = as.integer(rep(c(0, 15, 30, 45, 60, 75), 5)),
                 Model = factor(c(rep("A", 6),
                                  rep("B", 6),
                                  rep("C", 6),
                                  rep("D", 6),
                                  rep("E", 6)))) %>%
  pivot_wider(names_from = "Model", values_from = "GeoMean", names_prefix = "Model ") %>%
  mutate(`Model B` = log(`Model B`) - log(`Model A`),
         `Model C` = log(`Model C`) - log(`Model A`),
         `Model D` = log(`Model D`) - log(`Model A`),
         `Model E` = log(`Model E`) - log(`Model A`),
         `Model A` = 0) %>%
  pivot_longer(cols = starts_with("Model"), names_to = "Model", values_to = "GeoMean")

scores_no_reds %>%
  ggplot(aes(x = Minute, y = GeoMean, col = Model)) +
  geom_line() +
  geom_point() +
  scale_x_continuous(breaks = c(0, 15, 30, 45, 60, 75)) +
  theme_bw() +
  ggtitle("All predicted matches with no red cards in the first half") +
  ylab("Geometric mean for the scores")
```




```
scores_no_reds %>%
  pivot_wider(id_cols = "Model", values_from = "GeoMean", names_from = "Minute",
              names_prefix = "Minute ") %>%
  kable()
```

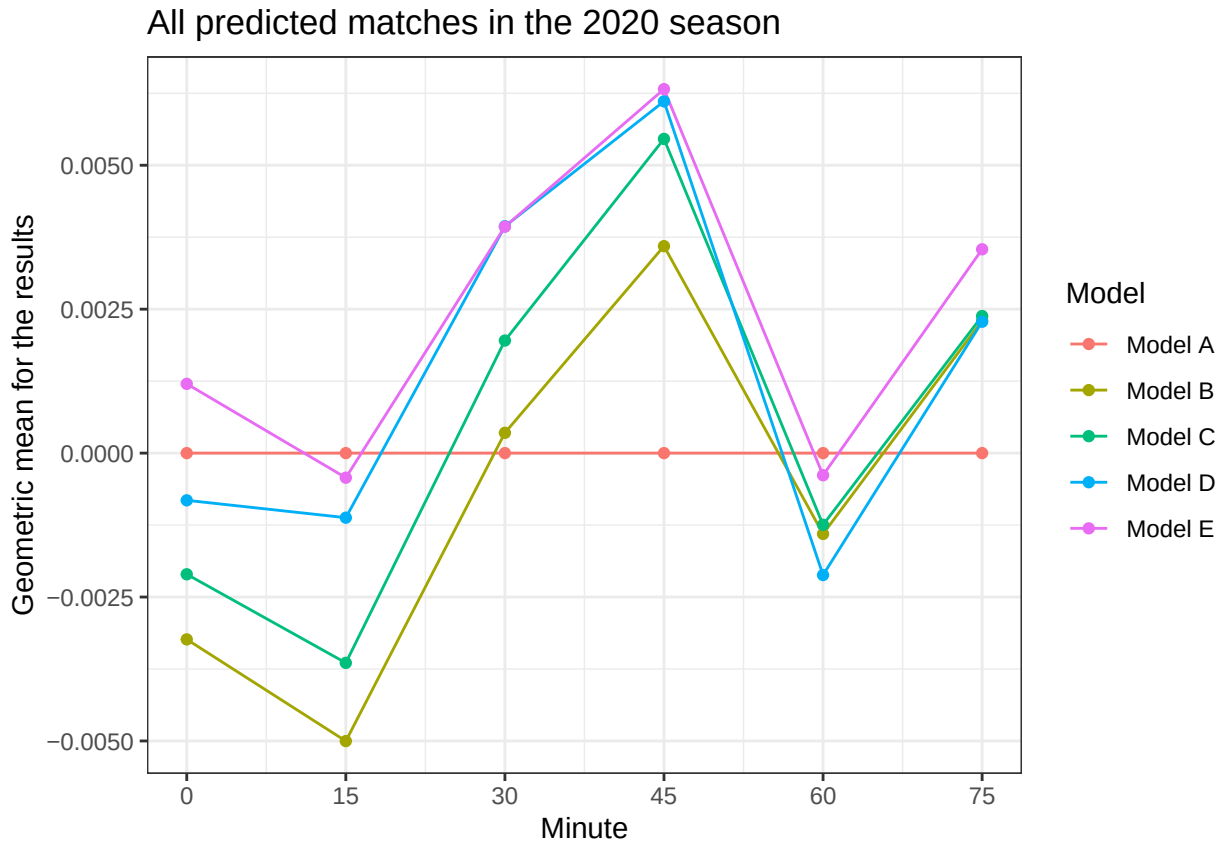
Model	Minute 0	Minute 15	Minute 30	Minute 45	Minute 60	Minute 75
Model A	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
Model B	0.0055947	0.0044703	0.0111232	0.0112016	0.0103060	0.0261170
Model C	0.0052202	0.0071933	0.0125732	0.0129769	0.0117883	0.0260817
Model D	0.0066361	0.0086818	0.0137765	0.0123813	0.0109543	0.0242092
Model E	0.0060774	0.0059181	0.0123008	0.0107087	0.0105242	0.0243578

```
HDA_2020 = HDA %>%
  filter(Season == 2020)

nrow(HDA_2020)
```

```
## [1] 376
```

```
results_2020 = tibble(GeoMean = apply(HDA_2020[,c(99:128)], 2, EnvStats::geoMean),
                 Minute = as.integer(rep(c(0, 15, 30, 45, 60, 75), 5)),
                 Model = factor(c(rep("A", 6),
                                  rep("B", 6),
```

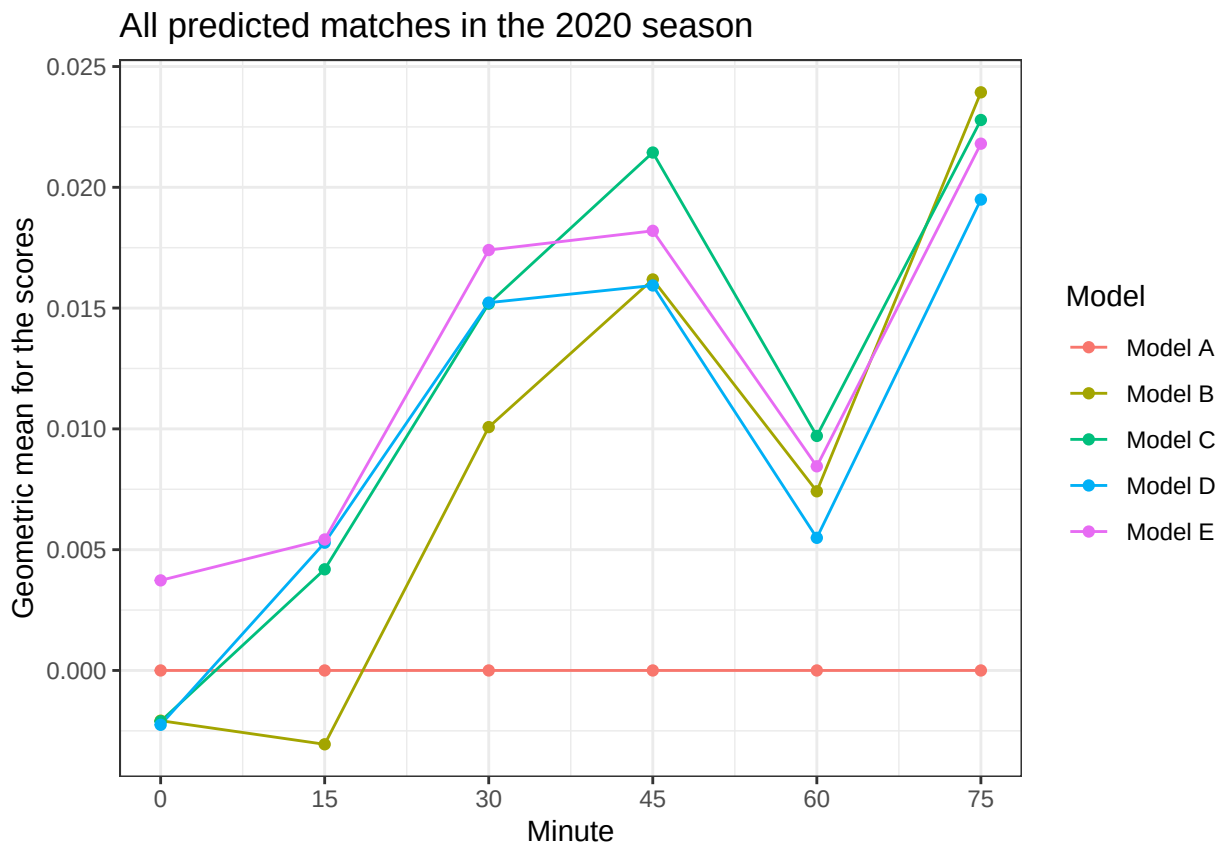
```
                                  rep("C", 6),
                                  rep("D", 6),
                                  rep("E", 6))) %>%
  pivot_wider(names_from = "Model", values_from = "GeoMean", names_prefix = "Model ") %>%
  mutate(`Model B` = log(`Model B`) - log(`Model A`),
         `Model C` = log(`Model C`) - log(`Model A`),
         `Model D` = log(`Model D`) - log(`Model A`),
         `Model E` = log(`Model E`) - log(`Model A`),
         `Model A` = 0) %>%
  pivot_longer(cols = starts_with("Model"), names_to = "Model", values_to = "GeoMean")

results_2020 %>%
  ggplot(aes(x = Minute, y = GeoMean, col = Model)) +
  geom_line() +
  geom_point() +
  scale_x_continuous(breaks = c(0, 15, 30, 45, 60, 75)) +
  theme_bw() +
  ggtitle("All predicted matches in the 2020 season") +
  ylab("Geometric mean for the results")
```




```
results_2020 %>%
  pivot_wider(id_cols = "Model", values_from = "GeoMean", names_from = "Minute",
              names_prefix = "Minute ") %>%
  kable()
```

Model	Minute 0	Minute 15	Minute 30	Minute 45	Minute 60	Minute 75
Model A	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
Model B	-0.0032359	-0.0050027	0.0003524	0.0035951	-0.0014039	0.0023000
Model C	-0.0021047	-0.0036441	0.0019553	0.0054578	-0.0012458	0.0023813
Model D	-0.0008202	-0.0011217	0.0039400	0.0061101	-0.0021181	0.0022818
Model E	0.0012030	-0.0004274	0.0039327	0.0063223	-0.0003843	0.0035402

```
scores_2020 = tibble(GeoMean = apply(HDA_2020[,c(129:158)], 2, EnvStats::geoMean),
                     Minute = as.integer(rep(c(0, 15, 30, 45, 60, 75), 5)),
                     Model = factor(c(rep("A", 6),
                                      rep("B", 6),
                                      rep("C", 6),
                                      rep("D", 6),
                                      rep("E", 6)))) %>%
  pivot_wider(names_from = "Model", values_from = "GeoMean", names_prefix = "Model ") %>%
  mutate(`Model B` = log(`Model B`) - log(`Model A`),
         `Model C` = log(`Model C`) - log(`Model A`),
         `Model D` = log(`Model D`) - log(`Model A`),
         `Model E` = log(`Model E`) - log(`Model A`),
         `Model A` = 0) %>%
  pivot_longer(cols = starts_with("Model"), names_to = "Model", values_to = "GeoMean")

scores_2020 %>%
  ggplot(aes(x = Minute, y = GeoMean, col = Model)) +
  geom_line() +
  geom_point() +
  scale_x_continuous(breaks = c(0, 15, 30, 45, 60, 75)) +
  theme_bw() +
  ggtitle("All predicted matches in the 2020 season") +
  ylab("Geometric mean for the scores")
```




```
scores_2020 %>%
  pivot_wider(id_cols = "Model", values_from = "GeoMean", names_from = "Minute",
              names_prefix = "Minute ") %>%
  kable()
```

Model	Minute 0	Minute 15	Minute 30	Minute 45	Minute 60	Minute 75
Model A	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
Model B	-0.0020800	-0.0030570	0.0100732	0.0161854	0.0074168	0.0239314
Model C	-0.0021044	0.0041864	0.0151830	0.0214392	0.0097079	0.0227856
Model D	-0.0022501	0.0052878	0.0152249	0.0159362	0.0054898	0.0194933
Model E	0.0037288	0.0054215	0.0174030	0.0181978	0.0084505	0.0218060

```
load("~/GitHub/soccer-live-predictions/soccer-live-predictions/scrape/data/results.RData")
load("~/GitHub/soccer-live-predictions/soccer-live-predictions/scrape/data/goals.RData")
```

```
at_45 = results %>%
  select(Season, Match) %>%
  filter(Season > 2015)
```

```
home_score_at_45 <- function(season, match) {
  goals %>%
    filter(Season == season,
           Match == match,
```

```
            Team == 1,
            Half == 1) %>%
    nrow()
}

away_score_at_45 <- function(season, match) {
  goals %>%
    filter(Season == season,
           Match == match,
           Team == 2,
           Half == 1) %>%
    nrow()
}
```

```
at_45 = at_45 %>%
  rowwise() %>%
  mutate(Home_Score = home_score_at_45(Season, Match),
         Away_Score = away_score_at_45(Season, Match),
         abs_dif = abs(Home_Score - Away_Score))
```

```
tmp_00 = at_45 %>%
  filter(abs_dif == 0) %>%
  select(Season, Match)

HDA_00 = HDA %>%
  inner_join(tmp_00)
```

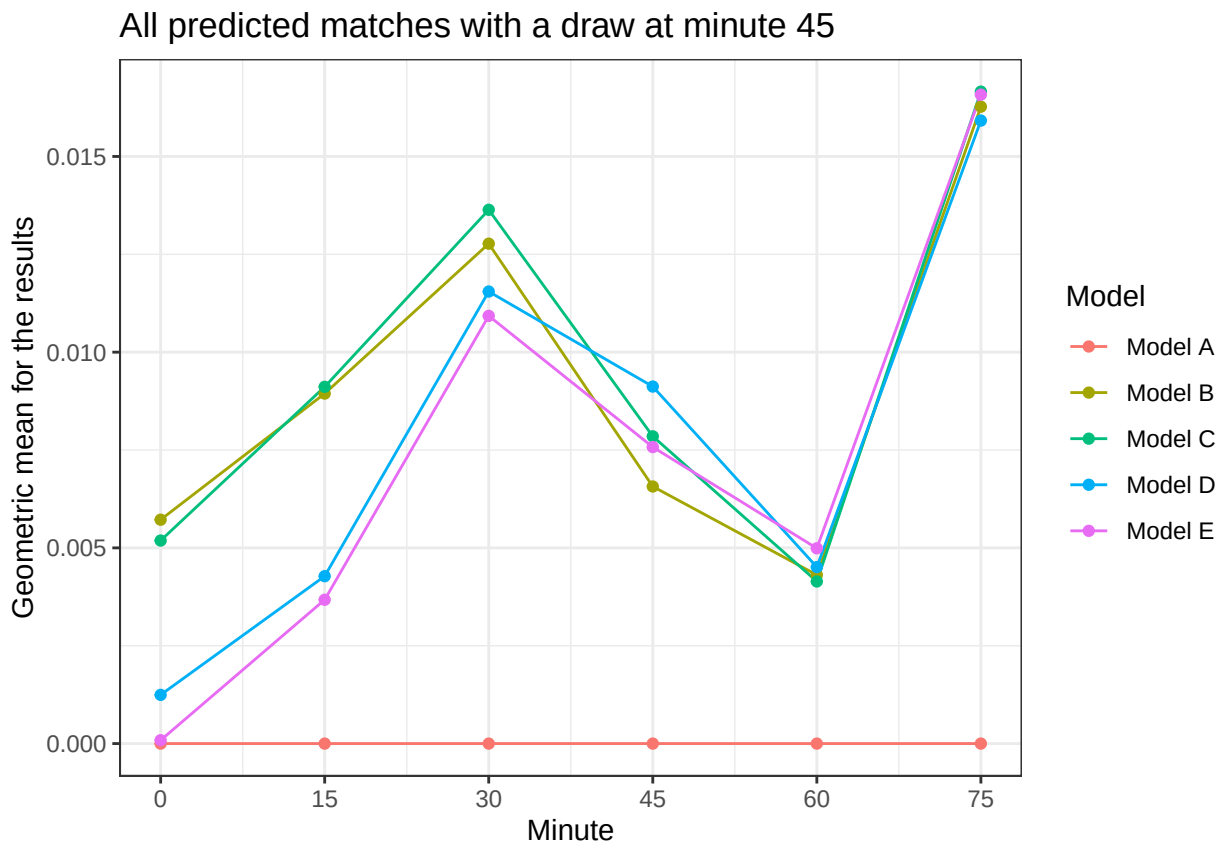
```
## Joining, by = c("Season", "Match")
```

```
nrow(HDA_00)
```

```
## [1] 838
```

```
results_00 = tibble(GeoMean = apply(HDA_00[,c(99:128)], 2, EnvStats::geoMean),
                Minute = as.integer(rep(c(0, 15, 30, 45, 60, 75), 5)),
                Model = factor(c(rep("A", 6),
                                 rep("B", 6),
                                 rep("C", 6),
                                 rep("D", 6),
                                 rep("E", 6)))) %>%
  pivot_wider(names_from = "Model", values_from = "GeoMean", names_prefix = "Model ") %>%
  mutate(`Model B` = log(`Model B`) - log(`Model A`),
         `Model C` = log(`Model C`) - log(`Model A`),
         `Model D` = log(`Model D`) - log(`Model A`),
         `Model E` = log(`Model E`) - log(`Model A`),
         `Model A` = 0) %>%
  pivot_longer(cols = starts_with("Model"), names_to = "Model", values_to = "GeoMean")

results_00 %>%
  ggplot(aes(x = Minute, y = GeoMean, col = Model)) +
  geom_line() +
```

```
  geom_point() +
  scale_x_continuous(breaks = c(0, 15, 30, 45, 60, 75)) +
  theme_bw() +
  ggtitle("All predicted matches with a draw at minute 45") +
  ylab("Geometric mean for the results")
```

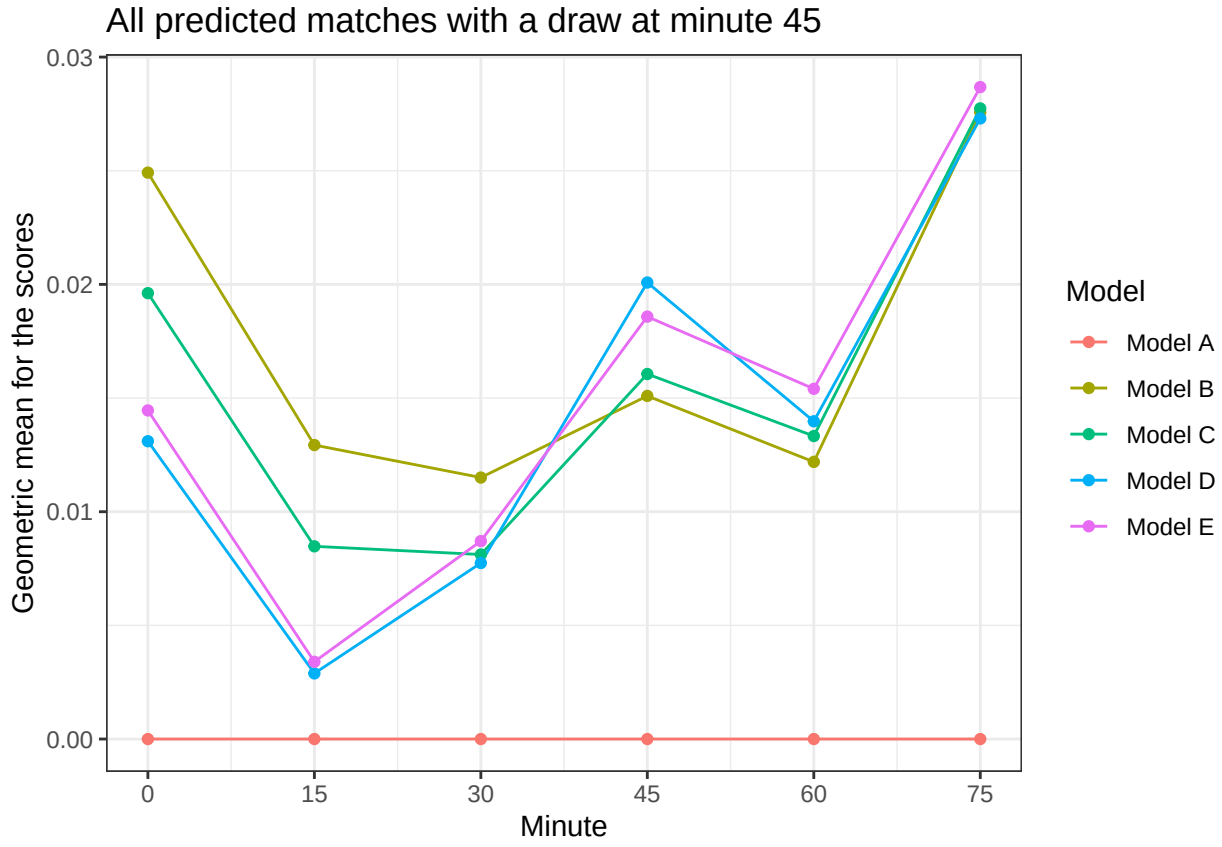



```
results_00 %>%
  pivot_wider(id_cols = "Model", values_from = "GeoMean", names_from = "Minute",
              names_prefix = "Minute ") %>%
  kable()
```

Model	Minute 0	Minute 15	Minute 30	Minute 45	Minute 60	Minute 75
Model A	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
Model B	0.0057201	0.0089421	0.0127716	0.0065683	0.0043080	0.0162713
Model C	0.0051874	0.0091147	0.0136369	0.0078501	0.0041422	0.0166561
Model D	0.0012447	0.0042770	0.0115487	0.0091217	0.0045096	0.0159167
Model E	0.0000862	0.0036750	0.0109277	0.0075754	0.0049892	0.0165797

```
scores_00 = tibble(GeoMean = apply(HDA_00[,c(129:158)], 2, EnvStats::geoMean),
                   Minute = as.integer(rep(c(0, 15, 30, 45, 60, 75), 5)),
                   Model = factor(c(rep("A", 6),
                                    rep("B", 6),
                                    rep("C", 6),
```

```
                                  rep("D", 6),
                                  rep("E", 6)))) %>%
  pivot_wider(names_from = "Model", values_from = "GeoMean", names_prefix = "Model ") %>%
  mutate(`Model B` = log(`Model B`) - log(`Model A`),
         `Model C` = log(`Model C`) - log(`Model A`),
         `Model D` = log(`Model D`) - log(`Model A`),
         `Model E` = log(`Model E`) - log(`Model A`),
         `Model A` = 0) %>%
  pivot_longer(cols = starts_with("Model"), names_to = "Model", values_to = "GeoMean")

scores_00 %>%
  ggplot(aes(x = Minute, y = GeoMean, col = Model)) +
  geom_line() +
  geom_point() +
  scale_x_continuous(breaks = c(0, 15, 30, 45, 60, 75)) +
  theme_bw() +
  ggtitle("All predicted matches with a draw at minute 45") +
  ylab("Geometric mean for the scores")
```




```
scores_00 %>%
  pivot_wider(id_cols = "Model", values_from = "GeoMean", names_from = "Minute",
              names_prefix = "Minute ") %>%
  kable()
```

Model	Minute 0	Minute 15	Minute 30	Minute 45	Minute 60	Minute 75
Model A	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
Model B	0.0249162	0.0129321	0.0115032	0.0150926	0.0121918	0.0275698
Model C	0.0196155	0.0084773	0.0081172	0.0160578	0.0133290	0.0277404
Model D	0.0131011	0.0028870	0.0077431	0.0200814	0.0139766	0.0273018
Model E	0.0144553	0.0033985	0.0087082	0.0185824	0.0154062	0.0286802

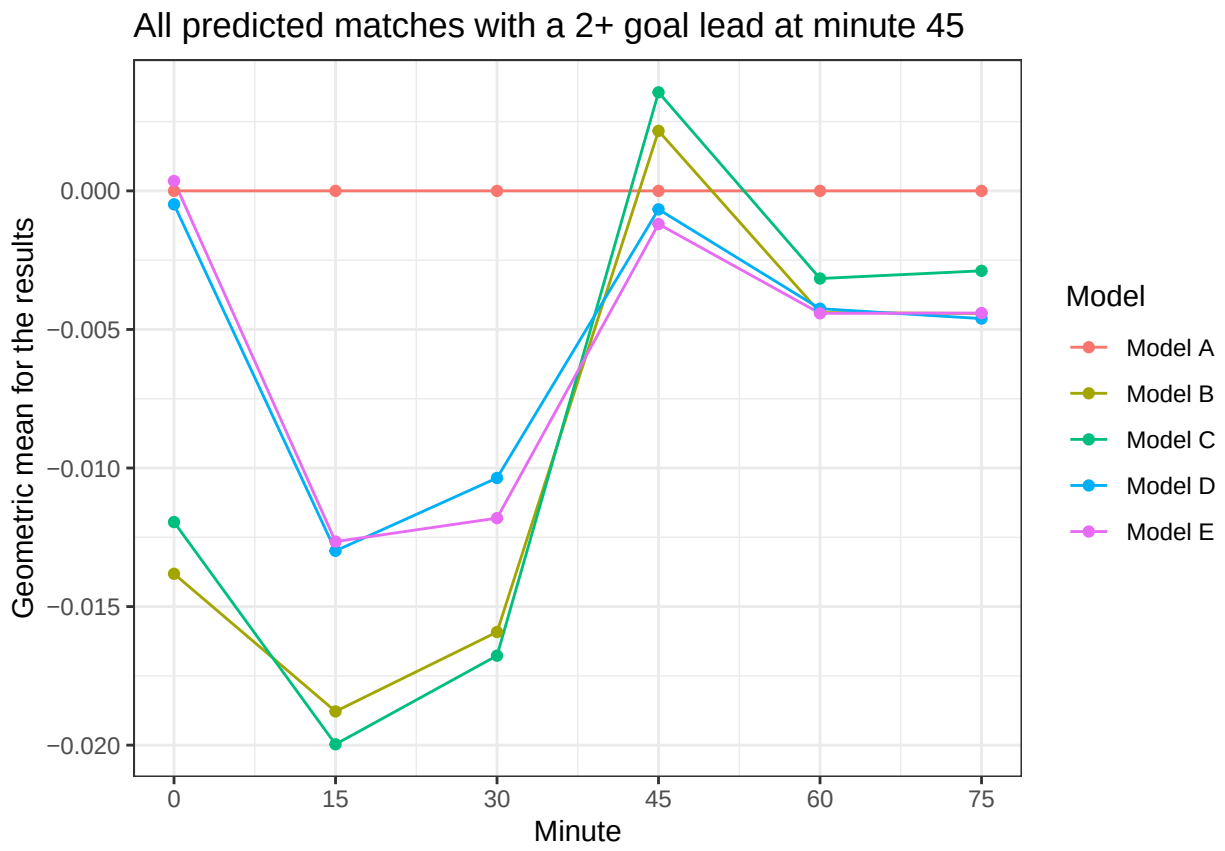
```
tmp_20 = at_45 %>%
  filter(abs_dif >= 2) %>%
  select(Season, Match)

HDA_20 = HDA %>%
  inner_join(tmp_20)
```

```
## Joining, by = c("Season", "Match")
```

```
nrow(HDA_20)
```

```
## [1] 211
```

```
results_20 = tibble(GeoMean = apply(HDA_20[,c(99:128)], 2, EnvStats::geoMean),
                    Minute = as.integer(rep(c(0, 15, 30, 45, 60, 75), 5)),
                    Model = factor(c(rep("A", 6),
                                     rep("B", 6),
                                     rep("C", 6),
                                     rep("D", 6),
                                     rep("E", 6)))) %>%
  pivot_wider(names_from = "Model", values_from = "GeoMean", names_prefix = "Model ") %>%
  mutate(`Model B` = log(`Model B`) - log(`Model A`),
         `Model C` = log(`Model C`) - log(`Model A`),
         `Model D` = log(`Model D`) - log(`Model A`),
         `Model E` = log(`Model E`) - log(`Model A`),
         `Model A` = 0) %>%
  pivot_longer(cols = starts_with("Model"), names_to = "Model", values_to = "GeoMean")

results_20 %>%
  ggplot(aes(x = Minute, y = GeoMean, col = Model)) +
  geom_line() +
  geom_point() +
  scale_x_continuous(breaks = c(0, 15, 30, 45, 60, 75)) +
  theme_bw() +
  ggtitle("All predicted matches with a 2+ goal lead at minute 45") +
  ylab("Geometric mean for the results")
```

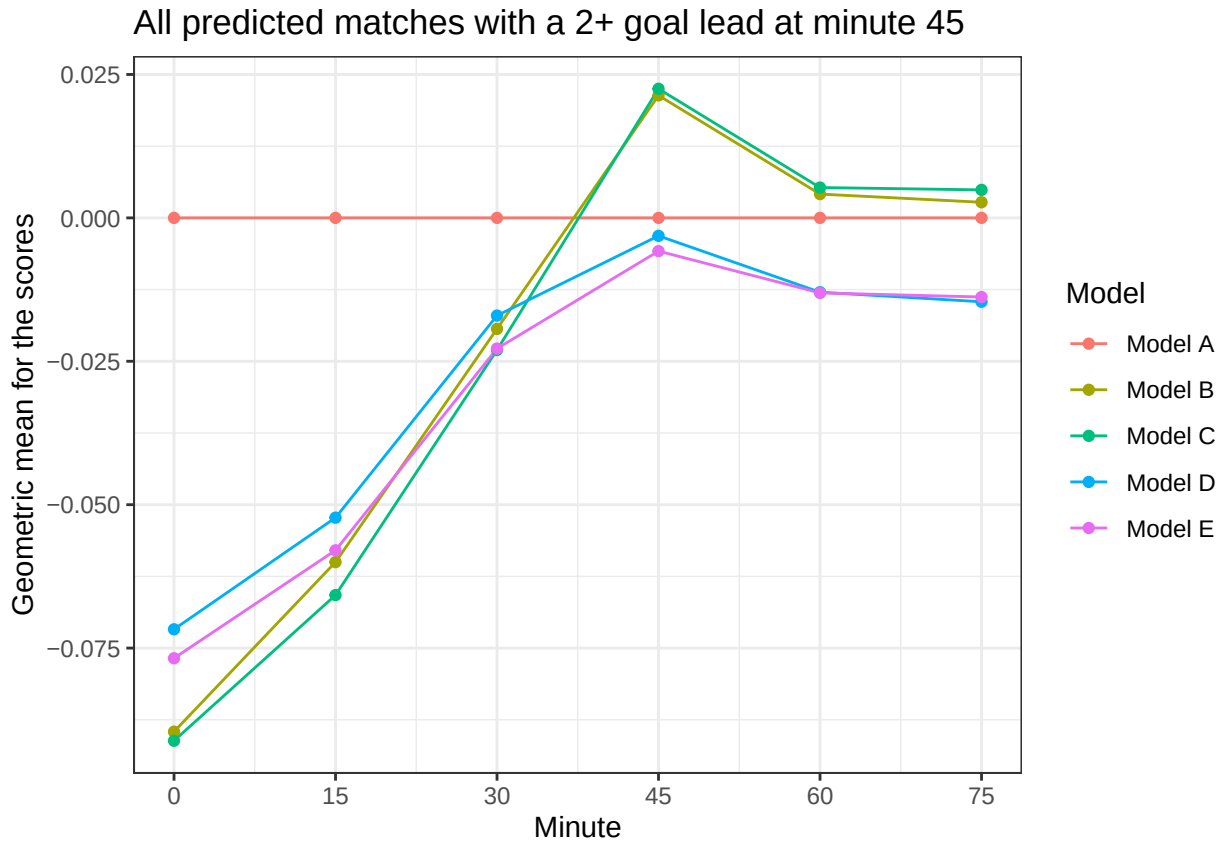



```
results_20 %>%
  pivot_wider(id_cols = "Model", values_from = "GeoMean", names_from = "Minute",
              names_prefix = "Minute ") %>%
  kable()
```

Model	Minute 0	Minute 15	Minute 30	Minute 45	Minute 60	Minute 75
Model A	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
Model B	-0.0138164	-0.0187832	-0.0159184	0.0021653	-0.0043905	-0.0044233
Model C	-0.0119503	-0.0199694	-0.0167705	0.0035561	-0.0031622	-0.0028845
Model D	-0.0004851	-0.0129888	-0.0103591	-0.0006657	-0.0042511	-0.0046076
Model E	0.0003560	-0.0126549	-0.0118085	-0.0011971	-0.0044189	-0.0044090

```
scores_20 = tibble(GeoMean = apply(HDA_20[,c(129:158)], 2, EnvStats::geoMean),
                   Minute = as.integer(rep(c(0, 15, 30, 45, 60, 75), 5)),
                   Model = factor(c(rep("A", 6),
                                    rep("B", 6),
                                    rep("C", 6),
                                    rep("D", 6),
                                    rep("E", 6)))) %>%
  pivot_wider(names_from = "Model", values_from = "GeoMean", names_prefix = "Model ") %>%
  mutate(`Model B` = log(`Model B`) - log(`Model A`),
         `Model C` = log(`Model C`) - log(`Model A`),
         `Model D` = log(`Model D`) - log(`Model A`),
```

```
         `Model E` = log(`Model E`) - log(`Model A`),
         `Model A` = 0) %>%
  pivot_longer(cols = starts_with("Model"), names_to = "Model", values_to = "GeoMean")

scores_20 %>%
  ggplot(aes(x = Minute, y = GeoMean, col = Model)) +
  geom_line() +
  geom_point() +
  scale_x_continuous(breaks = c(0, 15, 30, 45, 60, 75)) +
  theme_bw() +
  ggtitle("All predicted matches with a 2+ goal lead at minute 45") +
  ylab("Geometric mean for the scores")
```




```
scores_20 %>%
  pivot_wider(id_cols = "Model", values_from = "GeoMean", names_from = "Minute",
              names_prefix = "Minute ") %>%
  kable()
```

Model	Minute 0	Minute 15	Minute 30	Minute 45	Minute 60	Minute 75
Model A	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
Model B	-0.0895750	-0.0600116	-0.0193468	0.0213497	0.0041478	0.0027224
Model C	-0.0911574	-0.0657619	-0.0230302	0.0225048	0.0052877	0.0048798
Model D	-0.0717249	-0.0522621	-0.0170349	-0.0031316	-0.0129539	-0.0146241
Model E	-0.0767725	-0.0579532	-0.0227891	-0.0057950	-0.0130797	-0.0137969

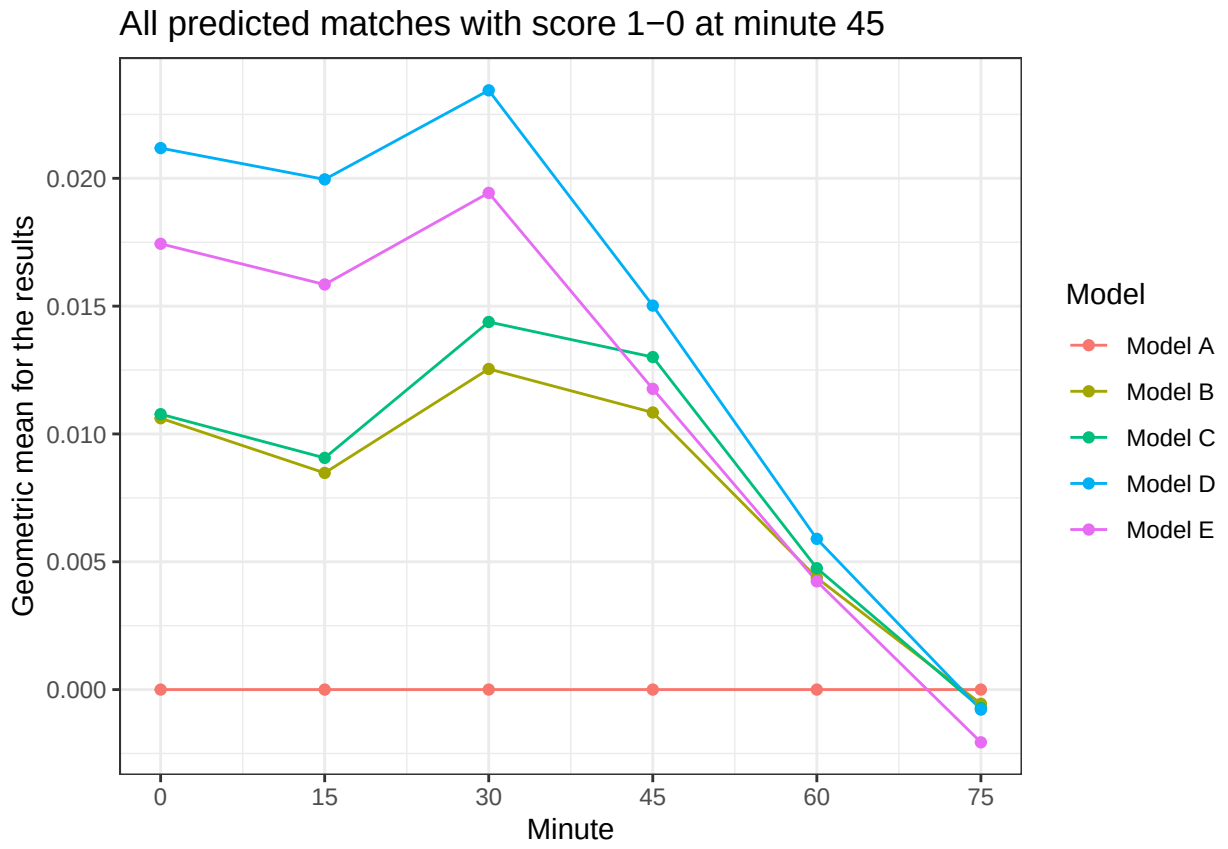
```r
tmp_10 = at_45 %>%
  filter(Home_Score == 1, Away_Score == 0) %>%
  select(Season, Match)

HDA_10 = HDA %>%
  inner_join(tmp_10)
```

```
## Joining, by = c("Season", "Match")
```

```r
nrow(HDA_10)
```

```
## [1] 435
```

```r
results_10 = tibble(GeoMean = apply(HDA_10[,c(99:128)], 2, EnvStats::geoMean),
                    Minute = as.integer(rep(c(0, 15, 30, 45, 60, 75), 5)),
                    Model = factor(c(rep("A", 6),
                                     rep("B", 6),
                                     rep("C", 6),
                                     rep("D", 6),
                                     rep("E", 6)))) %>%
  pivot_wider(names_from = "Model", values_from = "GeoMean", names_prefix = "Model ") %>%
  mutate(`Model B` = log(`Model B`) - log(`Model A`),
         `Model C` = log(`Model C`) - log(`Model A`),
         `Model D` = log(`Model D`) - log(`Model A`),
         `Model E` = log(`Model E`) - log(`Model A`),
         `Model A` = 0) %>%
  pivot_longer(cols = starts_with("Model"), names_to = "Model", values_to = "GeoMean")

results_10 %>%
  ggplot(aes(x = Minute, y = GeoMean, col = Model)) +
  geom_line() +
  geom_point() +
  scale_x_continuous(breaks = c(0, 15, 30, 45, 60, 75)) +
  theme_bw() +
  ggtitle("All predicted matches with score 1-0 at minute 45") +
  ylab("Geometric mean for the results")
```

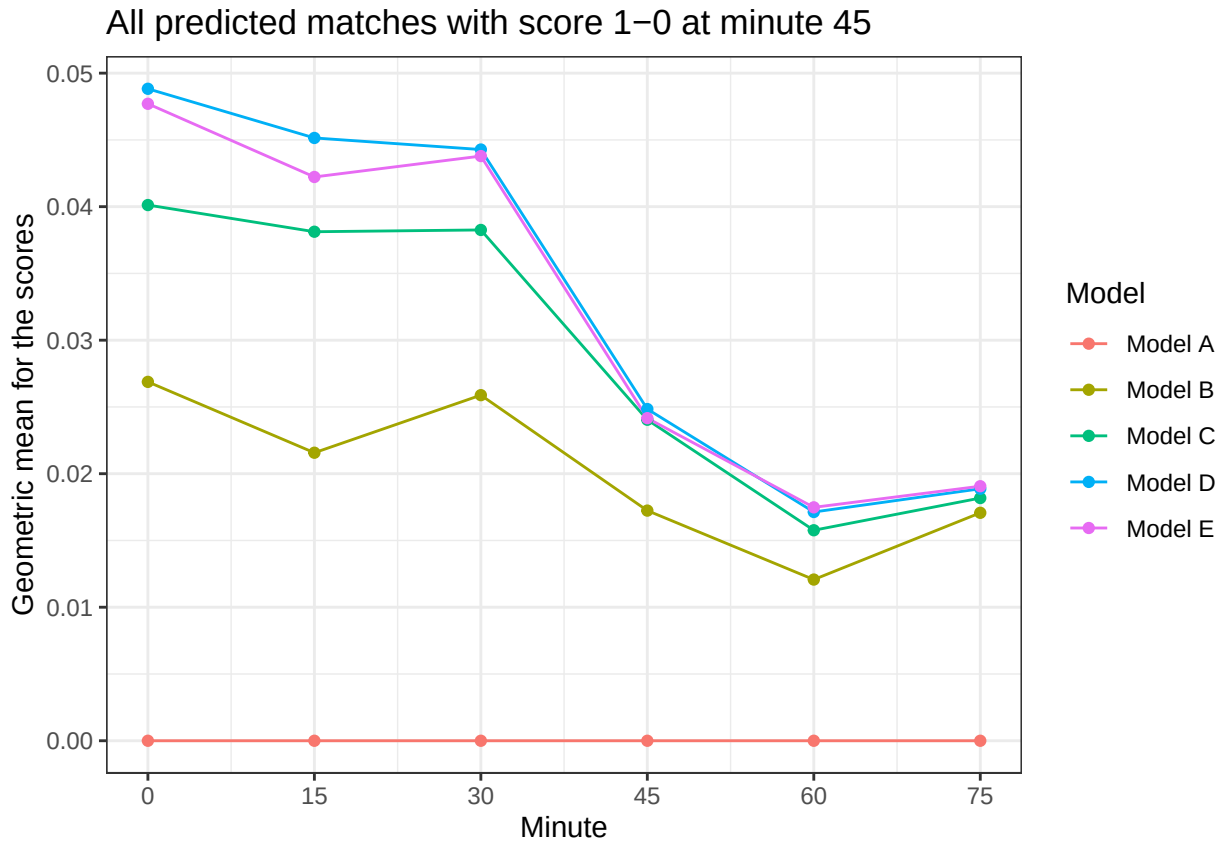



```
results_10 %>%
  pivot_wider(id_cols = "Model", values_from = "GeoMean", names_from = "Minute",
              names_prefix = "Minute ") %>%
  kable()
```

Model	Minute 0	Minute 15	Minute 30	Minute 45	Minute 60	Minute 75
Model A	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
Model B	0.0106138	0.0084752	0.0125425	0.0108392	0.0043706	-0.0005552
Model C	0.0107739	0.0090629	0.0143803	0.0130085	0.0047491	-0.0007233
Model D	0.0211822	0.0199586	0.0234416	0.0150208	0.0058963	-0.0007842
Model E	0.0174388	0.0158467	0.0194304	0.0117638	0.0042342	-0.0020619

```
scores_10 = tibble(GeoMean = apply(HDA_10[,c(129:158)], 2, EnvStats::geoMean),
                   Minute = as.integer(rep(c(0, 15, 30, 45, 60, 75), 5)),
                   Model = factor(c(rep("A", 6),
                                    rep("B", 6),
                                    rep("C", 6),
                                    rep("D", 6),
                                    rep("E", 6)))) %>%
  pivot_wider(names_from = "Model", values_from = "GeoMean", names_prefix = "Model ") %>%
  mutate(`Model B` = log(`Model B`) - log(`Model A`),
         `Model C` = log(`Model C`) - log(`Model A`),
         `Model D` = log(`Model D`) - log(`Model A`),
```

```
         `Model E` = log(`Model E`) - log(`Model A`),
         `Model A` = 0) %>%
  pivot_longer(cols = starts_with("Model"), names_to = "Model", values_to = "GeoMean")

scores_10 %>%
  ggplot(aes(x = Minute, y = GeoMean, col = Model)) +
  geom_line() +
  geom_point() +
  scale_x_continuous(breaks = c(0, 15, 30, 45, 60, 75)) +
  theme_bw() +
  ggtitle("All predicted matches with score 1-0 at minute 45") +
  ylab("Geometric mean for the scores")
```




```
scores_10 %>%
  pivot_wider(id_cols = "Model", values_from = "GeoMean", names_from = "Minute",
              names_prefix = "Minute ") %>%
  kable()
```

Model	Minute 0	Minute 15	Minute 30	Minute 45	Minute 60	Minute 75
Model A	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
Model B	0.0268790	0.0215709	0.0258845	0.0172387	0.0120726	0.0170711
Model C	0.0401219	0.0381234	0.0382565	0.0240422	0.0157654	0.0181781
Model D	0.0488289	0.0451475	0.0442797	0.0248422	0.0171337	0.0188732
Model E	0.0477052	0.0422251	0.0437901	0.0241611	0.0174806	0.0190636

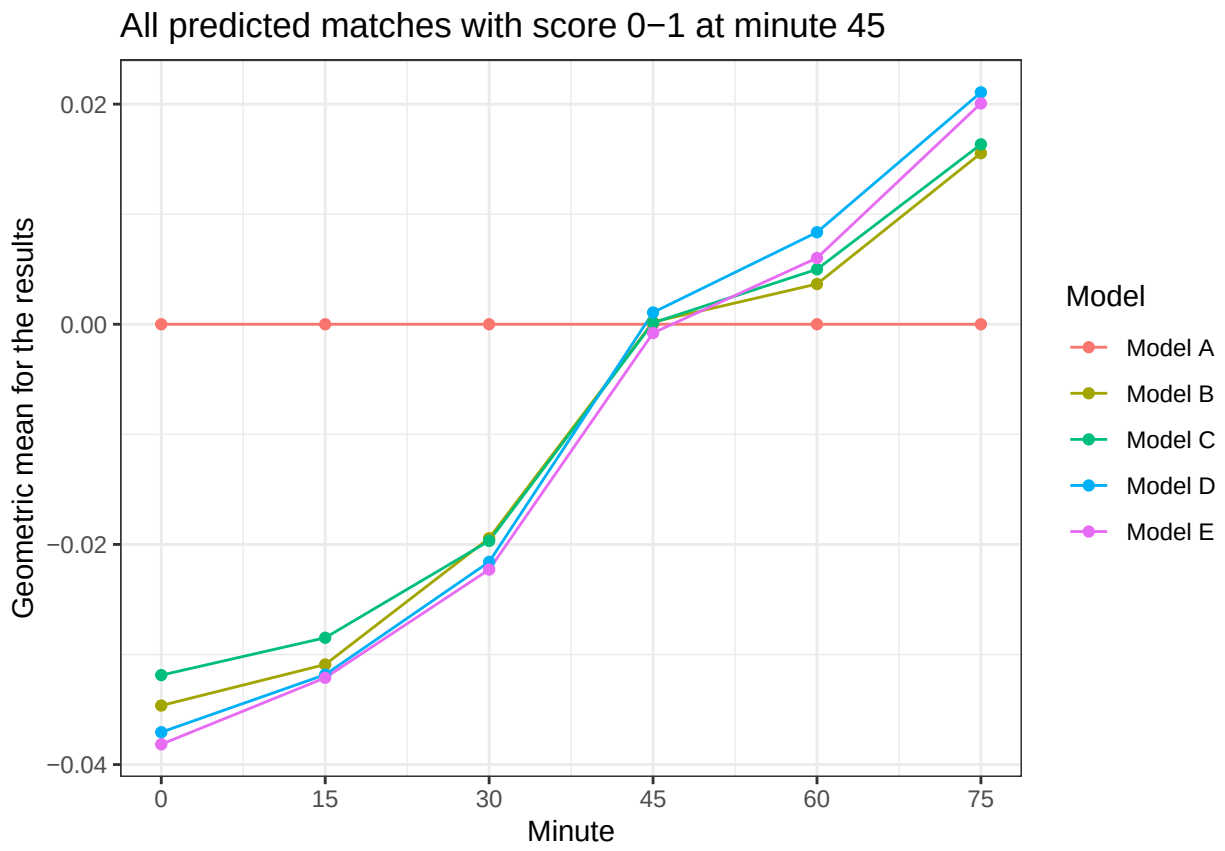
```
tmp_01 = at_45 %>%
  filter(Home_Score == 0, Away_Score == 1) %>%
  select(Season, Match)

HDA_01 = HDA %>%
  inner_join(tmp_01)
```

```
## Joining, by = c("Season", "Match")
```

```
nrow(HDA_01)
```

```
## [1] 275
```

```
results_01 = tibble(GeoMean = apply(HDA_01[,c(99:128)], 2, EnvStats::geoMean),
                    Minute = as.integer(rep(c(0, 15, 30, 45, 60, 75), 5)),
                    Model = factor(c(rep("A", 6),
                                     rep("B", 6),
                                     rep("C", 6),
                                     rep("D", 6),
                                     rep("E", 6)))) %>%
  pivot_wider(names_from = "Model", values_from = "GeoMean", names_prefix = "Model ") %>%
  mutate(`Model B` = log(`Model B`) - log(`Model A`),
         `Model C` = log(`Model C`) - log(`Model A`),
         `Model D` = log(`Model D`) - log(`Model A`),
         `Model E` = log(`Model E`) - log(`Model A`),
         `Model A` = 0) %>%
  pivot_longer(cols = starts_with("Model"), names_to = "Model", values_to = "GeoMean")

results_01 %>%
  ggplot(aes(x = Minute, y = GeoMean, col = Model)) +
  geom_line() +
  geom_point() +
  scale_x_continuous(breaks = c(0, 15, 30, 45, 60, 75)) +
  theme_bw() +
  ggtitle("All predicted matches with score 0-1 at minute 45") +
  ylab("Geometric mean for the results")
```

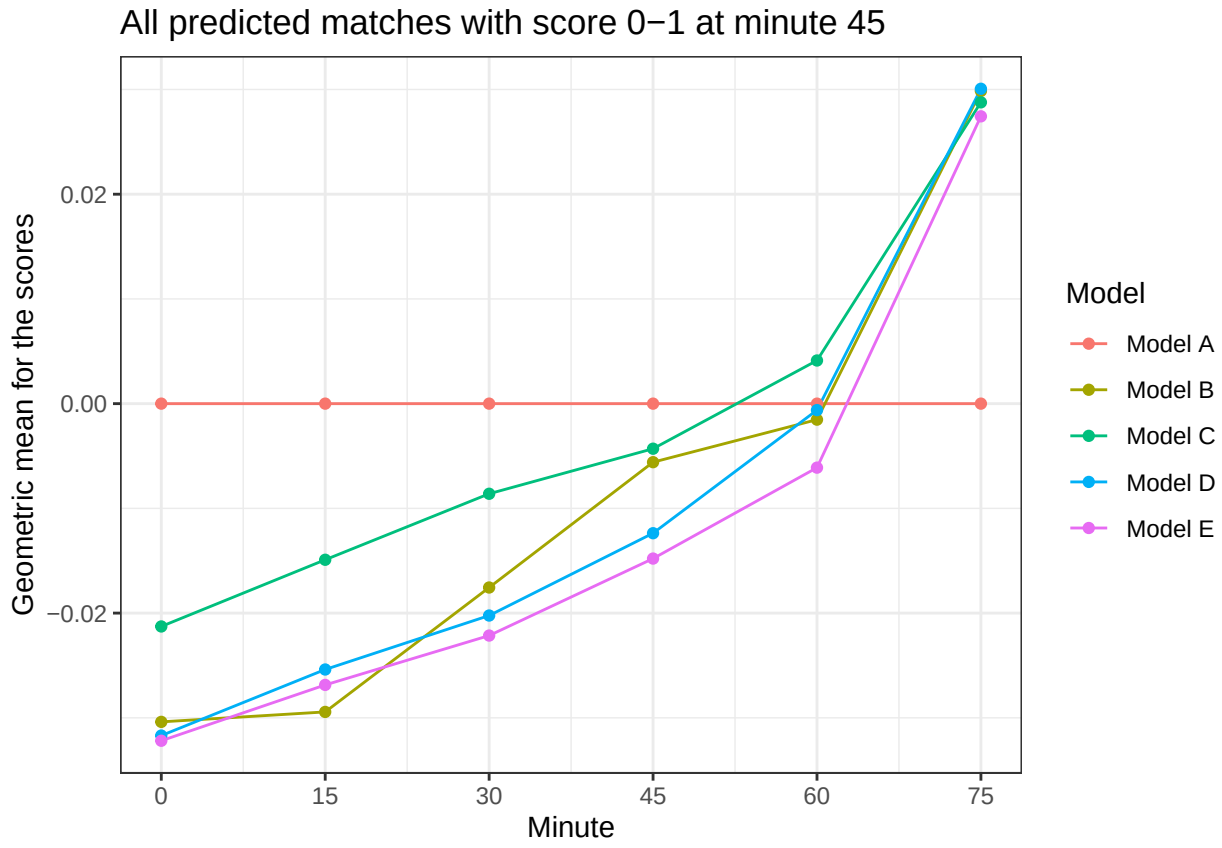



```
results_01 %>%
  pivot_wider(id_cols = "Model", values_from = "GeoMean", names_from = "Minute",
              names_prefix = "Minute ") %>%
  kable()
```

Model	Minute 0	Minute 15	Minute 30	Minute 45	Minute 60	Minute 75
Model A	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
Model B	-0.0346381	-0.0309120	-0.0194354	0.0002052	0.0036562	0.0155435
Model C	-0.0318755	-0.0284809	-0.0196795	0.0001321	0.0049957	0.0163456
Model D	-0.0370617	-0.0318315	-0.0215891	0.0010709	0.0083600	0.0210711
Model E	-0.0381594	-0.0321175	-0.0222853	-0.0007917	0.0060233	0.0200642

```
scores_01 = tibble(GeoMean = apply(HDA_01[,c(129:158)], 2, EnvStats::geoMean),
                   Minute = as.integer(rep(c(0, 15, 30, 45, 60, 75), 5)),
                   Model = factor(c(rep("A", 6),
                                    rep("B", 6),
                                    rep("C", 6),
                                    rep("D", 6),
                                    rep("E", 6)))) %>%
  pivot_wider(names_from = "Model", values_from = "GeoMean", names_prefix = "Model ") %>%
  mutate(`Model B` = log(`Model B`) - log(`Model A`),
         `Model C` = log(`Model C`) - log(`Model A`),
         `Model D` = log(`Model D`) - log(`Model A`),
```

```
         `Model E` = log(`Model E`) - log(`Model A`),
         `Model A` = 0) %>%
  pivot_longer(cols = starts_with("Model"), names_to = "Model", values_to = "GeoMean")

scores_01 %>%
  ggplot(aes(x = Minute, y = GeoMean, col = Model)) +
  geom_line() +
  geom_point() +
  scale_x_continuous(breaks = c(0, 15, 30, 45, 60, 75)) +
  theme_bw() +
  ggtitle("All predicted matches with score 0-1 at minute 45") +
  ylab("Geometric mean for the scores")
```




```
scores_01 %>%
  pivot_wider(id_cols = "Model", values_from = "GeoMean", names_from = "Minute",
              names_prefix = "Minute ") %>%
  kable()
```

Model	Minute 0	Minute 15	Minute 30	Minute 45	Minute 60	Minute 75
Model A	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
Model B	-0.0303868	-0.0294374	-0.0175572	-0.0055894	-0.0015256	0.0298584
Model C	-0.0212765	-0.0149073	-0.0086049	-0.0043029	0.0041221	0.0287738
Model D	-0.0316957	-0.0253860	-0.0202332	-0.0123716	-0.0006176	0.0300661
Model E	-0.0321905	-0.0268592	-0.0221450	-0.0147937	-0.0061110	0.0274365