All in the family: Anna and Lisa Hahner's finishing times in the 2016 women's Olympic marathon

David Cottrell[1] Michael C. Herron[2]

February 3, 2017

[1]Postdoctoral Research Fellow, Program in Quantitative Social Science, Dartmouth College, 6108 Silsby Hall, Hanover, NH 03755 (david.cottrell@dartmouth.edu).

[2]Visiting Scholar, Hertie School of Governance, Berlin, Germany, and Professor of Government, Dartmouth College, 6108 Silsby Hall, Hanover, NH 03755 (michael.c.herron@dartmouth.edu).

"I invested all I had and 300 meters before the finish line, I was next to Lisa. It was a magical moment that we could finish this marathon together. We did not think about what we were doing." – Anna Hahner

Introduction

At 9:30am on August 14, 2016, the Women's Olympic marathon kicked off in Rio de Janeiro, Brazil, when 156 runners from 80 countries across the world left the starting line en route to their destination 42.195 kilometers away. Two hours, twenty-four minutes, and four seconds later, Jemima Sumgong of Kenya would be the first to cross the finishline and take home gold; Sumgong was just three and one-half minutes slower than her prior personal best time in the marathon. Approximately 21 minutes later, twin marathoners from Germany, Anna and Lisa Hahner, would cross the finishline together, holding hands and celebrating a personal victory. Although the Hahners would finish 81st and 82nd, respectively, well behind the winners of the marathon, Anna Hahner would describe their joint finish as a "magical moment."

The media quickly picked up on the Hanher story as an image of the beaming twins finishing hand-in-hand captured a public audience. While many believed the moment was a reflection of the Olympic spirit, not everyone agreed with this rosy interpretation. The twins' happy facial expressions at the finish were portrayed as a bit contrived—smiling like "Honigkuchenpferde," cookies in the shape of a horse—and the sports director of the German Athletics Federation, Thomas Kurschilgen, stirred up controversy when he suggested that the Hahners' photo-finish was no coincidence. Kurschilgen averred that the twins slowed down so as to finish simultaneously and create a spectacle which would "generate media attention." Kurschilgen justified his charge with the fact that the twins ran in the Rio marathon at least 18 minutes slower their personal best times prior to the Olympics.[1] Not surprisingly, Kurschilgen's accusations were denied by the Hahner twins, who

[1] https://www.nytimes.com/2016/08/17/sports/olympics/twins-

claimed that their simultaneous finish was simply an unintended coincidence.

What happened in the women's Olympic marathon, and how might we develop a statistical approach that assesses whether the Hahner twin's finish was coincidental or intentional? These two interpretations are clearly at odds. If the former, then the Hahners are to be celebrated and their finish treated as an expression of the spirit behind the Olympic games. If the latter, though, then the twins may have violated this spirt by not trying hard enough. It is perhaps too easy for us to write such a glib sentence—neither of us can fathom being able to complete a marathon anywhere in the vicinity of two and a half hours—but we nonetheless want to know what the data from the Olympic marathon tell us. Was the Hahner finish in the 2016 women's Olympic marathon a lovely coincidence or something else?

Among female Olympic marathoners, the Hahner twins were not alone in their familial ties. The marathon also featured twins from North Korea, Kim Hye-song and Kim Hye-gyong, who posted identical times and finished 10th and 11th in the race, respectively. The Kim finish, unlike the Hahner finish, appears devoid of post-race controversy. Moreover, three triplets from Estonia competed in the Rio marathon, although only two, Lily Luik and Leila Luik, finished it, in 97th and 114th place, respectively. The third Estonia triplet, Liina Luik, recorded what is known as a DNF—an abbreviation that means did not finish. Although our focus here is the Hahner twins, we touch on the Kim twins and Luik triplets in the course of what follows.

Marathon data and our research design

For each participant who started the women's Olympic marathon, we know several things: personal best marathon time prior to the 2016 Olympic games; split times from the Rio marathon course at 5 kilometers, 10 kilometers, and so forth; and, finishing time. We cannot directly observe the

`finish-marathon-hand-in-hand-but-their-country-says-they-crossed-a-line.html` effort that an individual put into the race, and we do not know why some runners have DNF results. Some runners may have injured themselves on the course and accordingly dropped out, and others may have dropped out, uninjured, in anticipation of an unsatisfactory result. Of the 156 marathon starters, 133 completed the race and 23 DNFed at various locations throughout the course. The overall DNF rate was thus $\frac{23}{156} \approx 0.15$, and the relatively small sample size at our disposal means that a 95% confidence for this rate is fairly wide, namely, $(0.098, 0.22)$.

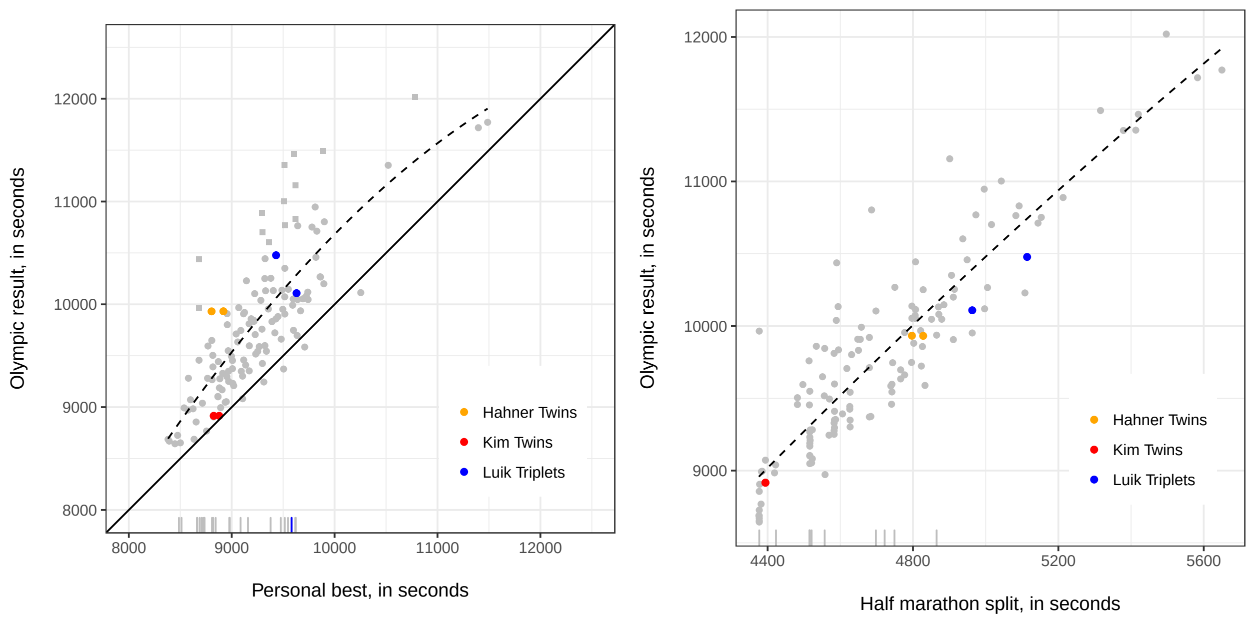
The Kurschilgen accusation against the Hahner twins has two components, that these two women ran slowly *and* that they finished simultaneously. We suspect that Kurschilgen would not have expressed ire at the Hahners had they finished in 1st and 2nd place in Rio, hand-in-hand with wide grins, but of course we do not know this. As such, our investigation of the charges that Kurschilgen offered will distinguish between the idea of slow finish versus a simultaneous finish.

Our research design is twofold. First, we present visualizations that describe various features of the Hahner twin's results, and an important element in our visualizations is the difference between a runner's Rio time and her prior personal best time in the marathon. Overall, our visualizations suggest that the Hahner twins' pace in the marathon was slow, but not excessively so, but that their simultaneous finish was quite unusual given the twins' differences in abilities (and similarly for the Kim twins). We then turn to a regression-based simulation of the marathon, which we develop based on relationships (excluding the Hahner and Kim twins and Luik triplets) between known personal best times and observed Rio finishing times. Our simulations show that the Hahner twins finished suspiciously close to each other given the disparty between their personal best times. Finally, we return to Kurschilgen claims about the Hahners and offer our thoughts about their validity.

Visualizing the Olympic marathon

Figure 4 contains two plots, both of which describe how finishing times from the Rio Olympic women's marathon varied as a function of athletes' personal best times and their Rio half marathon split times. The points in the plots are colored by twin/triplet status, and both plots contain second-order polynomial regression lines. We treat an athlete's prior personal best marathon time as a measure of the athlete's underlying marathon talent, and we interpret an athlete's half marathon split as a measure of the athlete's skill at marathon running on August 14, 2016, the date of the women's marathon in Rio.

Figure 1: Olympic results and measures of marathon ability



(a) Finishing and personal best times

(b) Finishing and half marathon split times

Figure 1b includes a solid 45-degree line as the axes in the figure correspond to full marathons; in addition, the figure's rug marks denote personal best times of runners who DNFed. Given the paucity of points (five of them) below the pictured 45-degree line, the vast majority of Olympic marathoners ran slower in Rio compared to their personal bests. Relative to personal best times, the Hahner twins were definitely on the slow side, but a number of runners had greater differences between their Olympic times and their personal bests than the Hahners. These runners are denoted with squares in Figure 4a, and there are 13 such symbols, highlighting approximately 10% of the racers who completed the marathon. The figure shows moreover that there were two women who had personal best times slightly faster than the Hahner's and yet finished after the German twins. Although Figure 4a suggests that the Hahner twins were slower than one would have expected given their previous best marathon times, it is not consistent with the accusation that they dramatically slowed down in the Rio marathon.

One limitation of using personal best times as indicators of underlying athletic talent is that these times are potentially confounded by the marathon courses at which they were set (some courses, like the Berlin marathon, are known for fast times) and race conditions like weather. In addition, personal best times may not capture raceday idiosyncrasies that might affect individual runners. With that in mind, Figure 1b plots marathon finishing times against half marathon splits. This figure has a regression line as before but no 45-degree line.

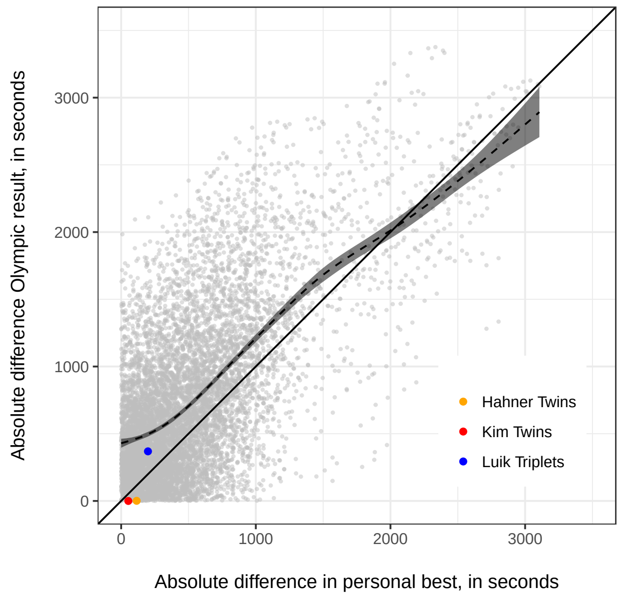
Figure 1b shows that there was nothing abnormal about the Hahner twins' overall finishing times, conditional on their half marathon split times. This is consistent with the previous Figure 4a and inconsistent with accusations made against the Hahner twins, at least the part of the accusation that focused on their overall pace.

If the Hahner twins did not slow down excessively, might they have run somewhat strategically at the end of the race in order to generate a simultaneous finish? We now offer a visualization that speaks to this question.

The personal best times of the Hahner twins were 115 seconds apart and their official finishing times were separated by one second. Is such a 115 to one compression typical among pairs of runners or, in contrast, was it unusual? Are there other pairs of marathoners who had a difference between personal best times of 115 seconds apart and, if so, how close where their finishing times?

Of the 133 marathon finishers, there are $\binom{133}{2} = 8,778$ pairs of runners. Of these and ignoring the Hahner twins, ten had exactly a 115 second gap in personal best times. Of these ten pairs, differences in finishing times, in seconds, are as follows: 36, 93, 172, 319, 379, 459, 552, 671, 675, and 739. In other words, of all pairs of runners in the Rio marathon who had a personal best difference that was equivalent to the Hahner twins' difference, the twins had the greatest compression based on finishing time.

Figure 2: Differences in personal best times and finishing times



We can generalize this result by looking at all pairs of runners in the marathon. For all 8,778 pairs of 133 finishers, Figure 2 plots differences in finishing times against differences in personal best times, and pairs of twins/triplets are identified by the same color scheme we used earlier.

Consider first the Hahner twins. They are effectively located on the horizontal axis because their difference in finishing times is one second. However, there are many points about the Hahner's orange dot, and this shows that, conditional on an approximate 115 second difference in personal best times, most marathoners did not have close finishing times like the Hahnes. Some pairs of runners with around 115 second personal best differences had finishing time differences of 1000 seconds, i.e., in excess of 15 minutes. Of course the points in Figure 2 are not independent, but they provide a sense of the dispersion in finishing time differences that one can expect conditional on differences in personal best times.

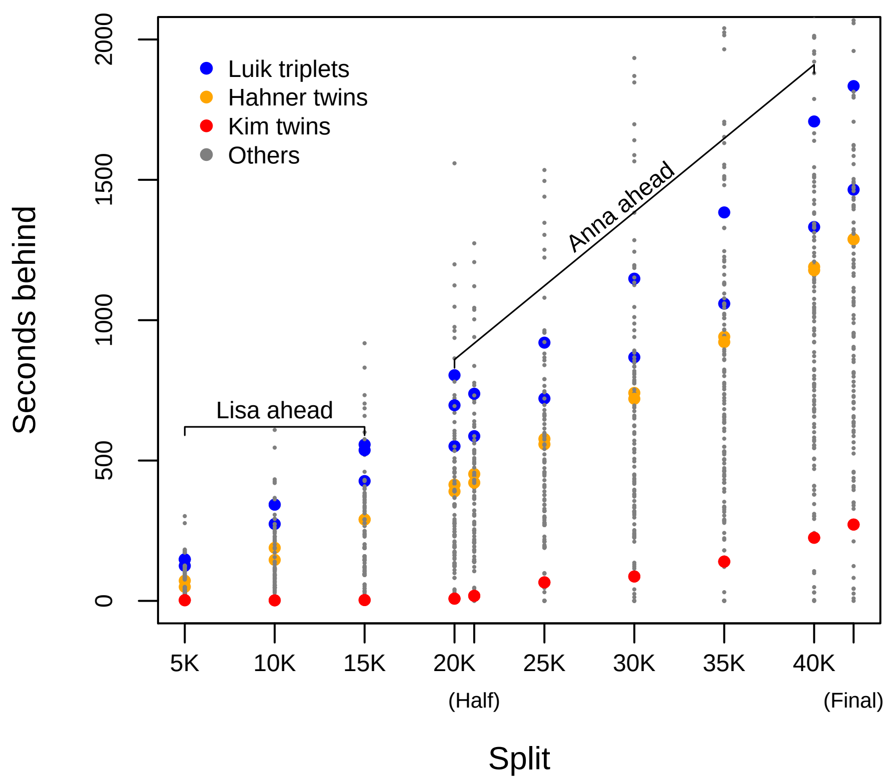
Thinking about the accusations leveled against the Hahner twins, Figure 2 suggests that Anna and Lisa Hahner did indeed run with an eye on each other. In fact, the same can be said of the Kim twins, who ran seemingly in lockstep throughout the entire Rio marathon. The North Korean twins had a personal best difference of 53 seconds and a finishing time difference of literally zero seconds. Beyond these twins, there were eight pairs of Rio runners with a 53 second personal best difference, and resulting finishing time differences are as follows: 9, 51, 228, 340, 352, 571, 662, and 751. As in the Hahner case, the Kim twins compressed their finishing times more than any other pair of runners with a similar personal best difference.

To get a sense of the lockstep nature of the Kim twins race, Figure 3 describes each runner's status at the various split times on the marathon course. Each dot in the figure depicts a recorded split and the number of seconds each runner was behind the race leader at the time. There are more dots at earlier splits due to the accumulation of DNFs. The usual color scheme applies here, and the Estonian DNF prior to the half marathon split is evident. The orange dots in the figure represent the Hahner twins, and as noted in the figure Lisa Hahner was ahead of her sister through 15 kilometers at which point Anna surged. Figure 3 contains two red dots representing the North Korean Kim twins, but this is not visually apparent because the Kim twins had identical split times during the entire race marathon.

The probability of an unintentional simultaneous finish

Thomas Kurschilgen's accusations of an intentional finish stems from two observations. First, the Hahner twins finished slower than he expected given respective personal bests. Second, the twins finished together at the exactly same time. Kurschilgen clearly believed that neither of these events would have occurred had both Hahner run the marathon independently, absent coordination. By his logic, the German twins should have run faster and not have finished simultaneously.

Figure 3: Runner status by split



Note: each dot represents one runner at a split. DNFs are not pictured, and splits are not to scale.

On the other hand, only a handful of runners completed the Rio marathon with times that were faster than their recorded best times. Thus, 18 minutes behind a personal record may not be the outlier that Kurschilgen claimed it to be. Moreover, there is good reason to suspect that an unintended simultaneous finish was more likely for the Hahner twin than it would be for any other set of runners. After all, these two women are twins with presumably similar abilities. Not only do they train together, but their pre-Rio best times are less than two minutes apart. While Anna may be slightly faster than Lisa measured by personal bests, we might expect the difference in their Rio result to be just as close as the differences in their recorded bests. And given random variation in finishing times, a simultaneous finish might not be out of the ordinary.

Even though they seem similar, finishing at very similar times and finishing together are different phenomena. If, for example, both Hahners were of similar ability to each other and also to many other runners, then we might expect similar finishing times yet not necessarily similar placements. The latter will be a function of the extent to which all runners on the Rio course have similar talent levels. This point is an important one and will be evident in the results that follow.

To test the assertion that the Hahner twins paced themselves to finish at the same time and with back-to-back placements, we need to compute the probability that such a result would have occurred unintentionally. Hence we need to know the probability distribution of Anna and Lisa's finishing times if their runs had been independent of each other. The challenge, of course, is that this distribution is unknown.

Modeling

The Hahner twins either acted independently I or coordinated their times C; these are the two possible states of the race $\psi = \{I, C\}$. Given that we observed Anna and Lisa's finishing times, $Y_A = y_A$ and $Y_L = y_L$, we want to know the probability that they ran independently, as they say they did. Hence, we are looking to determine,

$$P(\psi = I \mid y_A \cap y_L)$$

However, to determine this, we need to have some understanding of a likelihood function that specifies Rio finishing times. We want to know the likelihood of Anna's and Lisa's final times given independence $P(y_A \cap y_L \mid \psi = I)$. We can estimate this function with a few assumptions. First, we assume that under independence, any given runner's final time Y_i is conditional on his/her running ability plus noise. Specifically, we assume that the Y_i is a linear function of the runner's ability X_i plus a normally distributed error term $e_i \sim N(0, \sigma)$. Second, we assume that every runner shares the same linear relationship - meaning the slope and intercept remain constant across runners. Third, we assume that the error term is drawn from a common distribution across runners. Hence, luck and misfortune are drawn from the same distribution. Therefore,

$$Y_i \sim N(\beta_0 + X_i\beta_1 + e_i, \sigma)$$

We also assume that a runner's ability X_i can be measured precisely by her best marathon performance leading up to the Olympics.

If Anna and Lisa intentionally slowed down as a result of coordination then we would likely observe $y_A > E(Y_A \mid \psi = I)$ and $y_L > E(Y_L \mid \psi = I)$. In other words, the final times that Anna and Lisa recorded in the race would be greater than we would expect if they had run independently.

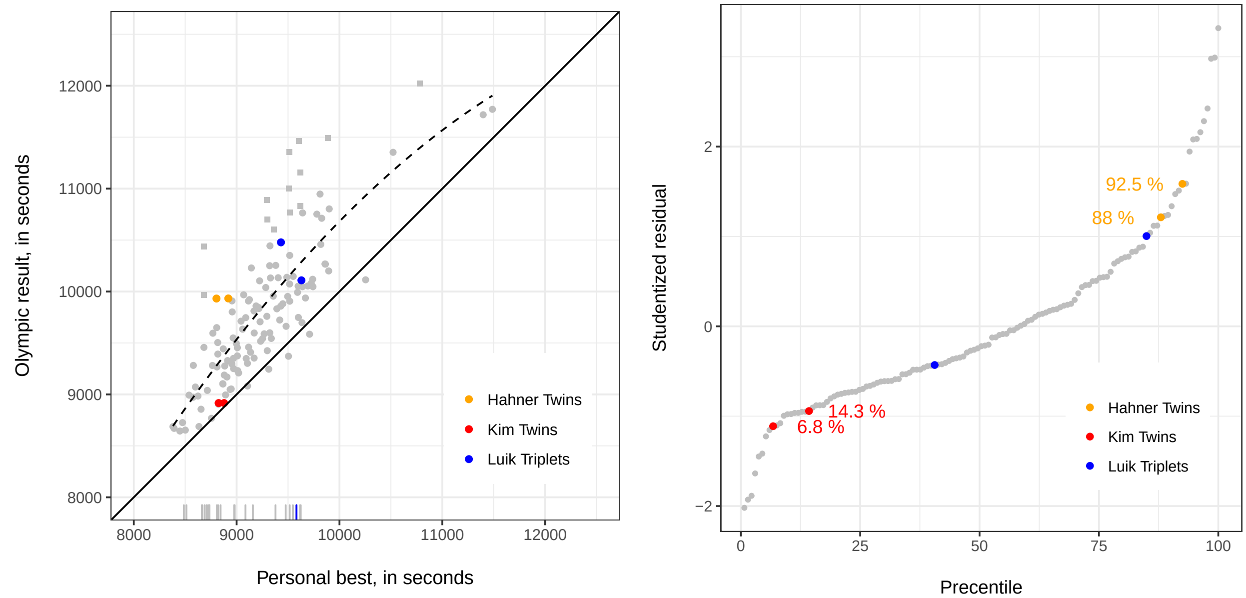
Moreover, if Anna and Lisa coordinated to finish simultaneously, then the difference between the two sisters' final times would be less that the expected difference had they run independently. Therefore, under a coordinated finish we would expect $|y_A - y_L| < E(|Y_A - Y_L| \mid \psi = I)$.

Did Anna and Lisa intentionally slow down?

According to Kurschilgen, Anna and Lisa underperformed in the Rio marathon. He claimed that because their goal was to finish simultaneously rather than finish at their fastest pace, their times were slower than they otherwise would have been. He claimed that the twins were simply trading speed for a photo-finish.

If this were the case, the function generating Anna and Lisa's final times would deviate from the function that generated everyone else's final times. Given everyone else would draw their times from the independent distribution $Y_i \sim N(\beta_0 + X_i\beta_1 + e, \sigma)$, Anna and Lisa would draw from a distribution of times that are slower in expectation. Hence, they would lie well-above the line that links a runner's performance in Rio to their previous best performance.

Figure 4: Relationship between Personal Best and Result – caption?



(a) Rio finishing times and personal best times

(b) Studentized residuals

Figure 4a displays the relationship between a runner's personal best (horizontal axis) and her final result (vertical axis). The 133 finishers are plotted as grey points, and the 23 DNFs are depicted below along the horizontal axis with tick-marks. The dashed black line with grey confidence intervals represent ordinarily least squares estimates of the linear relationship between racer's finishing and personal best times.

One can see in Figure 4 that there is a clear relationship between a runner's personal best time and her performance in the Olympics. While there is a significant amount of noise in the relationship, the Hahner twins' dots are well above the regression's fitted line notwithstanding this point. They finished the Rio marathon much slower than they should have, per their pre-Olympic personal best times. In fact, if we sort studentized residual, we can see in Figure 4b that the Hahner twins are in the tail-end of the residual distribution. Anna Hahner's residual deviation is greater than 91.7% of the runners who completed the marathon while Lisa Hahner's deviation is greater than 86.5%. Hence, the twins appear to have finished at a much slower pace than expected, which is what we might expect if they were coordinating their runs. The North Korean Kim twins are also in the tail of the residual distribution—but the left tail. These twins finished much faster than expected.

Did Anna and Lisa intentionally finish together?

Anna and Lisa finished the marathon at an unusually slow rate, which is what we might expect if the twins intentionally slowed their pace to finish simultaneously. However, the observation of a slow finish does not necessarily imply that it was an intentional result . It only suggests that Kurschilgen's claim regarding this issue holds up in the data.

Perhaps it would be easier to refute the claim if we could show that the Twins's back-to-back finish was not such an unusual event. Given that Anna and Lisa were so close in their personal best times, it may seem reasonable to expect that they would also be close in their finishing times. Hence, one way to determine if the twins's finish was intentional is to quantify the likelihood that such a close finish would occur by chance given the differences in their ability.

We can attempt to quantify this likelihood by looking at how differences in the personal best time of two runners translate into differences in Olympic times . If Anna and Lisa's finish was unintentional we'd expect that such a close finish would be unlikely given their difference in personal best times. Yet, how does the difference in personal best times relate to the difference in finishing times?

One way we can estimate this relationship is to look at all of the differences that we observe in the marathon. We can take the difference in personal best time for every combination of runners in the data and look at the difference in their final times. We should expect that compared to the difference in their personal best times, the difference in Anna and Lisa's final times should be commensurate with the conditional difference of all other runners.

Therefore, we find all combinations of runners in the marathon and calculate two quantities:

1) the difference in their personal best time and 2) the difference in their final time. We plot the relationship between the two quantities in Figure 2. All 8,778 dyadic combinations of runners are displayed by grey points in the scatter plot. The solid black line indicates a one-to-one relationship, where every point would fall if the difference in final times were systematically equivalent to the difference in personal bests.

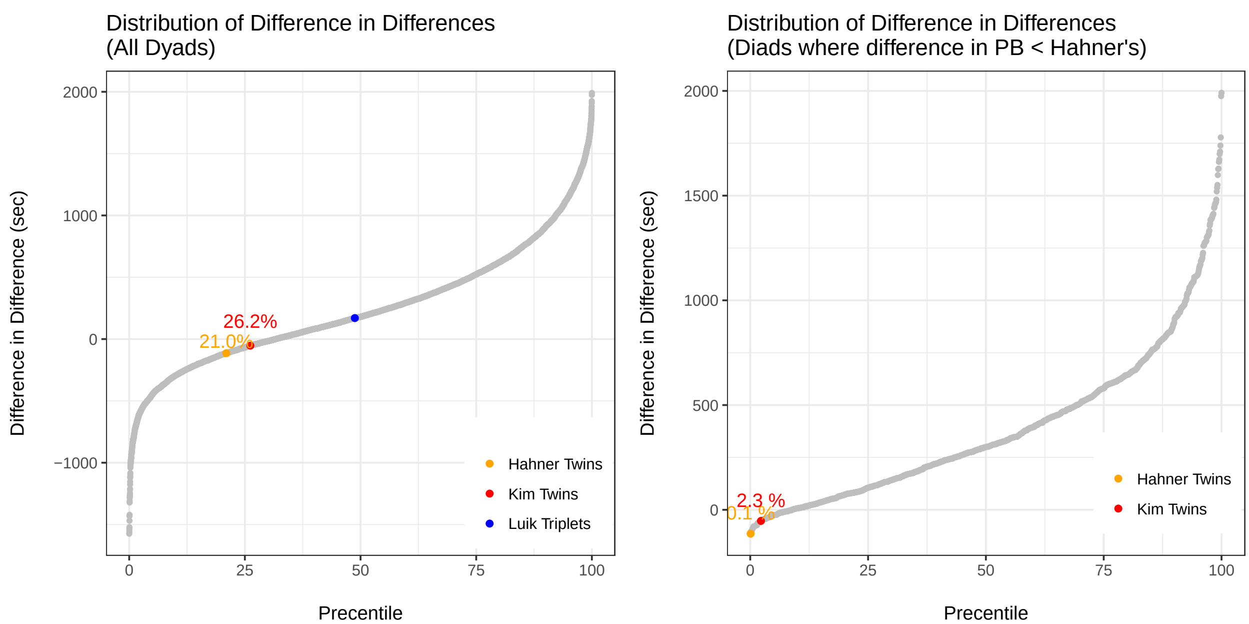
We can see from the yellow dot that the Hahner Twins are just below this one-to-one line. Hence, the twins' final times were closer to each other than their personal best times; but not by much. In fact, this might be an expected difference if we expected there to be a one-to-one relationship. However, although there is a clear relationship between the difference in final time and the difference in personal best, it is not a strong one-to-one relationship. We've added a dotted smoother to display a cubic splines estimate of the conditional averages. While the Hahner twins may be close to the one-to-one line, they are much further away from the average set of runners. Hence, we would expect for any pair of runners, the difference in their final results should be much greater than their difference in their personal best. However, the Hahner twins do not fit this expectation. Yet, how different from this expectation are they?

One way we can quantify the degree to which the twins are outliers is by comparing their difference in difference to every other pair of runner's difference-in-differences. We calculate the difference-in-differences for each pair of runners by simply subtracting their difference in result (y-axis) from their difference in personal best (x-axis). Therefore positive difference-in-differences mean that the runners ended up further apart than their difference in personal bests. Negative difference-in-differences mean that the runners end up closer together. Hence, the Hahner twins have a difference-in-difference of -114 seconds. They finished 114 seconds closer together than the distance between their personal bests.

We arrange all 8,778 difference-in-differences from small to large in the first plot in Figure 5. The Hahner twins are denoted by the yellow point. Using this plot we can visualize where the twins fall within the distribution of difference-in-differences. It is clear that they are closer to the lower tail of the distribution. However, compared to all other difference-in-differences, their result is not that extreme. Nearly 21% of the difference-in-dfferences were less than the that of the twins'. Hence, we'd expect at least one in five pairs of runners to have a reduction in difference at least as great as the twins.

However, if we limit the observations to only those runners who had personal best times that were at least as close to each other as the Hahner twins, we get a different perspective. Among those runners, the Hahner twins reduced the distance between them to the greatest extent. There was no other pair of runners whose reduced the difference in the final time relative to their difference in personal bests more than them.

Figure 5: How Hahner Twins Rank in Difference-in-Differences



Simulating the Marathon

The Hahner sisters claimed that they ran the marathon independent of each other. Their simultaneous finish was no more than a product of ability and luck. One might expect such a finish given that their personal best marathons are within 114 seconds of each other. Add a little bit of luck to the occasion and, as the twins have claimed all along, one might easily argue that the simultaneous finish is simply a joyous coincidence rather than a wholly implausible event.

This begs the question: if we take into consideration their similarities as runners as well as some natural variation in finishing a marathon, what are the odds that the twins unintentionally finished at the same time? To answer this question we would want to know the counterfactual distribution of potential finishing times that would have occurred had the twins independently and repeatedly run the marathon over again. Holding the twins' similarities constant, this would establish the set of potential outcomes that would have occurred as result of natural variation in the finishing the race. Then we could use this distribution of outcomes to determine the likelihood that a simultaneous finish (or at least a near-simultaneous finish) occurred by chance alone. If the twins rarely finished the race together in this counterfactual, then one should be skeptical that such a finish occurred without coordination.

Unfortunately, it is not possible to rerun the women's marathon in Rio to establish this counterfactual distribution. However, we can attempt to simulate it, by estimating the distribution of every other runner's results - conditioned on their prior abilities - and drawing from that distribution in order to observe the likelihood that Lisa and Anna finish simultaneously.

To estimate the conditional distribution of every runner's results, we must assume that each runner's time is distributed with a mean that is a quadratic function of X_i and β_j, such that

$$Y_i \mid X_i \sim N(\beta_0 + \beta_1 X_i + \beta_2 X_i^2 + e_i, \sigma).$$

This means that we are assuming that each runner's time is drawn from the normal distribution with a mean that is a quadratic function of their personal best and variance that is constant across all runners. We simply estimate β_j and σ using ordinary least squares. We then draw every runner's time from this estimated distribution, conditioned on their personal best outcome. Once a race is simulated, we locate Anna and Lisa and record the time between their finishes and the difference in their rank. We then simulate a new race - drawing a new set of finishing times - and record the same quantities. After repeating this procedure 10,000 times, we plot the distribution in Figure 6. The steps of the simulation are detailed below.

1. Ignoring twins and triplets, estimate with least squares a linear model that predicts a runner's finishing time (Y_i) based on her pre-Olympic personal best time (X_i) and her best time squared.

2. Extract coefficient vector and the covariance matrix from this model.

3. For each simulated rate, draw intercept and slope estimates β_0, β_1, and β_2, respectively, from a multivariate normal distribution with mean equal to the estimated coefficient vector and covariance equal to the estimated covariance matrix.

4. For each runner, draw an error (e) from a normal distribution with mean zero and a standard deviation equal to the standard deviation of the model's residuals (σ).

5. For each runner, predict the final result by combining the randomly generated beta coefficients and error terms, $\hat{y} = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + e$.

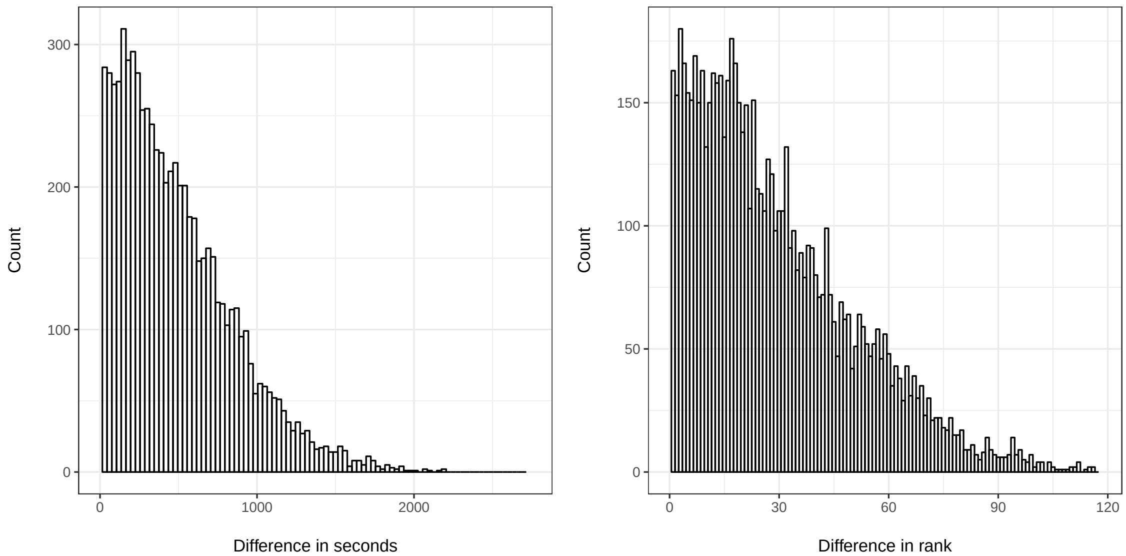
6. Eliminate each runner from the race with a probability equal to the fraction of runners who did not finish the marathon.

7. Calculate the difference in time and difference in ranking between Anna Hahner and Lisa Hahner, assuming both finished the race.

8. Repeat above steps ten thousand times.

9. Plot a histogram of the results in Figure 6.

Figure 6: Distribution of simulated results for the Hahner twins



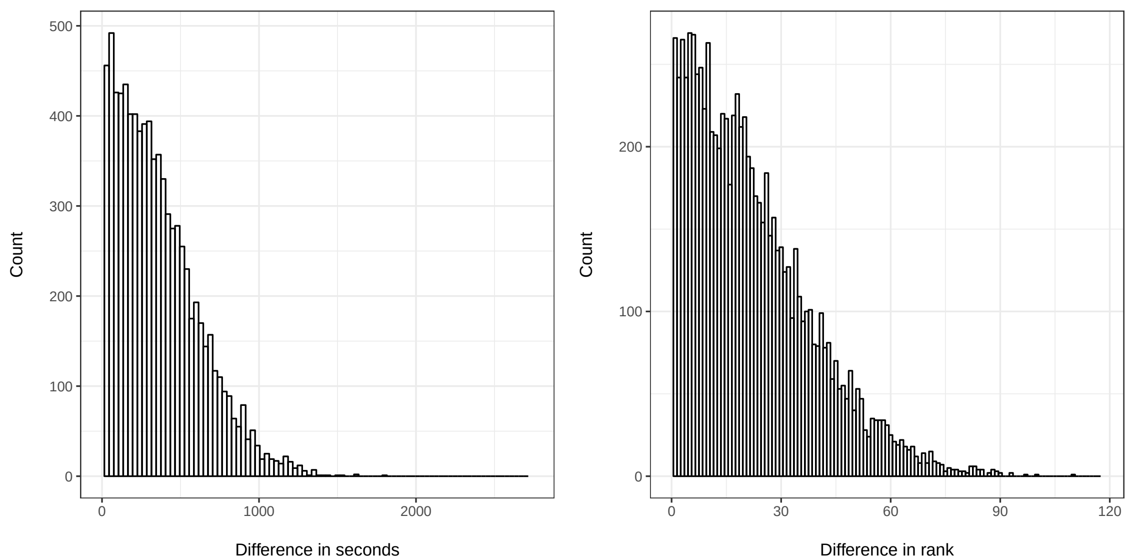
(a) Simulated differences in finishing times (b) Simulated differences in places

A histogram of the differences between Anna and Lisa's simulated finishes can be found in Figure 6. The first plot shows the distribution of the absolute differences in time. Differences are grouped in 30 second bins along the x-axis and the count for each bin is denoted by the vertical length of bar along the y-axis. The second plot shows the distribution of the absolute differences in ranking. Differences are grouped as single units ranging from no runner between the twins to nearly 120 runners between them.

The results plotted in the histograms question the credibility of Anna and Lisa's story. The results suggest that a simultaneous finish would be very rare if Anna and Lisa had run independently of each other. For example, in less than 300 of the 10,000 simulated races did Anna and Lisa finished within 30 seconds of each other . And they finished in consecutive rank in less than 175 of the 10,000 races. Therefore, the close finish that we observed, where Anna and Lisa crossed the finish line one-after-the-other, would be very unlikely if we were to believe the twins raced independently of each other.

The results holds even if you replace their personal best time with their time halfway through the race. While this reduces the variation in the predicted outcome of each runner and, therefore, reduces the expected distance between Anna and Lisa, it is still quite rare for Anna and Lisa to finish simultaneously or consecutively.

Figure 7: Distribution of simulated results for the Hahner twins (using halfway split as predictor)



(a) Simulated differences in finishing times

(b) Simulated differences in places

Conclusion

Further reading

- http://www.telegraph.co.uk/olympics/2016/08/17/german-twins-criticised-for-finishing-olympic-marathon-fun-run-h
- https://www.nytimes.com/2016/08/17/sports/olympics/twins-finish-marathon-hand-in-hand-but-their-country-says-they-crossed-a-line.html
- https://www.welt.de/sport/olympia/article157669264/Das-falsche-Laecheln-der-deutschen-Lauf-Zwillinge.html