All in the family:
German twin finishing times in
the 2016 women's Olympic marathon

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

February 24, 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 the Hanher story as an image of the beaming twins, finishing hand-in-hand, captured a public audience. While many believed the twin's near-simultaneous finish was a reflection of 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, was the description offered by one editorialist—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 the Rio marathon over 18 minutes slower their personal best times prior to the Olympics. Kurschilgen's accusations were denied by the Hahner twins, perhaps not surprisingly, who claimed that their joint finish was simply a fortuitous coincidence.

What happened in the women's Olympic marathon in Rio, and how might we develop a statistical approach that assesses whether the Hahner twin's finish in the race 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.

Among female Olympic marathoners, the Hahner twins were not alone in sharing familial ties. The Rio 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; age; split times from the Rio marathon course at 5 kilometers, 10 kilometers, and so forth; and, overall finishing time. We cannot directly observe the 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 yields a relatively wide 95% confidence for this rate, namely, $(0.098, 0.22)$.

Kurschilgen's 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. Thus, our investigation of the charges that Kurschilgen offered distinguishes between a slow finish and a simultaneous finish.

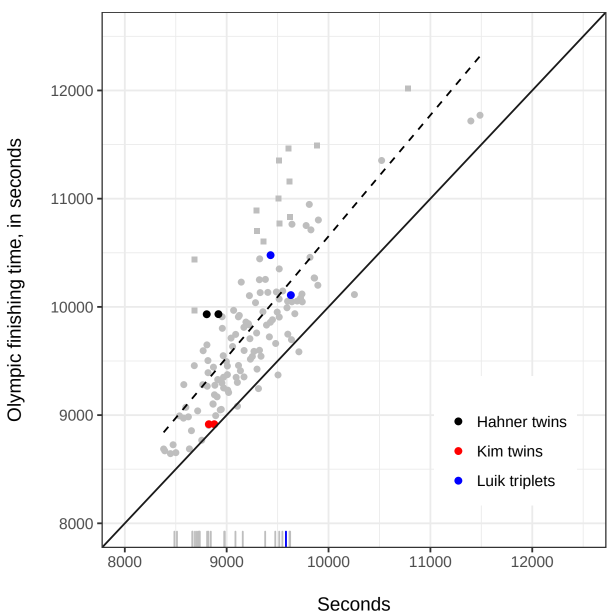
Our research design is twofold. First, we present visualizations that describe various features of the 2016 women's Olympic marathon, and an important element in our visualizations is the difference between runners' Rio times and their prior personal best marathon times. Our visualizations suggest that the Hahner twin's pace in the marathon was slow, albeit 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, and our simulations reinforce what we observed in prior visualizations, namely, that the Hahner twins did not run appreciably slowly yet finished suspiciously close to each other. We return in our conclusion to Kurschilgen's claims about the Hahners and offer thoughts about their validity.

Visualizing the Olympic marathon

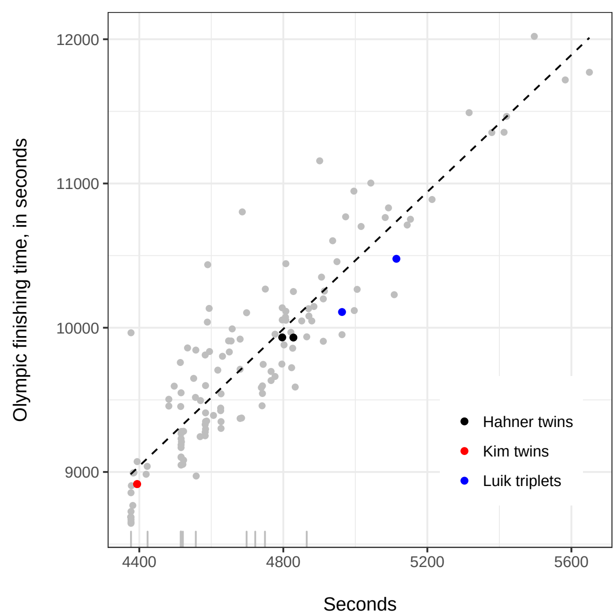
One way that we might assess whether a women marathoner's Rio finish was unusual—or, say, whether two finishes were jointly unusual—is by comparing a runner's observed Olympic finishing time on August 14, 2016, with a measure of her underlying marathon talent. For the latter we use prior personal best marathon times. This is a natural measure of marathon talent, we believe, but it is not entirely free of complications. For example, personal best times are potentially confounded by the marathon courses on which they were set; some courses, like the Berlin marathon, are known for relatively fast times. In addition, personal best times may be confounded by conditions, like weather, that varied across races. Finally, personal best times may not capture Olympic raceday idiosyncrasies that could affect individual runners. For example, a runner could have woken up with a minor cold on the morning of August 14, 2016. With these concerns in mind, we use a runner's half marathon split time from the Olympic marathon as a secondary measure of athletic skill in marathoning. Our two measures of marathon talent are featured in the visualizations below.

Figure 1 contains two plots, both of which describe how finishing times from the Rio women's marathon varied as a function of athletes' personal best times and half marathon split times. The points in the plots are colored by twin/triplet status, and both plots contain dashed lines representing least squares regression fits. The ticks along the horizontal axes in both plots indicate times, either personal bests or half marathon splits, of runners who earned DNF results.

Figure 1: Olympic finishing times as a function of underlying marathon talent



(a) Based on personal best



(b) Based on half marathon split

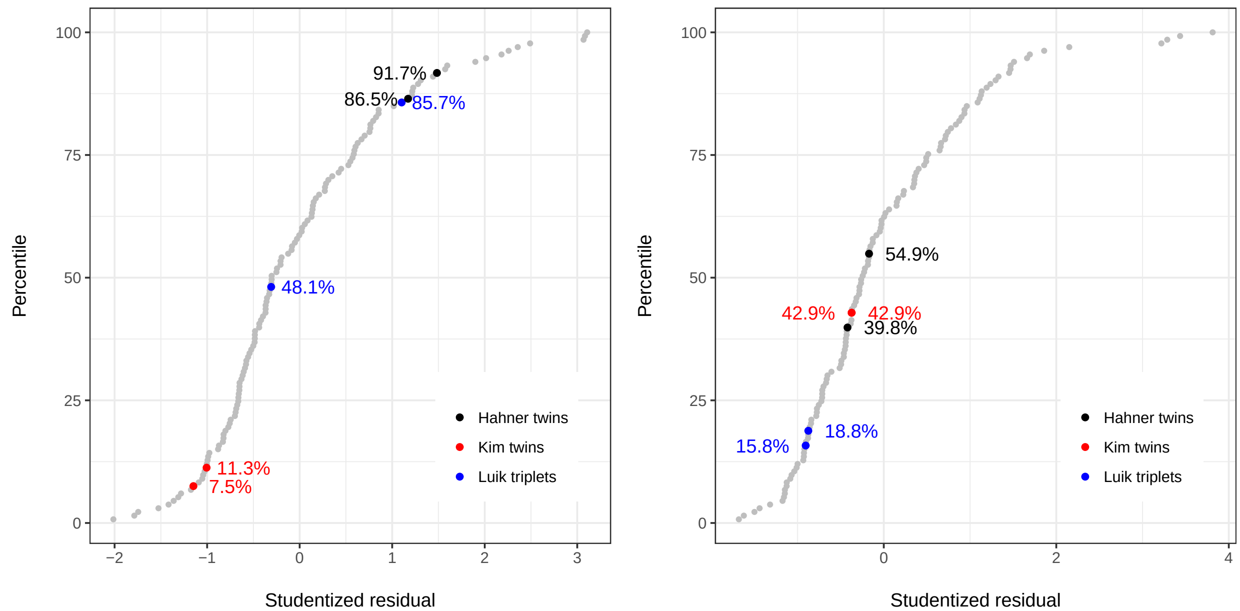
Considering first the relationship between Olympic finishing and personal best times, Figure 1a's solid 45-degree line is informative. Given the paucity of points (only five of them) below this line, it follows that the vast majority of Olympic marathoners ran slower in Rio compared to their prior personal bests. The Hahner twins were definitely on the slow side, well above the 45-degree line, but a number of runners had even greater differences between their Olympic times and their personal bests than the Hahners. These runners are denoted with squares in Figure 1a, and there are 13 such symbols, highlighting approximately 10% of the racers who completed the Rio marathon. Moreover, the figure shows that there were two women who had personal best times slightly faster than the Hahner's and yet finished after the German women. Although Figure 1a 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.

With respect to our second measure of marathon skill, Figure 1b shows that there was nothing abnormal about the Hahner twins' overall finishing times, conditional on their half marathon splits. As one might expect, a runner's time halfway through the course is a fairly strong predictor of her finishing time. Here we see that the relationship between the Hahner twin's half marathon split times and their finishing times is similar to that of other Rio runners. The two points representing the Hahners are in the middle of the distribution of points and therefore imply that the twins do not seem to have slowed down in any unusual way. This is consistent with the previous Figure 1a and inconsistent with accusations made against the Hahner twins, at least the part of the accusation that focused on their overall pace.

Another perspective on the extent to which the Hahners' overall marathon finishing times were not unusual can be gleaned from Figure 2. This figure presents cumulative distributions of the residuals from the regression models displayed in the two panels of Figure 1. In particular, we calculated the residuals from Figures 1a and 1b, Studentized them, and then arranged resulting Studentized residuals from least to most along the horizontal axes in Figures 2a and 2b.

The cumulative residuals depicted in the two panels of Figure 2 are consistent with our interpretation of the Hahner twin's Olympic finishing times as not particularly remarkable. With personal best time as a measure of marathon talent as in Figure 2a, the two Hahner residuals are in

Figure 2: Studentized residuals from Olympic finishing time regressions



(a) Based on personal best

(b) Based on half marathon split

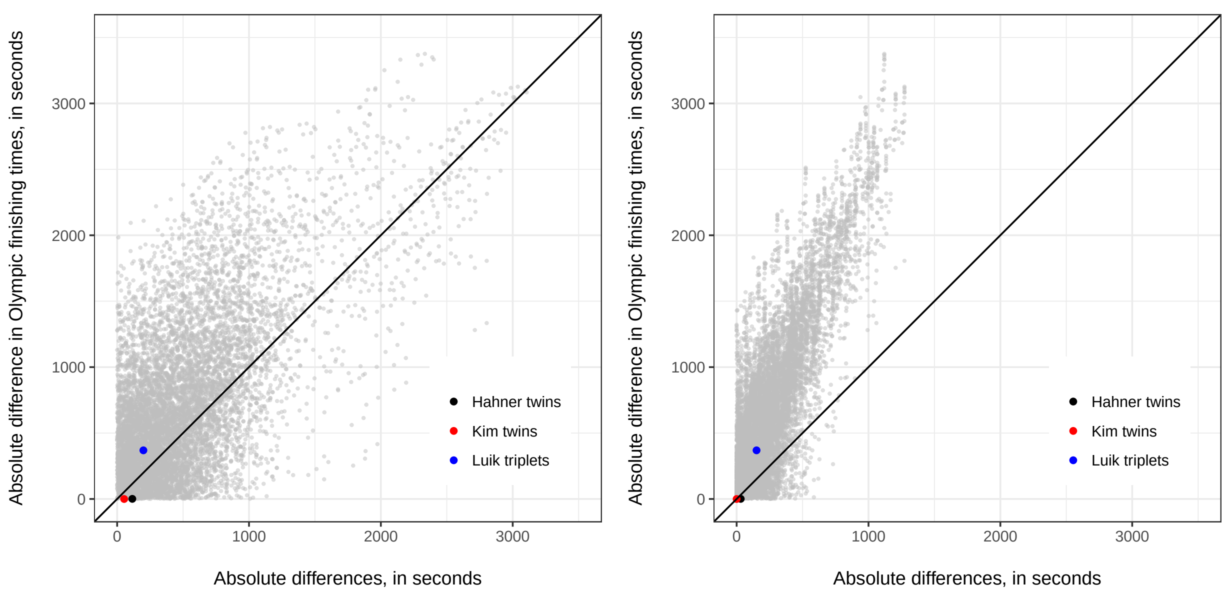
the right tail of the residual distribution; however, their locations are not extreme. One residual is located at the 87th percentile and the other at the 92rd. While these two residuals are in the tail end of the residual distribution, they are not major outliers that would lead us to think that the Hahner twin's marathon finishing times were extremely slow. Moreover, the residuals for the Kim twins are similarly not remarkable; these two residuals are in the left tail of the residual distribution, indicating that the Kims ran faster than one would have expected. Finally, if one relies on half marathon split times as measures of marathon talent, as in Figure 2b, similar conclusions follows. Neither Hahner twin had a finishing time that was particularly unusual given her half marathon split, and this applies to the Kim twins as well.

If the Hahner twins did not slow down excessively, might they have run somewhat strategically at the end of the Olympic marathon 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? Moreover, are there other pairs of Olympic marathoners who had a difference between personal best times of 115 seconds apart and, if so, how close were 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. Differences in finishing times of these ten pairs, 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 3: Pair-wise differences in Olympic finishing times and differences in marathon talent



(a) Based on differences in personal best times

(b) Based on differences in half marathon splits

We can generalize this result by looking at all pairs of runners in the marathon. For all 8,778 pairs of 133 finishers, Figure 3a 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. The two German women are effectively located on the horizontal axis because their difference in finishing times is one second. However, there are many points above the Hahner's black 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 Hahners. Some pairs of runners with around 115 second personal best differences had finishing time differences of 1,000 seconds, i.e., in excess of 15 minutes. The points in Figure 3a are not independent, but they provide a rough sense of the dispersion in finishing time differences between runners that one might expect conditional on differences in personal best times.

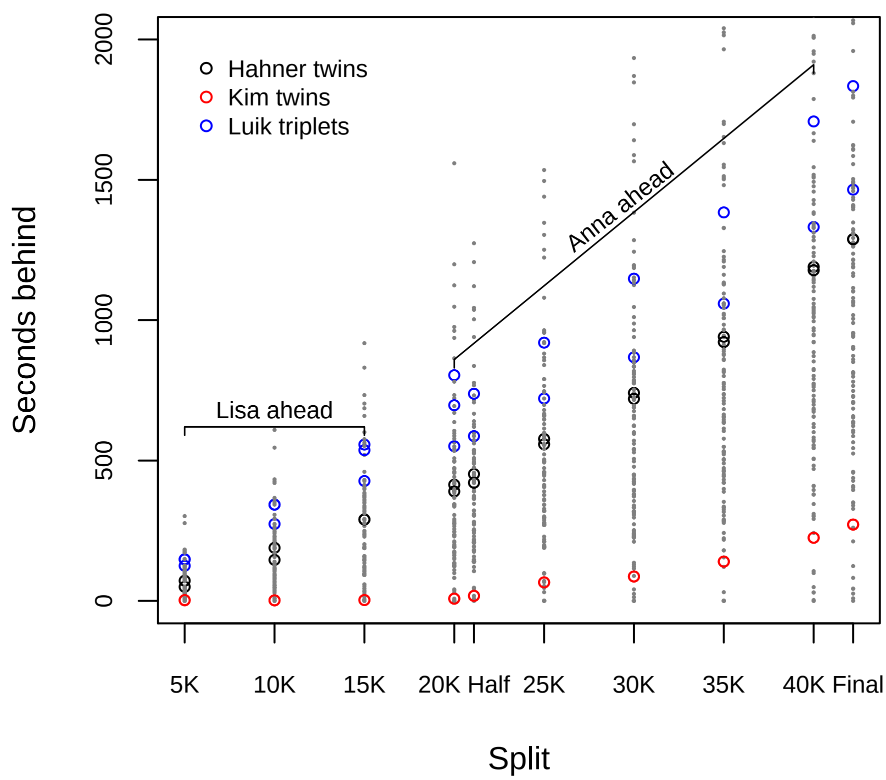
Thinking about the accusations leveled against the Hahner twins, Figure 3a suggests that Anna and Lisa Hahner did indeed run with an eye on each other. In fact, the same can be said of the Kim twins, who ran seemingly in lockstep throughout the entire Rio marathon. The North Korean twins had a personal 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—meaning that they finished with less time between them than the difference in their personal bests—more than any other pair of runners with similar personal best differences.

Similar conclusions follow from Figure 3b, which plots pair-wise differences between Olympic finishing times and half marathon split times. Namely, many pairs of runners had similar differences in half marathon times as the Hahner and Kim twins, but the vast majority of these pairs did not have close finishing times.

Figure 4 describes each Olympic runner's status at various split times on the marathon course. Each dot in the figure—color as before—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 subsequent DNFs.

The Estonian triplet DNF prior to the half marathon split is evident in Figure 4 , which also shows that Lisa Hahner was ahead of her sister through 15 kilometers. The figure contains two red

Figure 4: Runner status by split



dots representing the North Korean Kim twins, but this is not visually apparent because the Kim twins had identical split times during the entire marathon.

Simulating the Marathon

Our visualizations shed a fair bit of light on the Hahner twin's performance in the 2016 women's Olympic marathon. In the interest of increasing precision, we now pose the following question: if we take into consideration the twin's similarities in marathon talent as well as natural variation in marathon finishing times, what is the probability that they would finish the Rio race at roughly the same time and/or sequentially? To answer this question, we need to know the counterfactual distribution of potential marathon finishing times that would have occurred had the Hahner twins independently (in particular, of each other) and repeatedly run the Rio marathon, holding constant marathon conditions, the abilities of other runners, and so forth. Access to such a distribution would establish the set of potential outcomes that could have occurred on August 14, 2016, and we could in principle use this distribution to determine the likelihood that a simultaneous finish by the Hahners, or at least a near-simultaneous finish, occurred by chance alone. If these twins rarely finish the marathon together in such a counterfactual world, then one might be skeptical that their observed finish occurred without some degree of coordination.

Unfortunately for us but nonetheless fortunately for the race's participants, it is not possible to rerun the women's Olympic marathon to establish a distribution of potential race outcomes for the Hahner twins. However, we can attempt to simulate this distribution by estimating the distribution of every other runner's finishing time, conditional on marathon talent, and then drawing from this distribution in order to calculate the likelihood that, for example, Lisa and Anna Hahner finished the Rio race simultaneously.

To estimate the conditional distribution of each Rio final result, we assume that each runner's marathon time Y_i is distributed normally with a mean that is a linear function of the runner's marathon skill X_i and her age Z_i. Hence:

$$Y_i \mid X_i \sim N\left(\beta_0 + \beta_1 X_i + \beta_2 Z_i + e_i, \sigma\right),$$

where $N(\cdot, \cdot)$ denotes a normal distribution. Taking each runner's personal best marathon time as a proxy for her underlying marathon skill, we estimate β_0, β_1, β_2, and σ using ordinary least squares on finishing Rio marathoners. We exclude the Hahner/Kim twins and Luik triplets from the sample so that that our estimates are not affected by the twin/triplet finishes which, theoretically, could reflect runner coordination. For a simulated marathon, we draw a runner's time from an estimated distribution and condition on the runner's personal best marathon time and age. Once a race is simulated for all runners, twins and triplets included, we record both the time between Anna and Lisa Hahner's simulated finishes and the difference in their simulated ranks. We then simulate a new race—drawing a new set of finishing times—and record the same quantities. The steps of the simulation are as follows.

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 age Z_i.

2. Extract the resulting coefficient vector, estimated covariance matrix, and estimated regression variance from this model.

3. For each simulated rate, draw intercept and slope estimates $\tilde{\beta}_0$, $\tilde{\beta}_1$, and $\tilde{\beta}_2$, respectively, from a multivariate normal distribution with mean equal to the previously estimated coefficient vector and covariance equal to the previously estimated covariance matrix.

4. For each runner, draw an error $\tilde{e}$ from a normal distribution with mean zero and a standard deviation equal to the standard deviation of the original regression model's residuals.

5. Predict each runner's final result by combining the randomly generated beta coefficients and individual error terms, $\tilde{y} = \tilde{\beta}_0 + \tilde{\beta}_1 X_i + \tilde{\beta}_2 Z_i + \tilde{e}$.

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

7. Repeat above steps 10,000 times.

As these steps illustrate, our simulation repeatedly draws random coefficient vectors, and this captures variability in what we know about the relationship between runner talent and age and runner finish times. In addition, the simulation draws random disturbances for each runner, conditional on our original estimate of regression variance; these disturbances capture variability in runner finishing times, notwithstanding underlying marathon talent as proxied for by prior best finishing time. Importantly, the disturbances that we draw are independent across runners. Consequently, for each simulated race the finishing order among runners will vary.

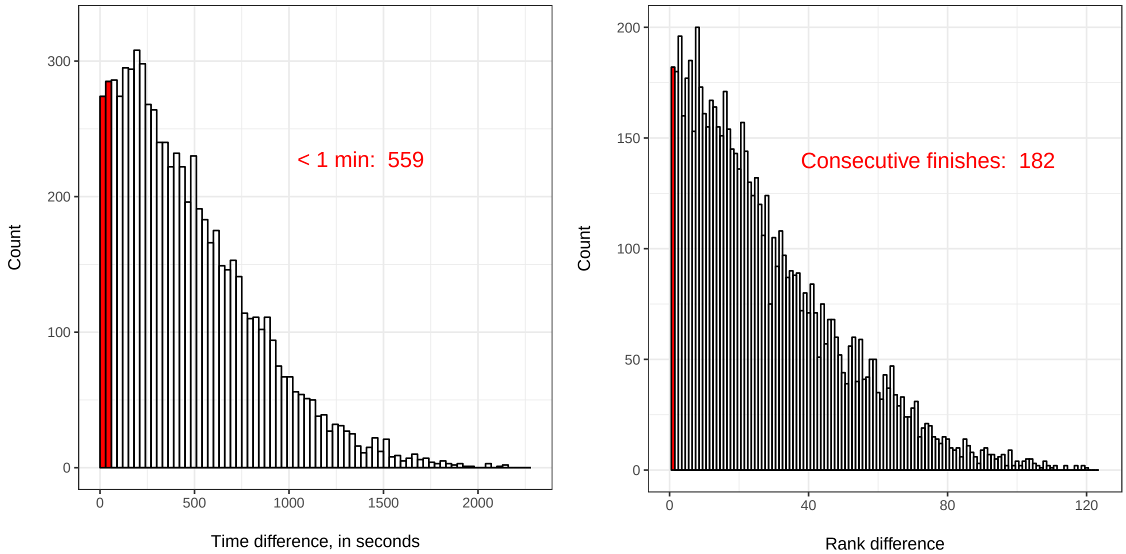
From our simulations we generate intervals which describe the extent of the variability in marathon finishing times. For example, in 95% of the simulations in which Anna Hahner completed the marathon, her finishing time was between 8,369 seconds and 9,999 seconds. This is consistent with Anna's observed finishing time in Rio, which was 9,932 seconds. In fact, our simulations estimate that Anna should have finished the Rio marathon in 9,181 seconds, which is slightly more than 12 minutes faster than her actual time. Lisa Hahner's corresponding 95% interval is 8,507 to 10,143 with an observed finishing time of 9,933 seconds.

The bottom line here is that our simulated marathon finishes are consistent with the Hahner twins' finishing times. And note that the regression model underlying the simulations was estimated without the Hahner twins (and same for the Kim twins and Luik triplets). According to our simulation, then, both Hahner twins did not run appreciably slowly, conditional on personal best times prior to the Rio Olympics and age.

With an eye on the matter of simultaneous finishing, Figure 5 contains two histograms based on simulated race results. Figure 5a is a histogram which shows the distribution of absolute differences in Hahner twin finishing time where differences are grouped in 30 second bins; counts for the various bins are denoted by the vertical lengths of the bars. Figure 5b is similar but depicts the distribution of the absolute differences in Hahner twin rankings. Differences are grouped as single units ranging from no runner between the twins to nearly 120 runners between them.

The histograms in Figure 5 raise questions about the credibility of Anna and Lisa Hahner's story and in particular suggest that a simultaneous finish in the Rio marathon would be very rare if Anna and Lisa had run independently. For example, in fewer than 300 of 10,000 simulated races did Anna and Lisa Hahner finish within 30 seconds of one another, and in fewer than 600 did the Hahner twins finish within a minute of each other. The histogram area associated with this result

Figure 5: Distribution of Hahner twin results in 10,000 simulated marathons, based on personal best times



(a) Differences in Hahner finishing times

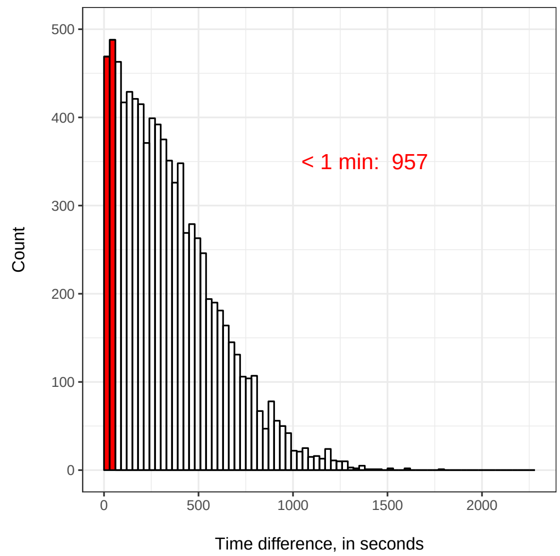
(b) Differences in Hahner twin places

is depicted in red in Figure 5a. Moreover, the Hahner twins finished in consecutive rank in fewer than 200 of 10,000 simulated races; the red zone in Figure 5b presents this visually. The close finish that was observed in Rio, where Anna and Lisa crossed the finish line one after the other, would have been highly unlikely if the two German twins had raced independently of each other.

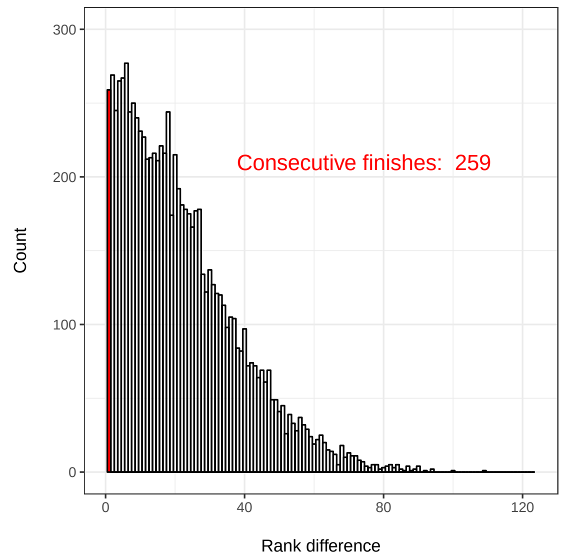
Parallel with our prior analyses, we repeated our simulations using half marathon splits as the predicting variable in our simulations; results are in Figure 6. While this use of marathon splits reduces the variation in the predicted outcome of each runner and therefore reduces the expected distance between Anna and Lisa Hahner, even with half marathon split as a measure of ability it is still quite rare for the twins to finish simultaneously or consecutively.

The way that we handled DNFs in our simulations is notable. As our description above indicates, we assumed that DNF probabilities are the same for all runners and that the likelihood of a DNF is not a function of a runner's anticipated marathon finishing time. The rug marks in Figure 1a suggest that runners with better personal best times may be more likely to DNF than other runners, all things equal. We suspect that this occurs because some better runners may ex-

Figure 6: Distribution of Hahner twin results in 10,000 simulated marathons, based on half marathon splits



(a) Simulated differences in finishing times



(b) Simulated differences in places

pend excessive energy trying to achieve a good result in the marathon and in so doing injure or exhaust themselves; lesser runners, in contrast, may be content to finish respectably. Regardless of the validity of this conjecture, Figure 1a shows that the Hahner twins are representative of the sort of runners who DNFed in Rio. Since our simulated Hahner statistics are conditioned on both Hahner twins finishing, it follows that they are conservative. The fact that both women finished the marathon was notable in and of itself, and by discounting the possibility of a Hahner DNF we are giving the benefit of the doubt to the Hahner twins.

Conclusion

We have studied Anna and Lisa Hahner's near-simultaneous finish in the 2016 women's Olympic marathon in Rio. This finish elicited a controversy insofar as the German twins were accused of deliberately slowing down and finishing next to each other so as to generate media attention. They denied this, not entirely surprisingly, We have offered two perspectives on the marathon, one based on visualizations and simple calculations and a second that draws on a simulation. Both perspectives have the same implications, and they are as follows. In a global sense, the Hahner twins did not slow down appreciably during the Rio marathon. Their times were not fast, but they were within reason for runners of the Hahners' abilities. Locally, though, we find that the Hahner's finish was in fact contrived. Their finish—one second between the two women—was an extremely low probability event. Compared to their differences in talent, the Hahner twins difference in finishing times was unusually compressed. This is evidence that their finishing had elements of intentionality—perhaps at the last minute but intentionality nonetheless.

Our goal is not to speak to the question of whether the Hahner twins should or should not have enjoyed a somewhat contrived moment of Olympic glory. Neither was in contention for a podium finish in Rio, and compared to the doping allegations that presently surround endurance sports in general, what the Hahners did seems relatively tame. Still, it might have behooved them to be a bit more open about their end-of-race tactics, and to this end we have shown here how a simple data analysis can shed light on claims about racing results. Should the 2020 Summer Olympics feature twins and triplets again, we look forward to a comparative analysis using the techniques illustrated here.

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