Peer Effect in Sports: A Case Study in Speed Skating and Swimming

Tue Nguyen

Advisor: Takao Kato

Abstract:

Peer effect is studied for its implication in the workplace, schools and sports. Sports provide nice revenues to study peer effect because results are completely observable. In this paper, I utilize similar methods to Yamane and Hayashi (2010) to examine international speed skaters and US swimmers. Using fixed effects estimations, I found positive and significant peer effect among speed skaters and swimmers. In speed skating there was no difference in magnitude of peer effects between male and female athletes and between losers and winners in their pairs. In swimming, the peer effect is found to be compositional instead of contemporaneous.

I. Introduction

Understanding peer effects and its consequences is important in efficiently directing people towards better results in sports, schools and work places. An individual's peers play an important part in molding that individual's behavior and motivations, contributing to the experience as well as offering a source of comparison and competition. It is natural for humans to observe those interacting with them and the social environment. Guryan et al. (2009) suggested two pathways for peer effect to influence us in the work place. The first is the "motivation effect", in which workers are motivated when they observe that their coworkers are doing well and working hard. The second is the "learning effect", in which workers learn from their coworkers how to perform a task in the best way. We learn from others and compare our own actions and efforts against them (Manski, 1993). This can also lead to the reflection problem, in which the peers who influence us are also affected by us (Manski, 1993). This problem could be overcome in estimations in sports using Yamane and Hayashi (2010) method of estimating with the peers' best record as a proxy for their ability. Peer effect estimations pick up the influence of our peers as a result of our comparative observations and the awareness of peer judgments. Knowing the extent of peer effects and its influence in various disciplines allows employers, coaches and teachers to effectively motivate the individual to maximize his or her contribution in a time and cost-saving manner.

There exist both negative and positive peer effects, which were extensively studied. Mas and Moretti (2009) find that supermarket cashiers are more productive, measured by number of item scans per second, when they are directly in line of view with more productive coworkers. In a study of randomly assigned roommates at Dartmouth College, Sacerdote (2000) concludes that a 1.0 increase in the GPA of a student's roommate is associated with a 0.11 increase in that individual's own GPA. In sports, Depken and Haglund (2007) find that peer effects have negative influence on NCAA relay teams as team members shirked, knowing that their teammates would pick up the slack. Guryan et al (2009) surprisingly find a lack of peer effects among golfers when they played in a playing group, which contradicts the famous Tiger Woods effect. Yamane and Hayashi (2010) find large peer effects among swimmers in Japan in competitions as swimmers can swim faster with fast, high-ability peers. They found that swimmers are aware of their peer who has a lower best record than theirs and that being chased improves swimmers' performance.

The purpose of this paper is to examine peer effects using data on international speed skaters in 500 m races and on swimmers in the United States. Speed skating data has many advantages for testing peer effects. The first advantage is that in speed skating, there are clearly defined lanes and athletes are drawn into pairs for each competition so it is easy to directly observe peer effect. The peer assignment for the first round in a competition is pseudo-random for the first competitions and depends on ranking for the following competitions. The peer assignment for the second round depends on the athletes' performance in the first round, which can vary because athletes can have different forms in different rounds and competition. This helps to eliminate the endogeneity of peer assignment and allows for the use of fixed effects regression based on the variation of opponents between rounds. Another advantage is that speed skating data is free from common shocks that can interfere with estimations of peer effects such as weather conditions in Guryan et al. (2009). The swimming data has advantages similar to Yamane and Hayashi (2010) data. The first advantage is that there are clearly defined rules for assigning peers in swimming, which takes away the endogeneity of peer assignment and allows for fixed effect estimation.

Similar to speed skating, swimming also has clearly defined lanes and peers so it is easy to observe peer effect.

Similar to Yamane and Hayashi (2010), my definition of peer effect is limited to the effect of being motivated by the other skater because the very short time in skating (less than 40 seconds) is too brief for learning to occur. This also applies to swimming. My speed skating dataset comprises of international skaters competing in international events so I investigate the best skaters and contribute to the literature by showing the existence of peer effect. My swimming dataset comprises of some of the best swimmers in America and can serve to compare with the Yamane and Hayashi (2010) dataset. Another contribution is the examination of the difference in terms of magnitude of peer effects for male and female athletes. I also investigate whether the "chased effect" mentioned in Yamane and Hayashi (2010) is present in speed skating by looking at the difference in magnitude of peer effects among direct winners and losers of the pairs and among expected winners and losers of the pairs.

There are monetary rewards for speed skating and incentives for speed skaters to try their best at every race. Skaters do not know their ranking and scores until everyone finishes racing so their optimal strategy is to obtain the best time possible. The scores of the two rounds for each competition are added together to give the final scores for the skaters. Results for the ISU World Cup competitions are tallied in a league format. Hence, every race matters. Good racing records mean higher ranking and better match ups, as well as fame and more selections to represent the country for the skaters. As such, speed skaters are highly motivated to perform as well as possible. In the case of swimming, I chose to work with only data from qualifying heats. As such, the swimmers have the incentive to swim as fast as possible in order to have a chance to compete in the finals. Swimmers are also motivated to swim as fast as possible regardless of their peers. The different strokes in swimming allow for determination of whether peer effect is compositional or contemporaneous in swimming, which will be exploited in this paper.

The rest of this paper is organized as follows. Section II describes the rules for drawing of skaters and swimmers. Section III discusses the dataset and the estimation models. Section IV shows the results and Section V concludes this paper.

II. Rules for drawing of skaters and swimmers:

i) Rules for drawing of skaters:

The International Skating Union (ISU) dictates the rules for speed skating competitions. There are two rounds in each competition. The rules for drawing of skaters in the first round are as follows:

1. For World Cup events:

a. Drawing at the first World Cup Competition:

Competitors are divided into groups of 4 based on their World Cup ranking in the previous season. There are 5 groups in Division A with descending ranking from group 1 to group 5 i.e. the skaters with the best rankings are in group 1. Pairs are drawn within each group. The pairs of Group 5 shall be drawn first, then the pairs of Group 4, etc. The two last drawn competitors from Group 1 will form the last pair of Division A.

b. Drawing after the first World Cup Competition:

Competitors, both in Division A and in Division B, shall be paired according to their individual World Cup rank. The starting order of the pairs shall be opposite to the ranking order of the competitors – so that the competitors with the highest ranks will start in the last pair, and so on. The better ranked competitor in a pair shall start in the inner lane.

2. For World Single Distances Championships and World Sprint Championships:

The drawing of pairs will be based on a ranking list of all participants. Competitors qualified by World Cup ranking shall be ranked first. The remaining participants shall be ranked according to their position in the Time ranking list, which is based on best times for that distance. When a complete ranking list of the participants has been established, the Competitors shall be placed in groups of 4 Competitors each. The group with the 4 best ranked Competitors shall be drawn to form the two last pairs, and so on.

The rules for pairing of skaters in the second round are as follows: The composition of pairs shall be based on the time ranking from the first race, so that the competitors with the best time from the first race will start in the last pair. In addition, the composition of pairs must take into account that the competitors shall change their starting lanes from the first race.

ii) Rules for swimmers:

The swimmers are assigned their lanes based on international rules. In my dataset, I only looked at competitions in 8-lane pools because those are the most common. The swimmers qualify for these events by submitting their best record for that event. The lanes of the swimmers according to their best records are shown in Figure 1.

Lane number	8	7	6	5	4	3	2	1
Order of best record	8th	6th	4th	2nd	1st	3rd	5th	7th

Figure 1: Lane assignment according to best records

III. Data and Model

1. Data Detail

Data for speed skaters was collected from www.isuresults.eu and focuses on international 500 m speed skating competitions organized by ISU over two seasons, 2011-2012 and 2012-2013. These competitions include ISU World Cup for Division A, ISU World Cup Final for Division A, ISU World Sprint Speed Skating Championships and World Single Distances Speed Skating Championships. The competitions have the best 500 m speed skaters in the world because the qualifying requirements for these competitions depend on world ranking and time ranking, which are held to very high standards. The countries participating in these events will only send their best athletes who qualify for the events. This allows us to examine peer effect in a highly talented pool, which is uncommon among previous studies. Pairs of skaters were removed from my data if one of them was disqualified from the race or if their finishing time was more than one minute as the usual skate time for these high-caliber skaters are less than 40 seconds. These longer race times may be due to injuries or false starts, which happen randomly and very rarely so their removal does not bias my regressions. Skaters who skated alone were also removed from dataset due to their lack of peers.

Data for swimmers were collected from www.omegatiming.com and www.usaswimming.org and focuses on 91 swimmers in different competitions over 3 seasons, 2010-2011, 2011-2012 and 2012-2013. These competitions include ConocoPhillips National Championships, U.S. Olympic Team Trials, Mutual of Omaha Swimvitational, Speedo Junior National Swimming Championships, Phillips National Championships & World Championships Trials and U.S. Open Swimming Championships. The athletes have to have sufficiently good qualifying time in order to qualify for these events. I only looked at swimmers with 2 peers adjacent to him or her in order to make sure that there are two peers for each of my observation.

All competitions take place indoor on two-lane ice tracks and the skaters are drawn into pairs following the aforementioned rules. During the races, the speed skaters have to transition to and finish in the lane they did not start in. As such, the skaters who start in the inner lane will finish in the outer lane and vice versa. This creates opportunities for a skater to observe his or her competitor throughout the race. The skaters have the incentives to race as fast as possible because the final rankings depend on their skate times and not on whether they beat their opponents or not. As such, the optimal strategy is to perform his or her best regardless of the opponent's performance.

As for swimmers, some strokes like freestyle allows swimmers to observe their peers while others do not. This creates an interesting case study to determine if the peer effects in swimmers are compositional or contemporaneous. Compositional peer effect happens before the race and during the race because the swimmers know each other's best records and can be motivated by that. Contemporaneous peer effect happens during the swim race when swimmers can observe each other and be motivated by their peers' performances. Strokes in which it is harder for swimmers to observe their peers such as backstroke and breaststroke can be used to determine the validity of my method of using peer best records as a proxy for peers' ability and to determine if peer effect is compositional or contemporaneous among swimmers. This is because the peer effects in these strokes highlight the compositional aspect of peer effect. The swimmers have the incentives to swim as fast as possible because they are competing in preliminary heats in order to qualify for the finals. Their qualification for the finals depends on their swim times and not on whether they beat their peers. Hence, their optimal strategy is to swim their best regardless of their peers' performance.

2. Estimation Models and Data Setting

For speed skating, I use panel data setting to follow 109 speed skaters through the different competitions. For the first round, peer assignment is not random and follows the mentioned rules based on prior rankings and best records. For the second round, peer assignment depends on performance in the first round of all the skaters, which can vary. There was no systematic difference between the first round and the second round records of a skater in the same event. Hence, endogeneity in peer assignment is not an issue when I incorporate peers' best records in my model. The panel identifier consists of athlete ID which is assigned to each skater in the dataset. The first estimation model is as follows:

$$R_{it} = \mu_{it} + \beta_1 B_{it} + \beta_2 B_{jt} + \beta_3 X_{it} + \beta_4 X_{jt} + \beta_5 Y_t + \varepsilon_{it} \qquad (1)$$

R_{it} denotes the skate time of skater i in competition t. B_{it} is skater i's best record before competition t. In the model, subscript j implies i's peer so B_{jt} denotes skater j's best record before competition t; hence, β_2 captures the peer effect. Model 1 deals with the reflection problem (Manski, 1993) because peer's best record is used to capture peer effect instead of the peer's skate time. All skaters are aware of the best times of their opponent before they race and infer the ability of their peers from their best time. X_{it} and X_{jt} are i and j's age variables with age and age squared, which are proxies for their experience. Y_t comprises of control variables such as lane, type of events (World Cup, World Cup Final, World Single Distances Championships or World Sprint Championships), location of the events and season. These parameters were estimated with fixed effects on i. If the skater follows the optimal strategy, which is to perform his or her best regardless of the competitor's performance, the coefficient of B_{jt} is 0. As such, if the athlete skates faster when matched up against a good skater, there is positive peer effect. Alternatively, if the athlete skates slower against a good skater, there is a negative peer effect. The dataset contains 109 speed skaters and 1382 observations after exclusion of the pairs and skaters who do not fit the dataset.

Secondly, I separate the male and female athletes into two different dataset with 694 observations in the male dataset and 684 observations in the female dataset. Model 1 is used to estimate the peer effect in the two datasets. In order to determine if there is significant difference in the peer effect for male and female athletes, the following estimation model is used on the general dataset containing both male and female athletes:

$$R_{it} = \mu_{it} + \beta_1 B_{it} + \beta_2 B_{jt} + \beta_3 X_{it} + \beta_4 X_{jt} + \beta_5 Y_t + \beta_6 B_{it}*S_i + \beta_7 B_{jt}*S_i + \beta_8 X_{it}*S_i + \beta_9 X_{jt}*S_i + \beta_{10} Y_t*S_i + \varepsilon_{it} \quad (2)$$

S_i denotes the sex of the athletes and takes the value of 1 if the skater is female and 0 if the skater is male. The relevant controls are interacted with S_i as shown in model 2. $B_{jt}*S_i$ is the interaction term for the best time of the opponent and their sex so β_7 captures the difference in peer effect between males and females. Peer effect is stronger among females if β_7 is positive and weaker if β_7 is negative. If there is no difference in peer effect between male and female athletes, β_7 is equal to 0. These parameters were estimated with fixed effects on i.

Thirdly, I separate the skaters who beat their direct opponent from the skaters who lost to their direct opponent to observe the difference in magnitude of peer effect among winners and losers of their pairs. The winner dataset contains 677 observations while the loser dataset contains 677 observations too. Ties were excluded from these datasets. After running regression using model 1 on these 2 datasets, I combine them and use the following model to estimate if there is any difference in terms of magnitude of peer effect between winners and losers in the races.

$R_{it}=\mu_{it} + \beta_1 B_{it} + \beta_2 B_{jt} + \beta_3 X_{it} + \beta_4 X_{jt} + \beta_5 Y_t + \beta_6 B_{it}*W_i + \beta_7 B_{jt}*W_i + \beta_8 X_{it}*W_i + \beta_9 X_{jt}*W_i + \beta_{10} Y_t*W_i + \varepsilon_{it}$ (3)

W_i denotes the direct result of the athletes and takes the value of 1 if the skater won and 0 if the skater lost. The relevant controls are interacted with W_i as shown in model 3. $B_{jt}*W_i$ is the interaction term for the best time of the opponent and the skater's direct race result so β_7 captures the difference in peer effect between winners and losers.

Similarly, I separate the skaters who are expected to beat their direct opponent from the skaters who is expected to lose to the direct opponent based on their best records to observe the difference in magnitude of peer effect among expected winners and losers of their pairs. The expected winner dataset contains 677 observations while the expected loser dataset contains 677 observations. Ties in best record were excluded from these datasets. Model 1 is used to run regressions on these datasets. The datasets are then combined; the following model is used to estimate if there is any difference in terms of magnitude of peer effect between expected winners and losers in the races.

$R_{it}=\mu_{it} + \beta_1 B_{it} + \beta_2 B_{jt} + \beta_3 X_{it} + \beta_4 X_{jt} + \beta_5 Y_t + \beta_6 B_{it}*E_i + \beta_7 B_{jt}*E_i + \beta_8 X_{it}*E_i + \beta_9 X_{jt}*E_i + \beta_{10} Y_t*E_i + \varepsilon_{it}$ (4)

E_i denotes the expected result of the athletes and takes the value of 1 if the skater was expected to win and 0 if the skater was expected to lose. The relevant controls are interacted with E_i as shown in model 4. $B_{jt}*E_i$ is the interaction term for the best time of the opponent and the skater's expected race result so β_7 captures the difference in peer effect between winners and losers.

For swimming, I use panel data setting to follow 91 swimmers through the different competitions. Peer assignment follows the rules as aforementioned. The swimmers have different peers for different meets so the different peers are the sources of variation for my estimation of peer effect. The peer best records are highly correlated to each other so I averaged their best records in order to observe peer effect. The panel identifier consists of athlete ID which is assigned to each swimmer in the dataset. The fifth estimation model is as follows:

$$R_{it}= \mu_{it} + \beta_1 B_{it} + \beta_2 B_{at} + \beta_3 X_{it} + \beta_4 X_{lt} + \beta_5 X_{rt} + \beta_6 Y_t + \varepsilon_{it} \qquad (5)$$

R_{it} denotes the swim time of swimmer i in competition t. B_{it} is swimmer i's best record before competition t. In the model, subscript a implies the average of i's peer so B_{at} denotes the average best record of the peers before competition t; hence, β_2 captures the peer effect. Subscript implies the peer on the left while subscript r implies the peer on the right. All swimmers are aware of the best times of their opponent before they race and infer the ability of their peers from their best time. X_{it}, X_{rt} and X_{lt} are i and j's age variables with age and age squared, which are proxies for their experience. Y_t comprises of control variables such as type of swimming events and season. These parameters were estimated with fixed effects on i. If the swimmer follows the optimal strategy, which is to perform his or her best regardless of the competitor's performance, the coefficient of B_{at} is 0. As such, if the athlete swim faster when matched up against two good swimmers, there is positive peer effect. Alternatively, if the athlete swims slower against two good swimmers, there is a negative peer effect. The dataset contains 91 swimmers and 439 observations.

I also use either the left lane peers or the right lane peers' best records to estimate the peer effect by individual swimmers. The sixth and seventh estimation models are as follows:

$$R_{it} = \mu_{it} + \beta_1 B_{it} + \beta_2 B_{jt} + \beta_3 X_{it} + \beta_4 X_{lt} + \beta_5 X_{rt} + \beta_6 Y_t + \varepsilon_{it} \qquad (6)$$

β_2 in the model captures the peer effect by the adjacent lanes. B_{jt} represents the best record of the peer in either left lane or the right lane.

In order to determine if peer effect is compositional or contemporaneous among swimmers, I estimate peer effects using model 5 and 6 in only breaststroke and backstroke events. There are 190 observations and 61 swimmers in my breaststroke and backstroke dataset.

IV. Results

1. Descriptive statistics

The descriptive statistics of key variables are shown in Table 1. Panels A and B show the descriptive statistics of the general skater dataset and dataset for male and female skaters respectively. The unit for skate time and best record is seconds. A smaller value of skate time or best record indicates a better performance. In the general dataset, the average skate time and best record are approximately 36.9 seconds and 36.2 seconds respectively. Male athletes on average have better skate times and best records than female athletes, with the mean of 35.2 seconds compared to a mean of 38.5 for female athletes. The winners have better skate times and best records than the losers. Ties were excluded from this dataset.

Panel E shows the descriptive statistics of the swimmer dataset. The unit for swim time and best records is seconds. A smaller value of swim time or best record indicates a better performance. Best records of the peers are also shown. Panel F shows the descriptive statistics of the breaststroke and backstroke dataset.

Table 1: Descriptive statistics

Panel A: General Dataset for speed skaters.

Variable	N	Mean	SD	Min	Max
skate time	1378	36.88672	1.721546	34.21	40.78
best record	1378	36.18475	1.642899	34.2	40.35
Age	1378	26.0096	3.458413	17.10062	37.31691

Panel B: Dataset for male and female skaters

Gender	Variable	N	Mean	SD	Min	Max
Male	skate time	694	35.25729	0.4235804	34.21	37.53
	best record	694	34.62553	0.4185382	34.2	37.49
	age	694	26.4794	3.389497	18.77344	37.31691
Female	skate time	684	38.53997	0.5984034	35.53	40.78
	best record	684	37.76677	0.5359178	36.8	40.35
	age	684	25.53293	3.464848	17.10062	35.23066

Panel C: Dataset for winners and losers in speed skating

Race result	Variable	N	Mean	SD	Min	Max
Winner	skate time	677	36.73195	1.687909	34.21	39.7
	best record	677	36.12041	1.611816	34.2	39.54
	age	677	26.22483	3.518432	17.10335	37.31691
Loser	skate time	677	37.04371	1.741729	34.42	40.78
	best record	677	36.25232	1.676055	34.2	40.35
	age	677	25.81195	3.410508	17.10062	37.31417
Combined	skate time	1354	36.88783	1.721473	34.21	40.78
	best record	1354	36.18637	1.644965	34.2	40.35
	age	1354	26.01839	3.469761	17.10062	37.31691

Panel D: Dataset for expected winners and losers in speed skating

Expected result	Variable	N	Mean	SD	Min	Max
Winner	skate time	678	36.84181	1.697169	34.21	40.21
	best record	678	36.03478	1.604197	34.2	39.47
	age	678	26.63439	3.473191	17.27858	37.31691
Loser	skate time	678	36.95912	1.748003	34.32	40.78
	best record	678	36.36119	1.663779	34.3	40.35
	age	678	25.36274	3.347085	17.10062	36.16427
Combined	skate time	1356	36.90046	1.723137	34.21	40.78
	best record	1356	36.19799	1.641795	34.2	40.35
	age	1356	25.99856	3.468286	17.10062	37.31691

Panel E: Dataset for swimmers

Variable	N	Mean	SD	Min	Max
swim time	439	110.779	70.42612	22.43	310.16
best record	439	108.242	69.30835	20.12	303.85
best record (left peer)	439	108.451	69.13716	22.04	307.04
best record (right peer)	439	108.776	69.80717	22.21	301.49
age	439	19.4021	2.945872	10.88569	27.70979

Panel F: Dataset for swimmers in breaststroke and backstroke events

Variable	N	Mean	SD	Min	Max
swim time	190	95.88237	38.36566	53.83	165.26
best record	190	93.12311	36.99332	47.22	160.76
best record (left peer)	190	93.72447	37.04174	47.22	160.71
best record (right peer)	190	93.74979	37.53244	47.47	161.02
age	190	19.76534	2.898443	11.87953	27.70979

2. Peer Effect Regressions

The regression results are shown in table 2. Column 1 on table 2 shows the estimated results of the regression using model 1 on the general dataset. In column 1, the coefficient of the opponent’s best records is highly significant and the sign is positive. Hence, the better the opponent is, the faster the skater skates. When the opponent’s record is better by 1 second the skater’s skating time decreases by 0.10 seconds. This represents the positive peer effect exerted on speed skaters by their opponents. This can come from the fact that speed skaters are more motivated when they race against good skaters because speed skaters have many opportunities to observe each other during a race, regardless of whether they win the race or not. The skaters also know each other's records before going into a race so the motivation can occur before the race as well. If the opponent's record is better by the standard deviations of male and female athletes, which are approximately 0.42 and 0.54 respectively, the skater skating time still decreases by 0.04 to 0.05 seconds, which can make a difference in their ranking.

Column 2 and 3 show the estimated results of the regression using model 1 on the male dataset and female dataset respectively. The coefficient of the opponent's best records is still significant and positive. Column 4 shows the estimated results of the regression using model 2 on the general dataset to determine if there is any difference in terms of peer effect between male and female athletes. The interaction term for opponent's best record and gender is not significant even at lower significance level. This shows that there is no difference between male and female skaters in terms of the magnitude of peer effect.

Table 2: Regression results for model 1 and model 2

The dependent variable is the skater's skate time, with individual fixed effects and relevant controls.

	(1) Model 1 on general dataset Coef.	(2) Model 1 on male dataset Coef.	(3) Model 1 on female dataset Coef.	(4) Model 2 on general dataset Coef.
Skater's best record	0.3788449***	-0.0780327	-0.0437119	-0.0780322
	(.0403805)	(.0760168)	(.068912)	(.0908134)
Opponent's best record	0.1048273***	0.0738447**	0.0833882**	0.0738445**
	(.0247124)	(.0315412)	(.0331324)	(.0376807)
Inside lane dummy	0.0281199*	0.0149477	0.038476*	0.0149477
	(.0156345)	(.0169867)	(.0232452)	(.0202932)
Gender and best record interaction				0.0095439
				(.0475923)
Constant	13.69144***	37.55038***	45.83732***	37.68615***
	(1.128264)	(5.579843)	(4.207042)	(6.512522)
Obs	1378	694	684	1378
R squared	0.6217	0.6347	0.5523	0.6851

Note: *** 1%, ** 5%, * 10% significance.

Table 3 shows the regression results of model 1 on winner and loser datasets and of model 3 on the combined dataset on column 1, 2 and 3, respectively. In column 1 and 2, the coefficient of the opponent's best records is still highly significant and the sign is positive. Thus, peer effect exists for both winners and losers. Column 3 shows the estimated results of the regression using model 3 on the combined dataset to determine if there is any difference in terms of peer effect between direct winners and losers. The interaction term for opponent's best record and the skater's direct results is not significant even at lower significance level. This shows that there is no difference between direct winners and losers in terms of the magnitude of peer effect.

Table 3: Regression results for model 1 and model 3

The dependent variable is the skater's skate time, with individual fixed effects and relevant controls.

	(1) Model 1 on winner dataset Coef.	(2) Model 1 on loser dataset Coef.	(3) Model 3 on combined dataset Coef.
Skater's best record	0.3047237*** (.0523868)	0.045657 (.0665067)	0.3178783*** (.0443398)
Opponent's best record	0.2224741*** (.0334171)	0.1686525*** (.0371333)	0.1942335*** (.0325683)
Inside lane dummy	0.0482649** (.02034)	-0.0124285 (.0218311)	-0.0163859 (.0211223)
Direct result and best record interaction			0.0180076 (.0409145)
Constant	17.29456*** (2.676247)	9.748004*** (2.815726)	13.29384*** (1.99776)
Obs	677	677	1354
R squared	0.6884	0.6927	0.6902

Note: *** 1%, ** 5%, * 10% significance.

Table 4 show the regression results of model 1 on the expected winner and loser datasets and of model 4 on the combined dataset on column 1, 2 and 3, respectively. In column 1 and 2 the coefficient of the opponent's best records is significant and the sign is positive. Hence, peer effect exists for both expected winners and losers. The interaction term for opponent's best record and the skater's expected results is not significant even at lower significance level. Thus, there is no difference between expected winners and losers in the magnitude of peer effect

Table 4: Regression results for model 1 and model 4

The dependent variable is the skater's skate time, with individual fixed effects and relevant controls.

	(1) Model 1 on expected winner dataset Coef.	(2) Model 1 on expected loser dataset Coef.	(3) Model 3 on combined dataset Coef.
Skater's best record	0.4842388*** (.0622038)	0.1159386 (.0814923)	0.3246214 (.059963)
Opponent's best record	0.1187725*** (.0421421)	0.1422717** (.0582515)	0.1673885 (.052771)
Inside lane dummy	0.0576807*** (.0219739)	0.010455 (.0233321)	0.0065224 (.0225801)
Expected result and best record interaction			-0.0532321 (.0672047)
Constant	15.90429*** (2.920128)	9.387393*** (3.16154)	13.29024
Obs	678	678	1356
R squared	0.6718	0.6197	0.6307

Note: *** 1%, ** 5%, * 10% significance.

The regression results for swimmers dataset is shown in table 5. In all three columns, the coefficients of the peers' average best records and the individual peers' best records are positive and highly significant. Hence, the better the opponents are, the faster the swimmer swims. When the opponents' average record is better by 1 second the swimmer swim time decreases by 0.18 seconds. This represents the positive peer effect exerted on swimmers by their opponents. This can come from the fact that swimmers are similar to speed skaters and are motivated during the race when swimming against good swimmers. The swimmers also know each other's records before going into a race so the motivation can occur before the race as well. In column 2, if the peer of the left's record is better by 1 second, the swimmer's time improves by 0.17 seconds. In column 3, if the peer of the right's record is better by 1 second, the swimmer timing improves by 0.11 seconds.

Table 5: Regression results for model 5 and 6. The dependent variable is the swimmers' swim time, with individual fixed effects and relevant controls.

	(1) Model 5 Coef.	(2) Model 6 on left lane peers Coef.	(3) Model 6 on right lane peers Coef.
Swimmer's best record	0.0536328 (.0349515)	0.0578788 (.0353012)	0.0848242** (.0326409)
Average Peer best record	0.1865628*** (.0481332)	-	-
Peer best record (left)	-	0.1703531*** (.0477827)	-
Peer best record (right)	-	-	0.1118345*** (.0363842)
Constant	12.77888*** (6.202727)	13.77452*** (6.199654)	13.60683*** (6.247712)
Obs	439	439	439
R squared	0.9978	0.9977	0.9977

Note: *** 1%, ** 5%, * 10% significance.

The regression results for the breaststroke and backstroke dataset is shown in table 6. In all three columns, the coefficients of the peers' average best records and the individual peers' best records are positive and highly significant. This represents the positive peer effect exerted on swimmers by their opponents in breaststroke and backstroke events, in which it is harder for the swimmers to observe their peers. Thus, these regression results show that the peer effect present among swimmers is compositional instead of contemporaneous. Even when the swimmers cannot observe their peers well, there is still peer effect due to the fact that swimmers know each other's best records and can be motivated by that. In this dataset, when the opponents' average record is better by 1 second the swimmer swim time decreases by 0.25 seconds. If the peer of the left's record is better by 1 second, the swimmer's time improves by 0.22 seconds. In column 3, if the peer of the right's record is better by 1 second, the swimmer timing improves by 0.16 seconds.

Table 6: Regression results for model 5 and 6 on the breaststroke and backstroke dataset. The dependent variable is the swimmers' swim time, with individual fixed effects and relevant controls.

	(1) Model 5 Coef.	(2) Model 6 on left lane peers Coef.	(3) Model 6 on right lane peers Coef.
Swimmer's best record	.033844 (.0315604)	.0374625 (.0314099)	.0440544 (.0318223)
Average Peer best record	.2457631*** (.0737978)	-	-
Peer best record (left)	-	.2172973*** (.0671751)	-
Peer best record (right)	-	-	.1554861** (.0621386)
Constant	67.64095*** (10.88692)	105.4712*** (12.79883)	75.13992 *** (10.52021)
Obs	190	190	190
R squared	0.9973	0.9973	0.9972

Note: *** 1%, ** 5%, * 10% significance.

V. Conclusion

In this study, I examined peer effect among international speed skaters in 500 m races and swimmers in various events and distances. The positive and significant coefficients for the opponents' best records show that skaters skate faster when racing against better skaters. The same positive and significant coefficient exists for swimmers. Hence peer effect exists in both swimming and speed skating. This is contrary to the optimal strategy, which is to skate or to swim as fast as possible regardless of the competition. The peer effect can come from observing the competitor or from being motivated by skating or swimming with good athletes. It is unlikely that they can get better during the race because the race is very short, but the speed skaters and swimmers might feel competitive when racing with good opponents and put in even more effort in their races. Though the head to head results generally do not matter as only the skate time matters, the desire to beat an opponent during a race can exist and may lead to a better performance for the skater. The result supports the existence of positive peer effect, in which people perform better when paired with productive partners. This correlates well with other studies on peer effect, particularly in sports. The study involves two very different sports, swimming and speed skating. As such, this provides external validity for this study as peer effect is not an isolated phenomenon in only one sport like swimming. The finding of peer effect has implications for the work place as well. The result shows us that people of lower ability should be paired with people with higher ability in the work place to improve their performance because of peer effect from the people with higher ability.

Male and female athletes are affected by peer effect by the same magnitude as there was no significant difference in terms of peer effect for male and female athletes. This contrasts with the finding of White (2012) among swimmers and may serve to dispute the claim that men are more competitive than women.

There was no difference in terms of magnitude of peer effect for winners and losers as the interaction term for direct race result and opponent's best record is insignificant. Similarly, there was no difference in magnitude of peer effect for expected winners and losers either. This contrasts with the finding of Yamane and Hayashi (2012) that the performance of a swimmer is improved by the slow-lane peer, aka the "chased effect" when the swimmer is only aware of the peer with the slower best records. An explanation for this is that the many opportunities for the skaters to observe each other during the race allow the "motivation effect" to push the skaters to perform better. Thus, the peer effect is similar in magnitude for both winners and losers in the pairs.

It is hard to estimate the compositional peer effect separately from the contemporaneous peer effect in speed skating because all races have the same setup. However, it is possible to separate the contemporaneous peer effect and compositional peer effect from each other in swimming because in certain strokes like breaststroke and backstroke, it is difficult for swimmers to observe their peer. The positive and significant peer effect coefficients for breaststroke and backstroke dataset show that even when the swimmers cannot observe each other well, there exists peer effect. As such, peer effect among swimmers is more likely to be compositional than contemporaneous, as it arises from the fact that swimmers know each other's best records prior to the races.

This study contributes to the growing literature on peer effect by confirming the existence of peer effects in both swimming and speed skating, and provides more insights into the nature of peer effect between different genders as well as the mechanism of peer effect. The insights learned from this paper can be applied to the work place, the school system and in sports, especially for situations that involve pairs or direct peers. This can lead to more efficient pairing between individuals with different abilities to maximize productivity at lowest social costs. One key insight from this paper is that the peers should be allowed to observe each other's effort, be it compositionally or contemporaneously, in order for peer effects to affect workers, students or athletes.

Acknowledgement: I would like to thank professor Kato, Maggie White and my classmates in ECON 442 and ECON 490 for helping me make this work possible.

References:

Depken, Craig A.II, and Lisa E. Haglund. "Peer Effects in Team Sports: Empirical Evidence from NCAA Relay Teams," Working Papers 0729, International Association of Sports Economists & North American Association of Sports Economists, 2007.

Guryan, Jonathan, Kory Kroft, and Matt Notowidigdo. Peer effects in the workplace: Evidence from random groupings in professional golf tournaments. No. w13422. National Bureau of Economic Research, 2007.

Kato, Takao& Shu, Pian. 2009. "Peer Effects, Social Networks, and Intergroup Competition in the Workplace," Working Papers 09-12, University of Aarhus.

Manski, Charles F. "Identification of endogenous social effects: The reflection problem." The review of economic studies 60.3 (1993): 531-542.

Mas, Alexandre, and Enrico Moretti. "Peers at work". No. w12508. National Bureau of Economic Research, 2006.

Sacerdote, Bruce. "Peer effects with random assignment: Results for Dartmouth roommates". No. w7469. National Bureau of Economic Research, 2000.

White, Maggie. "Peer effects among competitive swimmers: Does gender influence motivation?". Unpublished

Yamane, Shoko and Ryohei Hayashi. "Peer Effects of Swimmers". GCOE Discussion Paper Series No 236 (2012).

Appendix:

Table 7: Summary statistics for women’s 100 m freestyle

Variable	N	Mean	SD	Min	Max
swim time	21	57.5691	1.66266	54.39	60.95
best record	21	56.6595	2.177943	49.81	60.83
best record (left peer)	21	56.8952	1.41302	54.14	59.99
best record (right peer)	21	56.9367	1.277299	54.16	60.06

Table 8: Summary statistics for men’s 100 m freestyle

Variable	N	Mean	SD	Min	Max
swim time	21	51.7205	0.8972707	50.47	53.73
best record	21	51.1067	0.983404	49.43	53.59
best record (left peer)	21	51.2062	1.002065	49.55	53.52
best record (right peer)	21	50.8857	2.095976	43.35	53.85

Table 9: Summary statistics for women’s 200 m breaststroke

Variable	N	Mean	SD	Min	Max
swim time	18	156.444	5.072925	149.54	165.26
best record	18	151.948	6.459258	134.53	160.76
best record (left peer)	18	152.373	5.42231	144.46	160.71
best record (right peer)	18	153.519	5.356924	140.69	161.02

Table 10: Summary statistics for men’s 200 m breaststroke

Variable	N	Mean	SD	Min	Max
swim time	17	139.252	5.622962	129.16	149.38
best record	17	135.687	4.748785	120.1	140.7
best record (left peer)	17	136.981	3.439858	128.99	141.74
best record (right peer)	17	139.14	5.398528	129.72	151.8

Table 11: Summary statistics for women’s 400 m freestyle

Variable	N	Mean	SD	Min	Max
swim time	20	261.948	5.820068	250.79	274.95
best record	20	258.787	8.262329	249.83	289.99
best record (left peer)	20	257.863	8.433677	249.83	290.18
best record (right peer)	20	259.824	9.051052	245.19	289.99

Table 12: Summary statistics for men’s 400 m freestyle

Variable	N	Mean	SD	Min	Max
swim time	10	242.901	5.621282	234.69	252.05
best record	10	238.704	8.858183	232.38	261.07
best record (left peer)	10	238.973	8.114141	232.27	260.51
best record (right peer)	10	241.021	9.744935	229.96	261.79

Table 13: Summary statistics for women’s 100 m butterfly

Variable	N	Mean	SD	Min	Max
swim time	12	62.8475	1.242769	60.83	64.59
best record	12	61.8708	1.383477	59.54	64.17
best record (left peer)	12	61.8233	1.585963	58.79	64.13
best record (right peer)	12	61.9942	1.300556	59.91	64.23

Table 14: Summary statistics for men’s 100 m butterfly

Variable	N	Mean	SD	Min	Max
swim time	23	56.25	2.374302	52.5	62.73
best record	23	54.533	2.576302	47.55	57.84
best record (left peer)	23	54.5461	2.580054	47.76	57.98
best record (right peer)	23	54.6104	2.713291	47.72	57.95

Table 15: Summary statistics for women’s 400 m medley

Variable	N	Mean	SD	Min	Max
swim time	12	298.583	5.922687	289.75	310.16
best record	12	294.743	4.676497	287.94	303.85
best record (left peer)	12	294.518	6.670968	285.16	307.04
best record (right peer)	12	295.668	3.479039	289.1	301.49

Table 16: Summary statistics for men's 400 m medley

Variable	N	Mean	SD	Min	Max
swim time	11	273.376	8.037292	257.85	284.64
best record	11	268.1	5.402351	255.79	274.94
best record (left peer)	11	268.039	6.550601	249.92	275.01
best record (right peer)	11	270.74	4.858701	259.84	278.49

Table 17: Summary statistics for women's 200 m freestyle

Variable	N	Mean	SD	Min	Max
swim time	14	124.529	3.068522	119.18	131.06
best record	14	121.116	4.621228	106.6	125.6
best record (left peer)	14	121.143	4.566976	106.98	125.31
best record (right peer)	14	121.488	4.899613	105.97	124.98

Table 18: Summary statistics for men's 200 m freestyle

Variable	N	Mean	SD	Min	Max
swim time	20	113.281	2.508879	108.8	119.38
best record	20	112.071	1.408485	109.49	114.85
best record (left peer)	20	111.866	1.778768	107.94	114.5
best record (right peer)	20	112.117	2.268814	107.46	115.64

Table 19: Summary statistics for women's 200 m backstroke

Variable	N	Mean	SD	Min	Max
swim time	18	140.146	3.101654	135.53	147.3
best record	18	136.185	1.902711	132.83	139.41
best record (left peer)	18	136.082	2.157639	132.79	140.02
best record (right peer)	18	136.436	2.276595	132.57	139.78

Table 20: Summary statistics for men's 200 m backstroke

Variable	N	Mean	SD	Min	Max
swim time	29	126.0348	4.282315	118.38	134.4
best record	29	122.2893	5.965534	104.15	131.91
best record (left peer)	29	121.8303	5.757625	103.97	129.95
best record (right peer)	29	122.4117	6.128517	104.51	132.08

Table 21: Summary statistics for women's 50 m freestyle

Variable	N	Mean	SD	Min	Max
swim time	21	26.40476	0.5722554	25.16	27.66
best record	21	26.22381	0.5160085	25.2	27.63
best record (left peer)	21	26.22143	0.5277905	25.23	27.68
best record (right peer)	21	26.19286	0.621813	24.7	27.62

Table 22: Summary statistics for men's 50 m freestyle

Variable	N	Mean	SD	Min	Max
swim time	12	23.33083	0.4978583	22.43	24.01
best record	12	22.88583	0.9778033	20.12	23.67
best record (left peer)	12	23.145	0.4899814	22.04	23.59
best record (right peer)	12	23.0925	0.5399348	22.21	23.89

Table 23: Summary statistics for women's 200 m butterfly

Variable	N	Mean	SD	Min	Max
swim time	6	139.1267	4.430538	131.08	144.08
best record	6	136.0067	4.808483	131.12	144.13
best record (left peer)	6	134.4783	3.479387	128.81	138.53
best record (right peer)	6	136.6867	4.436569	132.57	143.62

Table 24: Summary statistics for men's 200 m butterfly

Variable	N	Mean	SD	Min	Max
swim time	11	127.7136	3.376449	122.64	132.64
best record	11	121.5118	5.627061	105.55	127.75
best record (left peer)	11	122.1118	6.112848	105.46	130.56
best record (right peer)	11	122.6191	5.768344	106.41	127.08

Table 25: Summary statistics for women's 100 m breaststroke

Variable	N	Mean	SD	Min	Max
swim time	26	71.86346	2.226218	67.72	76.87
best record	26	70.57808	1.794329	67.19	72.6
best record (left peer)	26	70.60231	2.077126	65.9	73.66
best record (right peer)	26	70.68731	2.123664	65.9	73.76

Table 26: Summary statistics for men's 100 m breaststroke

Variable	N	Mean	SD	Min	Max
swim time	19	63.62	1.700343	60.08	66.17
best record	19	63.52211	1.523704	60.47	66.91
best record (left peer)	19	63.26526	1.627211	59.76	66.08
best record (right peer)	19	63.78	1.578392	60.08	65.89

Table 27: Summary statistics for women's 100 m backstroke

Variable	N	Mean	SD	Min	Max
swim time	26	63.85154	1.883356	58.35	66.13
best record	26	62.44192	1.565097	58.48	64.03
best record (left peer)	26	62.91038	1.363981	58.52	65.56
best record (right peer)	26	62.75192	1.646694	58.49	64.8

Table 28: Summary statistics for men's 100 m backstroke

Variable	N	Mean	SD	Min	Max
swim time	37	57.28	1.618365	53.83	60.23
best record	37	55.77919	3.000036	47.22	60.62
best record (left peer)	37	56.22486	2.655778	47.22	61.19
best record (right peer)	37	56.00189	3.070723	47.47	60.64

Table 29: Summary statistics for women's 200 m medley

Variable	N	Mean	SD	Min	Max
swim time	16	143.1581	3.708421	138.07	153.09
best record	16	140.385	1.919635	138.39	144.88
best record (left peer)	16	140.2456	1.731323	138.4	144.43
best record (right peer)	16	140.4325	2.648843	134.13	145

Table 30: Summary statistics for men's 200 m medley

Variable	N	Mean	SD	Min	Max
swim time	19	128.3979	3.398711	120.46	137.34
best record	19	123.9605	5.961378	107.55	130.07
best record (left peer)	19	123.7021	6.013993	107.83	129.67
best record (right peer)	19	124.0542	6.290038	107.83	131.4