Mind the absent gap: Gender-specific competitiveness in non-professional sports

Anna Katharina Pikos* and Alexander Straub†

This version: March 26, 2019

Abstract

There is wide evidence for gender differences in competitiveness and performance under pressure from experimental economics and single-sex professional sports. We analyze these differences in a sport with direct gender competition. Our unique data consists of over 500,000 observations from around 11,000 German ninepin bowling games of which around 15% are from mixed-gender leagues. Men perform better against women on average but this is fully explained by differences in ability. Our results are robust to instrumenting for opposite gender using the sex composition of the opponent team. Surprisingly, gender differences in tight situations do not seem to play a role.

Keywords: gender; gender competition; sports economics

JEL Classification: J16, D90, Z22

*Corresponding author, Institute of Labour Economics, Leibniz Universität Hannover, Königsworther Platz 1, D-30167 Hannover, Germany, e-mail: pikos@aoek.uni-hannover.de, phone: +49 511 7625620.

†Institute of Labour Economics, Leibniz Universität Hannover, e-mail: straub@aoek.uni-hannover.de.

Man differs from woman in size, bodily strength, hairyness, &c., as well as in mind, in the same manner as do the two sexes of many mammals.

– Charles Darwin, *The Descent of Man, 1871, pp.13-14*

1 Introduction

Do men compete in the same way as women? Men earn higher wages, get more promotions, and hold the majority of leadership positions. This holds also when accounting for standard economic variables. For example, the wage gap persists for women with full-time employment history, without children, and without family plans (e.g Manning and Swaffield, 2008). Similarly, the promotion gap is not fully explained by worker performance and firm characteristics (e.g Blau and DeVaro, 2007). This does not necessarily prove discrimination as there might be differences in unobservable characteristics.

Personality differences received rising scholarly attention as a potential explanation for the persisting gender gaps in the labour market. The experimental literature finds that men are more prone to select into competitive environments (Niederle and Vesterlund, 2007; Dohmen and Falk, 2011), are less risk-averse (Croson and Gneezy, 2009), and women seem to perform worse when competing against men (Gneezy et al., 2003; Antonovics et al., 2009). The sports economics literature focuses on differences in competitive behavior of high performers and identifies risk behavior, environment, and stakes as important factors. Men take more risks when risky behavior might pay off (Böheim et al., 2016) and when it does not (Gerdes and Gränsmark, 2010), women perform better in a female environment (Booth and Yamamura, 2018), and men choke under pressure (Cohen-Zada et al., 2017). Incentives in professional sports are predominantly monetary. Insights about the importance of non-monetary rewards in gender-specific competitiveness are still missing but would help to better understand "intrinsic" motivation.

We aim at filling this gap by analyzing gender differences in a non-professional environment with non-monetary rewards. Our data comes from German ninepin bowling, a mixed-gender sport where individual players compete against one opponent to obtain points for their team. This data has three advantages. First, men and women compete against each other in the same environment. Second, unobserved group dynamics exert less influence in a one-against-one competition. Third, our panel data allows to control for past performance as a proxy for ability. This enables us to isolate the effect of competing against the opposite gender on performance. Our dataset consists of over 500,000 observations from more than 11,000 games. About 15% of the games are from mixed-gender leagues.

We answer two questions: first, do men and women compete differently against the opposite gender and second, are there systematic gender differences in performance under pressure? Descriptive statistics document that men perform slightly better and hence are more likely to win against women. Any differences in playing against the opposite gender are fully explained by ability or game characteristics. We confirm the OLS findings with fixed effects and IV. To rule out the possibility of non-random matching completely we instrument the probability of playing against the opposite gender with the gender composition of the opponent team. To address gender differences in performance under pressure, we analyze tight game situations. We do not find any evidence for a gender performance gap under pressure.

The remainder of the paper is organized as follows. Section 2 reviews the relevant literature. The data is presented in section 3, the estimation strategy in section 4. Section 5 contains the results, the last section concludes.

2 Literature

Competition is important for a variety of labor market outcomes such as wages and promotion. Card et al. (2016) find sorting and bargaining effects for premium pay among Portuguese workers. Women are less likely to sort into companies paying higher premiums and they only receive 90% of their male colleagues' premium. There is evidence that part of the gender gaps might be related to different performance in negotiations. When applying for new jobs, women are more inclined to accept lower wages and men are more likely to negotiate wages (Leibbrandt and List, 2015). The experimental literature supports gender differences in negotiations; e.g. women negotiate equally well as employers but make lower initial offers as employees (Dittrich et al., 2014).

Laboratory and field experiments shed light on how men and women behave in competitive environments. Gneezy and Rustichini (2004) for example show that boys and girls in Israel react differently to competition while running. Boys run faster when they face a direct opponent compared to running alone. This is not the case for girls whose performance does not increase under competition. According to Dreber et al. (2011), culture and task play a role, as the effect is less pronounced in Sweden and in rather gender neutral activities, such as skipping a rope. Laboratory experiments confirm that women shy away from competition if possible and do not increase their performance if forced to compete (e.g. Gneezy et al., 2003; Niederle and Vesterlund, 2007). The environment's gender composition is important but empirical evidence is mixed. Booth and Nolen (2012) document that a purely female environment may enhance women's competitive behavior, while Lee et al. (2014) find a reduction in competitiveness in single-sex environments (see also Niederle and Vesterlund, 2011, for an extensive literature review).

Numerous studies from sports economics confirm a gender gap in competitiveness. Men perform better in mixed-sex Japanese speedboat races, while the opposite is true for women (Booth and Yamamura, 2018). As speedboat races are group races, there might be spillover effects impacting individual behavior that might go in both directions, e.g. racing faster because somebody else is fast or slowing down because somebody cuts the lane. Men are less risk averse when risky behavior might pay off in professional basketball (Böheim et al., 2016) but they perform worse when stakes are very high in tennis (Cohen-Zada et al., 2017). A common problem with professional sports data is that men and women usually compete separately, e.g. National Basketball Association versus Women National Basketball Association, and Grand Slam tournaments. Different leagues have different characteristics (e.g. wage or quality of medical treatment in case of injury) which might impact players' behavior. A notable exception is Gerdes and Gränsmark (2010) who analyze international chess games. Men and women compete directly against each other which eliminates group spillover effects and league differences. They find that men not only are more risk-loving but also play more aggressively against women even if this reduces their probability to win the game.

We lack insights regarding the everyday work life of the general population. Professional sports are very selective and representative for top performers only, e.g. managers. Trophy money, advertising contracts, and prestige are at stake. To address competitiveness beyond the upper end of the ability distribution, non-professional sports are insightful. Analyzing situations when winning is not tied to monetary rewards enables a better understanding of gender differences underlying the "pure" motivation to win.

Differences in competitive behavior could be more or less pronounced in non-professional sports.[1]

3 Data and Descriptives

The data comes from ninepin bowling, a non-professional sport played in many European countries. Even world-class players do not get paid.[2] German ninepin bowling leagues ("Kegeln") on the national and federal state level are gender-separated. Lower leagues on county level can be mixed-gender to make it easier for small clubs to put together a team. Because there are no financial incentives, we refer to this sport as a low stake environment. Any differences in competitiveness would mostly be intrinsic, especially in the less selective lower leagues we use for this analysis. Reaching a higher league is not associated with financial incentives.

3.1 Data

This subsections outlines the rules for ninepin bowling leagues regulated by the German ninepin bowling association ("Deutscher Keglerbund Classic e.V."). Two teams of four or six players bowl against each other. Each player has a direct opponent against whom she bowls in four sets. Each player has 120 throws, i.e. 30 throws per set. For the first fifteen throws of each set the pins reset after each bowl. The sum of knocked down pins is recorded as *V-score*. For the last 15 throws in each set the pins reset only after the player knocked down all the 9 pins, i.e. if a player fails to knock down all of them, she uses up a next throw to fell the remaining pins. We refer to the sum of the cleared pins from the last 15 throws of a set as *A-score*. After the player and his opponent finished their 30 throws *V-score* and *A-score* are added up to the *score*. The player with the highest *score* receives a *point* for winning the set.[3] Thereupon the two players switch the lanes and the next set starts. After the four sets are completed, the player with the highest

[1] On the one hand, men choke under pressure (Cohen-Zada et al., 2017) and the highest-performing women are more prone to select into competition (Gneezy and Rustichini, 2004). Based on these results differences in non-professional sports could be more pronounced. On the other hand, women perform better in low stake environments (Ors et al., 2013). This could result in a lower gender gap.

[2] There might be sponsoring contracts for European and World championships, but this will hardly affect any of our results because we analyze leagues which are at least 6 levels lower.

[3] If both players have the same *score* they receive 0.5 set *points*.

amount of *points* receives a *team point*.[4] The alleys consist of four lanes where two home and two guest players bowl simultaneously and form a *pairing*. The players switch lanes after each set, so that every player bowls on every lane. When the game is finished, the team with the highest sum of *score* receives two additional *team points*. The team with the highest sum of *team points* wins the game. The losing team does rarely benefit from the earned *team points*.[5]

Our data comes from leagues in the Northern part of the federal state of Baden-Württemberg in Southern Germany. The local bowling association "Württembergischer Kegler- und Bowling-Verband e.V. (WKBV)" publishes game records online. Records contain information on league, location, date, start and end of the game. Player information is limited to name and player number. Performance measures include *points*, *V-score*, *A-score*, and *F-score* (mistakes, i.e. number of bowls not hitting any pin). We use game records from the seasons 2014/15 to 2017/18.

We exclude players playing fewer than five games, games with a predetermined winner[6] and players without opponents. Gender is coded according to Wikipedia lists for male and female given names.[7] We use record sheet information symmetrically from the home player's and the guest player's point of view, i.e. in mixed-gender competition, we have women vs. man and man vs. women. This leaves us with around 75,000 observations from about 2,000 games in mixed-gender leagues. The full sample with gender-separated leagues counts around 500,000 observations from more than 11,000 games.

3.2 Descriptives

Men bowl significantly better than women in mixed-gender leagues although the differences are not large (see appendix figure A.1). Table 1 shows the raw gender differences in outcomes in mixed-gender games (see appendix table A.1 for full sample descriptives). On average, men bowl 0.5 pins more than women; 0.3 pins in *V-score* and 0.2 pins in

[4]If both players obtained two set points, the one with the highest sum of knocked down pins (total *score* over four sets) gets the team point. If both knocked down the same amount of pins and both have two set points, each of them receives 0.5 team points.

[5]Team points are only important in the ranking if ranking points are the same for two teams. The winning team earns 2:0 ranking points, the losing team 0:2, and 1:1 for ties.

[6]If a team lacks two or more of the scheduled players, there is no possibility to win the game.

[7]About 250 of 3,500 players could not be matched because their names were either not listed on Wikipedia or for unisex first names such as e.g. Robin and Gabriele. We infer the gender for most of those players because they participated at least once in a strictly gender separated tournament or league. For the remaining players, we looked up the gender on the web-page of their club.

A-score. Men win their set slightly more often and make somewhat more mistakes but the latter difference is very small and significant only at the 10% level. Table 2 gives an overview over the number of observations for own and opponent's gender. There are 75,056 observations. Roughly 48% of the encounters are mixed-gender encounters. Male encounters make up around 31%, female encounters around 21%.

Table 1: Outcomes by gender

	men	women	difference	p-value
score	113.837	113.321	0.516	0.000
points	0.507	0.492	0.014	0.000
V-score	80.779	80.471	0.307	0.000
A-score	33.058	32.849	0.209	0.004
F-score	3.732	3.699	0.034	0.096
Observations	41252	33804		

Notes: Data source: WKBV. Results for t-test on difference of game outcomes by gender. See page 4 for a detailed discussion of the variables.

Table 2: Numbers of observations for own and opponent's gender

	opponent's gender					
own gender	male		female		Total	
	No.	%	No.	%	No.	%
male	23,338	31.1	17,914	23.9	41,252	55.0
female	17,914	23.9	15,890	21.2	33,804	45.0
Total	41,252	55.0	33,804	45.0	75,056	100.0

Notes: Data source: WKBV. Distribution of gender and opponent's gender in mixed sex leagues.

4 Empirical Strategy

This section describes our OLS, fixed effects, and IV models. The outcomes are performance measures y_{ijk} of player i against the opponent j in the environment k. The main explanatory variables are gender, playing against the opposite gender, and the interaction term ($female_i$, opp_gender_{ij}, and $female_i \cdot opp\_gender_{ij}$). Z'_k is a vector of "environmental" characteristics k containing dummy variables for *pairing*, *set*, and playing at *home*.

$ability'_{ij}$ is a vector of player i's, opponent j's and teams' ability measures.[8] ϵ_{ijk} is the error term clustered at players' level.

$$\begin{aligned} y_{ijk} =& \beta_0 + \beta_1 \cdot female_i + \beta_2 \cdot opp\_gender_{ij} + \beta_3 \cdot female_i \cdot opp\_gender_{ij} \\ &+ Z'_k \gamma + Ability'_{ij} \delta + \epsilon_{ijk} \end{aligned} \tag{1}$$

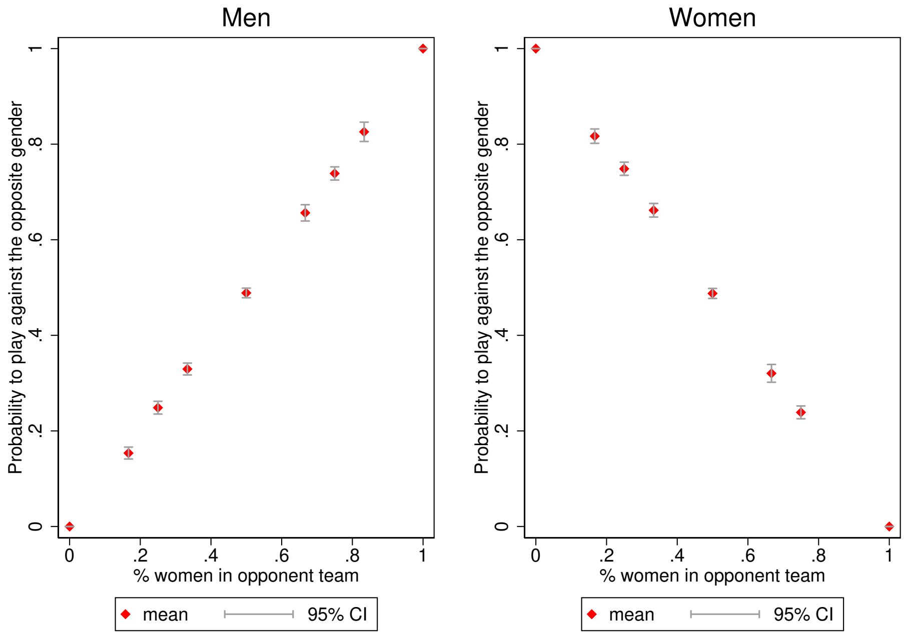
To account for systematic individual and lane differences, we estimate fixed effects models. We validate our results with player fixed effects in case our individual ability measures do not fully capture idiosyncratic differences, e.g. gender-specific returns to experience or panel attrition. There are two reasons to consider lane fixed effects. First, lane quality can vary substantially even within locations. Second, there are lane-specific differences in physical proximity to the other players. When playing on the inner lanes, the current score of the other players might be easier to assess, which could influence own performance.

OLS may yield biased estimates if individuals self-select into playing against opponents of a certain gender, i.e. if they have a preference for competing against men or women. To address the potential endogeneity of β_2 and β_3, we use the opponent team's female share as an instrument for playing against the opposite gender. Figure 1 displays the raw first stages for rounded values of opponent female share by own gender and confirms the relevance of the instrument. As the share of women in the opposite team increases, the probability to play against the opposite gender increases for men and decreases for women. All first stages are highly significant, the coefficient is virtually 1 in absolute terms, and F-statistics are large (5600 to 9200).

For the exclusion restriction to hold, the opponent team's female share must not have any direct effect on performance. The main concern is that players might have higher confidence when perceiving a predominantly female team as weaker. If indeed the team is weaker, this does not bias the results as long as we control for team and opponent team ability. If the team is not weaker, then players have a biased perception of the ability distribution of their opponents by gender. This means that opponent female share could lead to false overconfidence (or underconfidence) and thus has a direct effect

[8] We mostly use past performance as a proxy for ability. Alternative measures, such as individual fixed effects and different ability measures based on past and future performance, are used in robustness checks, see table A.2.

Figure 1: Probability to play against the opposite gender for men and women depending on opponent team composition



Rounded values of opponent female share. Excluding games with substitutions. Data source: WKBV.

on performance. Given that performance measures on player and team ability are easily available to all players,[9] there should not be any strong biases in perceived opponent ability. Even if players had gender-biased beliefs about performance, they would update their beliefs as soon as they gain experience in playing against both genders in their league. Hence, if such a direct effect existed, it is negligible.

5 Results

Men do not perform systematically better against women and vice versa (table 3). Not accounting for ability, men are 1.5 percentage points more likely to win the set against women.[10] This difference is significant at the 5% level. There are no significant gender

[9]Team and individual performance rankings are distributed online after each matchday. The rankings contain the weekly top 10 separated by home and guest games and cumulative home and guest performance of all players in the league.

[10]Due to the symmetry of our dataset, it is not possible to disentangle whether men perform better against women, whether women perform worse against men, or whether a combination of both is true.

differences for *score* and *F-score*. The models explain between 0.4% and 2% of the variation in the outcomes. Including ability controls in the right hand part of the table turns the opponent gender difference for points insignificant. The models explain larger shares of the variation in the outcome (between 18% for points and 36% for score). The results are robust to different ability measures (see appendix table A.2).

Table 3: OLS estimates for score and points and F-score

	without ability			with ability		
	score	points	F-score	score	points	F-score
female	-0.151	-0.000	-0.054	-0.078	0.000	-0.052
	(0.683)	(0.013)	(0.095)	(0.210)	(0.006)	(0.036)
opp. gender	0.160	0.015**	-0.025	-0.096	0.006	0.002
	(0.261)	(0.006)	(0.036)	(0.157)	(0.005)	(0.025)
female × opp. gender	-0.482	-0.030***	0.026	0.063	-0.012	-0.055
	(0.408)	(0.010)	(0.057)	(0.244)	(0.008)	(0.037)
2^{nd} set	0.783***		-0.102***	0.704***		-0.096***
	(0.129)		(0.021)	(0.136)		(0.022)
3^{rd} set	0.676***		-0.103***	0.664***		-0.093***
	(0.135)		(0.022)	(0.142)		(0.023)
4^{th} set	0.926***		-0.143***	0.913***		-0.152***
	(0.150)		(0.024)	(0.158)		(0.025)
2^{nd} pairing	1.037***		-0.167***	-0.159		-0.006
	(0.354)		(0.048)	(0.146)		(0.025)
3^{rd} pairing	6.149***		-0.906***	-0.142		-0.051
	(0.582)		(0.079)	(0.229)		(0.038)
home	1.676***	0.059***	-0.183***	1.565***	0.057***	-0.167***
	(0.174)	(0.004)	(0.024)	(0.153)	(0.004)	(0.021)
past ability				0.832***	-0.000	-0.110***
				(0.015)	(0.000)	(0.003)
difference ability				-0.004	0.015***	-0.001
				(0.009)	(0.000)	(0.001)
team ability				0.018	0.001***	-0.004
				(0.016)	(0.000)	(0.003)
opponent team ability				-0.047***	-0.001***	0.005**
				(0.014)	(0.000)	(0.002)
constant	110.904***	0.471***	4.121***	22.666***	0.473***	16.223***
	(0.504)	(0.008)	(0.073)	(1.400)	(0.038)	(0.252)
Observations	75056	75056	75042	64738	64738	64727
Adj. R^2	0.020	0.004	0.016	0.364	0.182	0.275

Notes: This table shows the relationship between player's gender and characteristics and the outcomes of interest in mixed gender leagues. The outcome *score* is the total score per lane; *points* are the set points obtained on one lane (0 if lost, 0.5 if tie, and 1 if won); *F-score* denotes the mistakes, i.e. how often the player did not hit any pin. *Female* and *opp. gender* are dummy variables if the player is female or plays against the opposite gender respectively. *Set* and *pairing* are a set of dummy variables that indicate the difference in *score* and *F-score* compared to 1^{st} set and 1^{st} pairing; these are omitted for *points* due to the symmetry of the data. *Past ability* is the average *score* of the player per lane if more than 8 lanes are observable from past data. *Difference* is the difference between *past ability* of the player and her opponent. *Team ability* and *opponent team ability* are measures for team's quality, they are calculated by the average of *past ability* of other players in the team. Robust standard errors clustered at the level of the player are in parentheses. ***, ** and * denote significance at the 1%, 5% and 10% level, respectively. Data source: WKBV.

To account for idiosyncratic differences across players and lanes, we run fixed effects regressions (see appendix table A.3). With individual fixed effects, there is only one significant gender difference: women score 0.3 fewer pins when playing against men (significant at the 10% level). There is no difference for points and F-score. All gender coefficients are insignificant with lane fixed effects. The same is true when combining individual and lane fixed effects.

Instrumenting playing against the opposite gender with opponent team's female share confirms our results. Table 4 displays the second stages; first stage t-statistics can be found at the bottom of the table. Men do not compete differently against women in any outcome but women make 0.14 fewer mistakes when playing against men compared to playing against women. This difference is significant at the 5% level. Hence, there is no evidence that our OLS results are driven by selection into competing against the opposite gender.

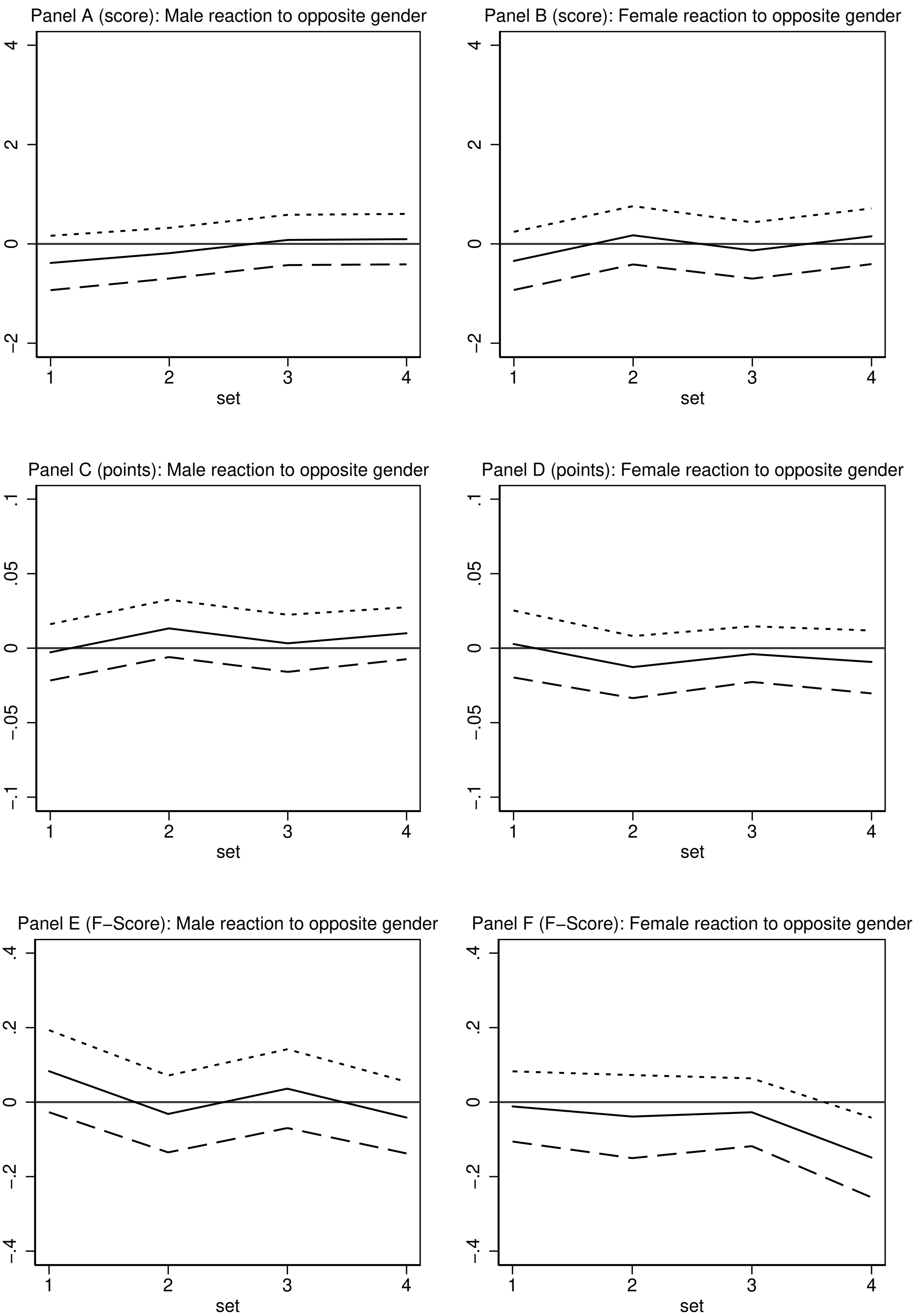
Table 4: Second stage IV estimates by gender

	score		points		F-score	
	women	men	women	men	women	men
opp. gender	0.640	-0.374	-0.016	0.008	-0.137**	0.070
	(0.418)	(0.315)	(0.011)	(0.010)	(0.061)	(0.052)
home	2.022***	1.187***	0.063***	0.052***	-0.209***	-0.133***
	(0.235)	(0.198)	(0.006)	(0.005)	(0.033)	(0.027)
past ability	0.819***	0.838***	0.000	-0.000	-0.110***	-0.110***
	(0.020)	(0.023)	(0.001)	(0.001)	(0.004)	(0.004)
difference ability	0.012	-0.014	0.016***	0.015***	-0.002	0.000
	(0.013)	(0.012)	(0.000)	(0.000)	(0.002)	(0.002)
team ability	0.010	0.025	0.001	0.001**	-0.001	-0.008**
	(0.024)	(0.022)	(0.001)	(0.001)	(0.004)	(0.003)
opponent team ability	-0.031	-0.056***	-0.001**	-0.001**	0.002	0.006**
	(0.020)	(0.019)	(0.001)	(0.001)	(0.003)	(0.003)
constant	22.562***	22.344***	0.489***	0.467***	16.132***	16.284***
	(2.049)	(1.879)	(0.062)	(0.047)	(0.403)	(0.322)
Observations	29146	35592	29146	35592	29142	35585
Adj. R^2	0.372	0.357	0.190	0.175	0.278	0.272
First stage t-statistic	-75.088	96.170	-74.586	96.004	-75.088	96.170

Notes: This table shows the relationship between player's gender and characteristics and the outcomes of interest in mixed gender leagues. The outcome *score* is the total score per lane; *points* are the set points obtained on one lane (0 if lost, 0.5 if tie, and 1 if won); *F-score* denotes the mistakes, i.e. how often the player did not hit any pin. Models control for *set* and *pairing* (except for points as dependent variable), and *home*. *Past ability* is the average *score* of the player per lane if more than 8 lanes are observable from past data. *Difference* is the difference between *past ability* of the player and her opponent. *Team ability* and *opponent team ability* are measures for team's quality, they are calculated by the average of *past ability* of other players in the team. Robust standard errors clustered at the level of the player are in parentheses. ***, ** and * denote significance at the 1%, 5% and 10% level, respectively. Data source: WKBV.

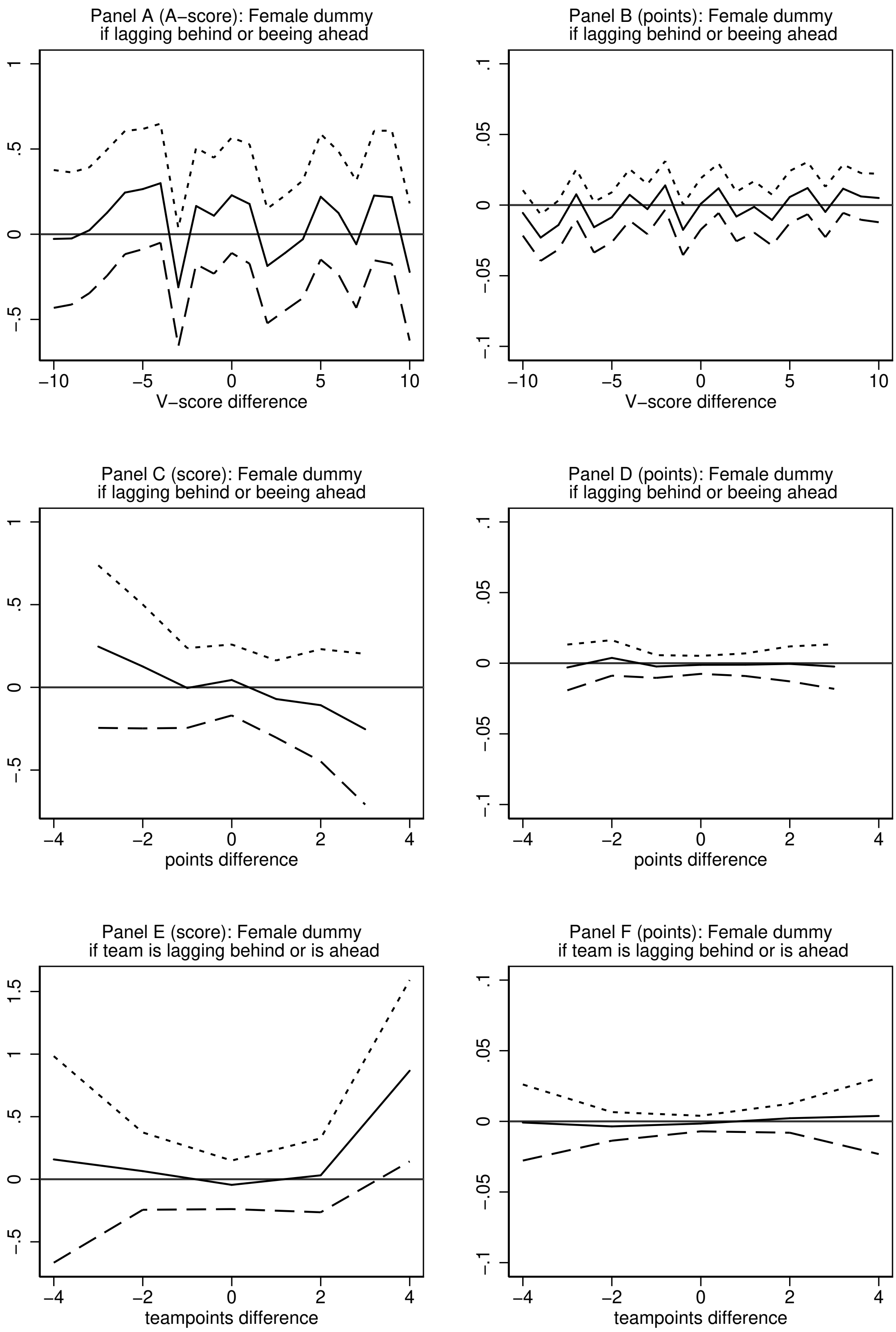
Reaction to opposite gender might differ over sets. The first set is important to get into the game and to make a good start, while the last set is the last opportunity to gain a set point. Gender-specific differences in physical and mental endurance could cancel each other out. To detect such differences we estimate equation (1) separately for each set. Figure 2 plots the reaction to opposite gender for men and women over the four sets (x-axis). There are no systematic gender differences in any outcome in any set. Effect sizes are very close to zero and confidence intervals almost always include zero.

Figure 2: Male and female reaction to opposite gender



Dotted lines: 95% confidence intervals. Data sources: WKBV.

Figure 3: Gender performance differences in tight situations



Dashed lines: 95% confidence intervals. Data sources: WKBV.

Our results could hide gender differences in performance under pressure. In some situations performance matters more than in others, e.g. turning around a tight game is more important than winning an already decided game. The overall estimate could be zero because one gender might perform better in tight situations but worse when pressure is low. We estimate equation (1) in subsamples to consider three different game situations. Figure 3 shows the estimated coefficients for the female dummy and their standard errors in three situations of the game. The x-axes are the differences between own and opponent's performance on individual or team level; negative (positive) values indicate that the player/team is lagging behind (leading).[11] First, we analyze performance in the second part of the set (*A-score*, panel A) and the probability to win the set (*points*, panel B) conditional on the difference in the first part of the set (*V-score*). Second, we depict *score* (panel C) and *points* (panel D) depending on the *points* difference from previous sets. Finally, we show *score* (panel E) and *points* (panel F) conditional on the difference in team *points* from previous players.[12]

There is no evidence for any systematic gender differences in playing under pressure. The performance of women lagging behind by up to 10 pins does not differ from their male counterparts (panels A and B). Similarly, we do not find systematic gender differences when lagging behind or leading individually (panels C and D) or as a team (panels E and F). Interestingly, there is no relationship between gender coefficient and the intensity of pressure, which varies along the x-axes, i.e. pressure is higher for smaller differences. For example, in panels C and D when the absolute points difference is three, turning around the game is impossible, while players should feel the highest pressure for absolute differences of one or zero. Overall, there is a reaction to intensity as can be seen by a significant effect of lagging behind or being ahead on subsequent performance (appendix figure A.2) but this reaction is not gender-specific.

[11] Note that in the analyses underlying figure 3, we do not limit the sample to mixed-gender leagues. As we do not find any (causal) effect (see the following), we are confident that this is not biasing the gender dummy we are interested in. The advantage of considering all leagues is that we have larger samples as we run separate regressions for every value on the x-axis of the graphs, e.g. the coefficients displayed in panel A stem from 21 different regressions.

[12] We run regressions only on even values of team point difference since the extremely low number of ties leads to a low number of observations for differences of 1 and 3 team points.

6 Conclusion

Using a rich panel dataset with direct competition, we analyze competitiveness in non-professional sports and do not find evidence for gender-specific differences. In our setting, participation is less selective, stakes are low, and motivation to win is mainly intrinsic. Our dataset is large ($n > 500,000$) and gives us precisely estimated coefficients with small standard errors. Men play better than women on average but their performance does not depend on their opponent's gender. This is robust across sets and hence not driven by differences in physical and mental endurance. Both genders react similarly to pressure.

There are three possible explanations why our results differ from the literature from professional sports: ability measurement, stakes, and environment. First, the significantly better male performance found in professional sports could stem from measurement error in ability. Our panel dataset allows to construct very detailed ability controls based on past performance. Ignoring gender differences in physical capacities would lead to omitted variable bias and might explain gender differences found in professional sports. As data availability increases, existing studies could be replicated using elaborated proxies for ability, e.g. running distance and speed in tennis and basketball.

Second, the discrepancy in results could arise from stakes. The previous literature exclusively considers high stake environments in which performance is tied to monetary rewards. In situations where mainly pure intrinsic motivation is decisive, men and women compete equally well. If monetary rewards induce men to perform better in mixed-gender settings, this finding could be one part behind the still unexplained gender pay and promotion gap. It also relates to the gender differences in occupational choices, i.e. women self-select into professions with lower expected returns (e.g. nurse, child care worker).

Third, previous works mostly study competitiveness in dynamic and gender-separated environments. In dynamic environments players react to their opponents' strategy and can put pressure on each other more easily, e.g. smashing in tennis. Analyzing individual behavior is more complex as it always depends on opponent behavior. Ninepin bowling is a more static environment in the sense that there are fewer opportunities to put pressure on opponents. This makes it easier to measure individual performance. Gender-separated environments complicate analyses because men and women, although doing the same sports, play in different leagues (except Gerdes and Gränsmark, 2010). Different leagues come with different characteristics which might impact behavior, e.g. wage or quality of medical treatment in case of injury. In ninepin bowling, children learn the sport in mixed-gender groups, and usually continue to train together as adults. Hence, men and women know from an early age on that they can perform equally well. This reduces false overconfidence from men and excessive underconfidence from women.

To conclude we do not find any gender gap in competitiveness in an environment where ability measures is easily observable, monetary incentives do not play a role, and men and women train together. This relates to Exley et al. (forthcoming) who document that women know their value and only enter competition when it pays off. In our setting, women assess their own performance correctly because they train in mixed-gender groups from the beginning. If the lack of appropriate information induces a difference in competitiveness between men and women, widely available performance measures might help. If the gender gap arises only in the presence of monetary rewards, this could reflect different preferences, e.g. women might value money less.

References

Antonovics, K., P. Arcidiacono, and R. Walsh (2009). The effects of gender interactions in the lab and in the field. *The Review of Economics and Statistics 91*(1), 152–162.

Blau, F. D. and J. DeVaro (2007). New evidence on gender differences in promotion rates: An empirical analysis of a sample of new hires. *Industrial Relations: A Journal of Economy and Society 46*(3), 511–550.

Böheim, R., C. Freudenthaler, and M. Lackner (2016). Gender differences in risk-taking: Evidence from professional basketball. IZA DP No. 10011.

Booth, A. and P. Nolen (2012). Choosing to compete: How different are girls and boys? *Journal of Economic Behavior & Organization 81*(2), 542–555.

Booth, A. and E. Yamamura (2018). Performance in mixed-sex and single-sex competitions: What we can learn from speedboat races in Japan. *Review of Economics and Statistics 100*(4), 581–593.

Card, D., A. R. Cardoso, and P. Kline (2016). Bargaining, sorting, and the gender wage gap: Quantifying the impact of firms on the relative pay of women. *The Quarterly Journal of Economics 131*(2), 633–686.

Cohen-Zada, D., A. Krumer, M. Rosenboim, and O. M. Shapir (2017). Choking under pressure and gender: Evidence from professional tennis. *Journal of Economic Psychology 61*, 176–190.

Croson, R. and U. Gneezy (2009). Gender differences in preferences. *Journal of Economic Literature 47*(2), 448–74.

Darwin, C. (1871). *The descent of man and selection in relation to sex*, Volume 1. Murray.

Dittrich, M., A. Knabe, and K. Leipold (2014). Gender differences in experimental wage negotiations. *Economic Inquiry 52*(2), 862–873.

Dohmen, T. and A. Falk (2011). Performance pay and multidimensional sorting: Productivity, preferences, and gender. *American Economic Review 101*(2), 556–90.

Dreber, A., E. von Essen, and E. Ranehill (2011). Outrunning the gender gap - boys and girls compete equally. *Experimental Economics 14*(4), 567–582.

Exley, C. L., M. Niederle, and L. Vesterlund (forthcoming). Knowing when to ask: The cost of leaning in. *Journal of Political Economy*.

Gerdes, C. and P. Gränsmark (2010). Strategic behavior across gender: a comparison of female and male expert chess players. *Labour Economics 17*(5), 766–775.

Gneezy, U., M. Niederle, and A. Rustichini (2003). Performance in competitive environments: Gender differences. *The Quarterly Journal of Economics 118*(3), 1049–1074.

Gneezy, U. and A. Rustichini (2004). Gender and competition at a young age. *The American Economic Review 94*(2), 377–381.

Lee, S., M. Niederle, and N. Kang (2014). Do single-sex schools make girls more competitive? *Economics Letters 124*(3), 474–477.

Leibbrandt, A. and J. A. List (2015). Do women avoid salary negotiations? Evidence from a large-scale natural field experiment. *Management Science 61*(9), 2016–2024.

Manning, A. and J. Swaffield (2008). The gender gap in early-career wage growth. *The Economic Journal 118*(530), 983–1024.

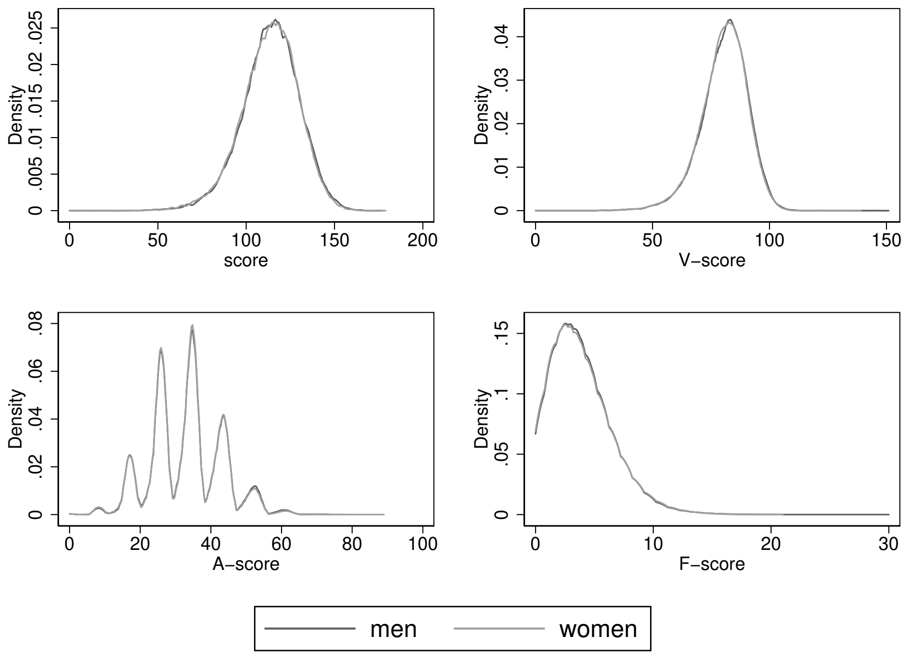
Niederle, M. and L. Vesterlund (2007). Do women shy away from competition? Do men compete too much? *The Quarterly Journal of Economics 122*(3), 1067–1101.

Niederle, M. and L. Vesterlund (2011). Gender and competition. *Annual Review of Economics 3*(1), 601–630.

Ors, E., F. Palomino, and E. Peyrache (2013). Performance gender gap: does competition matter? *Journal of Labor Economics 31*(3), 443–499.

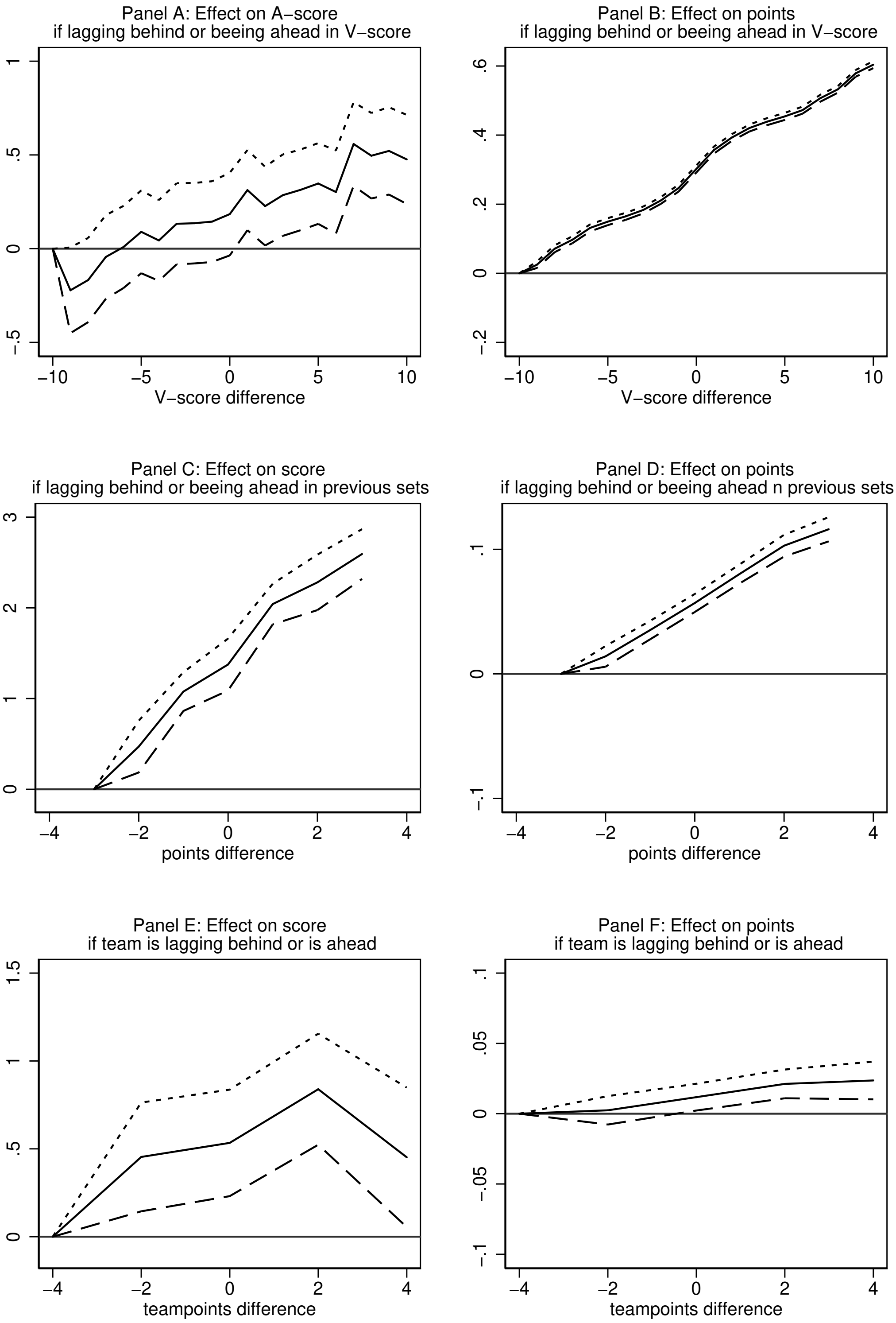
Appendix Figures

Figure A.1: Epanechnikov kernel density estimates by gender



Data source: WKBV. Kernel: epanechikov.

Figure A.2: Performance in tight situations



Full sample estimates, regressions also include control variables for playing at *home* and individual and team ability for player and opponent. Panel A, C, and E also include control variables for *pairing* and *set*. Data sources: WKBV.

Appendix Tables

Table A.1: Descriptive statistics (full sample)

	female				male			
	mean	sd	min	max	mean	sd	min	max
score	121.201	14.816	0.0	184.0	126.733	14.755	0.0	195.0
V-score	84.422	8.745	0.0	114.0	87.050	8.480	0.0	146.0
A-score	36.779	9.703	0.0	89.0	39.683	10.109	0.0	91.0
F-score	2.612	2.135	0.0	30.0	2.072	1.910	0.0	30.0
opposite gender	0.137	0.344	0.0	1.0	0.044	0.206	0.0	1.0
set 1	0.251	0.433	0.0	1.0	0.250	0.433	0.0	1.0
set 2	0.251	0.433	0.0	1.0	0.250	0.433	0.0	1.0
set 3	0.250	0.433	0.0	1.0	0.250	0.433	0.0	1.0
set 4	0.249	0.432	0.0	1.0	0.249	0.433	0.0	1.0
pairing 1	0.381	0.486	0.0	1.0	0.359	0.480	0.0	1.0
pairing 2	0.382	0.486	0.0	1.0	0.357	0.479	0.0	1.0
pairing 3	0.237	0.425	0.0	1.0	0.285	0.451	0.0	1.0
ability	120.414	9.740	52.0	149.5	126.288	9.016	50.2	154.4
ability difference	-0.100	9.938	-61.4	63.9	0.032	8.340	-71.9	71.9
team ability	120.456	7.677	73.0	137.4	126.204	7.853	60.3	147.6
opponent team ability	120.538	7.651	61.1	137.4	126.258	7.812	63.9	147.6
Observations	115180				357123			
unique players	965				2290			

Data sources: WKBV.

Table A.2: Robustness checks: different ability measures

	past performance			past & future performance			factors for past performance		
	score	points	F-score	score	points	F-score	score	points	F-score
female	-0.078	0.000	-0.052	0.221	0.000	-0.102***	-0.072	-0.000	-0.053
	(0.210)	(0.006)	(0.036)	(0.153)	(0.005)	(0.029)	(0.210)	(0.006)	(0.036)
opp. gender	-0.096	0.006	0.002	0.045	0.003	-0.008	-0.097	0.006	0.002
	(0.157)	(0.005)	(0.025)	(0.157)	(0.005)	(0.025)	(0.157)	(0.005)	(0.025)
female × opp. gender	0.063	-0.012	-0.055	-0.173	-0.006	-0.020	0.065	-0.011	-0.055
	(0.244)	(0.008)	(0.037)	(0.231)	(0.007)	(0.035)	(0.244)	(0.008)	(0.037)
home	1.565***	0.057***	-0.167***	1.515***	0.053***	-0.161***	1.570***	0.057***	-0.168***
	(0.153)	(0.004)	(0.021)	(0.143)	(0.004)	(0.019)	(0.154)	(0.004)	(0.021)
Constant	22.666***	0.473***	16.223***	-1.657	0.473***	19.667***	122.682***	0.472***	2.512***
	(1.400)	(0.038)	(0.252)	(1.199)	(0.037)	(0.247)	(0.251)	(0.006)	(0.044)
Ability measure	past performance			past & future performance			factors for past performance		
Observations	64738	64738	64727	75056	75056	75042	64738	64738	64727
Adj. R^2	0.364	0.182	0.275	0.388	0.193	0.298	0.364	0.182	0.275

Notes: This table shows the relationship between player's gender and characteristics and the outcomes of interest in mixed gender leagues by different types of ability measures. Models control for *pairing* and *set* (except for *points* as dependent variable). The outcome *score* is the total score per lane; *points* are the set points obtained on one lane (0 if lost, 0.5 if tie, and 1 if won); *F-score* denotes the mistakes, i.e. how often the player did not hit any pin. *Female* and *opp. gender* are dummy variables if the player is female or plays against the opposite gender, respectively. Ability measures include own ability, difference between own and opponent's ability, the team's average ability, and the opponent team's average ability. Ability measures are: *past performance* as the average of past scores, *past & future performance* as the average of past and future scores, and *factors for past performance* from a factor analysis of past *V-score* & *A-score*. Models control for *set* and *pairing* (except for points as dependent variable), and *home*. Robust standard errors clustered at the level of the player are in parentheses. ***, ** and * denote significance at the 1%, 5% and 10% level, respectively. Data source: WKBV.

Table A.3: Robustness checks: different fixed effects

	Individual FE			Lane FE			Individual × Lane FE		
	score	points	F-score	score	points	F-score	score	points	F-score
female × opp. gender	-0.323*	-0.009	-0.013	0.035	-0.012	-0.043	-0.139	-0.002	0.009
	(0.176)	(0.006)	(0.027)	(0.234)	(0.008)	(0.037)	(0.260)	(0.009)	(0.041)
opp. gender				-0.078	0.006	0.004			
				(0.152)	(0.005)	(0.025)			
home	1.557***	0.056***	-0.161***	1.559***	0.057***	-0.166***	0.529	0.012	-0.050
	(0.158)	(0.004)	(0.021)	(0.130)	(0.004)	(0.020)	(0.441)	(0.015)	(0.070)
past ability	0.462***	-0.008***	-0.071***	0.831***	-0.000	-0.110***	0.553***	-0.006***	-0.079***
	(0.048)	(0.001)	(0.007)	(0.015)	(0.000)	(0.003)	(0.061)	(0.002)	(0.010)
difference ability	-0.009	0.015***	0.000	-0.003	0.015***	-0.001	-0.008	0.015***	-0.000
	(0.009)	(0.000)	(0.001)	(0.008)	(0.000)	(0.001)	(0.013)	(0.000)	(0.002)
team ability	0.156***	0.003***	-0.019***	0.028*	0.001**	-0.006**	0.146***	0.002**	-0.019***
	(0.027)	(0.001)	(0.004)	(0.016)	(0.000)	(0.003)	(0.037)	(0.001)	(0.006)
opponent team ability	0.031**	-0.001*	-0.004	-0.035***	-0.001***	0.003	-0.010	-0.001*	-0.000
	(0.015)	(0.000)	(0.002)	(0.013)	(0.000)	(0.002)	(0.022)	(0.001)	(0.004)
female				-0.057	0.000	-0.061*			
				(0.213)	(0.006)	(0.035)			
Constant	39.926***	1.135***	14.358***	20.140***	0.472***	16.527***	35.970***	0.999***	14.889***
	(5.294)	(0.120)	(0.797)	(1.529)	(0.044)	(0.271)	(6.404)	(0.172)	(1.129)
Fixed effects	Individual FE			Lane FE			Individual × Lane FE		
Observations	64738	64738	64727	64046	64046	64035	64046	64046	64035
Adj. R^2	0.391	0.196	0.304	0.384	0.185	0.284	0.420	0.200	0.317

Notes: This table shows the relationship between player's gender and characteristics and the outcomes of interest in mixed gender leagues by different types of fixed effects. Models control for *pairing* and *set* (except for *points* as dependent variable). The outcome *score* is the total score per lane; *points* are the set points obtained on one lane (0 if lost, 0.5 if tie, and 1 if won); *F-score* denotes the mistakes, i.e. how often the player did not hit any pin. *Female* and *opp. gender* are dummy variables if the player is female or plays against the opposite gender respectively. *Past ability* is the average *score* of the player per lane if more than 8 lanes are observable from past data. *Difference* is the difference between *past ability* of the player and her opponent. *Team ability* and *opponent team ability* are measures for team's quality, they are calculated by the average of *past ability* of other players in the team. *Individual FE* are fixed effects for each player, *Lane FE* are fixed effects for each lane in each location, and *Individual x Lane FE* is a combination of both. Models control for *set* and *pairing* (except for points as dependent variable), and *home*. Robust standard errors clustered at the level of the player are in parentheses. ***, ** and * denote significance at the 1%, 5% and 10% level, respectively. Data source: WKBV.