Prof. Dr. Joachim Grammig

Department of Econometrics, Statistics and Empirical Economics, University of Tübingen

The Home Advantage and its Channels in German Professional Table Tennis: A Logistic Regression Approach

"Bachelor Seminar in Empirical Economics"

Summer Term 2022

Robin Gligorov
Grossgartacher Str. 110
74080 Heilbronn
robin.gligorov@student.uni-tuebingen.de
Student number: 4256850
Field of study: Economics and Business Administration
Date: 08.07.2022

Contents

List of Figures

List of Tables

1 Introduction

In all kinds of sports, home advantage is often named by experts as the possible cause, for the victory of the home team or player in a specific game. Phrases concerning the channels of home advantage such as "the spectators have pushed them to the win" or "the home player was better used to the local circumstances" are common in post-match analyses. However, there exists only limited literature analysing home advantage and its causal channels empirically, especially in table tennis. Klein-Soetebier et al. (2014) conducted a first analysis of the home advantage for the German Men's Table Tennis Bundesliga with space left for improvement. My bachelor thesis will try to improve upon previous papers with an extensive analysis of the home advantage in table tennis and its main channels. For this purpose, I will set up logistic regression designs to explain winning probabilities by several factors. Firstly, I will model the overall home advantage. Secondly, I will model the main channels of the home advantage, spectators and different ball brands used by the competing players in their home games. As team success may attract spectators, I will try to control for such a bias in the spectators channel. Further, I will model the different brands channel and the impact of the ball material change from celluloid to plastic on this channel. The argument is that plastic balls differ more from each other than celluloid balls do, so the away player has to make greater adjustments when not being familiar with the specific ball. The paper is structured as follows. There will be a short literature review and theory part before taking a detailed look at the econometric model specifications, the data and the results of the empirical analysis.

2 Main Part

2.1 Literature Overview

In most of the papers in academic sports literature, home advantage is defined as "the consistent finding that home teams in sports competitions win over 50% of the games played under a balanced home and away schedule" Courneya & Carron (1992), p. 1. This is rather a definition of what I call "home win share under a balanced schedule" than a causal advantage of playing at home. This definition only holds under specific econometric conditions, which I will discuss explicitly later on. Koning (2011) suggests a better definition including the causality that comes to one's mind when hearing "home advantage": "home advantage is the performance advantage of an athlete, team or country when they compete at a home ground compared to their performance under similar conditions at an away ground" Koning (2011), p. 2. As already emphasized, my thesis is mainly based on Klein-Soetebier et al. (2014). Their paper attempts to estimate the home advantage in the German Men's Table Tennis Bundesliga for the seasons 2008/2009 until 2012/2013. Using the definitions from above, the authors estimated the home win share rather than the home advantage. The estimated home win share in the paper is 51.48%. This interprets as a home advantage of $1.48\%p$, which is not statistically significant after a paired-sample t-test. This approach was also used in every other empirical paper concerning home advantage in table tennis. Klein-Soetebier et al. (2015) find a statistically significant effect of around 3% for the 1000 best German table tennis players. Paar et al. (2019a) estimate a statistically significant effect for the Austrian Bundesliga of around 5%. Paar et al. (2019b) do not find a significant estimate of the home advantage for Hungarian players in individual table tennis competitions. In essence, there are mixed findings in table tennis all using the same univariate estimation technique. Further, Klein-Soetebier et al. (2014) try to estimate an effect of spectators on the home advantage by running a what they call "correlation analysis"

Klein-Soetebier et al. (2014), p. 75 (in Sportwissenschaft, 44(2)) between the points won on the team-level of a table tennis match and the number of spectators. They state an R-squared of 16.2% and call it a "medium- large effect" Klein-Soetebier et al. (2014), p. 75 (in Sportwissenschaft, 44(2)). The authors rightly address the issue of possible omitted variable bias, given the argument that successful teams attract more spectators than less successful ones. In my bachelor thesis, I will control for the quality of the competing players and thus try to estimate an unbiased spectators effect. Regarding the influence of the ball material change from celluloid to plastic on the home advantage, no previous literature exists.

2.2 Theory on the Channels of Home Advantage

From a theoretical point of view, there are two main channels of home advantage in the table tennis Bundesliga: - Crowd support - Familiarity with local circumstances. Looking at the first main channel, it is not clear whether a large number of spectators is an advantage or a disadvantage for the home player. On the one hand, a large number of spectators could have a positive influence on the home player. The crowd could strengthen the confidence of the home player and simultaneously intimidate the opponent. On the other hand, home players could perform worse in front of the home crowd due to the pressure of disappointing the spectators in the case of a loss. Baumeister (1984) and Strauss (1997) describe this as the choking under pressure phenomenon. The second main channel is clearly an advantage following the argument that the home player is more familiar with the local circumstances, for instance with the ball or with the table. Regarding table tennis balls, something interesting happened in the summer of 2014. In the Bundesliga, the old celluloid balls were abolished and the new plastic balls were introduced. Since table tennis is a very sensitive sport, a change in the material of the balls could have a considerable impact on the game. Inaba et al. (2017) as well Goh & Lee (2022) found out that indeed

the game has changed with the introduction of the new balls. There is a significant reduce in spin and speed of the balls. Further and more importantly, I assume that on average plastic balls produced by two different brands differ more from each other than two celluloid balls produced by two different brands did. As every team can decide which ball brand to use, my hypothesis would be that the change from celluloid to plastic balls have led to an increase in the home advantage due to increasing differences between different ball brands. The away player would have to make more adjustments to the new ball (assuming the away player usually plays with a different ball brand). That being stated, the three main goals of my thesis are to:

1. estimate the causal home advantage instead of just observing the descriptive home win share.

2. estimate an unbiased spectators effect.

3. estimate the influence of the change in ball material on the home advantage via the different brands channel.

In the next section, I will explain the models that I derived to achieve my three main goals.

2.3 Model Specifications and Methods

2.3.1 Modelling the Overall Home Advantage

Based on the definition of home advantage by Koning(2011), I define home advantage as the causal increase in the probability of winning a match of the home player due to the fact he is playing at home. I aim to explain winning probabilities through the following models. My dependent variable W_{it} is a dummy variable, which takes the value 1 if the player won the match and 0 if he lost it. My first goal is to move from the home win share to a causal estimate for the home advantage. Admittedly, setting up

$$W_{it} = \beta_0 + \beta_1 H_{it} + \epsilon_{it} \tag{1}$$

will be a reasonable model for the overall home advantage, as there is arguably no omitted factor that could bias β_1 by being correlated with H_{it}. Noting that H_{it} is a dummy variable, which takes the value 1 if the player plays at home and 0 if he is not. By constructing conditional expected values, the probability of winning at home could be derived as:

$$E(W_{it}|H_{it} = 1) = P(W_{it} = 1|H_{it} = 1) = \beta_0 + \beta_1,$$

as well as the probability of winning away as:

$$E(W_{it}|H_{it} = 0) = P(W_{it} = 1|H_{it} = 1) = \beta_0.$$

The term $\beta_0 + \beta_1$ is the probability of winning at home and the home win share is its counterpart in a finite sample context. β_1 is the difference of the two winning probabilities. Further, $\beta_0 + \beta_1$ and β_0 lie symmetrically around 0.5 as we have a balanced home and away schedule. Due to the symmetry, $\frac{1}{2}\beta_1$ is the causal home advantage. Nevertheless, it would make sense to include further covariates in the estimation later on to ensure that there is no bias and to reduce the variance in the model. If I am able to reject the null hypothesis

$$H_0 : \frac{1}{2}\beta_1 \leq 0,$$

the interpretation would be that the estimated causal effect of playing at home $\frac{1}{2}\hat{\beta}_1$ is significantly larger than 0.

2.3.2 Modelling the Spectators Channel

The second model to set up concerns the influence of spectators on the home advantage. Thus,, the question is whether a higher number of spectators increases the probability of winning of the home player significantly. A first proposal to answer this question could be the model

$$W_{it} = \beta_0 + \beta_1 H_{it} + \beta_2 S_{it} + \beta_3 (H_{it} S_{it}) + \epsilon_{it}. \tag{2}$$

With S_{it} measuring the number of spectators. Conditioning the model on the two possible values of H_{it}, I am able to derive the split of the home advantage into a part contributed by the spectators and another part contributed by all other drivers of the home advantage:

$$E(W_{it}|H_{it}=1) = P(W_{it}=1|H_{it}=1) = \underbrace{\beta_0+\beta_1}_{\text{home constant}} + \underbrace{(\beta_2+\beta_3)S_{it}}_{\text{home spectators effect}}$$

$$E(W_{it}|H_{it}=1) = P(W_{it}=1|H_{it}=0) = \underbrace{\beta_0}_{\text{away constant}} + \underbrace{\beta_2 S_{it}}_{\text{away spectators effect}}$$

$\beta_0+\beta_1$ displays in this model the average winning probability of the away player without the spectators channel. $\beta_0+\beta_1-0.5$ is the home advantage without the effect of the spectators. Noting that β_0 and β_1 are not the identical parameters as in equation 1 due to the extraction of the spectators effect, $(\beta_2+\beta_3)S_{it}$ is the part of the home advantage contributed by the spectators. Analogously, β_0 is the average winning probability of the away player after extracting the spectators effect. Subtracting $\beta_0-0.5$, gives the disadvantage of the away player without the effect of the spectators. $\beta_2 S_{it}$ shows the spectators effect on the probability of winning of the away player. This would be the away disadvantage through the spectators channel. To test the significance of the channels a Wald Test is used. To check if there is still a significant home advantage when excluding the spectators channel, we have to test:

$$H_0 : \frac{1}{2}\beta_1 \leq 0,$$

to check if spectators increase the home advantage we have to test:

$$H_0 : \beta_2+\beta_3 \leq 0, \tag{3}$$

to check if spectators increase the away disadvantage we have to test:

$$H_0 : \beta_2 \geq 0. \tag{4}$$

Having a closer look at the model, it turns out that there is no possibility to distinguish between the advantage of the home player or the disadvantage for the away player. If the

home player won, the away player automatically lost. This leads us to the fact that the home spectators effect has to be equal to the away spectators effect multiplied by -1:

$$\underbrace{(\beta_2 + \beta_3)S_{it}}_{\text{home spectators effect}} = \underbrace{-\beta_2 S_{it}}_{\text{away spectators effect}}$$

Rearranging gives us:

$$\beta_2 = -\frac{1}{2}\beta_3.$$

Inserting this into our null hypothesis, we find out that the null hypothesis stated in equation 3 and the null hypothesis stated in equation 4 are redundant:

$$H_0 : \beta_2 + \beta_3 = -\frac{1}{2}\beta_3 + \beta_3 = \frac{1}{2}\beta_3 \leq 0$$

$$H_0 : \beta_2 = -\frac{1}{2}\beta_3 \geq 0 \Rightarrow \frac{1}{2}\beta_3 \leq 0$$

Thus, what we are searching for as the home spectators effect is $\frac{1}{2}\beta_3$.

The problem is that the spectators effect in the model is biased. As already stated, the reason is that more successful teams with more successful players attract on average a greater number of spectators. Obviously, successful players are more likely to win their matches. Thus the quality of the home player is correlated with his winning probability as well as with the amount of spectators attending. This leads to an overestimation of the spectators effect. In the model the error term and the spectators variable are correlated:

$$Cov(\epsilon_{it}, S_{it}) \neq 0.$$

Therefore the strict exogeneity assumption

$$E(\epsilon_{it}|H_{it}) \neq 0$$

is violated due to omitted variable bias. In order to get rid of the bias, a measure of the quality of the players has to be included in the model, ΔQ_{it}, which is the difference in the

quality of the two players. The proxy used for ΔQ_{it} will be explained explicitly in the data section of this paper. A better suited model is

$$W_{it} = \beta_0 + \beta_1 H_{it} + \beta_2 S_{it} + \beta_3 (H_{it} S_{it}) + \gamma \Delta Q_{it} + \epsilon_{it}. \tag{5}$$

After pulling the quality measure out of the error term, there should be no correlation left between ϵ_{it} and S_{it}. Therefore $\frac{1}{2}\beta_3$ is this time the unbiased spectators effect. The hypotheses and the interpretations of the coefficients are the same as in the previous model.

2.3.3 Modelling the Different Ball Brands Channel

The argument is that the home advantage increased due to the change from celluloid to plastic balls via the different ball brands channel assuming that plastic balls differ more between different brands. First, I set up a model concerning the influence of different ball brands used on the home advantage without a distinction between celluloid and plastic balls. Therefore I need to introduce B_{it} which is also a dummy variable taking the value 1 in the case both teams of the two players use the same ball brand in their home matches and 0 if they use different ball brands. The model looks as follows:

$$W_{it} = \beta_0 + \beta_1 H_{it} + \beta_2 B_{it} + \beta_3 (H_{it} B_{it}) + \epsilon_{it} \tag{6}$$

Setting up conditional expected values gives me:

$$E(W_{it}|H_{it}=1, B_{it}=1) = P(W_{it}=1|H_{it}=1, B_{it}=1) = \underbrace{\beta_0 + \beta_1}_{\text{home constant}} + \underbrace{(\beta_2 + \beta_3)}_{\text{home brands effect}}$$

$$E(W_{it}|H_{it}=1, B_{it}=1) = P(W_{it}=1|H_{it}=0, B_{it}=1) = \underbrace{\beta_0}_{\text{away constant}} + \underbrace{\beta_2}_{\text{away brands effect}}$$

Very similar to the model specification for the spectators effect, the relationship $\beta_2 = -\frac{1}{2}\beta_3$ holds here too for the same reason. The home advantage is split up into a part without the different ball brands effect, namely $\beta_0 + \beta_1 - 0.5$ or $\frac{1}{2}\beta_1$ and into the part which is explained by the use of different ball brands, namely $\beta_2 + \beta_3$ or $\frac{1}{2}\beta_3$. The hypotheses to

test are analogous to those of the spectators channel:

$$H_0 : \frac{1}{2}\beta_1 \leq 0,$$

$$H_0 : \frac{1}{2}\beta_3 \leq 0.$$

The next step would be to also account for the information about the used ball materials. Therefore, I need to introduce P_t which is also a dummy variable taking the value 1 in case the match took place in the plastic ball period and 0 if it took place in the celluloid ball period. To model the influence of the ball material change, I need to interact P_t with the model from above in the following way:

$$W_{it} = \beta_0 + \beta_1 H_{it} + \beta_2 B_{it} + \beta_3 P_t + \beta_4 (B_{it} P_t) + \beta_5 (H_{it} B_{it}) + \beta_6 (H_{it} P_t) + \beta_7 (H_{it} B_{it} P_t) + \epsilon_{it} \tag{7}$$

I need to include all possible interactions of the dummies as well as the dummies themselves. With the help of this model, one can answer many question concerning home advantage, different ball brands and different ball material. I am predominantly interested in testing my hypothesis. That is in testing whether the change from celluloid to plastic in the table tennis ball material increased the probability of winning of the home player via the different brands channel. When it comes to my hypothesis, one has to be careful not to switch things up. There is a small but crucial difference between an increase in the home advantage with the introduction of plastic balls given different ball brands versus due to different ball brands. It will become more clear after I have presented the model in full. I will now derive the change in the effect of the home advantage with the introduction of plastic balls given different ball brands. Then, I will derive the effects via the channels and finally I will explain the differences explicitly. To derive the effect given different brands, one has to construct the following conditional expected values and subtract them from each other:

$$E(W_{it}|H_{it} = 1, B_{it} = 1, P_t = 1) = \beta_0 + \beta_1 + \beta_2 + \beta_3 + \beta_4 + \beta_5 + \beta_6 + \beta_7 \tag{8}$$

$$E(W_{it}|H_{it}=1,B_{it}=1,P_t=0)=\beta_0+\beta_1+\beta_2+\beta_5 \tag{9}$$

$$E(W_{it}|H_{it}=1,B_{it}=1,P_t=1)-E(W_{it}|H_{it}=1,B_{it}=1,P_t=0)=\beta_3+\beta_4+\beta_6+\beta_7 \tag{10}$$

One has to model the probability of the home player given the away player uses a different ball brand in his home matches for the plastic ball period. That is equivalent to equation 8. The same has to be modeled for the celluloid ball period expressed by equation 9. The effect of the ball material change given $B_{it}=1$ is the difference in the probabilities of winning between these two periods, given $H_{it}=1$ and $B_{it}=1$. In other words, it is the additional probability of winning of the home player when changing the ball material to plastic for the case of the usual use of different ball brands by the two players. This effect is represented in equation 10 by $\beta_3+\beta_4+\beta_6+\beta_7$. Counter to what one could think initially, it does not represent the pure effect via the different brands channel, as β_3 and β_6 have no direct relation to the brand information. In the subtraction, we have to set P_t once equal to 1 and once equal to 0. The problem of doing so in this state of the model is that β_3 and β_6 jump in and stay.[1] In order to eliminate β_3 and β_6 we have to go even deeper.

This leads us to the pure effect of the change to plastic balls via different brands used by the two players. To follow this up, we have to specify the model once for each period.

For the plastic ball period:

$$E(W_{it}|P_t=1,H_{it}=h_{it},B_{it}=b_{it})=\underbrace{\beta_0+\beta_3+(\beta_1+\beta_6)}_{\text{home constant}}h_{it}+b_{it}\underbrace{(\beta_2+\beta_4)+(\beta_5+\beta_7)}_{\text{home different brands effect}}(h_{it}b_{it}) \tag{11}$$

For the celluloid ball period:

$$E(W_{it}|P_t=0,H_{it}=h_{it},B_{it}=b_{it})=\underbrace{\beta_0+\beta_1}_{\text{home constant}}h_{it}+\underbrace{(\beta_2+\beta_5)}_{\text{home different brands effect}}(h_{it}b_{it}) \tag{12}$$

[1] When constructing the expected value in equation 8, β_3 and β_6, as wells as all other coefficients of the model are present in that term. Subtracting equation 9 from equation 8, all here non-relevant coefficients drop out except β_3 and β_6. This is illustrated in equation 10.

Above the home advantage is again split up into a part which describes the different ball brand channel and into a part which describes all other drivers of the home advantage. This is the case for both periods. Thus $\beta_2 + \beta_4 + \beta_5 + \beta_7$ in equation 11 is the different ball brand effect for the plastic ball period and $\beta_2 + \beta_5$ in equation 12 is the one for the celluloid ball period. Taking the difference of these two terms brings us finally to the pure effect of the ball material change via the different brands channel on the home advantage, namely $\beta_4 + \beta_7$. Analogously to the cases above, there is no distinction possible between home advantage or away disadvantage. This leads to the following relationships of the coefficients in the model:

$$\beta_2 = -\frac{1}{2}\beta_5,$$

$$\beta_3 = -\frac{1}{2}\beta_6.$$

$$\beta_4 = -\frac{1}{2}\beta_7.$$

The hypothesis to test is for the remaining home advantage

$$H_0 : \frac{1}{2}\beta_1 \leq 0,$$

for the home advantage given $B_{it} = 1$ and $P_t = 1$

$$H_0 : \frac{1}{2}\beta_1 + \frac{1}{2}\beta_5 + \frac{1}{2}\beta_6 + \frac{1}{2}\beta_7 \leq 0,$$

for the home advantage given $B_{it} = 1$ and $P_t = 0$

$$H_0 : \frac{1}{2}\beta_1 + \frac{1}{2}\beta_5 \leq 0,$$

for the difference of the two above

$$H_0 : \frac{1}{2}\beta_6 + \frac{1}{2}\beta_7 \leq 0,$$

for the home advantage via the different brands channel given $P_t = 1$

$$H_0 : \frac{1}{2}\beta_5 + \frac{1}{2}\beta_7 \leq 0,$$

for the home advantage via the different brands channel given $P_t = 0$,

$$H_0 : \frac{1}{2}\beta_5 \leq 0.$$

and for testing if the home advantage via the different brands channel increased with the introduction of the plastic balls

$$H_0 : \frac{1}{2}\beta_7 \leq 0.$$

2.3.4 Estimation Techniques

As my dependent variable is binary, a logistic regression model is more suitable in my case than setting up a linear probability model. The advantage of a logistic regression model versus a linear probability model is that the logistic regression model accounts for the fact that the predicted probabilities cannot be smaller than 0 and larger than 1. The model will be estimated by the Maximum Likelihood Method. The Maximum Likelihood Estimator has the helpful properties of being consistent, asymptotically efficient and asymptotically normally distributed. Especially the last one ensures us to get reliable test statistics. Although logistic regression models assume by nature heteroscedasticity, it might be reasonable to also use robust standard errors due to the panel structure of the models. Of course, the exogeneity assumption has to hold for each particular model. As the results of logistic regression estimations are unintuitive odds ratios, I will compute marginal effects to interpret the results.

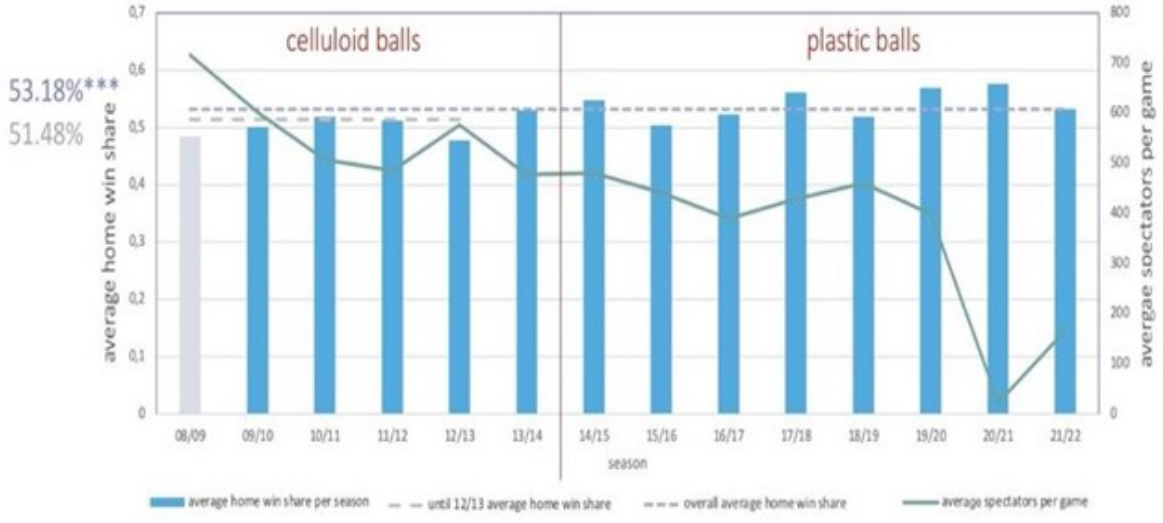
2.4 Data

The data I am using for my bachelor thesis was collected by myself by applying the technique of web scraping. It is an extension of the data used in Klein-Soetebier et al (2014). My algorithm automatically clicked itself through the databases of www.mytischtennis.de and collected the relevant information for my research. My data is an extension of the time dimension as well as of the number of variables. Klein-Soetebier et al. (2014) used data of

the German Men's Table Tennis Bundesliga for the seasons 2008/2009 until 2012/2013. I extend this to the season 2021/2022. I am using data on the singles level and not on the team level. Klein-Soetebier et al. (2014) included both levels, but the singles level allows for deeper insights and is better suited for my approaches. As there was the Covid-19 pandemic during the time span and as I am predominantly interested in estimating effects for normal seasons, I created G_t which takes the value 1 in the case of a so-called ghost game without any spectators attending, to be able to control for the effects of ghost games by only including observations where $G_t = 0$. The authors' data includes the result of the match and the number of spectators attending. I am of course using these two two variables as well. S_{it} is measuring the number of spectators and W_{it} the result of the match. Additionally, since I criticized the estimation of the spectators effect in Klein-Soetebier et al. (2014), I am using so called TTR values as a proxy for the quality of the home and the away player Q_{it}. TTR value stands for table tennis ranking value. Every table tennis player has a TTR value, which gets updated after each match. So the TTR value decreases after a loss and increases after a win. The size of the increase or decrease depends on the difference in the quality of the two opponents. Moreover, the difference in the TTR values ΔTTR_{it} is probably the best predictor for the outcome of the match, so it should definitely be included in the analysis. As already highlighted, I am also using data on ball material. B_{it} measures whether the clubs of the competing players use different balls in their home games or not. Unfortunately, data on the used ball brands were only available for the plastic ball period. I assumed that the clubs did not change the ball brand or their sponsor, in order to have ball branding data in the celluloid period too. P_t tells us whether the match was played in the plastic ball period or not. I have created further variables for efficiency in my estimation, which could all be relevant predictors of the winning probability of the player. First, I created the variable $nationality_{it}$ which takes the value 1 if the player we are looking at is German and the opponent not, 0 if either both are Germans or both

are foreigners and -1 if the opponent is German and the player not. The argument is that players could overperform relative to their playing level in their home country, as they could feel more comfortable than foreigners in the country and have probably on average more experience in the Bundesliga. I have created the other variables to maximize efficiency from my raw web scraped data. $momentum_{it}$ measures the team momentum, which could also have an influence on the performance of the player. It measures how many matches were won in a row by the team of the player we are looking at. The values are positive if the team was successful before and are negative if the teammates lost before. I have also created a variable measuring the difference in the lagged dependent variable W_{it} of the two players and named it ΔW_{it-1}. Further, $\Delta winstreak_{it}$ measures the difference in the winning or losing streaks of the two players. Lastly, $score_{it}$ measures the overall team score before the singles match. In total I have 4552 singles matches in my data set, after dropping those with missing information. The structure of the panel data set is very special. My raw data set in this state is from the perspective of the home player. In order to have variation in my home dummy H_{it}, each singles match shows up twice in the data set. Once from the perspective of the home player and once from the perspective of the away player. Therefore my final data set consists of 9104 rows. To get the away perspective, I have to multiply the variables ΔTTR_{it}, $nationality_{it}$, $momentum_{it}$, ΔW_{it-1}, $\Delta winstreak_{it}$ and $score_{it}$ by -1. The home dummy H_{it} and its interactions take the value 0 for the away perspective. The following graph gives a descriptive overview and raises all the main questions of my bachelor thesis.

Figure 1: Descriptive Overview



The graph shows the development of the home win shares, the number of spectators attending and the material of the ball over the seasons. Simultaneously, the home win shares rather increased, the number of spectators decreased and the plastic balls were introduced in the summer of 2014. Firstly, I replicated the result of Klein-Soetebier et al. (2014) and arrived at the identical home win share over the first six seasons in the data.[2] The home win share amounts to 51.48% and is statistically insignificant after a paired sample t-test. The home win share over the whole time span is equal to 53.18% and suddenly statistically significant at a 1% significance level. Is the reason for this the decline in the number of spectators and is this evidence for the choking under pressure phenomenon in German professional table tennis? Also, the fact that in the season when the pandemic hit one could observe the largest home win share is a hint for presence of this phenomenon. Or does my hypothesis hold and this is due to the change in the ball material? Or both? Descriptive statistics come to their end here and we need econometric methods to identify the causal effects here.

[2]The column for the season 2008/2009 is faded in figure 1, as there was only data on few matches available for this season

2.5 Results of the Empirical Analysis

Starting right away with the results, the logistic regression estimation for the marginal causal effect of the home advantage for the whole time span amounts to significant $2.787\%p$, when excluding ghost games. There is no spectators effect, as the significance and the size of it disappear when including TTR values in the model. There is a significant effect on the home advantage via the different brands channel in the plastic ball period of and no significant effect in the celluloid ball period. The difference of it is not significant, so I have to reject my hypothesis. Whereas, there is a significant marginal effect given different ball brands on the home advantage in the plastic ball period, in the celluloid ball period and in the difference between these two periods as well. It seems that the choking under pressure phenomenon is present for the Bundesliga as the home advantage causally increases in the case of a ghost match.

2.5.1 The Overall Home Advantage and the Spectators Channel

The following table shows logisitc regression results which are mainly based on the models set up in equation 1, equation 2 and equation 5.

Table 1: Regression Results: Overall Home Advantage & Spectators Channel

	(1)	(2)	(3)	(4)	(5)
H_{it}	0.0638***	0.0553***	0.0118	0.0573***	0.0514***
S_{it}			-0.00006**	-0.0000054	-0.0000036
$H_{it}S_{it}$			0.00012***	0.000011	0.000072
ΔTTR_{it}		0.0021***		0.0022***	0.0022***
G_{it}					-0.0535**
$H_{it}G_{it}$					0.1062***
$\frac{1}{2}H_{it}$	0.0319***	0.02765***	0.0059	0.02787***	0.0257***
$\frac{1}{2}H_{it}S_{it}$			0.00006***	0.0000054	0.0000036
$\frac{1}{2}H_{it}G_{it}$					0.0535***

Continued on next page...

... table 1 continued

Other Covariates	No	Yes	No	Yes	Yes
Ghost matches	No	No	No	No	Yes
Observations	4152	4152	4152	4152	4552
Pseudo-R^2	0.0029	0.1623	0.0044	0.1624	0.1734

This table reports average marginal effects of logistic regression estimations via Maximum Likelihood.
When computing marginal effects, the constant drops out.
Nevertheless, the constant is per construction equal to $-\frac{1}{2}\beta_1$.
Significance levels of 1%, 5% and 10% are indicated by ***, ** and * respectively.
In the second segment of the table, it is tested for significance with one-sided Wald tests.

As the coeffients alone are not able to answer the main questions, we have to test our hyptheses instead. Thus the second segment in the table is the relevant one. In column 1 and 2 we see that the overall home advantage estimation drops slightly from $3.19\%p$ to $2.77\%p$ when including covariates. It still remains statistically significant at a 1% level in both cases. Nevertheless, the very low Pseudo-R^2 of 0.29% in column 1 tells us that playing at home does only explain a very small part of the variance of W_{it}. The Pseudo-R^2 increases to 16.23% when including ΔTTR_{it} and the covariates explained in the data section. The analysis of the spectators channel is a perfect illustration of how omitted variable bias could lead to wrong conclusions. The average marginal spectators effect times the standard deviation of the spectators (=383.92) went from $2.30\%p$ down to $2.07\%p$, after controlling for ΔTTR_{it}. More importantly, the Wald Statisitc turned insignificant on all common significance levels. This means that we are not able to reject the null hypothesis that the spectators effect is ≤ 0. Further, we have a large marginal effect for ghost matches of 5.35%, which is statistically significant at a 1% significance level. This could be evidence of the choking under pressure phenomenon. The ghost matches also took place during the plastic ball period, which could lead to an overestimation of the ghost match effect here. This leads us to the effects of different brands and the ball material change on it.

2.5.2 The Different Brands Channel

The following table shows logisitc regression results which are mainly based on the models set up in equation 6 and equation 7.

Table 2: Regression Results: Different Brands Channel

	(1)	(2)	(3)
H_{it}	-0.0162	-0.0232	-0.0239
B_{it}	-0.0397*	-0.0274	-0.0273
P_t		-0.0056	-0.0104
$B_{it}P_t$		-0.0182	-0.0129
$H_{it}B_{it}$	0.0797**	0.0549	0.0549
$H_{it}P_t$		0.0113	0.0208
$H_{it}B_{it}P_t$		0.0365	0.0258
G_{it}			-0.0432*
$H_{it}G_{it}$			0.0859***
$\frac{1}{2}H_{it}$	-0.0081	-0.0116	-0.0120
$\frac{1}{2}H_{it} + \frac{1}{2}H_{it}B_{it} + \frac{1}{2}H_{it}P_t + \frac{1}{2}H_{it}B_{it}P_t$		0.0396***	0.0387***
$\frac{1}{2}H_{it} + \frac{1}{2}H_{it}B_{it}$		0.0158**	0.0154**
$\frac{1}{2}H_{it}P_t + \frac{1}{2}H_{it}B_{it}P_t$		0.0238**	0.0233**
$\frac{1}{2}H_{it}B_{it} + \frac{1}{2}H_{it}B_{it}P_t$		0.0456**	0.0402**
$\frac{1}{2}H_{it}B_{it}$	0.0397***	0.0274	0.0273
$\frac{1}{2}H_{it}B_{it}P_t$		0.0182	0.0129
$\frac{1}{2}H_{it}G_{it}$			0.0432***
Covariates	Yes	Yes	Yes
Ghost matches	No	No	Yes
Observations	4152	4152	4552
Pseudo-R^2	0.1629	0.1634	0.1742

This table reports average marginal effects of Logistic Regression Estimations via Maximum Likelihood.
When computing marginal effects, the constant drops out.
Nevertheless, the constant is per construction equal to $-\frac{1}{2}\beta_1$.
Significance levels of 1%, 5% and 10% are indicated by ***, ** and * respectively.
In the second segment of the table, it is tested for significance with one-sided Wald Tests.
ΔTTR_{it} is here included in the covariates.

Looking at the results in table 2, there is a significant home advantage via the different

brands channel of approximately $3.97\%p$. Splitting the effect up into the different ball material periods, but keeping the general ball material effects on the home advantage, the results from column 2 show a significant effect for both periods. 3.96% for the plastic ball period at a 1% significance level and $1.58\%p$ for the celluloid ball period at a 5% significance level. The difference of the two effects of 2.38% is also statistically significant at a 5% significance level. Estimating the home advantage via the different brands channel once for each period, we get a significant effect of $4.56\%p$ for the plastic ball period at a 5% significance level and an insignificant effect of $2.73\%p$. The difference of these two effects of $1.29\%p$ is insignificant. Thus, although there is a significant effect in the plastic ball period and no significant effect in the celluloid ball period, I am not able to reject the null hypothesis that the home advantage via the different brands channel decreased or stayed constant when changing the ball material from celluloid to plastic. I have found evidence for the home advantage via the different brands channel, but no evidence for the increase of it due to the change in the ball material. Also, there is still a significant ghost match effect on the home advantage of $4.32\%p$ after controlling for ball brands and ball material effects.

The results of this paper are internally valid under the assumption of exogeneity, such that no further unobserved exogenous shocks occurred or omitted variables exist correlated with any of my regressors as well as with the dependent variable. Moreover, the results are externally valid for any other professional table tennis league carried out under a similar system.

3 Conclusion

In this paper, I have tested for the presence of the home advantage in professional German men's table tennis and the impacts of spectators and ball brands on it. Home advantage

was defined as the causal increase in the winning probability of playing at home. The main findings are as follows. I have found a significant overall home advantage for the season 2008/2009 to 2021/20022 in the Bundesliga. Considering the home advantage through the spectators channels , it turns out that after controlling for the relative quality of the players to avoid bias since more successful teams attract more spectators, the number of spectators has no impact on the home advantage. There is even evidence for a choking under pressure phenomenon as the home advantage significantly increased during ghost matches in the Covid-19 pandemic. Further research could aim to find and implement a more precise measure for reducing the bias in the spectators effect. One could argue that the absolute success of the home and the away team are more strongly correlated with the number of spectators than the relative qualities of the players. Moreover, one could try to replace the total amount of spectators with the share of home and away supporters to get deeper insights there. In contrast to the spectators channel, the use of different ball brands of the two players in their home matches significantly increases the home advantage. Even though there is a significant effect on the home advantage via different ball brands in the plastic ball period and in the celluloid ball period not, I found no evidence for my hypothesis, that this effect via the different ball brands channel increased by the ball material change from celluloid to plastic in the summer of 2014.

References

[1] Baumeister, R. F. (1984). Choking under pressure: self-consciousness and paradoxical effects of incentives on skillful performance. Journal of personality and social psychology, 46(3), 610.

[2] Courneya, K. S., & Carron, A. V. (1992). The home advantage in sport competitions: a literature review. Journal of Sport & Exercise Psychology, 14(1).

[3] Goh, W. X., & Lee, M. J. (2022). Impact of ball material change from celluloid to plastic on game statistics in elite women table tennis. International Journal of Performance Analysis in Sport, 22(1), 174-182.

[4] Inaba, Y., Tamaki, S., Ikebukuro, H., Yamada, K., Ozaki, H., & Yoshida, K. (2017). Effect of changing table tennis ball material from celluloid to plastic on the post-collision ball trajectory. Journal of human kinetics, 55(1), 29-38.

[5] Klein-Soetebier, T., Senff, A., & Weigelt, M. (2014). Kein Heimvorteil in der 1. Tischtennisbundesliga der Herren. Sportwissenschaft, 44(2), 71-77.

[6] Klein-Soetebier, T. & Veenhuis, K. (2015). Der Heimvorteil im Individualsport - Eine Untersuchung der 1000 besten Tischtennisspieler /-innen Deutschlands, `https://www.mytischtennis.de/public/imagedata/Trainerbrief_Heimvorteil__der__1000__besten__TT-Spieler__Deutschlands.pdf`

[7] Koning, R. H. (2011). Home advantage in professional tennis. Journal of Sports Sciences, 29(1), 19-27.

[8] Paar, D., & Laczko, T.(2019b). Absence of home-advantage in table tennis individual competitions. In PROCEEDINGS BOOK OF THE 16th ITTF SPORTS SCIENCE CONGRESS, 325-332.

[9] Paar, D., Lindner, A. & Laczko, T.(2019a). Home-advantage in table tennis league competitions. In PROCEEDINGS BOOK OF THE 16th ITTF SPORTS SCIENCE CONGRESS, 315-324.

[10] Strauss, B. (1997). Choking under pressure: Positive public expectations and performance in a motor task. Zeitschrift fur Experimentelle Psychologie: Organ der Deutschen Gesellschaft fur Psychologie, 44(4), 636-655.

Eidesstattliche Erklärung

Ich versichere, dass ich die Arbeit ohne fremde Hilfe und ohne Benutzung anderer als der angegebenen Quellen angefertigt habe und dass die Arbeit in gleicher oder ähnlicher Form noch keiner anderen Prüfungsbehörde vorgelegen hat und von dieser als Teil einer Prüfungsleistung angenommen wurde. Alle Ausführungen, die wörtlich oder sinngemäß übernommen wurden, sind als solche gekennzeichnet.

Tübingen, den 08. Juli 2022