



## Why buy a ticket?

The demand for football in Finland and its relationship with social media activity

Jan Alexander Hölttä  
Master's Thesis  
University of Helsinki  
Faculty of Social Sciences  
Economics  
4/20



Tiedekunta – Fakultet – Faculty Faculty of Social Sciences		Koulutusohjelma – Utbildningsprogram – Degree Programme Department of Political and Economic Studies	
Tekijä – Författare – Author Jan Alexander Hölttä			
Työn nimi – Arbetets titel – Title Why buy a ticket? The demand for football in Finland and its relationship with social media activity			
Oppiaine/Opintosuunta – Läroämne/Studieinriktning – Subject/Study track Economics			
Työn laji – Arbetets art – Level Master's Thesis		Aika – Datum – Month and year April 2020	Sivumäärä – Sidoantal – Number of pages 44
Tiivistelmä – Referat – Abstract <p>Professional sports have been of interest to economists in growing proportions since the 1950s, due to the rather straightforward nature of the measurement of performance and the easy access to data. In Finland, football has only lately been making its way towards the limelight of large audiences. To react to this surge of interest, the national football authorities opted for a large revamp of the league system between the 2018 and the 2019 seasons. The reform was welcomed by both partakers and the public, and the Finnish football boom continued with the help of positive results achieved by the national team.</p> <p>This thesis seeks to determine the factors that form the total demand of football in the Finnish Veikkausliiga with a special emphasis on the possible relationship with social media activity. The research question of this thesis is the following: what factors are significant in determining the demand of football, and does social media play a role in it? The research question is approached from the point of view of previous literature and an empirical part. Data about the matches of the 2019 season and the factors that potentially had relationships with their demand has been gathered from various sources, following closely procedures used in the relevant literature. The data is modeled with three OLS regressions involving different choices of variables.</p> <p>The regression results support earlier findings in the field, implying statistically significant relationships between the home team's lagged attendance from the previous season, the home team's position in the league table, the uncertainty of the result, bad weather, and the attendances in the 2019 season. Additionally, the self-constructed variable – based on the existing literature – depicting social media activity is found to have a statistically significant relationship with attendance demand, although no conclusions can be drawn on causality.</p>			
Avainsanat – Nyckelord – Keywords Football, Attendance, Social Media, Demand of Sports, OLS			
Ohjaaja tai ohjaajat –Handledare – Supervisor or supervisors Roope Uusitalo			



Tiedekunta – Fakultet – Faculty Valtiotieteellinen tiedekunta		Koulutusohjelma – Utbildningsprogram – Degree Programme Taloustieteen maisteriohjelma	
Tekijä – Författare – Author Jan Alexander Hölttä			
Työn nimi – Arbetets titel – Title Why buy a ticket? The demand for football in Finland and its relationship with social media activity			
Oppiaine/Opintosuunta – Läroämne/Studieinriktning – Subject/Study track Taloustiede			
Työn laji – Arbetets art – Level Pro Gradu - tutkielma		Aika – Datum – Month and year Huhtikuu 2020	Sivumäärä – Sidoantal – Number of pages 44
Tiivistelmä – Referat – Abstract <p>Ammattiurheilu on ollut akateemisen tutkimuksen kohteena kasvavissa määrin 1950-luvulta alkaen, sillä sen alakohtaiset erityispiirteet sekä suoritusdatan avoimuus luovat ekonomisteille ystävällisen tutkimusympäristön. Suomessa jalkapallo on vasta viime vuosina tehnyt tuloaan suurimmalle estradille katsojien keskuudessa, ja kausien 2018 sekä 2019 välillä kotimaiset jalkapalloauktoriteetit päättivät mullistaa sarjajärjestelmän. Muutos koettiin yleisesti onnistuneeksi, ja jalkapallobuumi sai lisää nostetta alleen maajoukkueen menestyksen avittamana.</p> <p>Tämä tutkielma pyrkii selvittämään, mistä jalkapallon kysyntä koostuu Suomen Veikkausliigassa ja mitkä näiden muuttujien painoarvot ovat lopputuloksen kannalta. Tutkimuskysymyksen erityinen painopiste on sosiaalisen median mahdollisessa relevanssissa. Tutkimuskysymystä on tutkielmassa lähestytty sekä aiemman aihetta käsitelleen kirjallisuuden, että itse kehitetyn empiirisen osion avulla. Tutkimusta varten on kerätty dataa kauden 2019 otteluista ja niiden kysyntään mahdollisesti vaikuttaneista tekijöistä, siinä missä aiempi kirjallisuus on tälle antanut syytä. Dataa mallinnetaan kolmella OLS-regressiolla eri muuttujakokoonpanoilla.</p> <p>Regressiotulokset antavat tukea aiemmin esitettyjen teorioiden johtopäätöksille siitä, että merkittäviä vaikutuksia yleisömääriin on kotijoukkueen edelliskauden yleisökeskiarvolla, kotijoukkueen sijainnilla sarjataulukossa, ottelun tuloksen ennalta-arvaamattomuudella ja huonolla säällä. Tämän lisäksi aiempaan kirjallisuuteen pohjaava, mutta itse kehitetty sosiaalista mediaa kuvaava muuttuja todetaan tilastollisesti merkittäväksi, joskaan kausaalisuudesta ei voida vetää johtopäätöksiä.</p>			
Avainsanat – Nyckelord – Keywords Jalkapallo, Yleisömäärät, Sosiaalinen Media, Urheilun Kysyntä, OLS			
Ohjaaja tai ohjaajat – Handledare – Supervisor or supervisors Roope Uusitalo			

## Contents

<b>1 Introduction</b> .....	1
<b>2 Background</b> .....	5
2.1 The Finnish Football League Association.....	5
2.2 Football in Economics.....	6
2.3 Internet Media in Economics.....	10
<b>3 Literature</b> .....	12
3.1 Price and Income.....	12
3.2 The Superstar Effect. ....	16
3.3 Competitiveness and Success. ....	19
3.4 Social Media .....	23
<b>4 Empiria</b> .....	25
4.1 Data. ....	25
4.2 Variables.....	27
4.3 Model. ....	31
4.4 Results.....	32
<b>5 Conclusions</b> .....	38
<b>6 References</b> .....	40
<b>7 Appendix</b> .....	43

## Preface

I wrote this thesis for my Master's degree in Social Sciences with specialization in Economics at the University of Helsinki, Finland. The subject of the thesis is related to the demand of live football and the different parts from which it is constituted, with an emphasis on social media. It is a fascinating research topic as it blends together the importance of both sports and social networks in the everyday lives of people, and allows to study these phenomena from the point of view of multiple fields within academia.

The completion of this thesis marks the ending of a six-year period very significant in the life of a young adult through the ritual of graduation, which is widely recognized to be an important milestone and a great achievement in today's society. Upon entering a new phase in my life, one in which for the first time the eventual goal is not predetermined by an academic curriculum, I would like to extend a few words of gratitude to individuals that have played important roles in getting me this far.

First, I would like to thank my father Vesa, notably for introducing me to economics as an interesting career path due to his short workdays, frequent foreign assignments, and long vacations. Additionally, I thank him for teaching me the importance of physical activity and being present at all my football-related events, despite the occasional embarrassing shout of encouragement. Secondly, I would like to thank my mother Iwona, my sister Alicia and my girlfriend Sini for never doubting me and pushing me to continuously strive for more. Without their support, I probably would not have finished this thesis in due time. Thirdly, I would like to thank family friend Henrik Palmén and youth coach Sebastian Söderholm for giving me the initial nudges on the way to falling head over heels in love with football. This thesis is a testament to my never-ending attachment to the beautiful game.

“Without football, my life is worth nothing” – Cristiano Ronaldo

Finally, the six years spent at the University of Helsinki yielded many unforgettable memories, with many unforgettable friends. Out of them all, I would like to thank especially Veikko Walta, Jetro Anttonen, Jim James and Samuel Mueller for mentoring me during work on this thesis, Pyry Lehtonen for doing so with my bachelor's thesis, and Adam Saada for sharing a home and a family with me while motivating me through a difficult time in the studies. You are all extremely good friends and kind people, and I am privileged to have met you.

A last expression of gratitude goes out to SLF for believing in me. Without their financial input, it would not have been possible to take time off work and fully dedicate myself to the research. I hope you enjoy the end result.

Alexander Hölttä  
Helsinki, 20.4.2020

# 1. Introduction

Football has been a mainstay of interest within economic research since the mid-1950s (Dobson & Goddard 2001). It is the world's most followed sport, and viewership is continuously growing due to the opening of relatively new or previously unattained markets such as Asia or the Middle East. As of late, also Finnish football is going through a great surge of interest. The men's national team made its way to a first-ever appearance in a major European tournament and the national league, Veikkausliiga, after undergoing a change of league system, finally had a long-awaited tight and exciting ending which resulted in a championship for the Kuopio side KuPS.

Consumers all around the globe pay for entrance to a stadium or to purchase viewing rights from broadcasting services. In the consumer's mind, football is entertainment, and they tend to hold clear preferences on which entertainers they enjoy most (Rujig & Van Ophem 2014). From the point of view of the entertainer - a football club - they are in a business with a dual goal. Firstly, they must attain sporting success. Secondly, they must maximize financial revenues. These two goals are in football, as in many other sports, very much connected. According to Peter J. Sloane (1971) the utility function takes the following form:

$$U=(P,A,X,\pi R-\pi_0-T), \text{ when } \pi R \geq \pi_0+T,$$

In which P represents sporting success, A average attendance, X the league's wellbeing,  $\pi R$  the expected revenue,  $\pi_0$  the minimum acceptable net revenue, and T taxes. This differs significantly from what we are used to seeing in economics, especially when thinking about maximizing the X term. This translates in practice into one team not trying to acquire all the best players or most interesting entertainers of a league. Football would de facto not create the excitement amongst fans if the score was pre-determined. A competition in which one agent would monopolize winning would thus likely lower the average attendances, hence hurting the financial revenue (Scelles & Helley & Durand & Bonnal 2016).

Revenue is nowadays generated mainly through media rights such as broadcasting deals or sponsorships, but also through more traditional channels such as merchandise and ticket sales (Ruijg & Van Ophem 2014). It is worth noting that a great footballer from a sporting point of view might not be the optimal face for marketing and the name on the back of the most sold jerseys might not be the first on the teamsheet come Saturday.

Since sport is a form of entertainment by nature, it is hard to measure performer output concretely (Dobson & Goddard 2001). However, by the laws of supply and demand, one way to determine output is as the interest generated by an entertainer in the paying customer (Ruijg & Van Ophem 2014). How to measure this interest then? In addition to all the existing answers related to playing performance measurement and different aspects of pricing, part of the truth might be hidden in the relevance of social media, which is only beginning to be widely studied (Watanabe & Yan & Soebbing 2015).

This thesis attempts to measure the relationship that social media activity had with the demand for football match tickets in the Finnish Veikkausliiga season of 2019. The main hypothesis is that relatively high activity on social media resulted in higher ticket demand compared to low activity. The research focuses on Veikkausliiga due to the newly found relevance of football in Finland and the fact that in most countries football games are sold out regardless of social media performance. The specific choice of the season of 2019 is due foremost to the lack of some data points from earlier seasons, but also to the fact that the league system was changed substantially after the 2018 season. The subject is managed within a large framework of previous literature and empirical methods by firstly explaining what has been done before and then attempting to add something to top it off. More specifically, the aim is to decompose and identify the different factors that form the total demand of a football match in Finland. These include sporting success, competitive balance, ticket prices, star performers, the aforementioned social media activity and other nuances to be explained later on.

Sports economics are often misunderstood to be an arena for ‘geeks to play with data’, according to an article by Bryson, Frick, and Simmons (2015) in the National Institute Economic Review. It need not be so. The article states multiple views for the contrary: firstly, sports is a big and ever-growing business. Secondly, as any business accounting for as much as 3.7% of EU GDP, sports is a large employer – up to 5.4% of EU total labor. Thirdly, empirical work by multiple cited economists proves that sports can enhance the quality of life through a sense of belonging and physical activity. Fourth and perhaps foremost, the authors underline the non-negligible influence that sports and its economics study has had on questions such as racial bias, incentive analysis, effects of prizes and rewards, wealth inequality, taxation, community work, investment incentives, health, and labor market outcomes (Bryson, Frick & Simmons 2015). Upon closer inspection, there are really few aspects of economic life that cannot be studied with the help of sports.

The research question of this thesis is thus the following: what factors are significant in determining the demand of football, and does social media play an important role in it? The question was chosen for multiple purposes. Firstly, it is an attempt to add something new to the existing literature by using data intended for commercial purposes and thus unavailable to most researchers. Secondly, and quite importantly, it was chosen for the same reasons as many others have chosen to study sports or social media – the easy access to data, which is usually public by definition in both cases. Thirdly, the focus on attendance is important because it is one of the most important aspects of the entire business (Scelles & Helleu & Durand & Bonnal 2016). It affects directly the value of the club through revenue and the atmosphere generated by the presence of the crowd, which in turn prompts television viewership, which generates more revenue (Ruijg & Van Ophem 2014). Finally, the research question was formulated from the deep underlying interest in the study of football, social media, and Finnish culture by the Writer.

According to research by Boyd and Ellison (2007) - since their introduction in around 1997 - social networking sites, now in the form of applications and called social media, have attracted millions of users. There are multiple different forms and platforms, but the general idea is to get strangers to communicate with each other based on shared interests. Upon registration, individuals are to create profiles and



answer questions that help algorithms identify content that is relevant to the viewers. Social networks vary the way a profile is to be seen by others, in some it is completely public and others only for selected individuals by the user. This selected group of confirmed or binary connections is usually called friends, fans, or followers and its visibility to other users is crucial to how the networks work. To summarize: there exists in this world a variety of services to which individuals subscribe with their own personal data and voluntarily give information about their interests and social connections. It is no wonder that this information is often exploited by outsiders, companies, or individuals in what sometimes constitutes even breaches of legality. Nonetheless, this offers an excellent platform to capture and model consumer behavior, as has been done by many economists in the past and will be done in this very thesis (Watanabe & Yan & Soebbing 2015) (Scelles & Helleu & Durand & Bonnal 2016).

This thesis evolves around the following three chapters. In the second one, relevant background information is presented in order to navigate the subject with ease. This includes detailed information on how the Finnish Veikkausliiga is set up and how football and Internet media fit and are understood within economics. The third chapter is a review of previous literature in and around the research question of this thesis, aiming towards building a coherent foundation upon which to build the empirical part, located in the fourth chapter and performed by means of a regression analysis in order for it to be comparable to other similar work in the field. The fifth chapter will in turn discuss the relevant conclusions and implications.

## **2. Background**

This chapter presents the key concepts relevant to this thesis, and how they should be viewed and comprehended from an economics point of view. Firstly, we take a detailed look at the Finnish Veikkausliiga as a league system to understand exactly what is the product of which the demand is of interest. Secondly, we seek to understand how a football team and league functions within economic theory, and what do they try to achieve. This obviously holds great meaning as there is no demand without supply. Thirdly, we lay down measures by which Internet media metrics are to be evaluated from the point of view of economic value because they are of special interest to the research question and rarely utilized in the field.

### **2.1. The Finnish Football League Association**

The Finnish Football League Association, alias Veikkausliiga, is the top flight of Finnish men's football. It is the 40th most valuable football league in Europe when adding up the values of all the registered players (Transfermarkt, 2019). According to the Union of European Football Associations (UEFA) and its Club Licensing Benchmarking Report of the Financial Year 2018, it is one of 9 worst performers in Europe when looking at popularity on social media with less than 25,000 followers/likes compared to more than 2,000,000 for the top 5 performers. Giving slightly better context, CIES Football Observatory ranks the Finnish average attendances from 2013-2018 at 44th position in Europe, also noting a negative attendance evolution of up to 13% from 2003 to 2018. Safe to say, there is room for improvement in attracting customers to the product.

Veikkausliiga was founded in 1989 and hence boasts 30 seasons of existence as of the end of 2019. A season is played within a calendar year, contrary to many European league systems. Between the seasons 2018 and 2019 the league system was changed drastically to add competitiveness. The league consists of 12 teams, each playing the 11 others both home and away during the regular season. After this, the table is divided into a championship series and a relegation series, both of which are played as one-off games against each 5 competitors in the respective series. At the

end of the season, a five-team final tournament is played as a knockout mode of a total of 5 games to determine the third and final European competition berth – the first two going to the winner and runner-up of the championship series. The total of 167 games of domestic top tier football being played each year result in a champion, a relegated team, and a relegation play-off team (Veikkausliiga, 2019).

The system in which the Veikkausliiga is set up allows us to study the relationship of the championship battle, relegation battle, and the competition for European berths with the attendance levels, among other factors. This thesis will examine the 2019 season, which ended in the crowning of KuPS Kuopio as champions in a grande finale played in a sold-out 9,125-seated Veritas Stadium in Turku against runner-up team FC Inter. FC Honka obtained a bronze medal and the third European berth, while VPS Vaasa were relegated along with the relegation play-off team KPV. The season had an average attendance of 2,620 spectators per game, a 13,5% growth on the 2018 season (UEFA, 2019). It is especially this growth and underlying reasons for it that this thesis will attempt to comprehend.

## **2.2 Football in Economics**

As presented in the introduction, Sloane's paper from 1971 is about the football club as a utility maximizer. In the preface, Sloane states outlining the economic and organizational characteristics of professional football as his main goal, followed by the consideration of possible objectives of a football club's behavior. In 1971, not many previous studies existed even though already at the time, a major football club was economically comparable with a small or medium-sized business. Sloane's paper will thus be examined in this section as the writer intended, as a base of knowledge for how football fits within the economy and society and why it is a relevant field of study. Although Sloane's analysis is limited to the English Football League, he acknowledges that most of the features discussed also apply to other national leagues and as such much of the knowledge will be generalizable to this research.

What is truly noteworthy already in the first paragraphs is the figure of 200,000 sterling pounds, which Sloane states as an extravagant transfer fee for a star player.

A few sentences earlier, it had already been established that clubs were profiting from sponsorships by industrial firms. In comparison, a quick look into today's market values in the same context from the German provider Transfermarkt GmbH yields values of over 150 000,000 sterling pounds for the most appreciated footballers. The same figures for the Finnish Veikkausliiga barely reach one million euros, and market values should theoretically represent benchmarks of talent level, so moving forward we will have to mind this gap between the United Kingdom and Finland when making assumptions regarding the latter based on the literature on the former.

Sloane's next important notion in relation to this paper is in a paragraph in which he explains the league system and how its complexity affects the crowds. It is explained that in addition to competing with other English teams for the English Football League and additional domestic cup competitions, the highest-ranked teams from each season gain access to pan European competitions for the following season. Sloane states that since these European games attract large attendances - implying that they might attract larger attendances than the domestic ones - participation in them is of financial importance, which in turn implies that there is added demand from spectators when either the level of the football is higher due to increased level of the participating teams or there are performers on show that are not usual to the spectator. This section along with its underlying implications applies also to the spectators of the Finnish Veikkausliiga, of which the details of the league system are explained in its own section of this paper.

Following the explanation of the league system, Sloane moves on to point out the great paradox of competitiveness within the football industry. According to it, a football club's objective is to be the best during a season by achieving the highest possible rank through winning matches, but it has to do so while still maintaining competition - i.e. letting other clubs achieve success - because their collective success will result in higher popularity of the football. Sloane then goes on to reference one of the few pre-existing papers about sports in which the writer has claimed that the aforementioned competitiveness is the most important factor in determining the demand of sports, whereas the attendance figures from the English Football League compared to the sporting results do not seem to sustain that view –

likely due to the habit of people going to games out of commitment rather than entertainment. Taking this dilemma into the context of Finnish football, we are likely to see an outcome rather similar to Sloane's reference rather than Sloane's work, as Finns tend to have a reputation of following success rather than performance.

Another relevant point to take from Sloane's work is demand competition from other leagues, sports, and forms of leisure. Lower tier leagues are not to be considered in the case of Finnish football due to relatively low population and interest in football altogether, where sometimes they tend to overlap in interest with the top flight in England, thus effectively stealing potential customers from clubs. Sloane moves on to dismiss the impact of other sports on the demand, citing the often seasonal nature and careful arrangement of fixtures as reasons the sports should not collide. In the context of Finland, the situation is quite different as there exists a strong cultural divide between ice hockey and football up to the point of potential customers choosing one and completely disregarding the other even though the two would not be mutually exclusive. Finally, the effect of other pastimes is taken into account by Sloane and he does correctly acknowledge that there are a number of things each individual would rather do than watch football, resulting in high cross elasticities of demand. Particularly interesting here is the notion that television ownership is a close substitute and a competition, since potential customers may rather watch the game from the comfort of their own house. In 1971 showing football matches on television was closely regulated as there does not seem to have been the possibility of having customers buy pay-per-view, and mainly due to this limitation the television revenue would not have compensated the clubs for the no-show tickets. In today's world and market this seems rather gullible as television broadcasters are responsible for millions of revenue money globally. Whatever the impact of broadcasting is on the cash flow of clubs, television certainly affects the match-going crowds - possibly even in Finland – and this will be taken into account and researched within this paper.

On the market structure of football, Sloane states that the league is a monopoly that maximizes uncertainty which is related to maximizing revenue. He even goes on to argue that cross-subsidization, an idea often frowned upon by economists, might be in order to ensure the longevity of competition and thus fulfilling the dual goal of

maximizing sporting and financial goals. The idea has since been adopted by almost all of the American "Big Five" sports leagues in the form of drafting talent, but the European school of thought has developed differently, allowing for larger imbalances within leagues than in the American ones. Group equilibrium of the monopolistic league is defined by Sloane as the clubs earning sufficient profits to remain in the league and preserving the viability of the business side, paying rents, wages, and so forth. Regarding the profits, while presenting the utility function referenced in the introduction of this paper, Sloane states that the maximization of attendance can be thought of as a primary objective since part of the utility of football for both players and fanatics – thus the club as well - is derived from the entertainment provided by the crowd which supplies a very special form of belonging and atmospheric vocal participation. Yet another interesting point made in the paragraph is that historically speaking, when some clubs have had lesser attendances, they have gladly invested in new playing staff in order to re-attract the crowds to their stadiums even though it might have meant no direct return on the investment per se.

Indeed, football clubs are - according to Sloane - well known to behave very differently to traditional firms due to these aforementioned multi-dimensional goals. To set an example, ticket prices are often fixed irrespectively to the opposition and black-market activities arise before games of higher interest, even though it would maximize the revenue and profit of the club to regulate the prices themselves. In doing so, they would likely lose the sympathy of the most fanatic supporters, resulting in a worse atmosphere, in turn resulting in lower attendances, and so forth. The major outtake from Sloane's work is thus to always keep in mind the complexity of football as a field in which many actors hold many kinds of different interests. Secondly, it was implied that the demand for football is generated from competitiveness, the level of different players, the atmosphere of the crowd, and the lack of other pastimes.

## 2.3 Internet Media in Economics

As the introduction detailed, various social media platforms had millions of users already in 2007. Martin Peitz's and Markus Reisinger's chapter in the "Handbook of Media Economics" (2014) is important added context to this prior knowledge not only due to the more recent release date but thanks to the fact that the writers are economists and as such their relationship and way of seeing the issue at hand is more relevant to this thesis and economics as a whole. The purpose of their writing is to present the reader with pre-existing literature as to what Internet media is, what it brings to the society and economy. The text begins with the phrase "The Internet has changed the lives of most people, in both business and leisure activities", from which it follows that the shift from traditional media to the Internet has enabled immediate coverage of any event or connection globally. Due to this, the media-providing companies have become true powerhouses of business, which in turn calls for studying their behavior in the overall economy.

Peitz and Reisinger write that none of the biggest players in the market, such as Google or Facebook, started out as a news intermediary, but have become exactly that and more due to their ever-evolving algorithms that help present their customers with the most relevant and personal information possible. In order to be qualified as an Internet medium here, a site or service has to replace the old information with a new one or add to the existing one, which is the difference between media and, for example, encyclopedia. The key notion here is that the Internet media makes individuals behave otherwise than they would with the presence of only traditional media, according to the observation by Peitz and Reisinger. Additionally, the average consumer spends more time-consuming media online than offline, thus making the Internet the main source of media consumption in the society.

What is extremely important in Peitz and Reisinger's paper with regards to this thesis is the understanding of the movement of information on the Internet. Contrary to traditional media, information is not only sent from a sender to a receiver, but is created, curated, and spread by the users activating them in the sharing the visibility of a subject or matter. This is of great interest to this thesis especially due to the

relationship of internet media with advertising. According to Peitz and Reisinger, it is common that the financing of these Internet platforms is entirely built upon advertising, which in turn is easier on the Internet due to ad optimization through the search engines that the customers of the platforms use frequently.

To conceptualize why this is relevant concerning the research question of this thesis, we must create the idea of an individual choosing what to do with his time and money. His or her acquaintance might have gone to a football match some time ago and shared their experience about it in an Internet media. There might be discussion about football on some Internet platform or forum the individual frequently visits. The existence of this discussion might give the individual in question the initial reason to consider football as an option, and he or she might type something football-related to his preferred search engine. Due to the mechanism explained by Peitz and Reisinger, the platform will remember that interest and will start advertising football-relatedly, possibly affecting the choice of the individual regarding what to do with his or her money and time. By this logic, later on in this thesis when examining social media activity around the Finnish Veikkausliiga, we cannot limit ourselves to solely activity on accounts directly related to the league, teams, or players, but all individuals or institutions that have created, curated or shared related content.

Having presented multiple studies concerning who uses Internet media, why, when, and where, Peitz and Reisinger finally acknowledge that all and any problems to think of which concern the Internet have too many layers to be explicitly explained by a single framework. This is a very similar ending to a text as Sloane (1971) had concerning football economics, an intelligent way to handle such complex and intertwined subjects, and an important note to remember when reading this thesis.



### **3. Literature**

This chapter lays the foundation for the empirical part of the thesis by presenting related literature. The research question has not been studied per se, but this is rather a challenge than a limitation. The goal is to manage the research question within a framework of directly and indirectly related studies that will be presented in this chapter and commented on. Each subchapter adds a new dimension to the concept of demand of football, which will then be developed within empirical evidence in chapter four.

Let us continue with imagining an individual, who according to the very basic concepts of economics has a utility function and a budget restraint. In other words, the individual has to choose what he or she is willing to do with his or her time and the possible remuneration gained from investing that time into labor. What could persuade the individual to choose watching football over any other activity? He or she must get the impulse from somewhere to even consider whether or not football could be of interest. Then, he or she has to figure out whether the next game being played at an attainable venue is interesting enough to pay for. He or she has to think about whether he or she can afford to attend. According to this literature review, economists do not hold a clear view of which of those aforementioned factors has the most significance in the individual's final decision. Here are some of the most cited works in the field.

#### **3.1 Price and Income**

Peter J. W. N. Bird from the Department of Economics of the University of Stirling, UK, writes for the journal *Applied Economics* and researches the demand for league football. The introduction states estimating a demand equation as the purpose of the paper, and the research context is said to be a long-term decline of total attendances. Already at the very beginning of the paper a problem is presented: existing work fails to take apart all the nuances that make up the total interest power of a single match fixture. Additionally, Bird's study adds to existing work by applying a large sample

pool of seasons from 1948/49 to 1979/80, which is said to be done in order to eliminate uncorrected changes in customer demographics such as variations in income or ticket prices. The paper also recognizes – in line with cited work - geographic factors as relevant, which is likely to be the case in Finland as well and has to be taken into account in the empirical part of this thesis.

While reading about Bird's model and data, one stumbles upon an important idea also from the context of this thesis – that of fixed prices for each match in each football season. Bird explains this to lead to the possibility of using a single demand equation because all changes in ticket prices are due to administration and are not a product of the market equilibrium but rather a taken amount. The mentioned demand equation goes as follows:

$$\ln A_t = \alpha + \beta \ln \left[ \left( \frac{P_t}{R_t} \right) + \gamma \left( \frac{Q_t}{R_t} \right) \right] + \delta_t + zCHANGE_t + \theta \ln EXPEN_t + \varepsilon_t ,$$

in which  $A_t$  is total attendance at the league matches in a season starting from year  $t$ ,  $P_t$  the minimum ticket price for a league match starting from year  $t$ ,  $Q_t$  the level of total expenses on transport starting from year  $t$ ,  $R_t$  the retail price index starting from year  $t$ ,  $EXPEN_t$  the total consumer expenditure starting from year  $t$ ,  $CHANGE_t$  a dummy variable to navigate years. The term in the bracket hence represents the cost of leaving one's home and going to a league match, and the whole equation takes a rather financial form upon closer inspection. This might be necessary as it is an economic analysis, but it seemingly fails to capture any additional reasons for going to a game than the ones of financial nature.

Later on, in the paragraph it is explained that time trend proxies are however taken into account, although they might be hard to separate altogether from the income variable  $EXPEN$ . One of these is called  $GOALS$  and reflects simply on the number of goals scored in the league each year in order to obtain some kind of measurement of attractiveness of games. It is not a wrongful thought, but rather simplistic as true football fans will follow their team regardless of performance. Like it was stated earlier, the large sample pool of years is supposed to eliminate attractiveness differences between promotional and relegational battles each season. Even though

this is technically true, again, it seems to be a rather simple try but will have to do in the absence of harder data. The demand equation this thesis will look to build will have some of the same features than Bird's, but will obviously need to seek a broader perspective which would truly showcase the incredible nuances that make football an unmissable event each game day.

Equation	Estimation period	Estimation method	Constant	$\frac{P_t}{R_t} + \gamma$	$\frac{Q_t}{R_t}$	$\gamma^a$	EXPEN <sub>t</sub>	CHANGE <sub>t</sub>	$r$	HOOL <sub>t</sub>	ICE <sub>t</sub>	GOALS <sub>t</sub>	GOALS <sub>t-1</sub>	R <sup>2</sup> <sup>b</sup>	DW	$\rho^c$
1	1948-75	NLS	21.4	-0.24 (0.35)	0.65 (1.96)	-0.68 (0.37)	0.088 (0.038)	0.0022 (0.0106)	—	—	—	—	—	—	2.12	0.68 (0.15)
2	1948-75	CORC	21.8	-0.24 (0.10)	0.65 (1.86)	-0.68 (0.33)	0.089 (0.034)	0.0020 (0.0096)	—	—	—	—	—	0.74 (0.69)	2.12	0.67 (0.14)
3	1948-75	NLS	24.9	-0.23 (0.32)	0.60 (1.86)	-0.62 (0.15)	0.092 (0.035)	—	—	—	—	—	—	—	2.11	0.66 (0.15)
4	1948-75	CORC	25.0	-0.22 (0.09)	0.60 (1.86)	-0.62 (0.12)	0.093 (0.032)	—	—	—	—	—	—	0.75 (0.72)	2.11	0.66 (0.15)
5	1948-75	CORC	25.2	-0.23 (0.10)	0.60 (1.86)	-0.65 (0.15)	-3.39 (14.64)	—	—	0.0018 (0.0074)	—	—	—	0.76 (0.71)	2.09	0.65 (0.15)
6	1948-75	CORC	24.8	-0.23 (0.10)	0.60 (1.86)	-0.61 (0.13)	0.091 (0.033)	—	—	—	0.013 (0.025)	—	—	0.74 (0.69)	2.10	0.68 (0.14)
7	1948-75	CORC	28.9	-0.21 (0.11)	0.60 (1.86)	-0.80 (0.08)	0.111 (0.033)	—	—	—	—	-0.24 (0.12)	—	0.93 (0.92)	1.98	0.25 (0.19)
8	1948-75	CORC	28.5	-0.18 (0.10)	0.60 (1.86)	-0.77 (0.09)	0.096 (0.031)	—	—	—	—	—	-0.24 (0.10)	0.91 (0.90)	1.95	0.41 (0.16)
9	1948-79	CORC	24.6	-0.20 (0.09)	0.60 (1.86)	-0.60 (0.12)	0.093 (0.032)	—	—	—	—	—	—	0.77 (0.74)	2.10	0.63 (0.14)

Figure 1: Bird's (1982) regression results

Bird's results for NLS and CORC models – displayed in Figure 1 above - imply an income elasticity of -0.60 for football, which means that it is an inferior good for consumers and as their incomes increase they will likely purchase other leisure options than football. A second observation presented by Bird is that demand is price-inelastic, as across the time periods the prices of transportation and ticket prices effectively rose. These are results that have direct implications into this thesis, and they will be treated as assumptions from this chapter forward.

Some of the most cited works in the field, Rob Simmons has been published multiple times on his own and as a part of a team in the Journal of Applied Economics (1996). Simmons' solo work from 1996 is exactly what this thesis will be: an estimation of demand, albeit for English Football League clubs, which analyses economic determinants of the attendances, controlling for 'football-specific factors'. A major difference to Bird's (1982) work is the study of club-based data rather than aggregate one. A benefit of this should be that the data would understand better 'allegiances' to a certain team, preferences if you will. Indeed, as Simmons presents, spectators whose demand is not due to supporting one of the two teams are a minority at any

event. Due to this strength, the method could be used in this thesis, especially as the number of clubs in the Veikkausliiga is relatively low compared to other top flights. An important notion also seems to be a correction for season-ticket holders, who obviously do their best to attend every fixture and as such are not of much interest considering the economic implications of this thesis. However, it remains to be seen whether season-ticket data for Finnish football is public and thus applicable here.

Simmons has data for 30+ seasons, as did Bird (1982), specifically for seasons running from 1948 to 1991. The model is said to incorporate an income effect for the supporters proxied by log real earnings, which is plausible and not dissimilar to what Bird (1982) did and as such seems reasonable to use in this thesis. Simmons presents the model for each club in the form of:

$$a_{it} = b_0 + b_1 p_{it} + b_2 w_{it} ,$$

in which  $a$  represents the logarithm of attendance,  $p$  the logarithm of the real price, and  $w$  the logarithm of real earnings of the supporters. Simmons follows up this equation with speculation on whether unemployment should be included in it but chooses not to due to ambiguity of the issue. After proving that the demand equation should include lagged terms due to habit persistence, he reforms the model as:

$$\Delta a_{it} = \beta_{0i} + \delta_1 \Delta p_{it} + \delta_2 \Delta p_{it-1} + \delta_3 \Delta w_{it} + \delta_4 \Delta w_{it-1} + u_1 a_{it-1} + u_2 p_{it-1} + u_3 w_{it-1} + \varepsilon_{it} ,$$

where  $\Delta$  represents a difference operator and  $\varepsilon$  the standard error term. Citing previous work, the equation is augmented with variables specifically related to football such as success, log of goals per game, and cup performance. Most importantly, the empiria does not account for the uncertainty of outcome whatsoever. Additionally, transfer price to and from grounds used in Bird's (1982) are dismissed as not variant enough to provide any explanation. The variables used seem relevant enough for the context of this thesis, as is the estimation method, Ordinary Least Squares, which seems to feature regularly in the relevant literature.

	AR	TOT	WH	CH	AV	WV	WB	DY	NF	LC	LD	SU	SW	EV	LV	MC	MU	NC	SN
$aq_{t-1}$	-0.48 (3.42)*	-0.44 (3.11)*	-0.36 (2.18)*	-0.99 (6.08)*	-0.49 (3.51)*	-0.40 (5.15)*	-0.30 (3.20)*	-0.81 (4.76)*	-0.35 (3.30)*	-0.46 (3.09)*	-0.30 (3.40)*	-0.49 (3.75)*	-0.53 (2.81)*	-0.44 (2.90)*	-0.60 (5.68)*	-0.56 (4.03)*	-0.64 (3.88)*	-0.39 (2.81)*	-0.38 (3.14)*
$pg_{t-1}$	-0.87 (3.71)*	-0.19 (1.80)*	-0.41 (2.23)*	-0.41 (4.34)*	-0.44 (2.00)*	-0.28 (2.66)*	-0.27 (2.66)*	-0.33 (2.23)*	-0.26 (2.75)*	-0.41 (2.75)*	-0.26 (2.55)*	-0.66 (2.70)*	-0.45 (3.36)*	-0.75 (5.44)*	-0.27 (2.27)*	-0.46 (3.56)*	-0.25 (1.81)*	-0.19 (1.98)*	
$w_{t-1}$	1.03 (2.21)*		0.58 (1.56)		1.23 (2.18)*							1.39 (2.30)*		0.75 (3.19)*					
$s_{t-1}$								0.0057 (2.19)*				0.011 (4.08)*	0.0062 (1.80)*						
$\Delta pg_t$	-0.37 (1.81)*	-0.45 (2.03)*			-1.35 (4.27)*	-0.42 (2.77)*	-0.39 (2.96)*	-0.45 (1.68)		-0.62 (2.59)*			-0.72 (2.57)*	-0.70 (5.82)*	-0.58 (7.95)*			-0.65 (2.94)*	0.43 (2.13)*
$\Delta pg_{t-1}$		-1.07 (4.37)*	0.38 (1.88)*			0.42 (2.77)*	0.39 (2.96)*									0.85 (2.60)*			
$\Delta w_t$										-1.27 (1.88)*		2.59 (2.54)*		-1.37 (1.57)			-1.36 (1.67)		-1.67 (2.18)*
$\Delta w_{t-1}$		1.65 (1.75)*																	
$s_t$	0.016 (2.63)*			0.020 (4.74)*											0.016 (2.69)*				
$\Delta s_t$					0.0083 (1.63)	0.0089 (2.70)*		0.012 (3.68)*	0.015 (3.78)*	0.014 (4.30)*	0.012 (3.36)*		0.014 (3.10)*						
$g_t$	0.28 (2.01)*	0.24 (2.52)*	0.32 (2.76)*			0.37 (3.53)*		0.34 (2.26)*				0.87 (5.20)*				0.29 (2.84)*		0.40 (3.58)*	0.31 (3.11)*
$P21$		0.25 (1.78)*	-0.11 (1.98)*			0.19 (2.16)*	0.38 (4.54)*					0.29 (4.06)*							
$R12$			0.11 (1.98)*				-0.38 (4.54)*		-0.54 (3.15)*			-0.29 (4.06)*							-0.20 (2.97)*
$R23$						-0.37 (2.33)*													
$fa$		0.020 (1.92)*													0.023 (3.73)				0.060 (3.57)*
$lg$						0.042 (2.86)*		0.082 (3.27)*											
$d$										-0.87 (6.05)*				0.52 (3.98)*	-0.35 (5.87)*	0.33 (2.07)*	0.31 (2.68)*		
$R^2$	0.58	0.65	0.50	0.55	0.63	0.78	0.49	0.62	0.67	0.63	0.69	0.62	0.42	0.71	0.85	0.73	0.50	0.50	0.68
SE	0.108	0.101	0.103	0.174	0.147	0.113	0.136	0.138	0.156	0.116	0.118	0.138	0.194	0.118	0.054	0.116	0.111	0.148	0.117
LM	1.09	0.52	1.90	0.13	3.76*	0.00	1.13	0.00	1.86	0.00	0.01	3.35*	0.12	0.01	0.07	0.31	0.24	0.35	2.19
(1)																			
RESET	0.26	0.06	0.01	1.98	2.39	0.12	0.29	0.79	0.87	0.11	0.15	1.17	1.47	0.02	0.13	2.77	0.26	0.56	0.02
NORM	1.09	1.61	4.11	5.35*	1.33	0.57	1.09	1.03	3.91	0.03	1.36	0.82	1.56	0.37	0.46	0.47	1.00	0.49	2.05
HET	0.01	0.01	0.11	0.34	5.59*	0.36	0.84	0.74	0.25	0.00	0.80	0.90	0.31	1.54	0.59	0.71	0.66	0.45	2.35
CHOW	0.18	1.04	0.77	0.75	1.43	0.20	0.39	0.87	0.62	0.41	0.26(1)	1.20	0.70	0.14	0.01(2)	0.85	1.07	0.98	2.13
F	0.51	1.11	0.56	0.59	2.01	0.28	0.68	2.25	0.98	0.69	0.73	1.35	0.48	0.48	0.63	0.14	0.71	0.97	1.80
	(5.18)	(5.16)	(6.17)	(10.17)	(11.13)	(7.14)	(8.17)	(9.13)*	(10.15)	(9.15)	(11.14)	(8.14)	(10.15)	(7.17)	(5.17)	(6.15)	(10.15)	(10.15)	(9.13)
ECM	25.21	16.46	7.02	26.54	10.80	28.31	10.60	26.36	11.20	10.06	11.39	19.83	8.65	11.56	29.61	18.02	14.38	8.79	11.20
	(3)*	(2)*	(3)*	(2)*	(3)*	(2)*	(2)*	(3)*	(2)*	(2)*	(2)*	(4)*	(2)*	(2)*	(3)*	(2)*	(2)*	(2)*	(2)*

Figure 2: Simmons' (1996) regression results

Simmons' results imply a relatively larger price-elasticity for spectators who do not hold a season ticket and the league position, goals scored, and promotional or relegation battles to be important in determining the attendance. Considering the financial aspects from the point of view of the customer, Simmons does not replicate Bird's (1982) results of football as an inferior good, finding it instead to be a luxury good in some cases. As for the estimation equation – of which the results are displayed in Figure 2 – the results indicate that increase in real price has a relatively significant negative effect on demand. The football-specific effects, particularly league position, goals scored, promotion, and relegation are all found to have an important impact. Overall, Simmons' research holds many important notions concerning this thesis that will be incorporated accordingly.

### 3.2 The Superstar Effect

Sherwin Rosen (1981) writes for the journal *The American Economic Review* about *The Economics of Superstars*. This article is chosen in this literature review not on a footballing basis, but rather as a benchmark of understanding on why does society support and cherish a system in which certain individuals are able to earn such high

salaries by doing something that might seem to others as a mere hobby. Rosen uses in his work the professions of classical musicians and shoemakers to demonstrate the general idea – theory assumes that two musicians or two shoemakers make or produce undifferentiated outputs and that the only difference between what the two of each profession would sell would be the pricing. This of course is not at all in line with reality, as "box office appeal" - as Rosen calls it – enters the equation and enables the highly skilled to command large audiences and hence large salaries. This is a key notion with regard to this research, as it is highly likely that football fans hold clear preferences on which football teams or players have this "box office appeal" and are worthy of the investment of time and money in the ticket.

The way Rosen explains it, there exists an equilibrium of utility-maximizing buyers who can costlessly distinguish between the talents of the sellers, who in turn maximize profit and are assumed not to be able to earn larger amounts doing something else. The sellers are not cross-substitutable, and have a convex net revenue function  $R(q)$  - small differences in talent become magnified in larger earnings differences, accelerating the magnification towards the very top of the talent. One explanation for this phenomenon is said to be the scale economy of joint consumption, which allows for relatively few sellers to service the whole market – the better they are the more they can serve. The difference between this situation and a public good is the admission price, which obviously plays a very important role. Rosen ends this particular paragraph by stating that when the joint consumption is mixed with imperfect substitution of the sellers, there will always exist the possibility to attract large audiences and make large salaries. Translated into a football context, this is the explanation for why footballers earn high salaries – because they are highly skilled individuals who are not cross-substitutable and there exists an entrance fee to see them play.

Without going too deep into the specifics, the general idea behind Rosen's paper is that the phenomenon of the concentration of extraordinary incomes is due to technology which spreads extreme abilities of the sellers/performers to the eyes of the customers, who distinguish clearly whom they want to pay to perform. The demand of product is said to be a utility function formed by the appreciation of

talent, the indifference of using one's time into indulging in said activity, and lost wages. This line of thought will be used also in the empirical part of this thesis.

Lawson & Sheehan & Stephenson (2008) have written about the superstar effect in a more specific way, taking as a case study the effect of David Beckham's arrival in the United States and its Major League Soccer for the 2007 season. This study is included in this literary review especially for learning how to incorporate superstar-related variables in a regression model, since the effect is very intuitive – David Beckham's presence or expected presence did indeed grow attendances. For starters, the writers name a few previous studies which have found that in basketball, the presence of a star player on the visiting team has a positive effect on the home attendance. An interesting proposition follows: since the team in which the superstar plays do not reap the full reward, revenue that is, for paying the star player's contract, there is support for the argument of the league playing the part of the superstar's salary. While this is not of direct interest within the research question of this thesis, the magnitude of the proposal makes it an extremely interesting side note at the very least.

Interestingly, Lawson, Sheehan, and Stephenson use data from a single season of football with 13 teams in the league, coming very close with 195 observations to what we will deal with in this thesis with 167 games. In this case, the writers use as the dependent variable the percentage of stadium capacity and as explanatory variables of interest a dummy on whether David Beckham was included in the matchday squad or not. Other variables include information on whether it rained or not, what amount of points the adversary had achieved the previous season, what amount of points the home team had achieved the previous season, whether there was a player on any of the rosters that had been called up for the National Team, whether it was opening or closing day of the season, population, income per capita, and a few other – such as racial - aspects specific to the U.S. market that are of no importance in this thesis. The estimation was made with regards to calendar-specific dummies such as day-of-week, month-of-season, and holiday matches, and a Tobit model is used due to the cap of 100% capacity on the attendance.

Many of these variables are of great interest to this thesis and will be important information for the empirical part – if not per se, then at least adaptably to the dataset. It is however worth mentioning that the Finnish Veikkausliiga does not have such superstar power as did David Beckham, due to the lower financial standard. Likewise, the culture in Finland can differ greatly to the culture in the U.S., where it is not uncommon to perceive team sports as a combination of individuals rather than a team, meaning the value of a superstar to the average spectator could potentially be larger than in Finland. Nonetheless, teams and players are not financially equal on this side of the Atlantic either and it is worth discussing whether a high aggregate salary level of a team is an indication of the existence of superstar individuals for which to pay for entry to the stadium or rather a team built of lesser “known” individuals that happen to form a stronger footballing outfit.

### **3.3 Competitiveness and Success**

The number of points gained in the previous season, the position in the table, a "derby" game, competing for a European berth, fighting against relegation. Many of these factors have already been touted in this thesis and some are yet to be presented as variables of interest regarding why an individual would be interested to purchase an entrance ticket to a football ground. Babatunde Buraimo's work alone (2007), and with Rob Simmons (2015), David Forrest's work with Rob Simmons (2006), and Smith's work with Szymanski (1997) are presented in this section in order to further develop an understanding on which variables and results previously presented could be of use to us.

To start off with the most dated one, Smith and Szymanski (1997) is yet again one of the most cited studies in the field. It is actually a very large body of work, the title comprising English Football as a whole Industry, but the interest of this thesis is in the demand function presented in chapter 5. The dependent variable used here is log attendance, and it is attempted to be explained mainly with the price of admittance, dummies for the league position to determine whether or not the team was in a position to be promoted or relegated, number of cup games played in various cup competitions, and lagged attendance. The regression is made with an OLS model.



Interestingly, promotion and relegation show to be significant both positively and negatively in the following seasons from the occurrence, but due to Smith and Szymanski's study being done on seasonal rather than match-to-match data we cannot learn whether the significance is in play during the season. The major takeaway with regards to this thesis is the importance of the cup competitions, which will have to be incorporated in one way or another.

Moving onto Forrest and Simmons' research (2006), which is an attendance demand model with panel data on over four thousand fixtures, and the first study mentioned in this literature review so far to include football on television. In the introduction, it is stated to behave as a possible negative effect on attendance, which in turn is studied for the second level of English football due to the first level being economically very safe at the time of publication. Here, even though the attendance figures have risen through the years of data gathering, the clubs haven't made revenue accordingly due to a large chunk of it being spent on better players that require larger salaries. A few previous studies are criticized notably for using OLS, single-season data, or failing to control for capacity constraint or season ticket holders – which are the two main reasons for the failure of the OLS method, in turn validating it for the Finnish context. An important goal for Forrest's and Simmons' study is to understand the impact and importance of scheduling on attendance, which has according to them not been studied at least as a main subject of research, and is actually something of very strong interest also in the context of Veikkausliiga. The regression model takes the following form:

***LOG ATTEND<sub>ijkt</sub> = f (SUPPORT, FORM, PROMOTION CONTENTION, OUTCOME UNCERTAINTY, TELEVISION, SCHEDULE, OTHER DUMMIES).***

It is explained to be a cross-sectional time-series regression model in which  $i$  represents the home team,  $j$  the away team,  $k$  a note of divisional level, and  $t$  the home fixture number in a chronological order. *SUPPORT* includes lagged attendance for both home and away teams, to measure the persistence of following a club. Other nuances included as dummies in the variable are; whether the game is a derby match and what is the physical distance between the locations of the two

teams. The variable **FORM** is also made up of multiple measures, such as points per match up to the fixture at hand. The same type of a priori check was made for the **PROMOTION CONTENTION** variable, which measures whether the home or away team is in contention for the league title or relegation. Forrest's and Simmons' research is the first one mentioned in this thesis that includes a check for outcome uncertainty, a very interesting notion. The variable is built with a calculation of their own, but the text mentions that pre-match betting odds could be used although not unbiased. **TELEVISION** is not a relevant variable for the purposes of this thesis, but **SCHEDULE** is. It includes dummies for games being played on public holidays, weekdays, or even on consecutive weekends. Additionally, season months are grouped into month dummies due to attendances normally rising towards the end of the season irrespectively of the other presented factors. The authors speculate on models such as GLS and FGLS but end up settling with a Prais-Winsten regression, due to conservative inferences.

Babatunde Buraimo's solo effort from 2007 focuses especially on televised football and its relationship with revenue and stadium attendances. The importance of the relationship is according to Buraimo very underappreciated because the crowd at the stadium forms a part of the televised product. Similarly to Forrest and Simmons (2006), the regression model is presented as:

$$ATT_{ijt} = f(\text{HOME ATT, AWAY ATT, DERBY, DISTANCE, PERF HOME, PERF AWAY, HWAGE, AWAGE, UNCERTAINTY, SKY, ITV, TERRES, ITV DIGITAL, CHAMPS LEAGUE, UEFA, OTHER DUMMIES, HOME}).$$

The dependent variable **ATT** is here yet again the logarithm of match-day attendance. Many of the other variables have already been explained and detailed above, except for **HWAGE** and **AWAGE**, which are indicators of the wage level of the players on each team, therefore indicating the marginal revenue product, that is quality of play. Match uncertainty was performed the same way as Forrest and Simmons did it, by calculating *home advantage + home team's points per game – away team's points per game*. In this calculation, the *home advantage* is the difference in points per game won by all home teams and points per game won by all away teams in the previous season. One could argue that Buraimo has even used Forrest's and

Simmons' model almost entirely, merely adding new data and tweaking here and there with the variables. Even the same Prais-Winsten method was used.

Buraimo has collaborated with Simmons later on, notably in 2015 in order to understand whether it is the uncertainty of outcome or star quality that drives the demand for English Premier League football. The background of their work is the hypothesis that new, digital, match-goers actually prefer to see star quality players than uncertain outcomes. Wages are used throughout the empiria as a proxy for quality of play and as such also as uncertainty proxies. The dependent variable is yet again the natural logarithm of the audience. The context of measurement is however quite different, as there are for example no costs of travel and thus no necessity to consider the teams as home and away. The television setting of this research indeed takes us away from the main interests of this thesis, but we take a look at the variables below (Figure 3) in any case to be certain of not omitting something that might be important in determining the stadium attendance as well. The regression model used was notably an OLS with fixed effects, which according to the authors eliminates the need to create and measure variables related to price and income, for example. This appears to be a bit straightforward, and could even result in multicollinearity if for example the clubs of larger wealthier cities have also larger wage bills.

Overall, Buraimo, Forrest, Smith, Szymanski, and Simmons have used several measures that relate to competitiveness and success on the field. These variables and methods will be taken into account by this thesis and either replicated or developed in the empirical part.

Variable	Description
COMBINED RELATIVE WAGE	A team's relative wage is its wage bill for a give season divided by the mean wage bill for that season. COMBINED RELATIVE WAGE is the sum of the two teams' relative wages.
ABSOLUTE DIFFERENCE IN RELATIVE WAGE	This the absolute difference in the two teams' relative wage bill.
COMBINED LAG POINTS PER GAME	This is the combined points per game from the previous season of the teams involved in the match.
ABSOLUTE DIFFERENCE IN LAG POINTS PER GAME	This is the absolute difference in points per game from the previous season of the teams involved in the match.
COMBINED POINTS PER GAME	This is the sum of both teams' points per game prior to the match. Points per game is the ratio of the total number of points to the number of games played prior to the match.
ABSOLUTE DIFFERENCE IN POINTS PER GAME	This is the absolute difference in the teams' points per game prior to the match. Points per game is the ratio of the total number of points to the number of games played prior to the match.
CHAMPION CONTENTION	This is a dummy variable that takes the value of 1 if either of the teams in the match can win the championship if it were to win all its remaining games while others only take an average of one point from their remaining games and 0 otherwise.
EUROPE CONTENTION	This is a dummy variable that takes the value of 1 if either of the teams in the match can qualify for either the Champions League or the Europa Cup but not win the championship if it were to win all its remaining games while others only take an average of one point from their remaining games and 0 otherwise.
RELEGATION CONTENTION	This is a dummy variable that takes the value of 1 if either of the teams in the match can be relegated if it were to win all its remaining games while others only take an average of one point from their remaining games and 0 otherwise.
SETANTA	This dummy variable is 1 if the match was televised by the broadcaster Setanta and 0 otherwise. If 0, the match was televised by BSkyB.
OTHER MATCHES	This is a dummy variable that takes the value of 1 if there are other matches being broadcast at the same time and 0 otherwise.
DERBY	This is a dummy variable that takes the value of 1 if the match involves teams who are historical or local rivals and 0 otherwise.
WEEKDAY	This is a dummy variable that take the value of 1 if the game is televised on Monday to Friday inclusive and 0 otherwise.
FIRST HALF	This is a dummy variable that takes the value of 1 if the match is played between August and December inclusive and 0 otherwise.
SECOND HALF	This is a dummy variable that takes the value of 1 if the match is played between January and May inclusive and 0 otherwise.
OUTCOME UNCERTANTY	This is the absolute difference in home-win probability and away-win probability. Probabilities are derived from the bookmaker's odds and adjusted for over-round.

Figure 3: Description of the variables (Buraimo, Simmons 2015)

### 3.4 Social Media

Finally, Nicholas Watanabe's, Grace Yan's and Brian Soebbing's work from 2015 allows us to deepen the understanding of how to use data related to social media in economics, and more specifically the economics of sports. The authors reference

multiple studies about the growing role of social media in sports, underlining the importance of the issue. The main objective of Watanabe's, Yan's and Soebbing's work is to introduce demand theory to social media in the context of Major League Baseball. This is done by means of regression analysis of variables that are important to determine fan interest, represented by the change of Twitter followers as the dependent variable. Curiously, Twitter following is mentioned to be often utilized as a measure in the dependent variable in studies mixing sports and social media.

Watanabe, Yan, and Soebbing use an analytical software called NodeXL to obtain their data on Twitter activity, explaining that it works through typed commands that scrape the needed social media content and then exports it to an excel sheet. Data on follower account numbers, followed account numbers, and tweet numbers throughout the timeframe of the study were collected to form the variable of interest. The independent variables are not of straightforward interest to this thesis since they handle baseball, but it is worthy to state that they too were classified into variables that explain performance, scheduling of games, and economical market size of the team in question. The regression model and method is said to follow that of Bird (1982), further solidifying the methodology for the context of the Finnish Veikkausliiga.

## 4 Empiria

In this part of the thesis a regression analysis is performed in order to understand the composition of demand of Veikkausliiga football in Finland in the season of 2019. The goal of the regression analysis is to understand the relative weights of each factor that has a part to play in what makes a fan to buy a ticket to a game. The regression model is built especially upon the work by Buraimo (2007), Simmons (1996), Forrest and Simmons (2006), Lawson, Sheehan, and Stephenson (2008), and Watanabe, Yan, and Soebbing (2015) – but it includes bits and pieces from all of the studies mentioned in the literature review. Here, the dependent variable is the natural logarithm of attendances in the 2019 season of Veikkausliiga. The observations for these were extracted from the official site of the league, for all the relevant 167 match fixtures. In line with pre-existing work, this thesis uses (i) competition factors, (ii) financial factors, (iii) other attractiveness factors, including social media and superstars, as independent variables.

### 4.1 Data

The method used in this thesis is Ordinary Least Squares, OLS, as to fit with prior work in the field. The method has been used widely but also criticized for allowing multicollinearity and not taking into account the capacity limit at stadiums. (Forrest & Simmons, 2006). Other models presented in the literature such as Fixed Effects, GLS, CORC, Tobit, and Prais-Winsten were considered briefly, but none but Prais-Winsten and Tobit were found to have been used in more than one previous study, whereas OLS was used or considered in more cases. Nonetheless, the use of OLS is justified in this case where the capacity constraint is not binding (Lawson, Sheehan & Stephenson, 2008), and additionally it provides relatively simple results to interpret.

The data on football, and hence attendances, is extracted from the official internet site of the Veikkausliiga. The dataset contains observations on the date, time, location, participating teams, eventual score of each game, the league table after each

round, and how many viewers were in attendance at each event. Additionally, the Veikkausliiga webpage published the player budgets for each team in the 2019 season and provided dates for European and domestic cup games played by the participating teams. The data was extracted on 23.4.2020.

The data on income per capita for each city a team was extracted from the official tax records released by The Association of Finnish Municipalities. The records contain information on taxable income earned in each municipality and the exact number of residents. The data was extracted on 13.5.2020.

The data on betting odds for the 2019 season of Veikkausliiga was extracted from an internet site called Oddsportal, which stores recurring betting data for football globally. The odds displayed on the site are the “best” result of comparisons ran between over 10 betting sites for each match – the globally highest given odds for each result, that is. The data was extracted on 13.5.2020.

The data on weather conditions for each match day was extracted from an internet site called “Säähavaintoarkisto”, which stores recurring weather data from the open data delivered by the Finnish Meteorological Institute. The observations consist of information on how many millimeters of rain there was on each given day in each given municipality. The data was extracted on 13.5.2020.

The data of social media activity is provided by private company Meltwater and includes detailed observations on who has posted, what, when, and on which platform. The method of extracting the data is identical to that used by Watanabe, Yan, and Soebbing (2015) with NodeXL. The search words that described the content of the posts that were to be extracted were “Veikkausliiga”, the names of the participating teams and other keywords related to Finnish football used in searches made by football-related clients in the Meltwater portal. The data was extracted on 28.4.2020.

All the data were collected, checked, and formatted on excel in order to be compatible with RStudio.

## 4.2 Variables

The various variables used in this empirical model can be classified as i) competition ii) financial and iii) attractiveness – related factors. The variables related to competition follow closely the ones used in the model of Buraimo (2007), and include performance statistics from the ongoing season for each team for each matchday for the Veikkausliiga, the National cup competition called “Suomen Cup” and the European qualification matches played by the four teams that were eligible to participate in the summer of 2019. The goal of the competition variables is to understand whether the Finnish football fans go to games when their local team is doing well rather than support them through thick and thin. The financial factors such as the income and residential statistics for each city are part of the financial aspects, as per the work of Bird (1982) and Simmons (1996). The purpose of including these factors is to see whether it is more common to spend money on live football in areas of lower income or higher income. Finally, attractiveness factors – derived most notably from Lawson, Sheehan, and Stephenson (2008) - include such measures as the habit of going to a match, the wage budget of both teams, the amount of rain on the matchday, the weekday of the match and the variable of interest, the amount of social media activity. These variables are introduced to measure superstar quality, or rather the attractiveness of all the stars that form a team, fan persistence, the role of weather on the willingness to go out to the stadium, the role of the day of the week, and finally the “hype” for a matchweek generated by posts on social media about it.

One of the most common reasons to go to a football game is the sheer habit of doing so, giving one’s voice and energy towards supporting a certain team or club. Some fans are more committed to it than others, but generally there is a certain type of cut-off point at the number of fans – season ticket holders or one-off participants – that would have come regardless of other factors. Simmons (1996) was able to acquire data on season ticket holders, and Forrest and Simmons (2006) used the lagged average attendance from the previous season for both home and away teams. Since there was no data on season ticket holders in this dataset, a variable *SUPPORT* was



constructed on the basis of Forrest and Simmons (2006), following the exact same procedure.

Variables on form or performance, position, or contention for championship or relegation are the standard for the literature regarding the demand for live sports, used by Buraimo (2007) and Buraimo and Simmons (2015) for example. In this thesis, the competitiveness level and intensity of the two facing teams are measured by the variable **FORM**, which consists of matchday-to-matchday data of league table positions for each team. The variable, differentiated for the home and away teams, depicts the numerical standing of the team in the league prior to the match being played, thus taking values from 1 to 12. This measure was chosen because it should perform the best in capturing the importance of being in a position to win the league, qualify for the European cups, or get relegated. Here, there was also the option to use points-per-game data as some earlier studies have used, but the decision was made in favor of the standings due to perhaps more information being caught in them.

The Finnish Veikkausliiga publishes each year a list of all the teams' player budgets for the full season. For 2019, the highest budget was that of capital club HJK with 1,500,000€ spent on wages and the lowest that of Helsinki rival HIFK with a figure of 350,000€. The variable **BUDGET** was constructed upon these disclosed figures in order to have a variable similar to that of Buraimo (2007) or Buraimo & Simmons (2015), depicting the quality of the star performers on "stage", as Rosen (1981) explained the case to be. Again, theoretically, having a higher wage bill should result in higher attendances, because higher salaries have been found to correlate with higher abilities and thus higher quality of the product for the paying customer (Smith & Szymanski 1997). Unfortunately, there was no data available at this time that would have shed more light on how evenly those salaries were distributed within the playing squad, which would have likely given a better understanding of true superstar quality.

The uncertainty of outcome was one of the main themes in the research of Buraimo (2007) and touched upon in other studies as well. The most common way to analyze it was the result of a calculation, which was comprised of the probability of the home team to win a game that season and of points-per-game of both home and away

teams. Betting odds were even deemed too unreliable for the context, due to unknown bias in the bookmakers' estimations, but their inclusion in this research in the variable *UNCERTAINTY* should be relatively acceptable due to the odds not coming from one provider, but always from the highest provider out of many possible. Additionally, the calculations in the previous literature were extensive and this procedure should save up a lot of time.

The scheduling of a match was first brought up in this thesis through Forrest and Simmons (2006), although it figured also in some ways in Buraimo (2007) and Buraimo and Simmons (2015). In their works, the goal was to understand whether playing on a different day than the usual Saturdays or Sundays affected audiences, which has been a topic of discussion and discontent in Finland as well amongst fans and sometimes even players. For this need of information, the variable *SCHEDULE* was created within the dataset, acting as a dummy taking the value of 1 if the game was played on the weekend and 0 if it was scheduled on a weekday. A schedule measure could have been extended to another variable, a monthly one, which would have perhaps given insight into whether Finnish football fans prefer going to games towards the end of the season as in Forrest and Simmons (2006) or maybe perhaps the warm summer months. However, since the matchdays were often very unevenly distributed within the months the inclusion of such a variable was deemed unwise.

In England, a league often studied within sports demand economics, the clubs compete in the league, the FA Cup and the League Cup. The participation in these "additional" tournaments was found to have a significant impact on attendances by Simmons (1996) and Smith and Szymanski (1997), and thus had to be included in this thesis as well. For this purpose, the variable *CUP* was created, giving the exact number of games both the home and away team managed to stay in the "Suomen Cup" in the year 2019, ranging from 5 for the fewest and 9 for the finalists. As mentioned earlier in this thesis, also European participation is thought to affect attendances, and thus another dummy taking the value of 1 if a European tie involving one of the teams was played on the week before the game in question, and 0 otherwise.

Already the very first studies in the field of sports economics had in mind the importance of affordability of going to the game (Bird, 1982). Some omitted it due to methodology like fixed effects (Buraimo, Simmons 2015), and some kept it in their equations (Lawson, Sheehan & Stephenson 2008) yielding some interesting results. As unfortunately price data for games or even club aggregates was not available, the variable *INCOME* was built using the exact same process as Lawson, Sheehan, and Stephenson (2008), that is the total of taxable income in the municipality of question divided by the number of residents in the municipality. The purpose of this is to understand what kind of good football is for Finns, and also to see whether football is more popular in the large urban areas or smaller rural towns such as Kokkola.

Some studies, such as Lawson, Sheehan, and Stephenson (2008) have found relationships between bad weather and football attendances, which is a particularly interesting notion in the Finnish context where the weather conditions vary largely throughout the season. To be exact, previous work has studied precipitations, and this is exactly what the variable *WEATHER* is constructed for in this thesis, as a dummy for whether it rained on the matchday or not. An interesting extension of this could have been to control for temperatures and sunny conditions, as they tend to get Finns out of their homes during summers, but could also add into the interest of other pastimes. Whichever way this would have been set up, such data was not available at the time of gathering and the choice was made to proceed with precipitation.

Finally, the variable of interest, *SOCIALS*, is built upon all the social media posts that were extracted with the search words related to the Veikkausliiga season of 2019, exactly like it was done in Watanabe, Yan & Soebbing (2015). The major difference to their work comes from including all the Meltwater-accessible platforms such as Facebook, Twitter, and internet blogs, most notably, thus covering more ground and being more extensive. The number of posts was grouped for each matchweek, starting from the last day of the previous matchweek and ending on the last day of the ongoing matchweek – comprising the totality of 28 matchweeks of 6 or more matches. This grouping process was put in place because it would have been extremely hard to differentiate between posts made between matches from any given matchday, which are not uncommonly played at the same time. The process is

unfortunately not ideal but should give some insight into what is the magnitude of the relationship between social media "buzz" and attending a game.

Some notably omitted variables used in previous research include the aforementioned price, dummies for months of play, dummies for derby games, social media reach, the amount of national team players, and dummies for opening and closing days of the season. Months of play were not distributed well enough to consider, ticket prices were not disclosed, and there is no reliable source for national team call ups for each team for the whole season. The dummies for derby games and opening and closing days of the season were considered but eventually omitted due to the information already being present in the data in other variables thus possibly leading to multicollinearity. Social media reach would have been a very interesting variable, representing the total following numbers of all accounts that posted about a certain matchweek, but the data is behind a paywall and thus unattainable, leaving something worth exploring for future research around the subject. Social media reach could be used to understand whether “hype” around a matchday is generated by multiple accounts posting about it, or rather very influential accounts posting about it. A summary of all the 15 used variables and their descriptive statistics can be found in Table 1 in the appendix.

### **4.3 Model**

This subchapter presents the econometric model used to analyze the data. The regression model was run thrice on RStudio with the Ordinary Least Squares method and a linear regression model originally taking the following form:

$$\mathbf{LOG\ ATTENDANCE}_{ij} = f(\mathbf{SUPPORT, FORM, BUDGET, UNCERTAINTY, SCHEDULE, CUP, INCOME, WEATHER, SOCIALS})$$

To recapitulate, the goal is to explain the log attendance – a variable that represents the natural logarithm of actual attendances of the Veikkausliiga 2019 season. Most importantly, a thorough look at the data revealed that instead of using all the 167 games played in the 2019 season, games from matchweeks 1 and 28 had to be

removed from the set reducing the sample pool to 154 matches. Helas, gameweek 1 did not hold reliable information regarding form, due to the league table at the time being built solely upon the alphabet. Gameweek 28 – the mini cup tournament for the fourth European berth - was in turn completely played after the championship had already been completed, so it was deemed as competitively too different and thus problematic with relation to the plausibility of the data.

The independent variables are built upon measures as previously presented: persisting support, the form of the moment, wage budget, uncertainty, scheduling day, cup performance, income of the area, the weather of the day, and social media posts about the gameweek. These estimation results, along with the results from the other two regressions are presented in the next subchapter. For the second regression, only the statistically significant variables were used, which translates into the following form:

$$\mathbf{LOG\ ATTENDANCE}_{ij} = f(\mathbf{SUPPORT, FORM, UNCERTAINTY, WEATHER, SOCIALS})$$

In order to address some of the concerns regarding Ordinary Least Squares as a method, the last regression was run with the help of robust standard errors. This is a method first introduced by Newey and West (1987), which makes estimates robust to autocorrelation and heteroscedasticity. It can be performed on R with the help of the sandwich package containing the vcovHAC command. Finally, the third and last regression was run with the same statistically significant variables as the second one.

## 4.4 Results

This subchapter presents the results of the three regressions, also figuring in their entirety in Table 2 of the Appendix. All three regressions had the adjusted R squared of 0.403, thus accounting for around 40% of the variation of the dependent variable, which can be said to be successful enough for this kind of rather experimental study. The first regression consisted of a total of 14 independent variables, of which five resulted in a statistically significant relationship with the natural logarithm of

attendances. These five variables were then used as the explanatory variables of two more regressions, resulting again in statistically significant relationships with the natural logarithm of attendances for all of the variables, but on different confidence levels. Below, the most relevant findings of the three regressions are presented and discussed variable-by-variable.

	<i>Dependent variable:</i>		
	log(ATTENDANCE)		
	(1)	(2)	(3)
HOMESUPPORT	0.0003*** (0.0001)	0.0004*** (0.0001)	0.0004*** (0.0001)
AWAYSUPPORT	0.00001 (0.0001)		

The variable *SUPPORT* and specifically the part concerning the lagged aggregate attendance from the previous season of the home team was found to be statistically significant at a 99% confidence level in all three regressions, consistently with previous findings in the field (Smith, Szymanski 1997) (Forrest, Simmons 2006). The result follows intuition, as it is indeed hard to argue against the thought that people who tend to go to football games go there game after game, month after month, year after year. It is a habit, as the result largely confirms. In order to accommodate and exploit this behavior, football clubs have created the season ticket system, in which a supporter can sign up for attending the whole season of games at once. Unfortunately for this research, the data on the season tickets were not available and the effect could not be isolated more efficiently. This estimate implies that one additional viewer in the 2018 season's aggregate viewership for a team grew the attendance in the 2019 season by 0,03%, which is a hundred extra spectators in 2018 would have had a 3% added effect on 2019 attendances.

	<i>Dependent variable:</i>		
	log(ATTENDANCE)		
	(1)	(2)	(3)
HOMEPOSITION	-0.067*** (0.014)	-0.056*** (0.012)	-0.056*** (0.013)
AWAYPOSITION	0.012 (0.013)		

The variable **FORM**, consisting of the position in the table of both the home and away teams, was also found to have a statistically significant effect at the 99% confidence level on attendances for the part of the variable concerning the position in the league table of the home team. This finding is neither new nor surprising (Forrest & Simmons 2006, Buraimo 2007), and follows intuition like **SUPPORT**, due to Finns being often cited as success-crazed population. Moreover, the season of 2019 was special in the sense that the hegemony of Helsinki-based club HJK on the fields was terminated as they endured a rather mediocre season by their standards, leaving the door open for other teams to go up the table. The estimate - which has a negative coefficient due to the data taking the values from 1 to 12 - implies that a move up of one position in the table resulted in 6,7% more attendance according to the full regression model, and 5,6% in the smaller five-variable models.

	<i>Dependent variable:</i>		
	log(ATTENDANCE)		
	(1)	(2)	(3)
UNCERTAINTY	0.212*** (0.072)	0.162*** (0.043)	0.162*** (0.054)

The last of the variables significant at the 99% level, one unit worth of growth of **UNCERTAINTY** is estimated to have affected attendances by 21,2% and 16,2% in the full and smaller regressions, respectively. Interestingly, this result is contrary to the findings of Buraimo and Simmons (2015), but their study involved television-related data and very likely better players to measure the superstar status on. Based on this information, Finns could either be said to enjoy close contests or to respect the visiting team's strong chances of winning the game by coming to see it happen live. After all, distances in Finland are relatively long and for fans of weaker teams,

the presence of good quality football is more of an occasion than to the home fans of those better teams.

<i>Dependent variable:</i>			
log(ATTENDANCE)			
	(1)	(2)	(3)
<b>RAIN</b>	-0.132* (0.075)	-0.135* (0.072)	-0.135* (0.075)

The football season in Finland starts in April, includes a pause in mid-July, and extends all the way to early November. It is no wonder rainy days are experienced within this timeframe in this climate, and their effect was of interest in this study like it was in the research of Lawson, Sheehan, and Stephenson (2008) who did find it to be significantly negative. This regression estimate was in line with their results, the dummy *RAIN* showing statistically significant effect at the 90% confidence level. Indeed, a rainy day during the 2019 season affected the attendance negatively by 13,2% in the full model or 13,5% in the models made up of only the statistically significant variables. This was an interesting, although very intuitive result, due to most of the stadiums in Finland possessing at least one completely covered stand. Perhaps rain is simply an all-round mood killer, although it has been cited to be a preferred condition by many players.

<i>Dependent variable:</i>			
log(ATTENDANCE)			
	(1)	(2)	(3)
<b>SOCIALS</b>	0.0003* (0.0001)	0.0003** (0.0001)	0.0003* (0.0002)

Lastly, we analyze the relationship of *SOCIALS*, that is social media posts about a certain gameweek with the attendance levels. This relationship is the last of the five statistically significant ones in the first regression and interestingly jumps between significance at the 90% and the 95% confidence levels between regressions 1 and 2 before going back to 90% in regression 3 with the robust standard errors. The estimate behaves similarly to *SUPPORT*, with an increase of a hundred social media posts about a gameweek correlating with a 3% growth of audiences. Unfortunately, nothing can be said about causality in this case. Nonetheless, the gameweek-based



measurement can be said to have been successful for the purposes of this thesis and depicts the presence of discussion about football on the internet amongst individuals who may be interested in going to games or vice versa, presence of individuals at games who may be interested in discussing it on social media during the games.

Moving on to other variables that did not have statistically significant relationships with attendance levels, the *SCHEDULE* - or more simply the weekend dummy, found significant notably by Buraimo and Simmons (2015) - displayed a negative coefficient, meaning controversially that there would have been fewer viewers for the so-called prime time football than during the week. The implication is interesting, especially due to growing discontent amongst fans in the last few years about scheduling on many days of the week. The insignificance and the coefficient of the result could however be due to the few outliers, the Helsinki derbies, which are required to be held outside of the weekend due to safety concerns and had the highest attendances outside of the final round of the league.

*CUP* performance did not show statistical significance, in contrast with findings by Smith and Szymanski (1997). Buraimo (2007) and Buraimo and Simmons (2015) mentioned that some cup ties in midweek tend to decrease the attendance at the weekend, and it is indeed what happened in this research with negative coefficients in all cup-related variables, both domestic and European. It is possible that the negative European game dummy represents the letdown felt by supporters as Finnish teams crashed out of the competitions already after the preliminary rounds in the summer or simply the possibility that they had seen enough football for the week, but for the domestic cup, the explanation is not as straightforward. In the Finnish context, the domestic cup performance is perhaps not as important as the competition is largely played behind closed doors in the winter months before the season and often does not represent the true quality of the teams, and the early elimination of attendance “giants” such as HJK could have proven fatal to this variable.

The insignificance of the *BUDGET* and thus the possible superstar effect came as a surprise, as it would have made sense intuitively and based on the literary review (Buraimo 2007). Possibilities are many as to why this variable failed to perform, of which one of the most plausible was already presented in this thesis, the lack of true

superstar quality in Finnish football. Even the mean salary calculated based on the highest seasonal budget of HJK does not give a monthly wage much higher than that of a standard bank employee. Additionally, the lowest budgeted club HIFK is rather known for its fanatic fan base, implicating larger *SUPPORT*-based audiences than what its quality of players would normally result in. Overall, the variance between the budgets around the mean budget is perhaps not large enough in the Finnish context, and the audiences have other reasons to go than their team's high football skills.

Finally, as presented already in the sixties, seventies, and most notably eighties by Bird (1982), football is not a free good and the financial factors at play have to be considered. The *INCOME* variable, divided by capita, performed rather expectedly for the Finnish context of the 21st century. Up here in the North, we do not have many inequalities and the prices of admittance – although unaccounted for in the data - are not steep enough from preventing anyone from going to matches as many times as they would like. Previous literature is famously undecided on the type of good football is to consumers. However insignificant and small the relationship, the negative coefficient of the estimate in the results implies that when income per capita goes up by 100€, attendances suffer by 0,1%, making football possibly an inferior good in Finland, in line with Bird's (1982) work – deemed outdated by Simmons (1996). Could it be that our football is at the same stage of development as it was in England in the 1970s? A sorry line of thought.

## 5. Conclusions

The goal and research question setting of this thesis was to understand the demand of Veikkausliiga football in Finland and the different nuances that form the totality of it, with a special emphasis on the relevance of social media. In the literary review, many previous studies in and around the subject were presented in order to build a basis for the empirical part, which in turn combined them with fresh and relevant data to build an estimation modeling the attendance demand. In chapters 3.1, 3.2, 3.3, and 3.4, studies were presented with findings pointing towards the relevance of competition success, superstars, financial factors, and other attractiveness-related variables in the research of attendance demand. The regression results, displayed in chapter 4.4, implied statistically significant relationships of the attendances with the lagged home attendances from the previous season, the position in the league table of the home team, the uncertainty of the score, the presence of bad weather and the amount of social media activity. Based on both the literary and the empirical parts of this thesis, we can conclude that there are indeed relationships with these five variables, of different magnitudes, but nothing definite can be said about causality especially with regards to social media.

The Veikkausliiga decided to change the competition system before the 2019 season, which saw the league table being divided into a championship league of 6 teams and a relegation league of 6 teams after the preliminary season of playing twice against each team in the league. Based on the statistical significance of the variable about uncertainty, this change of system could be said to have been potentially successful with regards to attracting customers to the more uncertain matches involving teams of a more similar level. This, and the significance of the variables about lagged attendance and the position in the league table of the home team, imply that those Finns who do not go to games out of habit perhaps enjoy a close contest against a strong opposing team for important league positions. The significance of rain is in turn easily explained, and the relationship with social media activity follows intuition and the setting of this thesis.

Limitations of this thesis include the obviously small sample size of one season, the small variation within some of the variables, the relatively simple methodology, the relatively few previous studies about social media in economics, and thus insufficient guidelines on how to treat the social media activity as a variable. The setting of Finnish football is also quite unique in a cultural way, and thus the results of this thesis are not to be generalized to other football leagues with ease.

Home attendances from the previous season, home form, rain, and uncertainty could be treated here as effects, since none of the variables can be influenced in any way by the attendance levels, and the effect of the variables is determined prior to the actual match. Nothing can be unfortunately said about the causality of social media activity, but the finding and definition of the relationship is good enough for the purposes of this thesis. Further research around the topic could possibly focus on studying causality, and more work should definitely be conducted on the rise of football in Finland, perhaps with the help of studies about the MLS, Asia, or other newer markets.

As guidelines towards the Veikkausliiga - who presumably want to grow attendances and interest in their product further - this thesis suggests working with the variables which it can directly influence, that is maximizing the effect of uncertainty of outcome by backing the argument of Lawson, Sheehan, and Stephenson (2008) of the league taking measures to distribute player talent more equally within the league in its own self-interest. Additionally, since a relationship with social media activity does exist, the Veikkausliiga should seek to increase its own and the participating teams' presence on the platforms in order to maximize the potential interest generated through them.

## 6. References

- Bird, P. (1982).** “The demand for league football.” *Applied Economics*, Vol. 14, No 6.
- Boyd, D. & Ellison, N. (2007).** “Social Network Sites: Definition, History, and Scholarship”. *Journal of Computer-Mediated Communication*, Vol 1, pp 210-230.
- Bryson, A. & Frick, B. Simmons, R. (2015).** “Sports Economics: It May be Fun but What's the Point?.” *National Institute Economic Review*, Vol. 232, No 1, pp R1-R3.
- Buraimo, B. (2007).** “Stadium attendance and television audience demand in English league football.” *Managerial and Decision Economics*, Vol. 29, No 6, pp 513-523.
- Buraimo, B. & Simmons, R. (2015).** “Uncertainty of Outcome or Star Quality? Television Audience Demand for English Premier League Football”. *International Journal of the Economics of Business*, Vol. 22, No. 3, pp 449-469.
- Dobson, S. & Goddard, J. (2001).** “The Economics of Football.” *Cambridge University Press*. ISBN: 0521661587.
- Forrest, D. & Simmons, R. (2006).** “New Issues in Attendance Demand: The Case of the English Football League”, *Journal of Sports Economics*, Vol. 7, No 3, pp 247–266.
- Säähavaintoarkisto (2019).** “Päivätilastot”, *Ilmatieteen Laitoksen Sääasemien Arkisto*, viewed 13.5.2020, <[kilotavu.com/fmi-tilastot.php](http://kilotavu.com/fmi-tilastot.php)>
- Kuntaliitto (2019).** “Kuntien Verotiedot”, *Kuntatalouden Tilastot – Kuntien Verot ja Valtionosuudet*, viewed 13.5.2020, <[kuntaliitto.fi/talous/kuntien-verotiedot](http://kuntaliitto.fi/talous/kuntien-verotiedot)>

**Lawson, R. & Sheehan, K. & Stephenson, E. (2008).** “Vend it Like Beckham: David Beckham’s Effect on MLS Ticket Sales.” *International Journal of Sport Finance*, Vol. 3, No 4, pp. 189-195.

**Newey, W. & West, K. (1987).** “A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix”, *Econometrica*, Vol 55, No 3, pp 703-708.

**Oddsportal (2019).** “Veikkausliiga 2019 Results & Historical Odds”, *Odds Archive*, viewed 14.5.2020 <[oddsportal.com/soccer/finland/veikkausliiga-2019/results/](http://oddsportal.com/soccer/finland/veikkausliiga-2019/results/)>

**Peitz, M & Reisinger, M. (2014).** “Handbook of Media Economics”, *Elsevier*. ISBN: 9780444636911.

**Perry, S. (2019).** “Club Licensing Benchmarking Report: Financial Year 2018”, UEFA.

**Poli, R. & Ravenel, L & Besson, R. (2019).** “Attendances in football stadia (2003-2018)”, *CIES Football Observatory Monthly Report*, No 44.

**Rosen, S. (1981)** “The Economics of Superstars”. *The American Economic Review*, Vol. 71, No 5, pp 845-858.

**Ruijg, J. & Van Ophem, H. (2014).** “Determinants of Football Transfers.” *UvA-Econometrics Working Papers*, Vol. 14, No 1.

**Scelles, N. & Helleu, B & Durand, C & Bonnal, L. (2016).** “Professional Sports Firm Values: Bringing New Determinants to the Foreground? A Study of European Soccer, 2005-2013.” *Journal of Sports Economics*, Vol. 17, No 7, pp 688-715.

**Simmons, R. (1996).** “The demand for English league football: a club-level analysis.” *Applied Economics*, Vol. 28, No 2, pp 139-155.

**Sloane, P. J. (1971).** “The Economics of Professional Football: The Football Club as a Utility Maximizer.” *Scottish Journal of Political Economy*, Vol 44, pp. 121-146.

**Smith, R. & Szymanski, S. (1997).** “The English Football Industry: profit, performance and industrial structure.” *International Review of Applied Economics*, Vol. 11, No 1.

**Transfermarkt GmbH & Co. KG (2019).** “Veikkausliiga”, *Transfermarkt*, viewed 23 April 2020, <<https://www.transfermarkt.com/veikkausliiga/startseite/wettbewerb/FI1>>

**Transfermarkt GmbH & Co. KG (2019).** “Premier League”, *Transfermarkt*, viewed 04 April 2020, <<https://www.transfermarkt.com/premier-league/startseite/wettbewerb/GB1>>

**Veikkausliiga (2019).** ”Tilastot, Otteluohjelma”, *Jalkapalloliiga ry*, viewed 23 April 2020, <<http://www.veikkausliiga.com/tilastot/2019/veikkausliiga/ottelut/#stats-wrapper>>

**Watanabe, N. & Yan, G. & Soebbing, B (2015).** “Major League Baseball and Twitter Usage: The Economics of Social Media Use”. *Journal of Sport Management*, Vol 29, pp. 619-632.

## 7. Appendix

<i>VARIABLE</i>	<i>ATTENDANCE</i>	<i>WEEKEND (D)</i>	<i>EUROCUP (D)</i>	<i>INCOME/CAPITA</i>
<i>MIN.</i>	312	0	0	15 466
<i>1ST Q.</i>	1 598	0,25	0	16 060
<i>MEDIAN</i>	2 176	1	0	16 620
<i>MEAN</i>	2 626	0,7468	0,02	18 201
<i>3RD Q</i>	3 124	1	0	20 097
<i>MAX</i>	10 256	1	1	24 531
<i>VARIABLE</i>	<i>SOCIALS</i>	<i>HOME BUDGET</i>	<i>AWAY BUDGET</i>	<i>HOME SUPPORT</i>
<i>MIN.</i>	500	350 000	350 000	961
<i>1ST Q.</i>	673,5	500 000	500 000	1 611
<i>MEDIAN</i>	793	562 776	562 776	1 954
<i>MEAN</i>	820	684 569	684 569	2 144
<i>3RD Q</i>	861	850 000	850 000	2 618
<i>MAX</i>	1 625	1 550 500	1 550 500	3 779
<i>VARIABLE</i>	<i>AWAY SUPPORT</i>	<i>HOME POSITION</i>	<i>AWAY POSITION</i>	<i>HOME CUP</i>
<i>MIN.</i>	961	1	1	5
<i>1ST Q.</i>	1 611	3,25	3,25	5
<i>MEDIAN</i>	2 070	6,5	6,5	7
<i>MEAN</i>	2 162	6,4	6,4	6,8
<i>3RD Q</i>	2 623	9,75	9,75	8
<i>MAX</i>	3 779	12	12	9
<i>VARIABLE</i>	<i>AWAY CUP</i>	<i>UNCERTAINTY</i>	<i>RAIN</i>	
<i>MIN.</i>	5	1,33	0	
<i>1ST Q.</i>	5	1,85	0	
<i>MEDIAN</i>	7	2,38	0	
<i>MEAN</i>	6,8	2,56	0,4	
<i>3RD Q</i>	8	3,01	1	
<i>MAX</i>	9	5,77	1	

Table 1: Descriptive statistics of the variables.



	<i>Dependent variable:</i>		
	log(ATTENDANCE)		
	(1)	(2)	(3)
WEEKEND.D.	-0.008 (0.085)		
EUROCUP.D.	-0.317 (0.231)		
INCOMeCAPITA	-0.00001 (0.00001)		
SOCIALS	0.0003* (0.0001)	0.0003** (0.0001)	0.0003* (0.0002)
HOME BUDGET	0.00000 (0.00000)		
AWAY BUDGET	-0.00000 (0.00000)		
HOMESUPPORT	0.0003*** (0.0001)	0.0004*** (0.0001)	0.0004*** (0.0001)
AWAYSUPPORT	0.00001 (0.0001)		
HOMEPOSITION	-0.067*** (0.014)	-0.056*** (0.012)	-0.056*** (0.013)
AWAYPOSITION	0.012 (0.013)		
HOME CUP	-0.024 (0.029)		
AWAY CUP	-0.044 (0.029)		
UNCERTAINTY	0.212*** (0.072)	0.162*** (0.043)	0.162*** (0.054)
RAIN	-0.132* (0.075)	-0.135* (0.072)	-0.135* (0.075)
Constant	7.265*** (0.613)	6.596*** (0.222)	6.596*** (0.226)
Observations	154	154	154
R <sup>2</sup>	0.457	0.423	0.423
Adjusted R <sup>2</sup>	0.403	0.403	0.403
Residual Std. Error	0.435 (df = 139)	0.435 (df = 148)	0.435 (df = 148)
F Statistic	8.368*** (df = 14; 139)	21.676*** (df = 5; 148)	21.676*** (df = 5; 148)
<i>Note:</i>			*p<0.1; **p<0.05; ***p<0.01

Table 2: The full results of the regressions.