## Data and Methodology

## Data collection

The website: <http://www.football-data.co.uk/data.php> has weekly updated datasets for all the important leagues around the world. The datasets include statistics on full-time and half-time results, amount of shots and shots on targets of both home and away teams as well as the number of yellow and red cards and corners for each team. Furthermore, betting odds for the matches at major bookmakers are provided. This includes odds on the 3 main bet types in football betting. The classic 1x2 bets where you can bet on the final match outcome. Asian Handicap bets which allow you to give a team an imaginary handicap to increase or decrease payouts when betting on that team. Finally the over/under bets which revolve on the number of goals scored in the match. Our main interest lies in the influence of uncertainty in home advantage on betting margins so we decide to only consider the 1x2 odds in our subsequent empirical analysis. For each of the major 11 leagues in Europe there is historical data available up to the season 1994/1995. However the odds data per match is only available from the season 2000/2001 onwards. Therefore we only incorporate data from season 2000/2001 or later into our analysis. All the datasets are split per season, per country and per league level in countries with multiple tiers. We aggregate all these seperate datasets into a large final dataset since we are primarily interested in overall home advantage and bookmaker margins, not necessarily country level differences.

We need to include data on team strength in order to control for scheduling bias in our analysis. The most comprehensive index of team strength as fas as we know is the SPI index used by : project538. Their SPI index is constructed by ….. Their website provides weekly updated dataset on historical matches with SPI data, which can be merged with our previously constructed dataset with historical match outcomes and betting odds. Furthermore, we need to control for … and …. in our model, which can be done by including a measure of … into our analysis. Adding … and … into our dataset will aid in controlling for … and … . Finally a proxy for … will be incorporated into the analysis to increase the … of our analsyis.

Bookmaker margins as a percentage can be calculated as the sum of the inverse of each outcome odds and then minus 1. The calculation can be captured in the following formula: ((1/Oh + 1/Od + 1/Oa) -1) \* 100 where Oh represents the european odds value for a home win, Od the european odds value for a draw and Oa the european odds value for an away win. For example, suppose a fictional match between two teams with the following odds. Home team is favourites with odds of 1.8 for winning the match. A draw is priced at odds of 4.1 and an away win gives odds of 3.1. The corresponding bookmaker margin is ((1/1.8 + 1/4.1 + 1/3.1) -1) \*100 = 12.2%

## Variable operationalization

## Methods of analysis

## Descriptive statistics

Table … provides the descriptive statistics for the variables mentioned in section …

## Regression Equation

As mentioned in section 3.3, we deem … as most suitable method for analysis of our dataset. Therefore we use a … regression on our variables of interest. … is a statistical method that is based on the … estimation of .. . relationship between … and … … estimates the relationship between the .. Dependent variable(s) and … Independent variables by …

Our main model is illustrated by the following equation: