Charles University in Prague

Faculty of Social Sciences

Institute of Economic Studies

MASTER THESIS

Fair Bets in Sports Betting

Author: **Bc. Jiří Jansa**

Supervisor: **Mgr. Barbora Pertold-Gebická, Ph.D.**

Academic Year: **2011/2012**

Declaration of Authorship

The author hereby declares that he compiled this thesis independently, using only the listed resources and literature.

The author grants to Charles University permission to reproduce and to distribute copies of this thesis document in whole or in part.

Prague, January 9, 2012

Signature

Acknowledgments

The author is grateful to superviser Barbara Pertold-Gebická for her advices concerning the structure and aims of the work.

The author is grateful to his mother Alena Jansová and his girlfriend Miroslava Motyčková for their never-ending support.

Abstract

Market efficiency and existence of profitable strategy are the most frequent analysis in the research concerning betting on sport events. This thesis covers both these topics on the dataset (20 betting offices) of Czech ice-hockey league from 2004 to 2010. The theoretical part presents development of models of individual decision-making under risk and uncertainty, models of equilibrium on the betting market and several definitions of market efficiency (Fama and Sauer as authors of these concepts) on these markets. The statistical part is testing difference in margins of betting companies for 3 possible outcomes of game, convergence in quoted odds across betting offices, arbitrage opportunity and correspondence of quoted odds to the real probabilities (linear and non-linear). Simple model of perfect market might be by all these tests rejected, since there is no constant return from betting on all outcomes, betting offices differ in margins, quoted odds do not correspond to the real probabilities and arbitrage opportunity is not disappearing. Second empirical part is devoted to the search for profitable strategy. Using 14 explanatory variables and various statistical methods (linear probability model, logit model, multinomial logit model), author is trying to beat bias in odds and find long-term profitable betting strategy. Returns from strategies are usually positive, but for the second part of dataset are getting smaller. Still, hypothesis of efficient betting market might be rejected by this methodology as well. During the thesis, author suggests several improvements in analysis and potential properties of general model for this market.

JEL Classification	D01, D81, D80
Keywords	betting market, efficiency, profitable strategy, arbitrage

Author's e-mail	`jiri.jansa@centrum.cz`
Supervisor's e-mail	`bgebicka@cerge-ei.cz`

Abstrakt

Testování efektivity sázkových trhů a hledání ziskové strategie jsou nejčastější témata při analýze sázkových trhů. Obsahem této diplomové práce je obojí, použitým sázkovým trhem je česká hokejová extraliga v letech 2004-2010 s kurzy 20ti nejvýznamnějších sázkových kanceláří. Teoretická část shrnuje dosavadní vývoj modelů individuálního chování za rizika a nejistoty, prezentuje celkové modely sázkových trhů a několik definic efektivity sázkového trhu (autory těchto konceptů jsou Fama a Sauer). Ve statistické části jsou testovány rozdíly v zisku pro sázkové kanceláře ze 3 možných výsledků zápasu, konvergence kurzů napříč sázkovými kancelářemi, možnost arbitráže a vztah mezi vypsanou pravděpodobností výsledku a pravděpodobností reálnou (lineárně i nelineárně). Základní model sázkové trhu - model dokonalé konkurence s racionálními sázkaři - je všemi těmito testy zamítnut, jelikož zisky z možných výsledků zápasu jsou různé, konvergence kurzů je zamítnuta, možnost arbitráže se nezmenšuje a vypsané kurzy (pravděpodobnosti) neodpovídají realným pravděpodobnostem. Druhá část je věnována hledání ziskové strategie. Za použití 14ti vysvětlujících proměnných a různých statistických metod (lineární pravděpodobnostní model, logit model a multinomial logit model) se autor snaží porazit bookmakery a najít dlouhodobě ziskovou sázkovou strategii. Návratnost je ve vetšině případů pozitivní, ale pro nová data s klesající tendencí. I z výsledků hledání ziskové strategie může být efektivita českého sázkového trhu zamítnuta. V průběhu práce autor navrhuje několik vylepšení analýzy a potenciální vlastnosti obecného modelu tohoto sázkového trhu.

Klasifikace JEL	D01, D81, D80
Klíčová slova	sázkové trhy, efektivita, zisková strategie, arbitráž

E-mail autora	`jiri.jansa@centrum.cz`
E-mail vedoucího práce	`bgebicka@cerge-ei.cz`

Contents

List of Tables

List of Figures

Master Thesis Proposal

Author	Bc. Jiří Jansa
Supervisor	Mgr. Barbora Pertold-Gebická, Ph.D.
Fair bets in sports betting	Fair Bets in Sports Betting

Topic characteristics Using several statistical methods, I will construct models for prediction of sports (ice-hockey) results. Comparing results of these models with the odds offered by betting offices, I will search for the fair-bets, i.e. bets where the expected value is above 0. Searching for any pattern in these fair-bets situations might lead to interesting conclusions, especially when some of the betting strategies are included into the analysis. The dataset will cover ice-hockey leagues and several betting offices, used pre-game informations depending on availability of data and mainly on transformation of qualitative aspects into quantitative ones.

Hypotheses At this early stage of my research, there are several questions, which could be potentially answered in the thesis. First, I will be searching for certain characteristics of games, when the betting companies are more likely to offer fair-bets. Second, main emphasize will be given to several tests of market efficiency on the betting market. Third hypotheses might be that games with more volatile movements in bets offer lower EV for players.

Methodology Applied statistical work with binary and multinomial logit regressions. By using several explanatory variables (standing of the team, home ground, moves of the betting rate, difference of the betting rates between several betting offices) I will get estimation of the probability of each outcome (win, remis, lose) of the game. Using microeconomic theory of expected value, I will search for patterns in fair-bets and will test, whether betting market is the efficient one.

Core bibliography

1. Even, W.E. & N. Nobel (2006): "Testing Efficiency In Gambling Markets." *Applied Economics*, 1992.

2. Ben-Naim, E. & F. Vaquez (2006): "Parity and Predictability of Competitions." *Journal of Quantitave Analysis in Sports*, Volume 2, Issue 4

3. Goddard, J. & A. Asimakopoulos (2001): "Forecasting Football Results and the Efficiency of Fixed-Odds Betting." *Journal of Forecasting*, (23): pp. 51–66.

4. Havard, R. (1997): "Prediction and Retrospective Analysis of Soccer Matches in a League." *Statistics*, 10

5. Sauer, R. (1998): "The Economics of Wagering Markets." *Journal of Economic Literature*, 36(4): pp. 2021–2064.

6. Spann, M. & A. Skiera (2009): "Sports Forecasting: a Comparison of the Forecast Accuracy of Prediction Markets, Betting Odds and Tipsters." *Journal of Forecasting*, 28: pp. 55–72.

7. Vlastakis, N. & G. Dotsis, R. Markellos, (2009): "How Efficient is the European Football Betting Market ? Evidence From Arbitrage And Trading Strategies?" *Journal of Forecasting*, 444: pp. 426–444.

Jiri Jansa

Barbara Pertold-Gebicka

Chapter 1

Introduction

Betting markets for sport, political or other social or nature events (general name connecting these activities might be event betting) have been rapidly increasing on accessibility last years. The availability of internet betting leading to lower role of state permission[1], possibility of higher liquidity, live transactions and immediate re-betting, those were some of reasons why betting has grown from minor black market passion into large legal market. This development might be well documented by the size of market - in the European Union during 2010, around 10 billions EUR were bet in total. Leading Central European fixed-odds betting operator - Fortuna - received 384,2 millions EUR at stake, with profit being 92,8 millions EUR in 2010. And in the Czech Republic, around 8,8 billions CZK were at stake in total, from that around 1 billion CZK being bet on the Czech ice-hockey league - extraliga. The legality and accesability have made sport betting not only public entertainment and source of income for successful players, but rich source of sponsor contracts for sport federations, teams and single sportsmen.

The objective of this thesis is to present 2 hot topics in the field of betting - test of market efficiency and search for profitable betting strategy - and examine them on the dataset of Czech ice-hockey league. Since financial and betting markets are by its nature similar (prediction of future event, information available for general public), tests of market efficiency on the betting market are inspired by the tests on the financial markets, where the examination of this problematic has had over one hundred years long tradition and methods used for that are rich. The search for profitable strategy is in my thesis based on

[1]In some of the EU countries, on-line betting is restricted only to holders of national license (Czech Republic), in some countries, it is under the EU rules (Great Britain). But even companies without national license are thanks to internet offering on-line betting.

the use of statistical models - on the potential possibility of prediction from past values. This search for profitable strategy is as well used on the financial markets and is one of the test of market efficiency.

After the introduction, Chapter 2 presents historical difference between lottery and bet. These activities are usually put into one basket despite significant methodological difference. This difference might be well indicated by the historical origin of both activities. Chapter than continues with types of bets and similarities between financial and betting market. Since both these markets are devoted to the prediction of future events, there are similar fields of research and similar methods used.

Chapter 3 concerns models of individual gambling behavior and gambling markets. While development of models that are explaining individual behavior under risk and uncertainty has started 200 years ago, the models for general equilibriums on betting markets are from recent period. The models for individual betting behavior are trying to explain reasons why and under which conditions bettors are accepting bet, since the basic model of decisions based on expected value model (i.e. bettors are deciding according to the expected outcome obtained by weighting all potential outcomes of event by their corresponding probabilities) is not able to predict most of the empirical results. The reflection of these models in my empirical results is complicated, since my dataset does not contain amount of bets.

Chapter 4 presents several definitions of market efficiency and previous research in the field. Since the efficiency on the betting market is not one-common-definition concept, the previous research is rich, both in used methods and in results. Market efficiency is tested to find out, whether there is some general pattern in the "behavior" of market. The methodology is to define the market efficiency (but by several ways) and then - using statistical tests - check the potential correspondence of these definitions to the empirical reality. The practical use of testing market efficiency is for betting, since each ineffectiveness - if rightly discovered - leads to potential profitable betting strategy.

From the Chapter 5, the empirical part starts with several tests of efficiency concerning odds and ex-ante probabilities. Topics are various. First, convergence of odds of various betting offices is examined together with estimations of profit margins from all outcomes. In the model of perfect market and fully rational bettors, the betting offices should be forced to equalize their odds and margins. I am statistically testing if market behaves like that. Second, potential existence of arbitrage betting opportunity (i.e. situation when the bettor might earn profit without bearing any risk just by spreading bets between several betting offices) is examined. Last sections in this chapter concern the correspondence of quoted probabilities to ex-ante probabilities. I am searching potential bias in odds and intervals, on which - if on any - quoted odds do not correspond to the real ones.

In the Chapter 6, the profitable betting strategy is being searched for the dataset of Czech ice-hockey league Extraliga. The existence of betting strategy leading to profit is one of the indicator of market inefficiency, the results from this chapter are then interconnected to the tests of market efficiency from the previous chapter. For the search, I am using various statistical methods, from the linear probability model to the multinomial logit model. As explanatory variables, various indicators of potential form of both teams are used. In this chapter, I am as well commenting potential problems with this approach, since the results might be misleading.

The results of my thesis are summarized in Chapter 7. Special emphasis is given to the unique results of empirical research and to the potential extension of analysis.

Chapter 2

Betting

2.1 Betting vs. lottery

In this section I am presenting difference between betting and lottery (gamble). Both these activities are put into one basket despite significant methodological difference, which has consequences for my analysis and conclusions.

The difference between bet (for example on the prediction market[1] or sport betting and gamble (for example slot machine) lies in the existence/nonexistence of known probability distribution for outcomes. In the case of gambling on slot-machine, playing roulette or lottery, the result is outcome with ex-ante known probability (in the case of lottery publicly, in the case of slot machine privately by provider of the game) distribution[2].

Expected value of playing a lottery or roulette is for player without any insider information about algorithms of the machine or without the knowledge of winning lottery ticket negative. Playing such game is irrational from the point of view of Arrow´s expected value theory and as I will mention in the Chapter 3, literature and models trying to explain this phenomenon are rich.

In the case of betting (betting on sport, social or nature events), the probability distribution of potential results is not known publicly and (usually) even not privately. Each bet is then a duel between a player (bettor) and a bookmaker (or another player) in making opinion about this random distribution. The role of the bookmaker is more (or less) complicated by the fact that he should anticipate not only the probability distribution of each potential result

[1] In this thesis, I am using term prediction market for person-to-person betting, i.e. betting without bookmaker, where the odds are result of trading of several players.

[2] Playing electric roulette leads to expected loss between 5.1 to 7.9 percents, slot machine leads to loss 20-30 percents and lotteries around 50 percents.

of the game but also distribution of bets of players on all possible outcomes. As it is mentioned in the Chapter 3 in the research made by Levitt (2004) and Kuypers (2000), if bookmakers are better in judgment than most of the bettors then their most profitable strategy is to bias odds to attract even more bettors.

In the case of gambling or lottery, we can speak of fair bets only in the case of general equilibriums with utility functions of players subsidizing expected loss by the joy of playing or by the special shape of utility function. While the decision to gamble (or to buy lottery) is then without some private information irrational (from the point of view of Arrow´s expected value theory), decision to bet might be rational and profitable. Each player has a chance to make better judgment than bookmaker about the probabilistic distribution of results of game and might use this knowledge for profitable strategy[3].

To even strengthen the difference, I might go into the history. The lottery started as a tool for state to finance projects - tool to collect money from public. As Thomas Jefferson once said: "Lottery is wonderful thing: It lays taxation only on the willing." Contrary to that, bet originated as a contract between 2 gentlemen discussing the future of some event. The lottery (gamble) is more similar to game, bet is to prevent cheap talk about future events - to take responsibility for own opinion.

Since the probability distribution of outcomes of each game is not known in advance, the conclusion from this section is that any statistical analysis of betting uses the assumption of the possibility of judging future from the past (the past patterns are somehow reflected in the future patterns). As will become evident in the summary of previous empirical research, this assumption does not have to always hold. One pattern (for example betting on home underdogs being the profitable strategy) changes for another with the new data coming. Task of any researcher is not only estimation of probability distribution function of results, but moreover anticipation and explanation of any change in followed patterns.

2.2 Types of Bets

This section briefly presents types of bets, recalculations between them and new form of betting, so called prediction markets. Since the possibilities of

[3]During 1990s, there were at least 2 Czech betting offices that went bankruptcy.

betting are various, I am presenting on which kind of bets I am aiming and what it means for my analysis.

Sauer (1998) mentions 3 main types of wagering markets (markets where players are betting on one outcome out of several possibilities): pari-mutuel odds, odds offered by bookmakers and point spreads offered by bookmakers. I can add the 4th form, which is nowadays becoming more common - prediction markets. The pari-mutuel odds are used in sports where the competitors finish in ranked order (for example horse racing). The bookmaker collects money from bettors and after subtraction of his provision, he divides the rest of money between bettors according to the results of competition. This division has various rules (one example - winners take the full bank). This makes the bookmaker risk free, since he does not face any position in the betting situation. His role is just provision of betting opportunity and connection of various bettors with different views on outcomes. The fact, that no player knows the potential win when placing his bet, makes this kind of betting specific and with its own models.

In Europe, the most common form of betting system for sport events - usually 2 teams playing against each other - is the odds system. Bookmakers publish their odds (number according to which bettor receives his income after recalculation) for each of possible outcomes depending on their beliefs and bettor decides whether to accept the odds offered (and bet) or not. In England, the most common form of fixed-odds betting are so called fractional odds (for example x:1), which means that from betting 1 unit, bettor receives profit x units (i.e. x + his 1 unit back). In continental Europe and for on-line betting, more common form of citing the odds is decimal odds (for example 1.25). The pay-off for player is obtained by multiplication of his bet by decimal odds, and obtaining by that the reward from bet. Fractional odds might be easily recalculated to the decimal form by formula:

$$\begin{gathered} odds_{fractional} = odds_{decimal} + 1 \\ 1.25 = 1 : 4 + 1 \end{gathered} \tag{2.1}$$

The way it works is the same for both systems. Bettors should recalculate quoted odds to the estimated probability of outcome and decide according to their view whether to bet or not.

All quoted odds might be recalculated to ex-ante probabilities by following formula:

$$p = 1 \div (odds) \tag{2.2}$$

The lower the estimated probability of outcome, the higher the quoted odds. As will become evident from the theoretical models or from my empirical analysis, p (ex-ante probability) from Equation 2.2 does not have to correspond to ex-post probability.

The third form of betting are point spreads. It is used for sport events where the followed result is not just in the form winner/looser. Bettor can be not only betting on the win/tie/lose outcome but as well on the margin of goals or points of winner to looser. These point spreads are nowadays offered in decimal odds as well, but because of limited data availability, I am not concentrating in my work on this kind of betting.

Most rapidly growing betting markets are because of IT development predictions markets (to name some: Intrade.com, Betfair or Bet2give). There is no role for bookmaker, or better to say - each player can be bookmaker for others. Players are offering odds for other players. To make the picture more concrete - on the prediction market Intrade.com, the players are trading the events with just 2 outcomes, true or false (one example of betting occasion: Will Barack Obama be re-elected President of US in 2012 ?). Player that is in the position of bookmaker offers minimal sum (on the scale 0 to 1, in the example of Barack Obama current price is 0.52) that he is willing to accept to make a deal. Those players on the side of bettor might decide, whether to accept the price or not. If the event is true, the bookmaker pays 1 to bettor, if the event is not true, the bookmaker receives the payment 0.52 from the bettor. The more probable the event is believed to be, the higher portion of 1 the bookmaker demands for trade. These betting markets are together with live betting most similar to financial markets since each new information leads to change in price offered.

In my work, I am using decimal odds for the final result, since it is the most common form of betting in the Czech Republic. The odds from prediction market are presented by odds of company BetFair. Prediction markets are from one point of view - correspondence of quoted probabilities to the real probabilities - said to be more effective than common market with pair bettor bookmaker, but as will be evident from the statistical analysis, it is not the case for the Czech ice-hockey league.

2.3 Financial and Betting Markets

As I have already stated in the introduction, the topics of market efficiency and of search for profitable strategy are inspired by the same topics from the financial markets. The reason for that is that both markets are in several attributes similar. Between similarities we can count large number of investors (open for everybody) and free access to rich information sets. On the internet, there are specialized websites with live information about injuries, potential forms of players and expected line-ups. Another attribute - fast repetition of trials (you can bet immediately again) - contributes to good learning opportunities and because of that - according to Williams (2005) - judgment of bettors should be improving faster than of investors on financial markets.

The most important difference between these 2 markets is in the clear ending point of event. On the betting (prediction) markets, the price terminates, i.e. game has clear ending point, when the outcome is realized. Contrary to that, financial markets are running infinitely, the price does not terminate. Especially this point makes betting market even better for testing market efficiency (that will be defined later on) than financial markets, since it is easier to separate which information the odds contain (Thaler & Ziemba (1988)) and with this knowledge test market efficiency.

Despite similarities mentioned above, there are technical differences that make betting market not fully comparable to financial one. Mainly, since the odds are cited before the beginning of the game, their values should correspond to all available information before the game. But each outcome of sport event consists only partly from the predictable part, the second part is not possible to anticipate (influence of randomness). Thus the test of efficiency of betting market similar to the financial market should be based on the live data (odds, information about actual form of each player) during the game, not on the data before the game as is common in most of the research (there are exceptions, for example Smith *et al.* (2006), Wolfers & Zitzewitz (2004) or Manski (2004) studying prediction markets, where the prices are moving on-line as a response to new information coming).

In my view, the betting market has grown to the size when it is interesting to know its properties per se, not with strict aim of comparing this market to the financial one. Market and potential financial rewards from profitable strategy have increased so rapidly that the qualities of betting market are useful information without the prior use as proxy for financial market.

Chapter 3

Theoretical Models

This chapter summarizes development of theoretical models concerning decision-making when facing betting situations. The first section - Section 3.1 - presents models of individual decision-making, the second section - Section 3.2 - concerns aggregate view - models of general equilibrium on the betting market.

3.1 Theoretical Models of Individual Gambling Behavior

In this section, I am briefly presenting models of individual behavior under risk and uncertainty. The term risk is used for the outcome from known probability distribution, term uncertainty is used for outcome from unknown probability distribution. This corresponds to the difference between lottery and bet from the previous chapter. Despite this methodological difference, the models for individual behavior of gambler and bettor are similar.

The summary of development of models of individual gambling behavior might be found in Sauer (1998), Diecidue *et al.* (2004) and Peel (2008). The standard theory proposed to describe decision-making of player was model of expected outcome (value), which might be dated back to Pascal and De Fermat. In the model, player is linearly weighting all potential outcomes by known probabilities and decides according to the expected value about the advantage/disadvantage of playing. But such model of decision-making leads to strange results - for example, there are games when player would have been willing to pay infinity to play the game. Example of such game is so called St.

Peterburgs paradox[1]. The expected value model as general model of decision-making for betting was then by empirics falsified.

Bernoulli (and after him, more rigorously Neumann and von Morgenstern) came with the idea to change the theory of expected value to the theory of expected utility. In this model, the player is deciding according to the expected utility from wealth. By suitable shape of utility function (marginal utility of wealth has to be decreasing (for example natural logarithmic utility function)), he managed to model St. Peterbourgs paradox with results corresponding to empirics. But even this approach leads to predictions not corresponding to the reality - bets with expected negative return or even fair bets should not be accepted by players at all. Since the same people are obtaining lottery tickets with expected negative return and together with that insurance against loss (people are risk-seeking and risk averse at the same time), the model of Bernoulli was by empirics falsified.

To model the empirical results contradicting Bernoulli´s result - concave individual utility function leading to no bets - Friedman & Savage (1948) come with the concept of so called S-shaped utility function of wealth. The utility function of wealth for the low income group has first concave part that after some value of wealth changes to convex shape. This shape of utility function explains logical contradictions derived/arising from the concave (risk-averse) utility function on the whole interval of wealth. The division of people into different subgroups (Friedman and Savage are suggesting these subgroups to be representing different social classes) depending on wealth can explain anomalies. Extension of model was done for example by Bhattacharya & Garrett (2008).

Markowitz (1952) or for example Hirshleifer (1966) come with the critique of the S-shaped utility curve. Such model leads to other strange conclusion - too risky behavior of people on the concave-convex border (the most optimal behavior leads to betting until one‘s is totally poor or totally rich) and no bets by poor people. Again, the model of Friedman and Savage was contradicting empirics, which led to its falsification. The model Markowitz suggests instead is reversible to the S-shaped curve and concerns changes in wealth, not in its absolute value. First, for low changes in wealth, the utility curve is convex (marginal utility of wealth is increasing), since the potential losses are too low

[1]The coin is thrown and if on n-th trial the head comes, player receives 2^n outcome. Since the expected value of such game for n going to infinity is infinity, the player should pay any sum to participate. In reality, it is not the case and according to the research, the average sum paid is 10.

for player. After some value (for each player different), the utility function becomes concave since the sum to lose is high. The reverse than holds for losses - for small ones, player is risk-seeking, for big ones, player is risk-averse (and obtain insurance). Each player has his own utility function based on the initial wealth but the convex/concave shape is the same for all. St. Peterbourg´s paradox is solved by upper-bound of bet by each player. As will become evident after introducing Cumulative Prospect Theory (CPT), the predictions of Markowitz´s model are similar to this theory as well.

As a response to the critique that gambling is not wealth oriented activity but mainly pleasure (for example Samuelson (1952)), one branch of research and models have been devoted to this quality. Fishburn (1980) and Conlisk (1993) model and add into the utility function the pleasure from gambling suitable for explaining small pay-off gambles and large-prize lotteries.

With connection of previous research (nonlinear utility functions, nonlinear weighting of probabilities) and with similar results to Markowitz model of convex/concave utility function come Kahneman & Tversky (1979). The theory called Cumulative Prospect Theory (CPT) was mainly trying to explain empirical facts contradicting expected value and expected utility models - for the small probability outcomes, the player is usually risk-seeking, for the medium to large probabilities, the player is usually risk-averse. According to the CPT, the value function depends on changes in wealth (this fact is used as well by Markowitz) and they (Kahneman and Tversky) are suggesting use of probability weighting function that overweights small probabilities and underweights medium and large probabilities. This nonlinear valuation of probabilities corresponds to the empirical research done (summary of this empirical research might be found in Wu *et al.* (2004)).

In 1980s, the so called regret theory was introduced by Loomes & Sugden (1982). By adding into the utility function the variable for regret (which is the difference between outcome received and outcome that would have been received under some other choice), they were able to give similar predictions to the CPT or Markowitz´s model of convex/concave utility function. Together with models mentioned above, there were other fields of research connected to betting - neuron networks (Breiter), psychological reasons as for example emotions (Brandstätter) or inter temporal decisions-making (Sagistrano).

In the empirical research concerning the sport´s betting, only low number of these models has been used so far. Most often is the market efficiency (which will be defined in the next chapter) tested with the most primitive model -

expected value model. In such environment, bettors should not accept any bet if the expected value of that bet was negative. Since bettors are betting, it is evidently not the case and there might be several explanations for that.

First, expected value model is not the right model for the prediction of behavior of bettor. After at last 50 years of model development, the variety is rich. For most of the bettors, the model of expected utility function with the S shape suggested by Markowitz or the expected utility function with the variable adding joy from playing, will explain a lot of observable phenomena. Bettor is then betting on low probability outcomes, since he does not face huge losses but has a chance for huge profit. Players are combining games into 1 bet and by that, they are increasing profit and at the same time increasing risk, that some of the games included will not finish with success (and will spoil ticket). Sport-betting becomes similar to lottery - player despite being aware of negative expected value is still betting since he is on the risk-seeking part of his utility function. As will become evident from results of nonparametric regression for Czech betting offices, this explanation is at least for part of market improbable, since odds of Fortuna and Tipsport for low-probability outcomes are in the case of Extraliga league biased upwards.

Second explanation is simple, bettors might be just wrong in estimating probabilities of outcomes. This bias might not be linear and might depend on the ex-ante probability. Such bias might be for example favorite-longshot bias (bias in favor or against favorites). Since various betting strategies across various leagues lead to different losses (as will become evident in the summary of empirical research), this potential bias does not seem to have general pattern and might differ year-to-year (and from league to league).

My dataset does not allow me to test and answer which model of individual decision-making and with which characteristics is the most suitable for explaining betting behavior. Still, from the results of empirical analysis in next chapters, I can get picture about empirical facts that the right model should explain.

3.2 Models of Betting Markets

In the previous section, I have presented models of individual behavior. In this section, I am presenting models of betting markets. This section adds aggregative view, the topic concerns existence of general equilibrium on betting market. This topic is connected to the following chapter, where I am presenting

Sauer´s definition of market efficiency. One (of three) of the points states that if the empirics corresponds to the prediction from explicitly stated model, market might be called efficient. And in this section, such models are presented.

While there is a rich literature concerning testing of market efficiency from the view of prices (odds) on the market, the models of whole market (with amount of bets, maximization of profit by bookmakers and preferences of bettors) are rare. One example of such model is in Kuypers (2000), who is modeling quoted odds by bookmaker as a response to the bias in bettors expectations. The result of his model is that under the assumption of bookmaker maximizing profit, the quoted odds are not efficient (do not correspond to the objective probabilities). The second result of this model is that bookmaker, when maximizing profit, is risk-taker (and can choose to have lower profit with less risky behavior). The odds quoted by such bookmaker are (from one of the definition) not efficient.

Levitt (2004) presents similar model to Kuypers with the same outcome - bookmaker biasing odds to increase profit. His dataset consists of amount of bets. Using them, he tests the results of Kuypers and his model of biased odds as being the most profitable for bookmaker. The odds are not quoted to balance amount of bettors on all outcomes, but are biased to maximize profit of bookmaker.

Another model of equilibrium might be found in Woodland & Woodland (2001). Authors are trying to explain why point-spread betting for American football is more common than odds system. Their derivation finish in another equilibrium, when the use of point-spread betting is the result of risk-averse bettors and profit-maximizing bookmaker. Under the condition of risk-averse bettors, bookmaker´s profit is higher in the system of point-spread betting than in common odds system.

For the horse betting market, general equilibriums on the betting market might be found in Sauer (1998) or in Quandt (1986). Since the nature of these markets is different (bookmaker is not risk-taker), derivations are different and equilibriums have different properties then in the previous cases.

3.3 Summary of Chapter 3

From the first section, I can conclude that there does not exist any general model suitable for explaining all the empirical facts concerning betting. The predictions from models are various and sometimes in contrast to each other.

But from their - still particle but growing - success in giving right predictions, I can conclude, that the research program for betting behavior might be called progressive according to the methodology of science of logician Imre Lakatos. Still, none of the models is able to explain all empirical facts that will follow from the statistical analysis.

Models of general equilibrium on the betting markets are so far rare in the literature. The reason for that might lie in the difficulty when obtaining dataset with amount of bets and betting history of each bettor. Such dataset allows discovering potential biases in betting behavior of players and use of this knowledge for profitable strategy. Without these information, any analysis of betting market is not complete, since one restricts himself only to several models of betting market. More about that follows in the Section 4.1.

Chapter 4

Efficiency

In this chapter, I am presenting various definitions of market efficiency, ways of testing and previous research in the field. Definition and testing of market efficiency on the financial markets can be dated back to 1900 to the work of Bachelier. Summary of methods used and obtained from 1900 to the year 2004 might be found in Williams (2005). The standardization of various topics in the field comes from Fama (1969) for financial markets and from Sauer (1998) for betting market.

Fama (1969) defines, as summarized in Figure 4.1, three types of market efficiency on financial markets - weak-form, semi-strong and strong form.

Figure 4.1: definition of market efficiency

- weak form of market efficiency - current prices reflect all weak (past) information, no chance for prediction from past price movements
- semi-strong form of market efficiency - weak form + new public information impacts on security prices instantaneously and in an unbiased way
- strong form of market efficiency - semi-strong form + new private information impacts on security prices instantaneously and in an unbiased way

In the betting-market literature, the border between weak form and semi-strong form is not sharp. In some of the work, the weak form of market efficiency is defined just by impossibility of prediction from past prices (but not from past information). The semi-strong form than by including past public information into analysis.

Sauer (1998) comes with different approach of market efficiency on the betting market. He examines horse betting market and uses three different conditions, which needs to be satisfied for the betting market to be called efficient.

Figure 4.2: Sauer´s definition of market efficiency

- constant (and same) returns from betting on each outcome
- absence of a profit opportunity (arbitrage, profitable strategy)
- existence of equilibrium of an explicitly stated model

First, under the condition of fully informed agents maximizing wealth with identical risk-neutral preferences, returns from various betting strategies should be constant. Sauer derives this rule for horse betting market, when the bettors are reacting to win/losses by shifting to another - more profitable - betting strategy.

Second point states that bettors should not have the possibility of profitable strategy. This point is similar to the Fama´s definition of market efficiency, i.e. non-existence of profitable prediction of future from past values. Despite its relative simplicity, it is the most common test of market efficiency.

The third point presented by Sauer is unique compared to Fama and concerns existence of equilibrium (but different than in the previous cases) on the betting market. I have devoted next section - Section 4.1 - to that.

It is enough if just one of these points is fulfilled and market might be called efficient. The most restrictive from these Sauer´s points is the first one - constant returns from all betting strategies. In the literature, such kind of test is usually performed together with the search for profitable strategy. The third of Sauer´s point is least restrictive, since it allows nearly any empirical result. It is enough if this empirics is repeatedly predicted by explicitly stated model.

In the previous research, there is no clear united methodology for testing efficiency on the betting market. Based on that, my work than consists of more tests than only of one general test giving clear answer about the efficiency on the betting market.

4.1 Equilibrium of Model

Sauer´s last point states that if there exists explicitly stated model and prediction from this model corresponds to the real data, then the market is in the equilibrium and is said to be effective. Market should behave according to this pattern repeatedly. Such equilibrium models for odds system for sport betting and for horse betting markets were presented in the previous chapter. Levitt and Kuypers are deriving equilibrium where the quoted odds do not correspond to the ex-ante probabilities. Bookmaker using the bias in the bettor´s judgment sets the odds to maximize his profit. From derivations follows that potential risk of bookmaker is not minimized (as a price for profit maximization). Levitt´s paper contains dataset with amounts of bets for each outcome. He is testing various hypotheses with the use of this dataset with the results corresponding to the theoretical derivations. This result of Levitt suggests that most of the tests of market efficiency are not complete, since the empirical research with dataset containing amount of bets and information about bettors is rare. Because my dataset does not contain amount of bets on each possible outcome, I leave this test of market efficiency apart as well.

In the following sections, I am presenting several different ways (together with previous empirical research) how to test market efficiency. But as is evident from this section and from previous chapter, these tests are just concentrating on one potential model of betting market (model of perfect market), not on all of potential models.

4.2 Convergence of Betting Offices

The convergence of odds of various betting offices is first broadly used indicator of market efficiency. Under several assumptions (perfect competition, no frictions, fully informed bettors), the odds of various betting companies should converge. Profits of betting companies should converge as well. The estimation of this convergence (of odds and of implicit fees) is one of the way, how to examine one form of market efficiency. In the literature, there were several methods used. Pope & Peel (1989) calculate the size and distribution of margins of various betting offices, i.e. the average profit ("average probability from quoted odds - outcome frequency") from betting on each occasion. Then they compare these margins across betting offices using t-test of equality of means. The result is that average profit of betting offices is not equal across betting offices. I am following their methodology for my dataset, tables with margins are presented in Subsection 5.2.1, discussion of results is in the Section 5.2. The second method used is OLS regression of odds of one company to the odds of another company. The results from this method corresponds to previous case, companies statistically differ in quoted odds, market does not force companies to equalize them. I am performing OLS test in the Section 5.2.

4.3 Arbitrage Opportunity

One branch of research concerns the examination of arbitrage opportunity - i.e. situation when the bettor is using different odds across various betting offices and by betting in the most optimal way on all 3 possible outcomes of game is obtains profit without bearing any risk. It is one of tests of market efficiency, since under condition of perfect market, the arbitrage opportunity should not exist at all (or should disappear immediately). In the decimal odds system, the arbitrage possibility for Premier League during years 1981-1982 was examined in Pope & Peel (1989) with the result of existence of arbitrage opportunity, but in restricted amount. The same league and similar result (dataset 1993-1996) present Dixon & Pope (2004). The most comprehensive dataset is in Vlastakis *et al.* (2009) with 26 leagues for the period 2002-2004, with the result of less than 1 percent of games with arbitrage opportunity. Such low number of arbitrage opportunities might be caused by the use of closing odds, not of the pre-game odds, as Vlastakis notes. For my dataset, I am testing arbitrage opportunity in the Subsection 5.2.3.

4.4 Quoted Odds (Ex-ante Probabilities) Corresponding to Ex-post Probabilities

Next topic examined will be relationship between quoted probability (recalculated from quoted odds) and real probability. The research tries to answer basic questions - do the quoted odds correspond to the real probabilities or is there a bias ? And is the relationship between ex-ante (odds) and ex-post (result) probability linear or nonlinear ? Under the assumption of fully rational bettors, the market efficiency should mean correspondence of quoted odds to ex-post probabilities. Pope & Peel (1989) are estimating the shape of ex-post probabilities depending on ex-ante probabilities using WLS method. For their dataset (4 betting offices, Premier League 1981-1982), the ex-post probabilities correspond to the quoted odds for the outcomes win of home team and win of away win, not for the outcome tie. In Woodland & Woodland (1994), t-statistic is used for the comparison of quoted probability of spread in goals and its relative amount in major baseball league. Their result is that market is efficient (quoted odds correspond to the probabilities) for separate spreads but inefficient generally.

In some of the research, the assumption of linear probability was left by the use of nonparametric regression (for example in Goodwin (1996) for dog racing). I am testing potential nonlinearity in the Subsection 5.3.3.

4.5 Existence of Profitable Strategy

The most common research done is search for profitable strategy. The reason for that is its simplicity and potential profitability. Under Fama´s definition of weak form of market efficiency, the current price on the market should reflect all public information and each profitable strategy should disappear in time.

The sport leagues examined and methods used are rich, from testing simple strategies as "bet on all home teams" or "bet on home favorites" to complicated statistical models (for example Poisson distribution used for prediction of number of scored goals by each team). But the task is simple and clear - to check, whether the strategy was profitable and whether this profitability persisted on the newer dataset. Some of examined leagues and sports were following - football (Premier League, Australian league, World Cup, Bundesliga, research across all leagues), ice-hockey (NHL), American football (NFL) or basketball (NBA). In the following subsections, I am presenting previous research for each of them.

4.5.1 Football

In the case of Premier league, some of the strategies were profitable, some not. The oldest dataset was used in Pope & Peel (1989) (1981-1982, 2 strategies - based on odds and based on tips from experts) with no profit strategy, dataset in Cain *et al.* (2000) (1991-1992, Poisson and negative binomial regression) and the longest dataset (1999-2000, ordered probit model) in Goddard & Asimakopoulos (2004) were without any profitable strategy. Dixon & Coles (1997) found in 1992-1995 profit using Poisson distribution for modeling number of scored goals. In Dixon & Pope (2004), the dataset 1993-1996 was with profitable strategy, profitable method was modeling number of scored goals by Poisson variable. From the results, it seems that odds for Premier league are slightly inefficient.

In the case of German Bundesliga, the strategies examined in Spann & Skiera (2009) with dataset 1999-2000 (strategies - prediction market, odds, tipsters) were not profitable. The other national league examined was Australian football league (1987-1995, strategies - home win, probit) in Brailsford *et al.* (1995). For these years, there was favorite-long shot bias, but no bias during 2001-2004 in Schnytzer & Weinberg (2008). Football leagues around the world between years 2002-2004 were examined in Vlastakis *et al.* (2009) (simple rules, regression), with profitability of betting on away-favorites.

4.5.2 American Football

The American National Football League (NFL) was examined in Rodney & Weinbach (2002) with the dataset 1979-2000 and there was profitable strategy (betting underdogs) that persisted. In Kochman & Goodwin (2004), profitable was betting on underdogs during 1999-2003. And in Vergin & Sosik (1999), the profitable strategy was found in 1981-1996 (betting on home teams, betting on underdogs).

4.5.3 Basketball

The American National Basketball League (NBA) was as well deeply examined. In most of the research, there were profitable strategies (1985-1997 (strategies based on moves in odds) in Gandar *et al.* (2000), 1991-1998 (regressions) in Osborne (2001) and 1995-2002 (betting on underdogs) in Rodney & Weinbach (2005)). Other strategies were tested on the dataset 1984-1999 (simple strategies) and market was generally efficient.

4.5.4 Ice-hockey

The literature concerning ice-hockey is rare and only league examined so far in this sport was NHL. In the work of Gandar *et al.* (2004) and Woodland & Woodland (2001), there was profitable strategy during 1990-1996- betting on long shots. This work is not fully suitable for my work since datasets used were containing odds for betting on goal differences, not directly of win/loss outcome.

4.5.5 Summary

In most of the cases and leagues, there was some profitable strategy, that in some of the cases persisted. These strategies were various, from simple rules - bet on away favorites, bet on home teams when away team is favorite - to complicated statistical models (Poisson variable used for prediction of number of scored goals by both teams). No general betting strategy was profitable across all sports and leagues. Betting based on logit models and betting on underdogs seem to be the strategy that was most often leading to profit. From these results, I can conclude that one of definition of market efficiency - constant returns from betting on each of outcome - was never fulfilled. The existence of profitable strategy was not rare and in some of the cases, it did not disappear. My own estimates for the Czech ice-hockey league are in the Section 6.1.

4.6 Other Topics

While the literature concerning search for profitable strategy using mechanical rules is rich, it is not the case for two stronger versions of market efficiency. The semi-strong form of market efficiency in NFL on point-spread market is tested in Zuber *et al.* (1985) and Sauer (1998) with the use of closing odds as the parameter describing public information. Sauer recalculates the work of Zuber and rejects inefficiency. With unique strategies (or with their summary from previous works) comes Gandar *et al.* (1988), who is testing semi-strong form of market efficiency using various information from market, mainly moves of odds, and is setting strategies against and in direction of public opinion. His conclusion is that market is semi-strongly inefficient. To the similar results come Jakobsson & Karlsson (2007) on the decimal odds system for Swedish horse racing. The summary of strong form of market efficiency might then be found in chapter 3 in Williams (2005).

Another interesting field of research concerns the quality of tips given by tipsters. Usually, it goes about some public person in media (expert) giving prediction of results. These tips are evaluated for example in Andersson *et al.* (2005) with no difference in success of prediction on WC of football between experts and general public. With comparison of experts to some mechanical betting rule come Forrest & Simmons (2000). Their result is that tipsters (especially when their judgment is put together into one prediction) outperform mechanical betting rules. To similar results come Amor & Griffiths (2003) or Spann & Skiera (2009).

From the results, it does not seem that tipsters are generally more successful in prediction than are some mechanical rules or odds. The complexity of each game and part of results given by randomness seem to be stronger than the strength of experts judgment.

4.7 Summary of Chapter 4

In this chapter, I have summarized the term "market efficiency" for the betting market. Since it is not the case that there is one united methodology and one common definition, I have presented 2 most often used - Fama´s one borrowed from financial markets and Sauer´s one. Together with that, I have presented results from the previous research and methods used. Constant returns from betting on each of potential outcome were mostly absent. The search for profitable strategy was very often successful, but without any strategy common for all datasets. Together with that, tips given by individual tipsters are not information leading to profit. In the following chapter, I am testing efficiency using methods mentioned in this chapter for the dataset of Czech ice-hockey league 2004-2010.

Chapter 5

Testing market efficiency

This part of my thesis is devoted to the empirical tests of various hypotheses concerning the quoted odds and results of games, probabilities of outcomes and various betting offices. These tests have been so far used for various football leagues (mainly Premier league), basketball league NBA or ice-hockey league NHL, summary of previous research was in the previous chapter.

In the Section 5.1, I am presenting dataset that was available to me. Margins as a proxy for profitability from each outcome are presented in Section 5.2. After that, test of convergence of odds of various betting offices follows in Section 5.2. Potential difference in quoted odds is used for the test of potential arbitrage opportunities in Subsection 5.2.3. One form of market efficiency is then tested in Section 5.3, when the correspondence of ex-ante to ex-post probability is examined. This subsection is followed by the same topic in Subsection 5.3.3, but this time with the assumption of nonlinear relationship between these 2 variables and use of nonparametric regression.

5.1 Dataset

The dataset for my work is fully devoted to the Czech highest ice-hockey league - Extraliga. I have obtained dataset from the owners of software for professional betting - Trefik[1].

Since the betting market is under big changes in recent years (new companies have been entering market, internet operations have became easier enabling online betting), dataset is different for each year. The longest dataset available

[1]The website of the software is www.trefik.cz. The software receives odds from most of the betting offices on the market. According to the owner of the program, the delay between change in odds and an announcement about it in software is 5 minutes at most.

was for the Czech biggest betting offices - Tipsport, Chance, Sazka, Fortuna, MaxiTip and Synot Tip. For these companies, the dataset starts from 2004. During time, the availability

In total, the dataset contains 2866 games (from season 2004/2005 to the middle of the season 2010/2011, i.e. 6 and half seasons). Since the rates of betting companies are changing as a response to new information or to the amount of bets on each outcome, the dataset covers these changes as well. In total, it consists of 62 628 triples of betting rates. Not for all betting agencies the dataset is available for the whole time series (and some of the games are not included). The Table 5.1 contains information about the start of betting office in database, origin of betting office, number of games, ratio describing changes and fee (in percents) taken by office for each bet.

Table 5.1: betting offices

number	company	range	betting office	n. of games	changes	fee
1	Tipsport	2004-2011	CZ	2857	1,66	5
2	Fortuna	2004-2011	CZ	2846	1,63	5
3	Chance	2004-2011	CZ	2842	1,47	5
4	Maxitip	2004-2011	CZ	2849	1,25	10
5	Sazka	2004-2011	CZ	2850	1,18	0
7	Gamebookers	2004-2011	INT	2721	1,19	0
8	STS	2004-2011	INT	2787	1,17	0
10	Expect	2004-2011	SVK	2673	1,04	0
14	Nike	2004-2011	SVK	2656	1,37	0
15	Synot Tip	2004-2011	CZ	2851	1,58	0
25	Pinnacle Sports	2010-2011	INT	375	3,64	0
26	Bwin	2005-2011	INT	2437	1,41	0
27	Sporting Bet	2005-2011	INT	2002	1,08	0
28	Eurobet	2005-2011	INT	1654	1,35	0
32	William Hill	2008-2011	INT	1109	1,05	0
38	Startip	2006-2011	CZ	2045	1,26	0
40	Bet-at-Home	2006-2011	INT	1905	1,00	0
41	Betway	2007-2011	INT	1687	1,05	0
43	Unibet	2007-2011	INT	1564	1	0
44	Betfair	2007-2011	player to player	1445	4,52	0

The variable changes (average number of changes in odds by betting office for each game) might be partly used as proxy for liquidity (amount of bets) at each office (except of Betfair, where bettors are trading with each other, so changes are more common). The higher the figure, the higher amount of bets is betting office supposed to receive (and because of that is reacting by changes in odds). I can expect the most bets on the Czech league to be placed at the Czech offices and the higher rate of changes compared to the rest of companies corresponds to that.

For each company and each game, I can calculate level of fairness. It is potential reward from betting on all 3 possible outcomes quoted by one company. Usually, this level of fairness is around 90 percents (i.e. from 100 on stake, player receives 90 back) when betting at the Czech betting offices and more at the internet and international offices. Level of fairness across betting offices (i.e. bettor chooses the best odds for all three outcomes across the whole market) is one of the indicator of effectiveness on the market. If there is arbitrage opportunity, the market (betting offices) should react on bettors using this arbitrage and the arbitrage opportunity should disappear. One of my hypothesis is concerning this topic, since I am testing, whether the arbitrage possibility is - because of the openness of internet betting market - getting shorter (in time) or not. This test is in the Subsection 5.2.3.

5.2 Betting Offices

In this part, I am putting together betting offices and their odds and use 2 tests to examine, whether the odds between companies differ statistically. If the odds are set in the environment of perfect market, all the odds (and the profits leading from them) should be similar or should converge (at least in time) to each other. Or it might be the reverse, since there are still a lot of restrictions or old-fashioned bettors, the market is not perfect and odds do not converge[2].

5.2.1 Margins of Betting Offices

Betting agencies might differ in the margins they are demanding for their service. In the ice-hockey betting, contrary to some of horse betting, whole margin for bookmaker is not known in advance and is hidden in the odds. Some companies are charging fee from each bet they accept, some of them are earning only on odds. The summary of the fees used by different betting agencies is in the Table 5.1. The odds of companies that were charging fee are recalculated to contain this information.

There are 2 main ways, how to estimate the hidden margins.

[2]As I have already mentioned in the introduction, there are still law restrictions leading to potential higher risks when betting abroad. Or some companies are aiming at special consumer groups (betting company Maxitip having its terminals mainly in the Czech pubs), which might lead to low (or to not any convergence at all) convergence of odds. Moreover,

Figure 5.1: different measures of implict fees

- Average margin - the average margin is calculated as a difference between average odds and average result for each possibility (method comes from Pope & Peel (1989)). Again, the fee for playing game has to be taken into account since it means clear income for agency.
- Level of fairness - by placing bets in the most optimal way on all three outcomes (i.e. without bearing any risk) at one company, how much the bettor will receive back.

Tables for margins for home win (ordered by size) are in the Table B.15, for tie in the Table B.16 and for loss in the Table B.17. From the results for all 3 outcomes it follows, that average margin for each outcome significantly differ, which is one of the indicator of non convergence of odds. If the average margin might be used as a proxy for potential profit of betting offices, I can conclude that Czech betting companies have from this statistic highest profit from all betting offices for betting on home team wins and higher than average for the outcome loss of home team. The odds for home team win seem to be mostly biased downwards compared to the rest of betting offices.

So far, I have compared betting offices to each other (more robust test of convergence of odds - OLS regression - follows in the next section). But the results have other interpretation as well. The most restrictive of Sauer´s point concerning the market efficiency was the hypothesis of constant returns from betting on all three outcomes. From the results, I can conclude, that profit of each betting company (loss of players) for outcome home win is low compared to the rest of outcomes. The condition of constant returns is not fulfilled, which suggests bias in odds (upward biased odds for home win compared to the rest of outcomes). Explanation for that might lie in the fact that betting on the home team is the most common betting strategy (empirical confirmation of this fact was in Levitt (2004)).

each company has as well its own policy concerning payments (hidden fees) which might contribute to differences in odds.

5.2.2 Convergence Across Betting Offices

In this subsection, I am using OLS regression for comparing odds across betting offices. This time, I am not comparing just mean values as in the previous case, but all the observations. By regressing odds of one company on the odds of another company, I will obtain linear estimate of differences in quoted odds. The functional form is in Equation 5.1. In the case of similar odds of 2 betting companies, the constant term should be close to zero and the coefficient at odds should be close to 1. I am testing this similarity using F-test as the test of linear restrictions in OLS framework.

$$\begin{aligned} odds1_{office1} &= const + \alpha * odds1_{office2} + \epsilon \\ const &= 0 \\ alpha &= 1 \end{aligned} \tag{5.1}$$

I have divided dataset into 2 parts, till 2009 and from 2009, to check, whether the entrance of Czech betting offices[3] to the internet market has lead to convergence of odds or not. To simplify my work, I have decreased the number of betting offices included into this analysis. In the first basket - Czech companies - there are Tipsport, Fortuna, Chance, Maxitip, Sazka and Synot Tip. In the second basket - internet or international offices - there are Bwin, Eurobet, Unibet, Bet-at-Home, Sporting Bet and Betway. My hypotheses are, that the odds of internet betting offices are (statistically) similar. And second - the odds of Czech companies should converge to each other after their entrance to the internet betting market.

From the total number of 264 OLS regressions and tests of linear restrictions mentioned in Equation 5.1, in only 4 cases I can't reject hypothesis about difference in quoted odds. There is only one case before 2009, when the odds of one company were statistically (level of significance 1 percent) similar to the odds of other company - odds for tie quoted by companies Sporting Bet and Betway (p-value 0.20964). For the second dataset (from 2009), there were 3 cases of statistically not significant difference in odds - odds for tie quoted by companies Bwin and Bet-at-home (p-value 0.046), odds for tie quoted by companies Bwin and Betway (p-value 0.336) and odds for win of home team quoted by Eurobet and Betway (p-value 0.011).

These results lead to conclusion that there is nearly no convergence of odds.

[3]From 2009, it has been possible for companies under Czech license to officially offer online betting.

Odds are quoted using different models, using different expected amount of bets or each company might be earning on different games. Despite the size of market and increase in number of betting offices in the Czech Republic, odds of main Czech firms are not converging. These results suggest that market contains a lot of frictions and is far away from one of definition of market efficiency - the one of perfect market and fully rational bettors shifting from one company to another. This knowledge of statistically different odds might be as well used in search for arbitrage opportunity.

5.2.3 Arbitrage Opportunity

In this section, I am interested in the arbitrage possibilities. As was examined in the previous section, the odds of various betting companies differ. This difference in odds might be used for betting strategy based on diversifying bets between 2-3 betting offices. Bettor, by placing amount of bets in the most optimal way, may find profitable strategy without bearing any risk. Since there are costs connected to this strategy (mainly running 20 betting accounts in one time, but hidden fees for transferring money as well), I understand as the arbitrage possibility strategy with the profit higher than 5 percent. For each time and for each moment (each change in odds) in time, I have calculated using Visual Basic the most optimal bet. When the profit exceeded 5 percents, I measured how long it had lasted before the arbitrage possibility disappeared. If the hypothesis of effective market holds, these arbitrage possibilities should disappear in time.

Since my dataset consists of 20 betting companies, the arbitrage possibilities during the 7 and half year of examination were rich. Using Visual Basic calculations, I have found 367 situations when the bettor could had earned more than 5 percents without bearing any risk. The average length of arbitrage possibility of more than 5 percents was 260 minutes (s.d. of 214.64). My hypothesis is, that because of online betting, the length of arbitrage possibility has been decreasing in time - market has become more effective. To test this hypothesis, I am running OLS regression of the length of arbitrage (in minutes) on time (measured by minutes from the beginning of measurement). The summary of the regression is in the Table 5.2.

To control for heteroscedasticity, I am running the regression with robust standard errors. By F-test, I am testing null hypothesis about no relationship between length of arbitrage and time. The result contradicts my hypothesis -

Table 5.2: length of arbitrage possibilities

variable	coefficient	st. deviation
const	261.895^{***}	32.0380
time	$-0.3766 * 10^{-6}$	0.00104
R-squared	0.00004	

I do not reject the null hypothesis of coefficient at variable time (p-value being 0.97). From the results of regression, I can conclude that arbitrage possibilities are not disappearing in time and its length did not shorten.

There might be various explanations for that. First possibility is that since my dataset does not contain all the odds for the whole period of time (with 2 main reasons - data were not available, or the company did not exist during the whole period of time), the most current period than contains full spectrum of betting offices, while the oldest part of dataset contains only half of companies. By that, the amount of arbitrage opportunities increased. I expect this explanation to be the most probable.

Second, the Czech ice-hockey league is minor betting market compared to the international tennis or football. Most of the bets on the Czech ice-hockey league are placed at the Czech companies (with lower odds than the rest of betting offices) but the arbitrage possibility generally appears at the international offices.

5.3 Do Quoted Odds Predict Ex-post Probabilities

I have already devoted Chapter 4 to the definition of efficiency on the betting market. There are 2 main benchmark papers - Fama´s and Sauer´s one. Fama´s weak-form of market efficiency is tested in the next chapter. In the following section, I am interested in another form of market efficiency/inefficiency - testing more detailed hypotheses, whether the odds correspond to the ex-ante probabilities or not. The methods used are OLS and SUR. Generally, the posted odds can be written as a function of estimated probability and the estimated amount of bets on each occasion.

$$odds = f(p, bets), \tag{5.2}$$

where the probability estimate is based on the true probability plus a random variable.

$$p = p^* + \epsilon \tag{5.3}$$

One part of literature is then testing effectiveness using probability recalculated from odds (as if p was the only factor influencing odds) and real probability (p^*) - testing whether real probability corresponds to the expected (quoted) one. In the empirical research, I have not met any test except of work of Levitt (2004) working with the functional form from Equation 5.2. The explanation for that is simple - amount of bets is not public and it is not easy to obtain such dataset[4].

5.3.1 OLS Test of Market Efficiency

The method for estimation was suggested by Pope & Peel (1989). The question that estimation tries to answer is whether the odds correspond to the real probabilities. On the "ideal" betting market (full information for bettors, perfect market) the variables used for estimation are inFigure 5.2, and the estimation form looks in this way - Equation 5.4.

Figure 5.2: variables in efficiency testing

- ϕ_{ij} stands for quoted odds
- p_{ij} stands for true probability
- z_{ij} stands for impact of new information on the odds
- b_{ij} stands for brokerage fee
- f is binary outcome (1 - event happened/ 0 - did not happen)

$$\phi_{ij} = p_{ij} + z_{ij} + b_{ij} \tag{5.4}$$

Since the probabilities and margins are not explicitly seen, Pope is estimating the market efficiency by regression of ex-post probability (given by relative

[4] The only exception for that is betting at BetFair, where bettors (players) are publicly quoting odds and amounts of bets.

amount) of outcome on ex-ante probability (recalculated from quoted odds). My estimation form is then in the Equation 5.5.

$$f_{ij} = \alpha_{ij} + \beta_{ij}\phi_{ij} + v_{ij} \tag{5.5}$$

The f_{ij} is binary outcome (0 - event occurred, 1 - event did not occur), ϕ_{ij} stands for quoted probabilities. The dataset for these estimations contains all the betting companies during the whole period (i.e. 2004-2011). I am running regression Equation 5.5 and then - using F-test, I am testing following linear restrictions:

$$\begin{aligned} \alpha &= 0 \\ \beta &= 1 \end{aligned} \tag{5.6}$$

The results for the test of efficiency of quoted odds for the outcome "win of home team" are summarized in the Table B.18. From the total number of 20 betting offices, in 8 cases we cannot reject the null hypothesis about zero constant term and unit coefficient for quoted probability. It means that odds of these companies statistically correspond to the ex-post calculated probability of results. According to this result, all the Czech betting offices (Tipsport, Fortuna, Chance, Maxitip, Sazka and Synot Tip) do not quote odds corresponding to the real probabilities of outcome. From the estimate of constant term - from 0.136 (Fortuna) to 0.157 (Sazka) - I can judge that for low probability outcomes, quoted odds are biased upwards by 13-15 percentage points than should correspond to the ex-ante probability of outcome win (quoted odds are higher than is real probability). The slope (for most of the offices around 0.7) is significantly lower than unity, which means that increase in quoted probability is not followed by the same increase in real probability. The odds for highly probable outcomes are too low.

At the majority of international betting offices, the null hypothesis (i.e. correspondence of their odds to the ex-ante probability) cannot be rejected. These companies seem to be operating on more competitive - online - market and with less legislative restrictions, which leads to the odds more corresponding to the real results.

The results for the test of efficiency of quoted odds for outcome "tie" are summarized in the Table B.19. For all the odds quoted by Czech companies, it holds that the test of effectiveness in the form of Equation 5.5 leads to the rejection of null hypothesis. This time, from the negative constant terms from estimations, it seems that the odds are more likely to be biased downwards, at least for some interval of probabilities. The case of international offices is similar to the previous estimations of odds for home win - the effectiveness of odds can not be rejected at 5 percents level of significance.

The outcome for "loss of home team" is summarized in the Table B.20. Only odds quoted by Pinnacle Sports statistically correspond to the ex-post probabilities. Czech betting companies present underestimated probabilities (constant term is closed to zero, slope around 0.7). Contrary to the case of home-win results, this underestimation seems to be on the whole interval of quoted probabilities.

Generally, results might be interpreted as rejection of one form of market efficiency. Odds - especially for 2 main cases mentioned above - do not correspond to the ex-ante probabilities. The only exception seems to be betting on low probability home win, where the odds of Czech betting offices seem to be overestimated. More of this analysis will follow in the Subsection 5.3.3 devoted to nonlinear estimations.

5.3.2 SUR test

Williams (2005) comes with another way of testing efficiency. The estimate of all possible outcomes is through the linear probability model, but the estimation method is SUR regression. SUR method takes into account correlation between disturbance terms of each estimated equation. All equations are than estimated as system. My methodology is a bit different, since I have recalculated odds as if the sum of quoted probabilities of each outcome was 1. The SUR estimation is than different to the OLS estimation from the previous section.

From three separate equations, I have obtained following system of equation:

$$\begin{aligned} loss &= \alpha_1 + \beta_1 \phi_{loss} + v_{ij} \\ tie &= \alpha_2 + \beta_2 (1 - phi_{loss} - phi_{win}) + v_{ij} \\ win &= \alpha_3 + \beta_3 \phi_{win} + v_{ij} \end{aligned} \tag{5.7}$$

On this system of equations, I am using SUR estimation. Results of the SUR regression are in the Table B.21 for loss, Table B.22 for tie and Table B.23 for win. After that, I have used linear restrictions to test the market efficiency. Since the equation for tie was recalculated, the restrictions are in different format compared to OLS, see Equation 5.8:

$$\begin{aligned} \alpha_1 &= 0 \\ \beta_1 &= 1 \\ \alpha_2 &= 1 \\ \beta_2 &= -1 \\ \alpha_3 &= 0 \\ \beta_3 &= 1 \end{aligned} \tag{5.8}$$

Table 5.3 contains the results of the test of efficiency - test of linear restrictions. From the total number of 20 betting offices, in 2 cases I do not reject the null hypothesis about the odds corresponding to the real probabilities. Both cases are traditional betting companies from Great Britain - William Hill and Pinnacle Sport. Compared to the previous test (i.e. each equation step by step), the null hypothesis is rejected at most of the international betting offices. The explanation is following - to use SUR framework, I have recalculated odds as if the provision for betting office from each outcome was the same. But since it is usually not the case (depending on the amount of bets received by the betting office), the odds used for SUR estimation are biased and because of that, linear restrictions do not hold.

For 2 betting offices - Pinnacle and William Hill - at least some of OLS and SUR estimations do not lead to the rejection of null hypothesis of biased odds. The results might be interpreted that at both these companies, the odds correspond to the real probabilities and that potential margin (which was not confirmed for Pinnacle in the previous section) is spread equally for all three outcomes. For the rest of companies, the potential gain in efficiency of estimation from the use of SUR framework was lost by the standardization (recalculations) of odds.

5.3.3 Kernel Regression

In the Subsection 5.3.1 devoted to the testing of market efficiency, I am using assumption of linear relationship between quoted probability and ex-post probability of outcome. In this section, I am leaving this assumption. On some of the intervals of predicted probability, the market might behave as under perfect market efficiency hypothesis (i.e. quoted odds corresponding to the ex-ante probabilities), but on the other part, it does not have to be the case. Some of the factors (which I will examine in more details later on) might contribute to

Table 5.3: efficiency of quoted odds - SUR estimation - test of effectivity

company	rejection	p-value
Tipsport	rejected	
Fortuna	rejected	
Chance	rejected	
Maxitip	rejected	
Sazka	rejected	
Gamebookers	rejected	
STS	rejected	
Expect	rejected	
Nike	rejected	
Synot Tip	rejected	
Pinnacle Sports	not rejected	0.9755
Bwin	rejected	
Sporting Bet	rejected	
Eurobet	rejected	
William Hill	not rejected	0.266
Startip	rejected	
Bet-at-Home	rejected	
Betway	rejected	
Unibet	rejected	
Betfair	rejected	

the nonlinear relationship between these 2 variables and to jumps in the fitted curve.

The method of nonparametric regression is described in the following way. I have two variables, y as explained variable, x as explanatory variable.

$$y = f(x) + \epsilon, \tag{5.9}$$

where f is some unknown smooth function. I am searching for such form of f that will be best - that will fit well explained variables and still will keep smoothness. The Equation 5.9 might be rewritten into the matrix form, where the fitted values $\widehat{y_{ker}}$ are recalculated from the real observations y_j by the formula:

$$\widehat{y_{ker}} = \sum_{j=1}^{n} w_{ij(ker)} \times y_j \tag{5.10}$$

where

$$w_{ij(ker)} = K * [(x_i - x_j)/h] / \sum_{j=1}^{n} K[(x_i - x_j)]/h \tag{5.11}$$

where K is some decreasing function of distance between x_0 and x_i and h stands for bandwidth. While in the moving average method, all the observations are weighted by the same weight, in the nonparametric method, the closest points to the "firm" point are weighted the most. The most distanced observations are then weighted the least. The estimations are done using R software, which uses Gaussian kernel for weighting observations. The details for computation and iteration are in R package np. The outcomes of the nonparametric regression are in the graphs in Appendix A, summary follows in following subsections.

Kernel Regression for Home Win

In the case of betting market operating under the perfect competition model, the quoted probabilities (recalculated from odds) should correspond to the ex-post probability. And this relationship should be linear on the whole interval of quoted probabilities. As is evident from the graphs - Figure A.1, Figure A.2 and Figure A.3 - it is not the case. Generally, in the case of win, companies might be divided into 2 main groups. The first group consists of companies that offer the same ratio "ex-ante probability"/"quoted probability" on the whole interval (i.e. the slope is constant). These companies are mainly that international betting offices (7,8,10,25,26,27,28,32,40,41,43,44), but surprisingly even the Czech ones (3,5). The second group of companies is more interesting since there is nonlinearity, which I was searching for. All these companies (1,2,14,15) are underestimating the outcomes with lowest probability - the quoted odds are higher (less probable) than should correspond to the ex-ante probability. The approximate border for this change in pattern is in the interval approximately 0.2 to 0.3 (in the odds 5 to 3.33). This result corresponds to the OLS estimations, since for some of companies the constant term was around 0.17. The quoted low probability outcomes at some of betting offices are in reality not so improbable as one gets from the odds.

This nonlinearity might have several reasons. First explanation might be that in the games, where the away team is strong favorite, the bookmakers are overestimating its strength compared to the home team. This might be caused by underestimation of home-field advantage or by long-term overestimation of favorites. I.e. the bookmakers have biased judgment. I do not expect this to be the right explanation and prefer the second one.

Since the most common betting strategy is betting on favorite, these companies are purposefully overestimating the favorite away team to attract bettors (by quoting that this team is higher favorite than it really is) and by that underestimating home outsiders. Since most of the bets for Czech ice-hockey league are placed at Tipsport or Fortuna (and nonparametric estimation is for these companies similar), I expect this explanation to be the most probable. Betting offices are purposefully offering biased odds to attract bettors.

The underestimation of home team win is for the quoted odds around 3.5 to 5. These odds are suspicious and the knowledge of results from kernel regression might be used for more informed betting.

Kernel Regression for Tie

In the results of kernel regression for tie, there is no pattern common for all or part of betting offices. In some of the cases, the quoted odds are not significant variable for ex-ante probability at all, in some of the cases, there are some nonlinearities, but its explanation is unclear and these results might be just coincidence. In general, it is not possible to discover any rule with the exception that the quoted odds are not significant for prediction of tie (i.e. betting offices are not able to predict tie (and quote the odds with the accordance to that) well).

Kernel Regression for Win of Away Team/Loss of Home Team

Similar to the results of kernel regression for win - but with the reverse order - are the results of kernel regression for the loss of home team (win of away team). The companies might be generally divided into 3 subgroups - those overestimating the probability of high-probable outcomes (i.e. quoted probabilities are too high compared to the ex-ante probabilities), those with the same slope on the whole interval and the company (companies) with no general rule concerning nonlinearity. The tables are to be found here - Figure A.7, Figure A.8 and Figure A.9. Most of the companies are overestimating (either intentionally or not) the strength of away favorite team. This result corresponds to the estimates for home win and explanation is probably the same as in that case - betting offices are attracting by lower odds players whose main strategy is betting on winner (and using odds as the main indicator for their decision-making). The odds lower than approximately 1.55 (probability higher than 0.6) for the playing away are suspicious, teams with these odds at some of the betting offices seem to be overvalued to attract bettors.

Summary of Kernel Regression

In this part of my research, I have left the assumption of linear relationship between quoted and ex-ante probability. By that, I have realized that the nonlinearity is not the same for all three outcomes and that it differs for the betting offices. For each outcome, I have found the intervals, where - at least for some of the betting offices - the quoted odds are suspicious. Home team outsiders/away team favorites seem to be undervalued/overvalued. This result corresponds to my expectations.

5.4 Summary of Chapter 5

In this chapter, I have tested several properties of betting market that should correspond to the market efficiency hypothesis. Generally, it might be concluded that results do not correspond to the model of perfect competition. First, condition of constant returns from betting on all three outcomes is not fulfilled. Second, betting offices are not forced by market to quote equal odds and these odds are even not converging. Because of that, arbitrage opportunities did not disappear and are not getting shorter. Quoted odds at most of the betting offices (the main exception is company Pinnacle) do not correspond to the real probabilities. The SUR estimation after recalculation adds to this analysis another estimation of margins from each of outcome, when only Pinnacle and William Hill seem to have similar margins from all three outcomes. All these results are contradicting the properties of perfect betting market with no frictions, with rational bettors and with betting offices operating under the perfect competition. But as I have already stated several times, these results are contradicting just one model of market efficiency, not all of them.

Results might be as well interpreted for the potential model of Czech betting market. By empirical analysis, I have found out that the "right" model for Czech ice-hockey betting market should explain (predict) following empirical facts. First, lower margins for betting offices from home team win outcome compared to the rest of outcomes (i.e. bias in bettors preferences in the direction home-away). Still, these margins are highest at the Czech betting offices (this bias overweighted by some property on the Czech betting market). Second, generally more downwards biased odds of Czech betting offices compared to the internet betting offices (reflecting different structure of bettors betting at the Czech companies then at the internet companies). Third, upwards biased odds of Czech betting offices for home team outsiders (reflecting bias in bettors preferences, this time in direction low probability outcome, outsider-highly probable outcome, favorite). Fourth, no convergence of odds and potential arbitrage opportunity (frictions on the market). None of the models mentioned in the Chapter 3 is able to capture all these empirical results.

Chapter 6

Search for Profitable Strategy

In the previous chapter, I was examining several tests of market efficiency. It consisted of comparing quoted probabilities to the real probabilities and of comparing odds of several betting offices and its convergence. In the following sections, I am interested in the most common test of market efficiency/inefficiency in the literature - the existence of profitable strategy. Previous research in this topic was summarized in the Section 4.5 in Chapter 4.

6.1 Prediction of Match Results

In this section, I am interested in predicting the results of matches. To simplify the analysis, I am looking at the results from the perspective of home team. The literature concerning the topic of outcome prediction is broad and methods used are rich. The task is always the same - to evaluate the potential of each team and using this estimate obtain an opinion about the potential result of the game (or about the probabilities of each outcome). This evaluation might be done using intuition (i.e. some expert makes a judgment about the form of both teams and according to it, he predicts the result. The success of these experts was mentioned in chapter 4), using the statistical model based on past values, or using just the observed bets.

Except of the mathematical model, the statistical model is the only way how to add figures into the prediction. The variables used in statistical methods are very often betting rates, number of goals scored in previous games, short and long-term performance of each team, sudden injuries e.t.c. In my estimations, I am using similar variables. The long-term performance of both teams in the game is captured by variables concerning the average scored or received goals and by the relative difference in gained points. The current form of both teams is estimated by variables concerning points gained in last 2, last 4 and last 6 games. Both the current form and the long-term form should be captured in the quoted odds as well.

The next variable included in the dataset is distance, which the away team has to travel for each game. In total, I am using 14 variables to estimate the forms of both teams for each game. For the odds, I am using median values from the companies that had quoted odds for the game.

Summary of variables is presented in the Figure 6.1. The signs in parentheses (+,-,-) reflect my expected influence of the variable on the probability of each outcome (+/- for increase/decrease in probabilities, the 1st sign for the home win, the 2nd sign for tie, the 3rd sign for the win of away team). I have chosen these variables as summary of the variables used in the previous research.

6.1.1 Prediction of Match Results Using OLS

Prediction of the Difference in Goals

In the papers of Osborne (2001), Rodney & Weinbach (2002) and Brailsford *et al.* (1995), the authors are using OLS regression for prediction of the game outcome. As a dependent variable, the result of the game or some indicator of results (difference in goals) is used. As explanatory variables, the authors use different variables describing the current form of both teams (as a proxy for the ideal variable "performance of the team in the game").

Osborne´s approach is estimation of the difference between scored goals of both teams. As an explanatory variable, he uses the sum of scored goals during the season by both teams. Originally, Osborne was using this method to test efficiency of spread betting[1].

In the paper of Rodney & Weinbach (2002), authors are estimating the margin in goals using betting rates (posted closings on totals) to estimate the sum of the goals scored.

I will merge these 2 functional forms into one general equation and estimate the influence of each variable on the match result - Equation 6.1.3. To obtain more complex result, I am adding other variables, for their summary, see Table 6.1.

[1]Spread betting is a bet where the bettor bets not only on the result of the game but as well on the margin between winning and losing team

Figure 6.1: variables

- H_{scored} stands for average goals scored by home team (+,-,-)
- $H_{received}$ stands for average goals received by home team (-,+,+)
- A_{scored} stands for average goals scored by away team (-,+,+)
- $A_{received}$ stands for average goals received by away team (+,-,-)
- $odds_1$ stands for quoted odds on home win[a] (-,.,.)
- $odds_2$ stands for quoted odds on away win (.,.,-)
- $odds_0$ stands for quoted odds on tie (.,-,.)
- $points2$ stands for gained points in last 2 games (+,-,-)
- $points4$ stands for gained points in last 4 games (+,-,-)
- $points6$ stands for gained points in last 6 games (+,-,-)
- $a - points2$ stands for gained points in last 2 games (-,-,+)
- $a - points4$ stands for gained points in last 4 games (-,-,+)
- $a - points6$ stands for gained points in last 6 games (-,-,+)
- $distance$ stands for distance between 2 cities (+,-,-)
- $diff_{points}$ stands for relative difference in points [b] (+,-,-)

[a] In all 3 outcomes, it goes about median values from the whole spectrum of betting offices.

[b] The calculation is $points_{home}/points_{away}$.

$$\begin{aligned}\Delta goals = const + \alpha H_{scored} + \beta H_{received} + \gamma A_{scored} + \delta A_{received}\\ +\phi odds_{home} + \varphi odds_{away} + \gamma points2 + \eta points4\\ +\iota points6 + \kappa o - points2\\ +\lambda o - points4 + \mu o - points6 + \nu distance + \o diff_{points} + \epsilon\end{aligned} \tag{6.1}$$

First, I am running regression with all the variables (model 1), than with the reduced amount of explanatory variables (model 2). Since the linear probability model estimation leads to heteroscedastic error term, I am using robust calculation of standard errors.

Table 6.1: prediction of scored goals

variable	model 1	model 2
const	-0.3871	-0.2863
	(0.6114)	(0.5623)
H_{scored}	0.3513***	0.3356***
	(0.128)	(0.1248)
$H_{received}$	-0.3652 ***	-0.3489***
	(0.1243)	(0.1180)
A_{scored}	-0.4235 ***	-0.3839***
	(0.1176)	(0.0902)
$A_{received}$	0.3479 ***	0.3181***
	(0.1031)	(0.0815)
$odds_1$	0.1094 *	0.106*
	(0.06040)	(0.06)
$odds_0$	0.151 **	0.1468**
	(0.0637)	(0.0629)
points2	0.0209	
	(0.0371)	
points4	-0.0563	
	(0.0399)	
points6	0.06551 **	0.0337*
	(0.02865)	(0.0176)
a-points2	-0.0127	
	(0.0416)	
a-points4	0.0395	
	(0.0413)	
a-points6	-0.0147	
	(0.0307)	
distance	0.0002	
	(0.00004)	
$diff_{points}$	-0.0194	
	(0.0974)	
R-squared	0.056	0.0523

First to notice is that the power of model in explaining the variation of explanatory variable is low. One of the indicator of the power of model - R-squared - is only slightly above 0.05. From this very low amount of variation in scored goals predicted by my explanatory variables, I can conclude that the goal difference is from the big part caused by other variables than are included (and is partly unpredictable).

In both models, the values of estimated coefficients are similar. The sign at the significant variable describing current form of both teams correspond to my expectations, the average scored goals of home team and average received goals of away team with the positive sign, the reverse with the negative sign. The other significant variable is another variable describing current form of both teams - points6 - with the expected positive sign. The sign at the variable odds1 does not correspond to my (or general) expectations. The common reasoning is that the higher the odds, the less is the home-team favorite. The explanation why the coefficient is positive instead of negative might lie in the biased odds, as was suggested in the previous chapter.

The insignificance of the variable distance can be explained by relatively short travel distance between Czech and Moravian cities, where the longest journey is only around 500 km. This number is low compared to the NHL or NBA league. The relative difference in points (diffpoints) is not significant variable, the information contained in this variable might be part of other variables.

I will use this knowledge for the prediction of results and for one of the tests of market efficiency.

Prediction of the Result of Game by Linear Probability Model

In the previous section, I have estimated the outcome of game through the score difference. In this section, I am predicting the outcome directly. The method used is the linear probability model (LPM). The dependent variable is binary (0 or 1). The result of the estimation is the estimated probability of each outcome. There are several issues that concerns estimations by the LPM. First, the error term is heteroscedastic. For this reason, I am using robust standard errors counted with the use of White matrix. Second, the fitted values (the estimates of the probability for each combination of explanatory variables) will run out of the interval (0,1). Third, the LPM model uses assumption that the relationship between the dependent variable (the probability) and explanatory variable is constant. In the next section, I leave this assumption and will estimate each outcome using logit models.

My expected influence of each variable on the estimated probability is similar to the previous case. This time, the estimation will have clear outcome, the coefficient of each explanatory variable is the increase in estimated probability of win after one-unit increase in each of explanatory variables.

The summary of results of the estimation are in the following Table 6.2.

Table 6.2: prediction of win - LPM

variable	model 1	model 2
const	0.5691 (0.1595)***	0.4215 (0.0818)***
H_{scored}	0.0219772 (0.0258093)	
$H_{received}$	-0.0786067 (0.0247136)***	-0.053 (0.0157)***
A_{scored}	-0.0853706 (0.0245797)***	-0.0387 (0.014)***
$A_{received}$	0.0617935 (0.0221840)***	0.038 (0.0132)***
$odds_1$	0.0224741 (0.0115697)*	0.0314 (0.0114)***
$odds_0$	0.0178422 (0.0117261)	0.0387(0.0108)***
points2	0.0119059 (0.00795797)	
points4	-0.0125145 (0.00821563)	
points6	0.00947796 (0.00614268)	
a-points2	-0.00676182 (0.00870080)	
a-points4	0.0104801 (0.00873117)	
a-points6	0.000340248 (0.00643172)	
distance	0 (0.0000086)	
$diff_{points}$	0.0135135 (0.0133871)	
R-squared	0.0392	0.0277

The results of the regression are similar to the previous estimations with several exceptions. The value of constant term around 0.57 corresponds to the amount of games won by the home team. The significant variables to reflect current form of both team are variables describing long-term performance of team. Surprisingly but again, the coefficients at the quoted odds do not correspond to the expected ones. Again, the explanation might lie in the fact that information about the potential form of both teams is reflected in the rest of variables and the coefficients correspond to the bias that is hidden in odds.

The same estimation is done for tie, the assumptions for this model are the same as in the previous case - Table 6.3.

The coefficient at quoted odds for tie corresponds to my expectations - the rest of the variables did not cover the information contained in this variable. The lower values of R-squared and only 2 significant variables correspond to my expectation that to predict tie is harder than to predict home win.

The last model using LPM method was prediction of loss of home team. The summarizing table is following - Table 6.4

The results of estimations are similar to the case of home win, the same

Table 6.3: prediction of tie - LPM

variable	model 1	model 2
const	0.0733 (0.105)	0.125 (0.0503)**
H_{scored}	0.04 (0.0213)*	0.036 (0.0185)**
$H_{received}$	0.0124 (0.029)	
A_{scored}	0.0092 (0.0205)	
$A_{received}$	0.0055 (0.0183)	
$odds_1$	-0.0053 (0.0099)	
$odds_0$	-0.0168 (0.0101)*	-0.0194 (0.0078) **
points2	-0.0072 (0.0064)	
points4	0.0057 (0.0067)	0.0052 (0.0033)
points6	0.0029 (0.005)	
a-points2	0.0037 (0.007)	
a-points4	-0.0116 (0.007)*	
a-points6	0.0049 (0.0051)	
distance	0 (0.0000086)	
$diff_{points}$	-0.0082 (0.01567)	
R-squared	0.011	0.0066

variables are included with the reverse signs. The significant variable - odds for the home win - has again the reverse sign than expected, the reason for that might be found in the bias of odds. Generally, from these estimations, I can conclude that the coefficients of odds for home win or of home loss do not correspond to the expected values, the information that is reflected by them can be gained by different variables (average scored goals of home or away team, average received goals of home or away team). The rest of information is (based on my estimations of LPM) biased in the disadvantageous direction for the bettor.

6.1.2 Prediction of Match Results Using Binary Choice Models

In the previous section, I have estimated LPM with the assumption of linear shape of the probability curve depending on the explanatory variables. In this section, I am leaving this assumption and will estimate the model using logit model, obtaining S-shaped probability of each outcome. At the same time, I am getting rid of the fitted values above 1 and under 0. This approach was used in the work of Golec & Tamarkin (1995). As explanatory variables both previous cases are used - odds and previous goals scored during season. As a dependent variable, they use the binary outcome (1 - outcome j occurred, 0 - outcome j does not occur). By this functional form, authors are estimating

Table 6.4: prediction of loss - LPM

variable	model 1	model 2
const	0.4338(0.112)***	0.334 (0.085)***
H_{scored}	-0.0621(0.0234)***	-0.063 (0.0203)***
$H_{received}$	0.0661 (0.023)***	0.081 (0.0195)***
A_{scored}	0.0762 (0.023)***	0.064 (0.013)***
$A_{received}$	-0.067 (0.021)***	-0.057 (0.0123)***
$odds_1$	-0.017 (0.0099)*	-0.0206 (0.01)**
$odds_0$	-0.001 (0.011)	
points2	-0.005 (0.007)	
points4	0.0068 (0.0073)	-0.008 (0.004)**
points6	-0.0124 (0.0055)***	
a-points2	0.0031 (0.0079)	
a-points4	0.0011 (0.0077)	
a-points6	-0.0052 (0.0058)	
distance	0.0000004 (0.000078)	
$diff_{points}$	-0.0053 (0.0121)	
R-squared	0.0425	0.0389

only 2 outcomes, but for both teams. The functional form (in my case just for the home team) is following:

$$P(j = 1) = \Phi(X * \beta) + \epsilon \tag{6.2}$$

where X stands for explanatory variables.

I am using the logit function from the several reasons. First, I do not expect the underlying distribution of probability to be normal. Second, the prediction power (number of cases correctly predicted) is slightly better for the logit model than for the probit model. And third, I expect the S-shaped curve to be more gentle than less.

The functional form of logit function is

$$\Phi(X) = \frac{e^{X*\beta}}{1 + e^{X*\beta}} \tag{6.3}$$

I am searching for such beta's to maximize likelihood function of outcome, for the details about method, see Baltagi (2007).

The results for the logit estimations of win are in Table 6.5.

The coefficients are corresponding to the expected ones with - again - one exception. The odds for home win seem again to be undervalued for bettor.

The tie estimations are summarized in the Table 6.6.

The prediction power of model is zero, since from the 385 tie games that appeared during the examined period, the model did not predict any tie. The

Table 6.5: prediction of win - logit

variable	model 1	model 2
const	-0.1338(0.5486)	0.237 (0.284)
H_{scored}	0.077(0.112)	0.139 (0.0689)**
$H_{received}$	-0.3152 (0.109)***	-0.296 (0.071)***
A_{scored}	-0.334 (0.109)***	-0.258 (0.0532)***
$A_{received}$	0.239 (0.0977)***	0.248 (0.050)***
$odds_1$	0.103 (0.0523)**	0.086 (0.045)*
$odds_0$	0.084 (0.0523)	
points2	0.051 (0.0333)	
points4	-0.053 (0.034)	
points6	0.0383 (0.0257)	
a-points2	-0.0284 (0.036)	
a-points4	0.0435 (0.036)	
a-points6	0.002 (0.0266)	
distance	0.00005 (0.00036)	
$diff_{points}$	0.0957 (0.0965)	
Success	0.574	0.559

noticeable result is than the fact that odds for tie are significant variable with the negative sign - i.e. in the case of tie, quoted odds reflect the real probability. The model is - as in the previous case for tie - very weak, not predicting any successful tie result.

The last estimation of logit is for the outcome loss - Table 6.7.

The last estimation corresponds to the previous ones. Main variables reflecting current form of both teams are average goals scored and received. Contrary to the common sense, the coefficient at the quoted odds for the home win is negative (i.e. the higher the odds, the lower the probability of reverse event). This might again reflect the biased odds, when the expected information is contained in the rest of variables.

Using logit models, I have obtained similar results to LPM, but with better prediction power. The significant variables - reflecting the forms of both teams - are again average scored and received goals by both teams. Tie is, by the use of logit model, not predictable at all. And the odds (median value on the market) for home win seem to be bias, coefficients are with the reverse sign.

6.1.3 Prediction of Match Results Using Multiresponse Models

Next method used for prediction of results of sport competition are so called multiresponse models - in my case ordered logit. The theory behind these

Table 6.6: prediction of tie - logit

variable	model 1	model 2
const	-2,094(0.654)***	-1.72 (0.336)***
H_{scored}	0.275(0.141)*	0.329 (0.112)***
$H_{received}$	0.7662(0.13)	
A_{scored}	-0.035(0.129)	
$A_{received}$	0.0393 (0.117)	
$odds_1$	-0.0451 (0.068)	
$odds_0$	-0.122 (0.0712)*	-0.145 (0.055)***
points2	-0.048 (0.0413)	
points4	0.038 (0.043)	
points6	0.0198 (0.0337)	
a-points2	0.024 (0.043)	
a-points4	-0.074 (0.043)*	-0.026 (0.0209)
a-points6	0.031 (0.031)	
distance	-0.00008 (0.00036)	
$diff_{points}$	-0.0794 (0.0942)	
Success	0.81 (no)	0.81 (no)

models might be found in Baltagi (2007) or Greene (2002). This statistical method was used for example in Goddard & Asimakopoulos (2004). The result obtained for each combination of explanatory variables with its coefficients is the discrete value corresponding to the potential result of the game (in my case 1 for tie, 0 for loss and 2 for win, all from the perspective of home team) and by maximum likelihood method obtain estimates of coefficients β_k and threshold values for each of outcomes.

$$P(y_i = k) = \frac{exp(X_i\beta_k)}{1+\sum_J^{j=1} exp(X_i\beta_j)} \tag{6.4}$$

Estimation is summarized in the Table 6.8.

The signs of coefficients correspond to the expected ones. Except of the traditional variables describing current form of both teams, it is for the first time in my estimations, when the variable difference in points is significant. The prediction power of the model is 48 percents (46 percents respectively) of right answers. This number is low because of bias towards outcome tie, which is predicted in 3/4 of cases. The prediction of home win is successful in 55 percents of cases and prediction of loss of home team is successful in 62.5 percents. These high figures of success in prediction of home win and

Table 6.7: prediction of loss - logit

variable	model 1	model 2
const	0.0474(0.652)	-0.438 (0.479)
H_{scored}	-0.272(0.131)**	-0.327 (0.116)***
$H_{received}$	0.293(0.13)**	0.358 (0.112)***
A_{scored}	0.337(0.129)***	0.372 (0.078)***
$A_{received}$	-0.300 (0.112)***	-0.325 (0.073)***
$odds_1$	-0.104 (0.06)*	-0.094 (0.054)*
$odds_0$	-0.03 (0.069)*	
points2	-0.0255 (0.0367)	
points4	0.036 (0.038)	
points6	-0.061 (0.028) **	-0.047 (0.018)***
a-points2	0.014 (0.040)	
a-points4	0.0075 (0.0399)	
a-points6	-0.028 (0.297)	
distance	-0.00002 (0.0004)	
$diff_{points}$	-0.12 (0.168)	
Success	0.729	0.724

prediction of home loss can be interpreted that the information contained in the explanatory variables is suitable for the prediction of around 1/4 of games, the rest (those ones that were predicted tie outcome despite the relative scarcity of this outcome in the real results) needs different variables not contained among my ones. The odds have the expected coefficients, which contradicts previous estimations from LPM.

6.1.4 Summary of the Prediction of Results

I have estimated by various statistical methods influence of 16 variables on the result of the ice-hockey games. The main statistically significant variables are those ones concerning average scored or received goals. Since in most of the estimations, these coefficients are similar (but with reverse sign), I can conclude that they are important with the similar strength. Still, the goals scored are in some estimations slightly more important variable than those ones received by the team. Long-term performance of the team is the most important variable among my variables.

The variables describing current form of the team - points gained in several last games - do not seem to play generally important role and its use for the description of the current form of each team is limited.

Distance (length of traveling of away team) is not important at all.

The quoted odds are in some cases statistically significant variable, but the

Table 6.8: prediction of results - multinomial logit

variable	model 1	model 2
H_{scored}	0.321(0.107)***	0.417 (0.093)***
$H_{received}$	0.329(0.104)***	0.237 (0.092)**
A_{scored}	-0.255(0.1069)**	-0.211 (0.095)**
$A_{received}$	0.33 (0.094)***	0.26 (0.084)***
$odds_1$	-0.301 (0.049)***	-0.2918 (0.048)***
$odds_0$	-0.181 (0.052)***	-0.171 (0.05)***
points2	-0.026 (0.0312)	
points4	0.041 (0.032)	
points6	0.017 (0.023)	
a-points2	0.023 (0.0343)	
a-points4	-0.057 (0.034)	-0.021 (0.017)
a-points6	0.026 (0.025)	
distance	-0.00002 (0.0004)	
$diff_{points}$	0.139 (0.080)	0.14 (0.083)*
Success	0.48	0.46

explanation, why it is like that, is unclear. The odds quoted for tie seem to correspond more to ex-post probability than is the case for the rest of outcomes. In several estimations, from the reverse sign of quoted odds, it seems that the odds are biased to cheat the bettor. This result was in more detail examined in Subsection 5.3.3.

6.2 Efficiency as Non-existence of Long-term Profitable Strategy

One of the definition of efficiency concerns the nonexistence of long-term profitable strategy. With such definition of efficiency comes Fama with weak-form of market efficiency. But even this simple rule obtained critique, see for example Williams (2005). First, the existence of profitable strategy might be just coincidence and can disappear in time. To control this development, I have divided my dataset into 2 parts. For the data from the first part (i.e. 2004-2010), I was - using various statistical methods - searching for the profitable strategy. For the second part of data (2010), I checked whether the so far profitable strategy persisted profitable or not. If the profit disappears in time, the existence of profitable strategy was probably just coincidence. But if it persists, than according to the definition of weak-form of market efficiency, the betting market might be seen as inefficient.

If there are several profitable strategies and some of them disappear, how the bettor can in advance realize which strategy will finish with profit ? Or is it enough that the bettor stays in profit even for the second dataset ? One possible solution is that the efficiency should be examined continuously since the past values do not have to be connected to future. Second solution is, that researcher finds model according to which is betting market operating (for example shape of utility function of bettors) and is able to anticipate change in followed patterns. Since the second case is demanding specific not available data, I am using the first solution - splitting dataset into 2 parts.

For the prediction, I have used reduced forms of models, i.e. models obtained after elimination of insignificant variables. After that, I am counting profit obtained by betting on the games, where the model predicted success. The odds used are the highest closing odds on market. The first prediction used for testing market efficiency concerns prediction of final results through goal difference. After that, I am using linear probability model, logit models and multinomial logit model.

6.2.1 Prediction Model for Goal Difference

First, I ran OLS regression with the old dataset (2004-2010). Then, the same estimates with the dataset consisting of 3/4 of season 2011. After that, I have used various borders to decide, from which predicted values to bet. The summary of results is in the Table 6.9.

Table 6.9: prediction through difference in goals

	seasons 2004/2010		seasons 2010/2011	
border	number of bets	profit	number of bets	profit
0	2037	8.9	325	4.8
0.5	1623	8.2	250	2.8
1	694	10.8	100	1.5
1.5	146	22.9	16	1.2
2	26	18	no bet	
2.5	5	-5	no bet	

All the betting strategies (except of betting on total favorites - i.e. the games where the model predicts goal difference more than 2.5) were profitable. But the profit for the newer dataset decreases significantly. The results correspond to the Fama´s definition of non effectiveness of betting market. Even by simple methods, it was possible to find profitable strategy that persisted in time. But because of the significant decrease in this profit for the second part of dataset, it is necessary to follow the results further on.

6.2.2 Linear Probability Model for Home Win

I have obtained the same results by direct estimates of results of game - outcome win. Again, strategy based on linear probability model brings profit, in one case even more than 20 percents, but the profit diminishes for the new dataset. As in the estimations through goals, betting on most probable outcomes is not profitable at all. Summary of results is in the Table 6.10.

Table 6.10: LPM model - home win

	seasons 2004/2010		seasons 2010/2011	
border	number of bets	profit	number of bets	profit
0	2182	7.9	341	5
0.4	2049	9	324	5.2
1	694	10.8	100	1.5
1.5	146	22.9	16	1.2
2	26	18	no bet	
2.5	5	-5	no bet	

The estimations for tie are in the Table 6.11. There is no profitable strategy that persisted through time. It corresponds to the hypothesis that to predict tie is more complex problem than to predict the rest of outcomes. Despite the use of several variables, I did not manage to find suitable combination and coefficients to successfully predict tie.

Table 6.11: LPM model - tie

	seasons 2004/2010		seasons 2010/2011	
border	number of bets	profit	number of bets	profit
0	206	-23		
0.01	4	35	no bets	
0.011	4	-50	no bets	

The results for prediction of loss are in the Table 6.12. Some of the betting strategies stayed profitable, some not. The market efficiency is in this case fulfilled.

6.2.3 Logit Models Used for Prediction

In this subsection, I am using results of logit estimates for search of profitable strategy. Most of the profitable strategies disappear for the new dataset. The

Table 6.12: LPM model - loss

	seasons 2004/2010		seasons 2010/2011	
border	number of bets	profit	number of bets	profit
0.4	279	9.4	43	-4.2
0.5	66	30.3	10	5.7
0.6	17	13.3	4	23
0.7	12	44	2	-1

main exception is betting on home win, where the profit of more than 27.5 persisted. But since the amount of bets is low (only 2 bets), it is hardly significant result. The results of home win strategy are summarized in Table 6.13. All the strategies were profitable for the old dataset. For the new data, it is not the case. Model gives only a few prediction "to bet" for the new dataset which means high volatility in potential profit.

Table 6.13: logit model - home win

	seasons 2004/2010		seasons 2010/2011	
border	number of bets	profit	number of bets	profit
0.45	1792	7.2	no bets	
0.5	1383	9.2	10	-0.4
0.55	888	15.5	5	1
0.6	484	19.8	4	-36.3
0.65	259	27.4	2	27.5
0.7	119	31.2	2	27.5

The results of tie strategy are summarized in Table 6.14. All the strategies were leading to losses with the old dataset. These losses persisted with one exception. As in the case of LPM estimations - it is hard to predict tie.

Table 6.14: logit model - tie

	seasons 2004/2010		seasons 2010/2011	
border	number of bets	profit	number of bets	profit
0.2	241	-1.4	48	10.7
0.18	671	-1	123	-8.4
0.16	1307	-5.6	234	-3

Last logit table is for the prediction of loss - Table 6.15. The profitable strategies disappeared with the new data coming.

Table 6.15: logit model - loss

	seasons 2004/2010		seasons 2010/2011	
border	number of bets	profit	number of bets	profit
0.4	1599	-2.2	53	-23
0.5	108	19.4	13	-0.4
0.6	30	18.7	5	-1.6

To summarize estimations from logit model - it seems that this statistical model with variables used is not suitable for profitable betting. Most of the successful strategies disappeared in time and the number of predicted games predicted for betting was low. Market efficiency cannot be rejected.

6.2.4 Multinomial Logit Used for Prediction

The main model I have used for prediction was multinomial logit. In this case, I have not divided the dataset into 2 parts. First, I am running multinomial logit regression with the old dataset - from 2004 to 2010. After that, I am running regression with the full dataset - from 2004 to 2011. The model is exaggerating amount of ties (in both cases, betting according to such model leads to losses around 59 percents). The profitable betting strategy on home win persisted with profit of more than 6 percents. For the betting on loss, the profit increased, but it is caused by decrease in number of bets.

Table 6.16: multinomial logit

	seasons 2004/2010		2004/2010 + 2010/2011	
result	number of bets	profit	number of bets	profit
WIN	305	7.7	442	6.6
TIE	1557	-59	1742	-58
LOSS	138	17.8	129	24

6.2.5 Summary of Results of Profitable Strategies

In total, I have used 4 statistical methods to search for profitable strategy. In some of the cases, the profit persisted even with the new data coming, in some of the cases, there was no profit at all. If I follow strict definition of market efficiency, than the betting market on the Czech ice-hockey league was not efficient. The profitable strategies were randomly spread (betting through difference in scored goals brings profit from home win, betting through logit does not bring profit at all, betting through multinomial logit brings highest profit for win of away team and low profit for home win). Surprisingly, more advanced statistical method does not mean higher profit. The results from simple linear probability model are comparable even to multinomial logit results and it is one of the reason why I am more skeptic than optimistic in the use of these models for profitable betting.

The analysis might become more exact with the separate examination of dataset season by season. But the problem with this approach lies in the low number of observations within each season predicted as suitable for betting. Second improvement might be use of more specific variables - for example previous games played by teams against each other, more detailed variables concerning performances in the games prior to the betting event or more qualitative information about each team (injuries, form of goalkeeper).

Chapter 7

Conclusion

In my thesis, I have examined the effectiveness on the betting market for Czech ice-hockey league during years 2004-2010. By summarizing previous research in the field, I have found out that there is no united methodology and definition of efficiency. Testing market efficiency then means testing various properties of betting market.

First, I have examined betting offices, their margins and quoted odds. By using simple statistics, I have found out that margins of betting offices differ, it means that firms are operating under different models when quoting odds. Profit from outcome home win is for all betting offices lower then for the rest of outcomes.

Second, odds of various betting offices are not converging, statistically similar odds of 2 betting offices are rare. The odds of internet betting offices are more close than odds of Czech companies. The introduction of official internet betting from 2007 did not change anything on this result. This knowledge of different odds was used for search of arbitrage opportunity. Using Visual Basic script controlling best odds on the market at each moment of time, I have found out that arbitrage opportunity exists on the market and the length of such opportunity is not getting shorter.

Third result is that quoted odds (quoted probabilities) do not correspond to the ex-post probabilities of outcomes. The only exception is British betting office Pinnacle Sport. The odds of Czech betting offices are generally biased downwards. Results of SUR regression might be interpreted in the similar way as the first point in this conclusion - there are different margins of betting offices for each of outcome. Using nonparametric regression, I have found out the interval of quoted probabilities, when the odds of Czech betting offices are biased nonlinearly. Odds for the home outsider team seem to be biased upwards (outcome is quoted as being less probable than corresponds to the reality), in favor to bettors. Suspicious games are around probability 0,2.

From these three results, I have rejected one of the model of market efficiency - model of perfect market with fully rational, risk neutral, bettors. Still, it does not have to be the case that betting market is generally inefficient, I have only rejected one model of market. Other potential tests are restricted by the nonpublic dataset with amount of bets and utility functions of bettors. In the Chapter 4, I have presented summary of models of individual behavior under risk and uncertainity and models of whole betting market. There is no model that would have predicted all the empirical facts. The extension of analysis then should contain such model that repeatedly predicts several empirical observations summarized in Chapter 5.

The second empirical part examines the possibility of profitable prediction of results using various statistical methods. I am using LPM (linear probability model), OLS, logit and multinomial logit model for the prediction of results of each game and I have used 14 variables as explanatory variables. For most of the statistical methods, there was profitable strategy that persisted even for the new dataset. But since the profitability for both datasets differ significantly, it is hard to make any firm conclusion. The use of examined methods for betting might be helpful. At least, player might follow long-term statistics and use this knowledge for correction of his own judgment. From the variables used, only a few were statistically significant. Average scored goals as a proxy for long-term performance of team was among variables appearing as most statistically significant in all the estimations. Contrary to other leagues (NFL, NHL), geographical distance between teams is not significant variable. And another interesting result is that profitable prediction of tie is scarce compared to the rest of outcomes. Potential extension of search for profitable strategy should contain more detailed variables, since my results might be interpreted in the way that most of the information helping successful prediction was already contained in one variable - average scored goals. Interesting suggestion might be use of some proxy for current form of goalkeeper, injuries of top-players or potential changes in lines.

One form of market efficiency on the betting market of Czech ice-hockey league might be rejected. Especially arbitrage opportunities and quoted odds not corresponding to the real probabilities suggest that there is a space for bettors for profitable betting.

Bibliography

Amor, M. & W. Griffiths (2003): "Modelling the Behaviour and Performance of Australian Football Tipsters." University of Melbourne Working Papers, 871.

Andersson, P., J. Edman, & M. Ekman (2005): "Predicting the World Cup 2002 in soccer: Performance and confidence of experts and non-experts." *International Journal of Forecasting* **21**: pp. 565–576.

Baltagi, B. H. (2007): *Econometrics*, volume 4. Springer.

Bhattacharya, N. & T. Garrett (2008): "Why people choose negative expected return assets - an empirical examination of a utility theoretic explanation." *Applied Economics* **40**: pp. 27–34.

Brailsford, T., S. Easton, P. Gray, & S. Gray (1995): "The Efficiency of Australian Football Betting Markets." *Australian Journal of Management* **20(2)**: p. 168.

Cain, M., D. Law, & D. Peel (2000): "The Favourite-longshot Bias And Market Efficiency In Uk Football Betting." *Scottish Journal of Political Economy* **47(1)**: p. 25.

Conlisk, J. (1993): "The Utility of Gambling." *Journal of Risk and Uncertainty* **6**: pp. 255–275.

Diecidue, E., U. Schmidt, & P. Wakker (2004): "The Utility of Gambling Reconsidered." *The Journal of Risk and Uncertainty* **29(3)**: pp. 241–259.

Dixon, M. & S. Coles (1997): "Modelling Association Football Scores and Inefficiencies in the Football Betting Market." *Applied Statistics* **46**: pp. 265–280.

DIXON, M. & P. F. POPE (2004): "The value of statistical forecasts in the UK association football betting market." *International Journal of Forecasting* **20**: pp. 697–711.

FAMA, E. (1969): "Efficient Capital Markets: A Review of Theory and Empirical Work." *The Journal of Finance* **25(2)**: pp. 383–417.

FISHBURN, P. (1980): "A Simple Model for the Utility of Gambling." *Psychometrika* **45**: pp. 435–448.

FORREST, D. & R. SIMMONS (2000): "Forecasting sport: the behaviour and performance of football tipsters." *Inte* **16**: pp. 317–331.

FRIEDMAN, M. & L. SAVAGE (1948): "The Utility Analysis of Choices Involving Risk." *The Journal of Political Economy* **56(4)**: pp. 279–304.

GANDAR, J., R. ZUBER, T. O'BRIEN, & B. RUSSO (1988): "Testing Rationality in the Point Spread Betting Market." *The Journal of Finance* **43**: pp. 995–1008.

GANDAR, J. M., R. A. ZUBER, & W. H. DARE (2000): "The Search for Informed Traders in the Totals Betting Market for National Basketball Association Games." *Journal of Sports Economics* **1**: p. 177.

GANDAR, J. M., R. A. ZUBER, & R. JOHNSON (2004): "A Reexamination of the Efficiency of the Betting Market on National Hockey League Games." *Journal of Sports Economics* **5**: p. 152.

GODDARD, J. & I. ASIMAKOPOULOS (2004): "Forecasting Football Results and the Efficiency of Fixed-odds Betting." *Journal of Forecasting* **23**: pp. 51–66.

GOLEC, J. & M. TAMARKIN (1995): "Do Bettors Prefer Long Shots Because They Are Risk-Lovers, or Are They Just Overconfident ?" *Journal of Risk and Uncertainty* **11**: pp. 51–64.

GOODWIN, B. (1996): "Semiparametric (Distribution-Free) Testing of the Expectations Hypothesis in a Parimutuel Gambling Market." *Journal of Business & Economic Statistics* **14**: p. 487.

GREENE, W. H. (2002): *Econometric Analysis*, volume 5. New York University.

Hirshleifer, J. (1966): "Investment Decision Under Uncertainty: Applications of the State-Preference Approach." *Quaterly Journal of Economics* **80**: pp. 252–277.

Jakobsson, R. & N. Karlsson (2007): "Testing Market Efficiency in a Fixed Odds Betting Market." *Orebro University Working Paper Series* **12**: pp. –.

Kahneman, D. & A. Tversky (1979): "Prospect Theory: An Analysis of Decision under Risk." *Econometrica* **47(2)**: pp. 263–292.

Kochman, L. & R. Goodwin (2004): "Underdogs are Man's Best Friend : A Test of Football Market Efficiency." *Journal of Sports Economics* **5**: p. 387.

Kuypers, T. (2000): "Information and Efficiency: an empirical study of a fixed odds betting market." *Applied Economics* **32**: pp. 1353–1363.

Levitt, S. D. (2004): "Why are gambling markets organised so differently from financial markets ?" *The Economic Journal* **114**: pp. 223–246.

Loomes, G. & R. Sugden (1982): "Regret Theory: An Alterantive Theory Of Rational Choice Under Uncertainity." *The Economic Journal* **92**: pp. 805–824.

Manski, C. F. (2004): "Interpreting the Predictions of Prediction Markets." *NBER Working Paper* **10359**: pp. –.

Markowitz, H. (1952): "The Utility of Wealth." *The Journal of Political Economy* **60(2)**: pp. 151–158.

Osborne, E. (2001): "Efficient Markets? Don't Bet on It." *Journal of Sports Economics* **2**: p. 50.

Peel, D. (2008): "Introduction: economics of betting markets." *Applied Economics* **40**: pp. 1–3.

Pope, P. F. & D. Peel (1989): "Information, Prices and Efficiency in a Fixed-Odds Betting Market." *Economica* **56**: pp. 323–341.

Quandt, R. E. (1986): "Betting and Equilibrium." *The Quarterly Journal of Economics* **101**: pp. 201–208.

Rodney, J. P. & A. Weinbach (2002): "Market Efficiency and a Profitable Betting Rule : Evidence From Totals on Professional Football." *Journal of Sports Economics* **3**: p. 256.

RODNEY, J. P. & A. WEINBACH (2005): "Bettor Misperceptions in the NBA : The Overbetting of Large Favorites and the "Hot Hand"." *Journal of Sports Economics* **6**: p. 390.

SAMUELSON, P. (1952): "Utility, and the Independence Axiom." *Econometrica* **20(4)**: pp. 670–678.

SAUER, R. D. (1998): "The Economics of Wagering Markets." *Journal of Economic Literature* **36**: pp. 2021–2064.

SCHNYTZER, A. & G. WEINBERG (2008): "Testing for Home Team and Favorite Biases in the Australian Rules Football Fixed-Odds and Point Spread Betting Markets." *Journal of Sports Economics* **9**: p. 173.

SMITH, M. A., D. PATON, & L. WILLIAMS (2006): "Market Efficiency in Person-to-Person Betting." *Economica* **73**: pp. 673–689.

SPANN, M. & B. SKIERA (2009): "Sports Forecasting: A Comparison of the Forecast Accuracy of Prediction Markets, Betting Odds and Tipsters." *Journal of Forecasting* **28**: pp. 55 – 72.

THALER, R. H. & W. ZIEMBA (1988): "Parimutuel Betting Markets: Racetracks and Lotteries." *Journal of Economic Perspectives* **2(2)**: pp. 161–174.

VERGIN, R. & J. J. SOSIK (1999): "No Place Like Home: An Examination of the Home Field Advantage in Gambling Strategies in NFL Football." *Journal of Economics and Business* **51**: pp. 21–31.

VLASTAKIS, N., G. DOTSIS, & R. N. MARKELLOS (2009): "How Efficient is the European Football Betting Market? Evidence from Arbitrage and Trading Strategies." *Journal of Forecasting* **28**: pp. 426–444.

WILLIAMS, L. V. (2005): *Information Efficiency in Financial and Betting Markets*. Cambridge University Press.

WOLFERS, J. & E. ZITZEWITZ (2004): "Prediction Markets." *Journal of Economic Perspectives* **18**: pp. 107–126.

WOODLAND, B. M. & L. WOODLAND (2001): "Market Efficiency and Profitable Wagering in the National Hockey League: Can Bettors Score on Longshots?" *Southern Economic Journal* **67(4)**: pp. 983–955.

WOODLAND, B. M. & L. M. WOODLAND (1994): "Market efficiency and the favorite-longshot bias: the baseball betting market." *The Journal of Finance* **49**: pp. 269–279.

WU, G., J. ZHANG, & R. GONZALEZ (2004): *Blackwell Handbook of Judgment and Decision Making, chapter 20.* Wiley-Blackwell.

ZUBER, R., J. GANDAR, & B. BOWERS (1985): "Beating the Spread: Testing the Efficiency of the Gambling Market for National Football League Games." *Journal of Political Economy* **93(4)**: pp. –.

Appendix A

Results of Nonparametric Regression

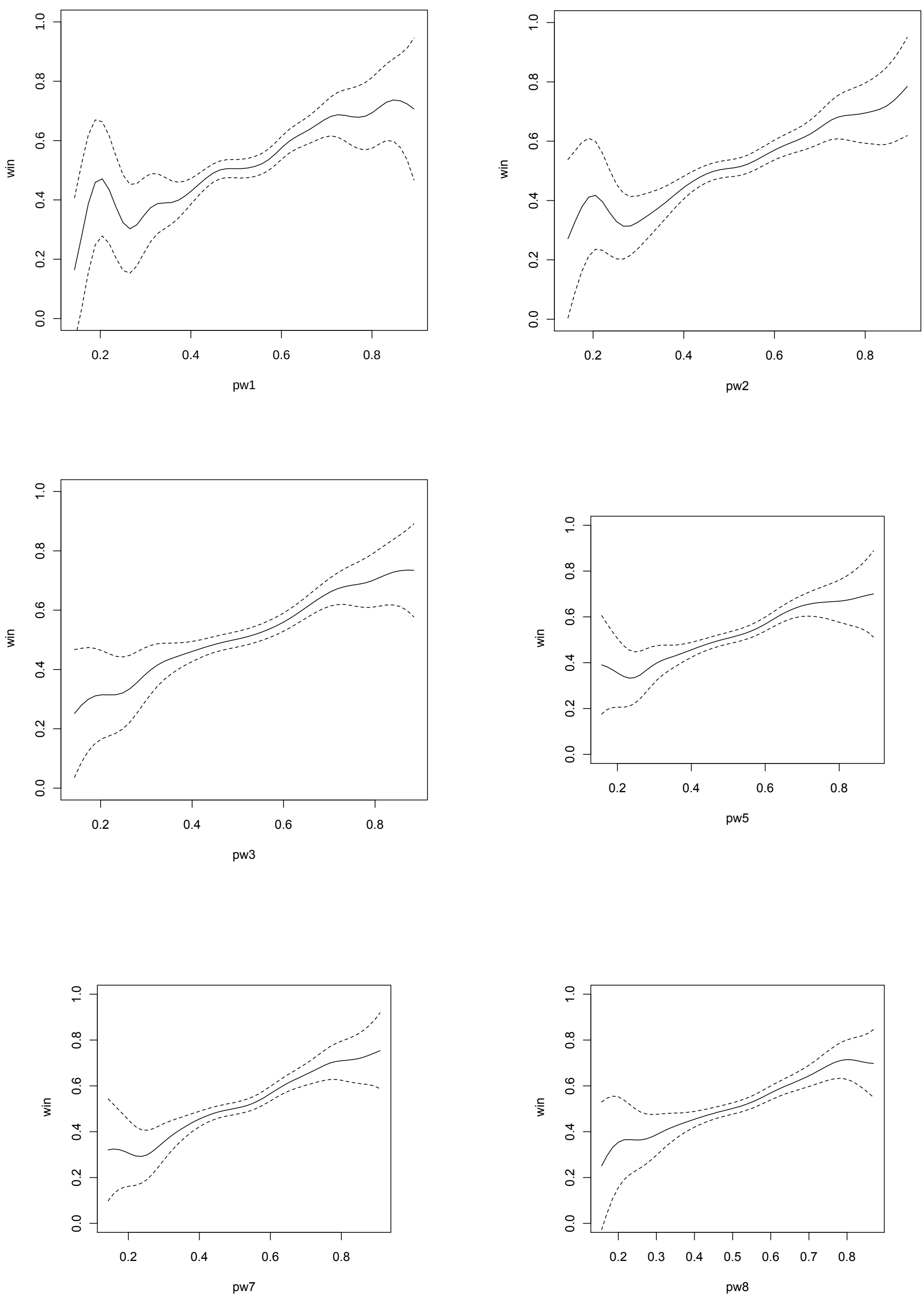

Figure A.1: nonparametric regression for home win - offices 1,2,3,5,7,8

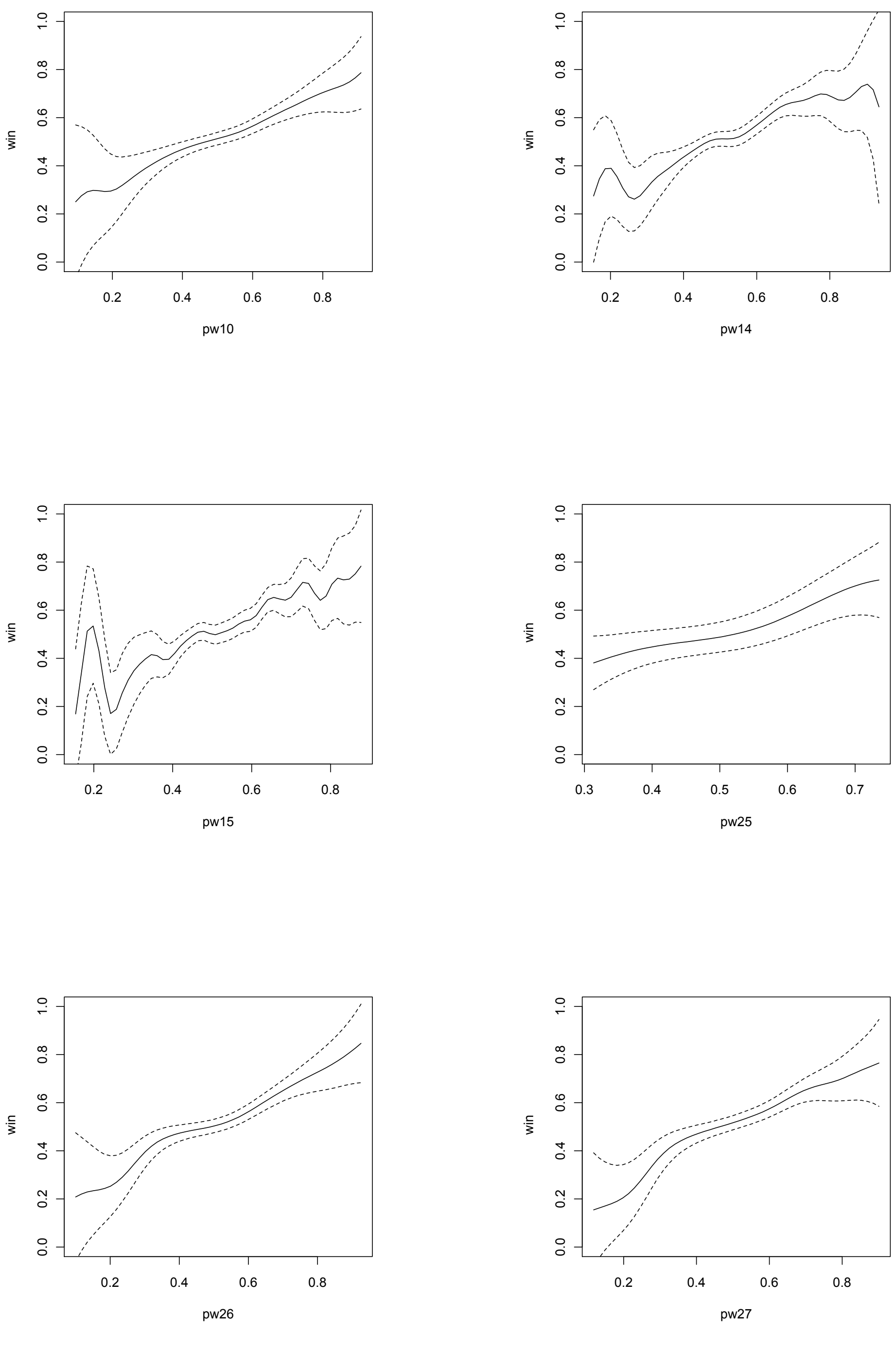

Figure A.2: nonparametric regression for home win - offices 10,14,15,25,26,27

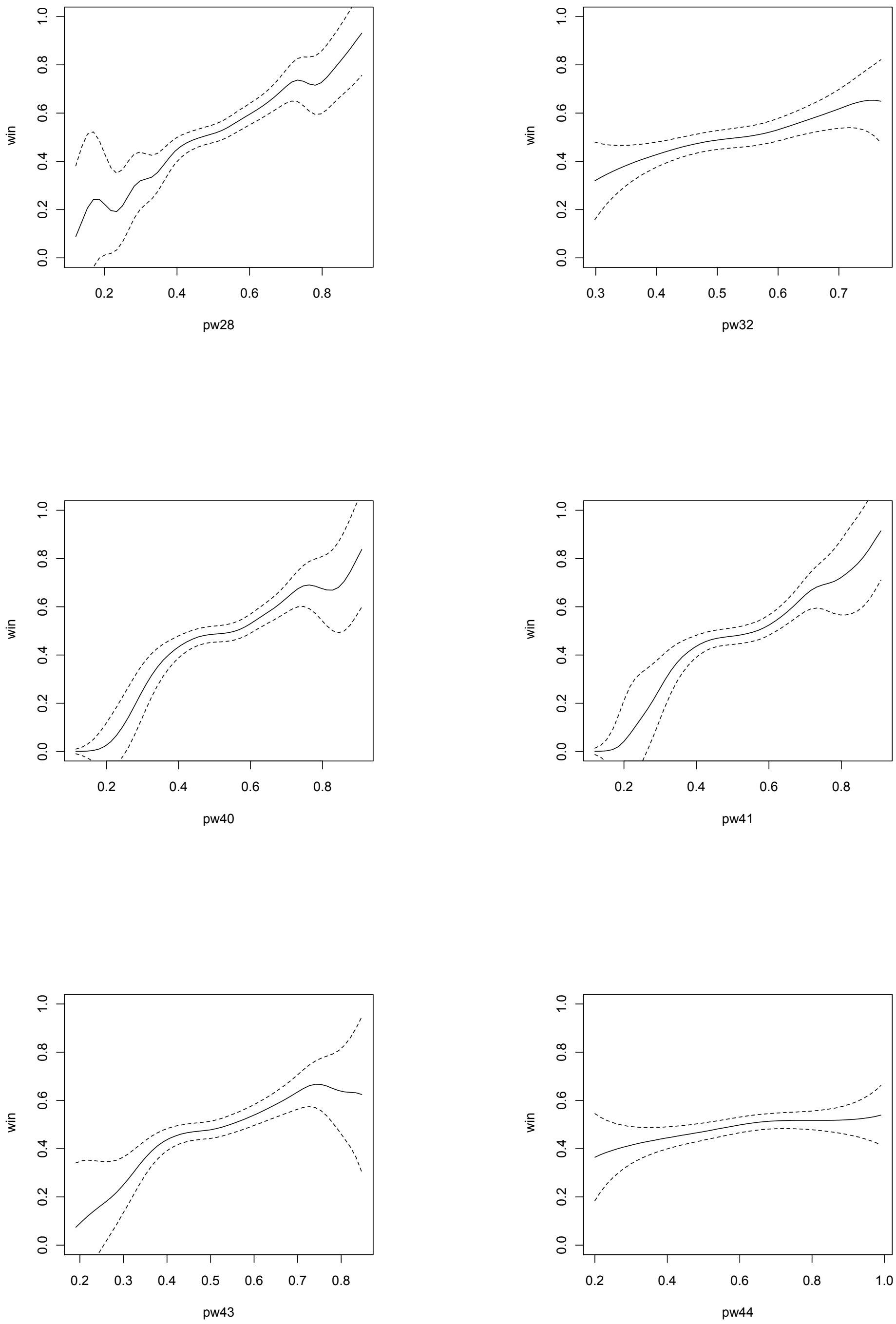

Figure A.3: nonparametric regression for home win - offices 28,32,40,41,43,44

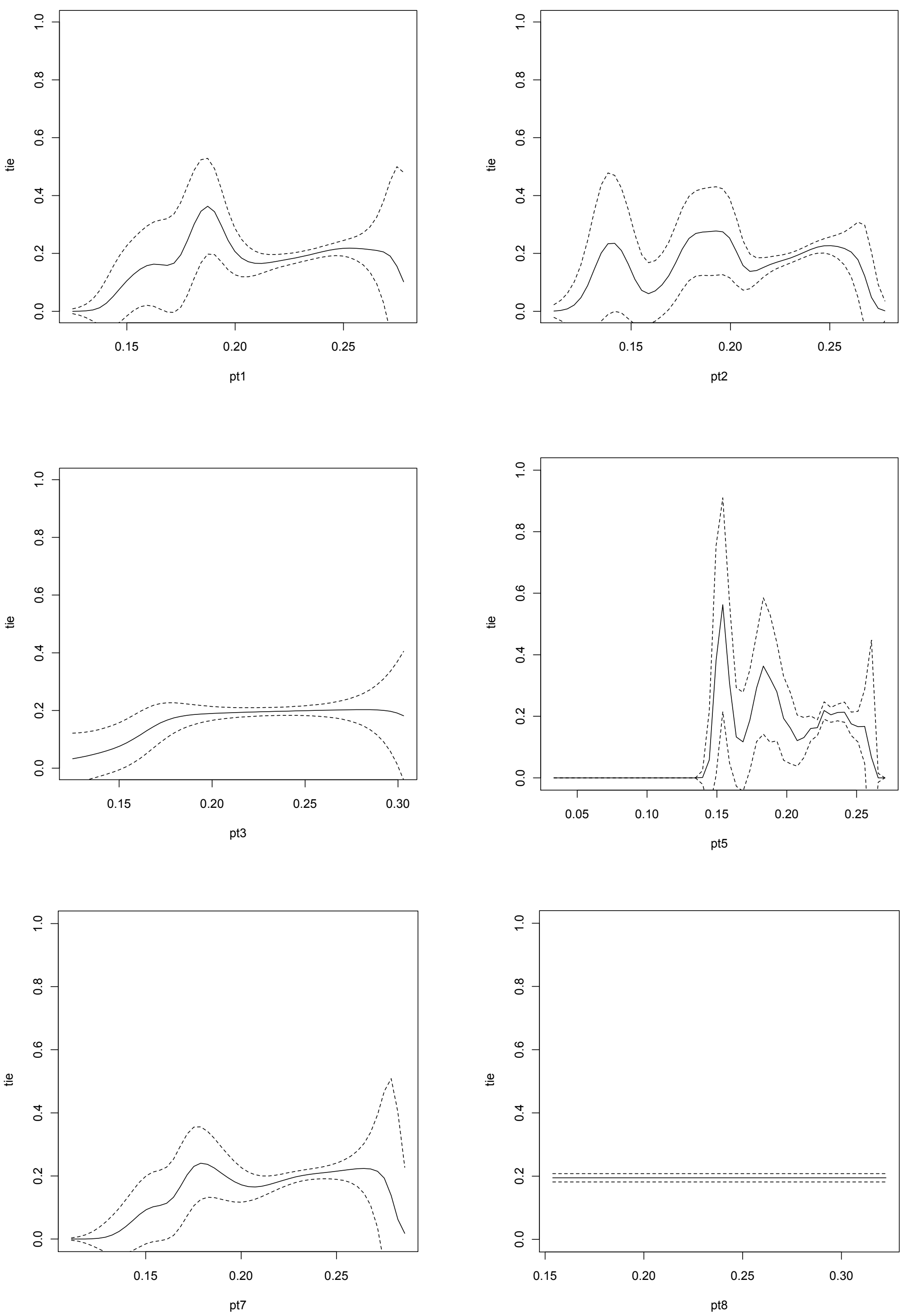

Figure A.4: nonparametric regression for tie - offices 1,2,3,5,7,8

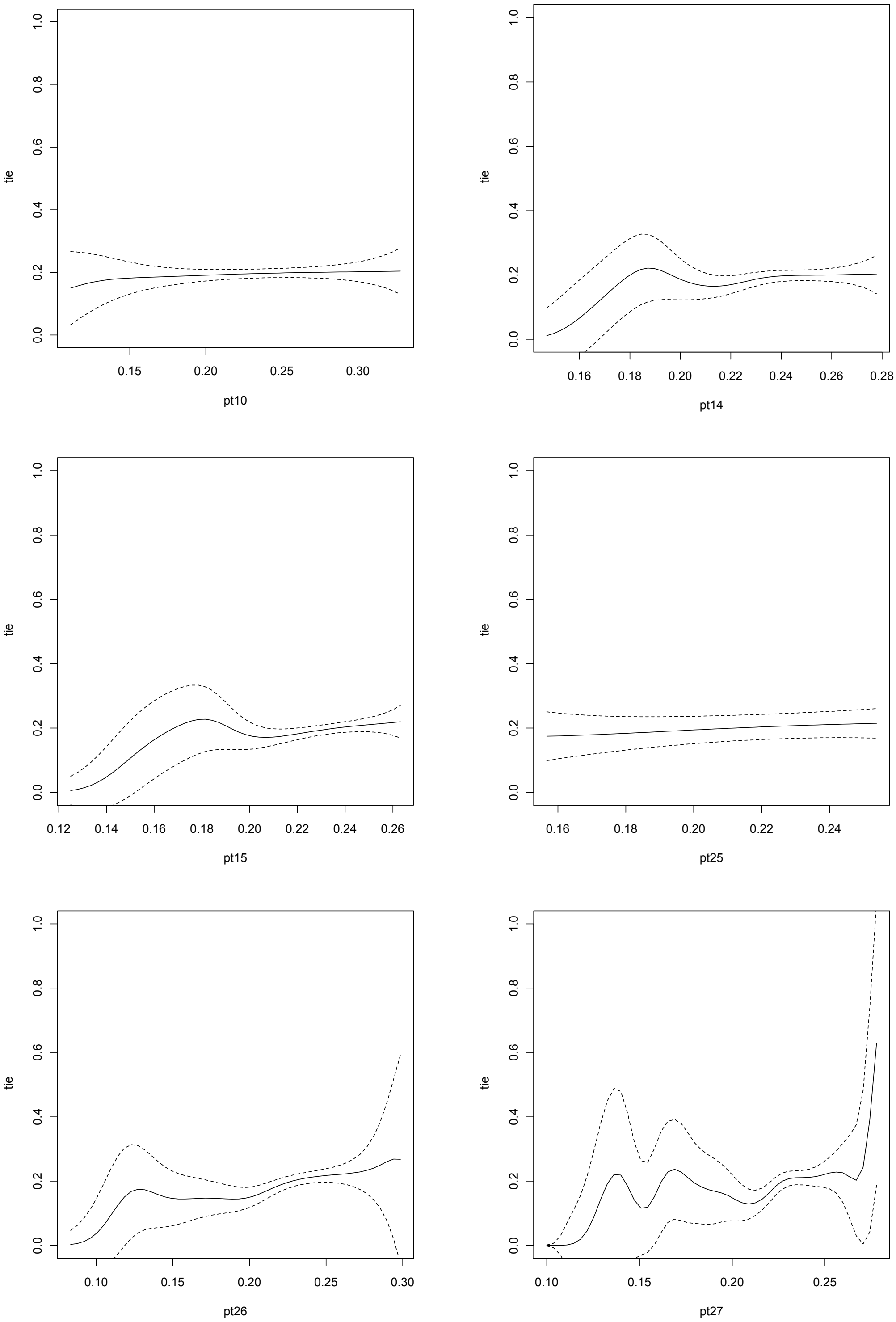

Figure A.5: nonparametric regression for tie - offices 10,14,15,25,26,27

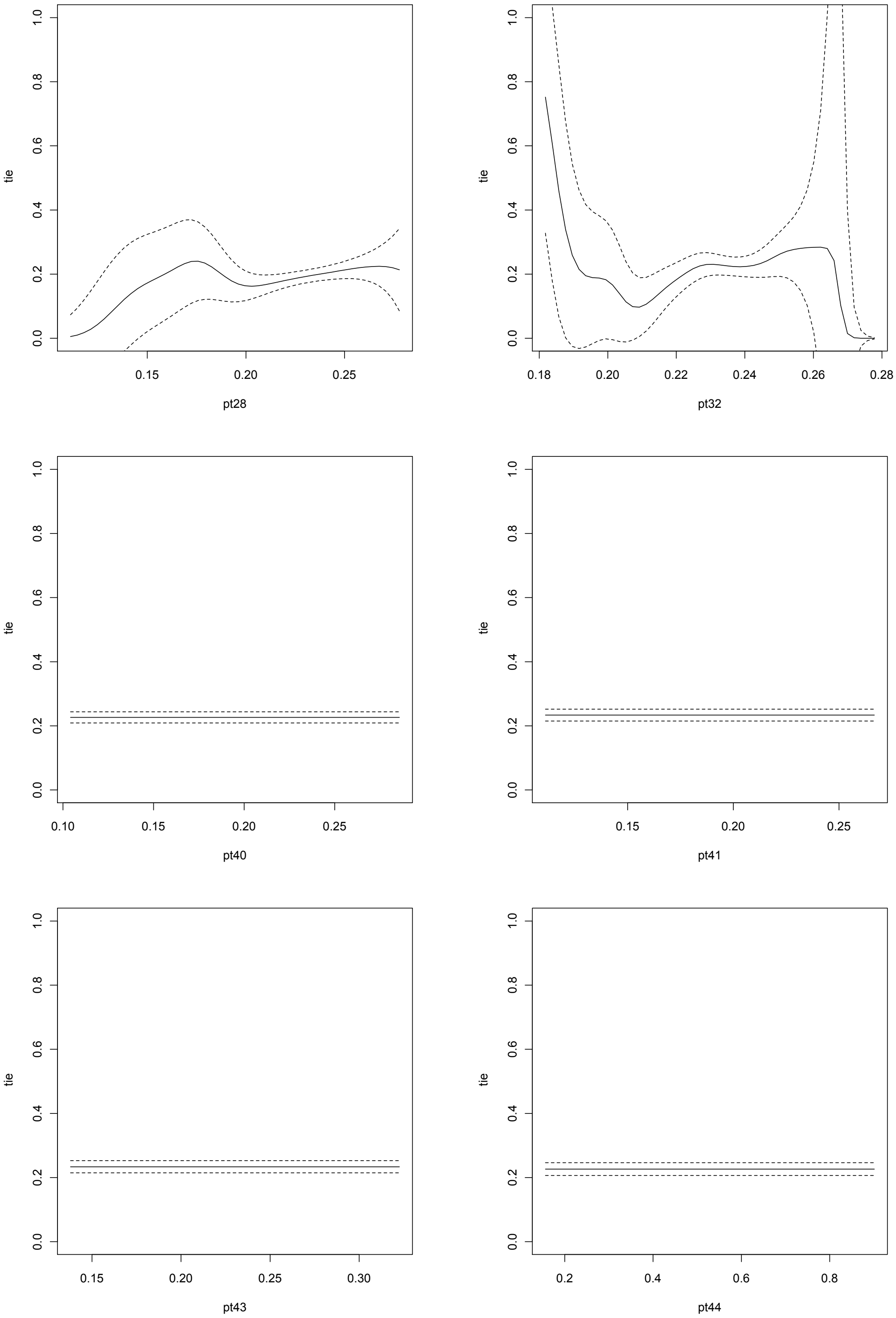

Figure A.6: nonparametric regression for tie - offices 28,32,40,41,43,44

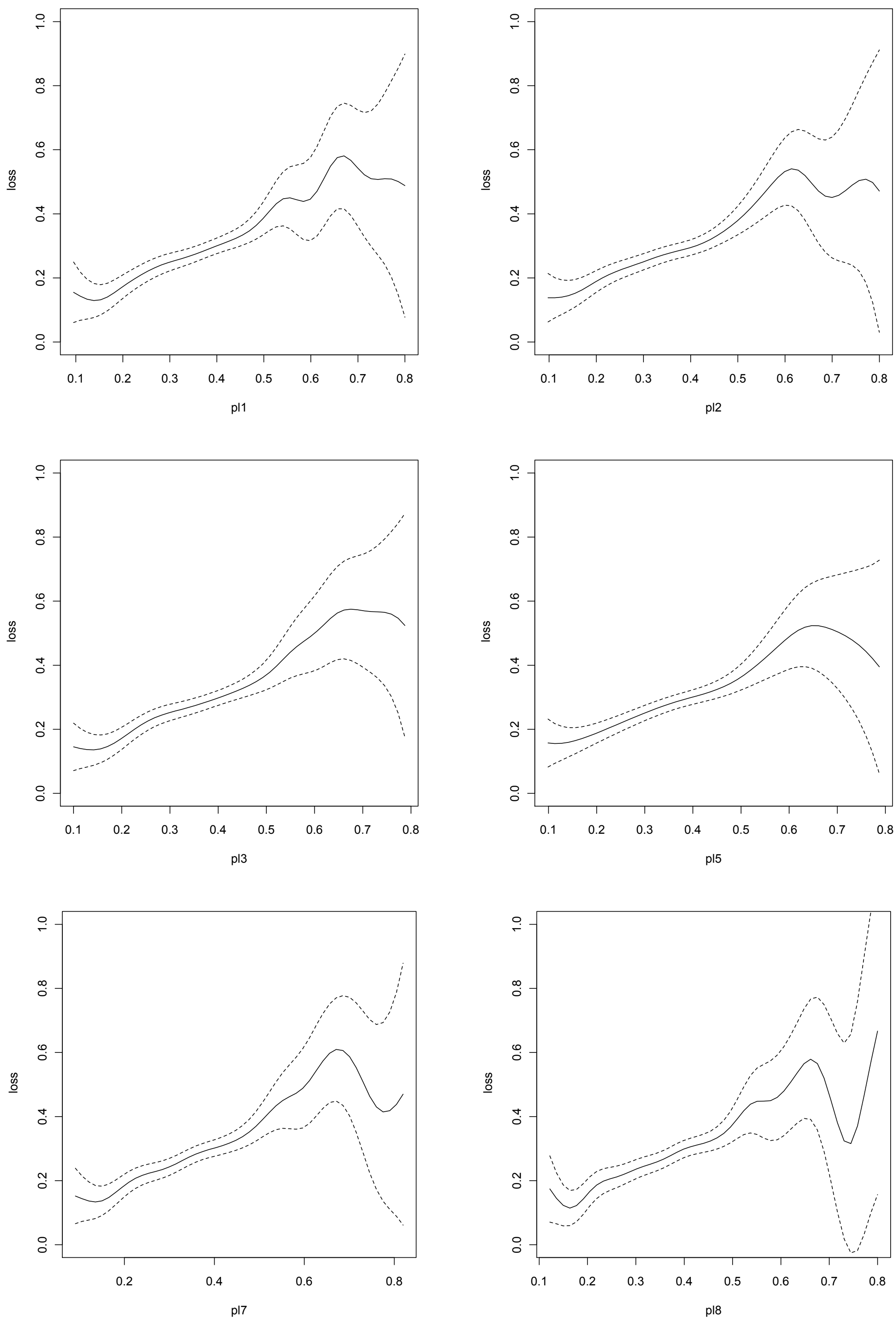

Figure A.7: nonparametric regression for loss - offices 1,2,3,5,7,8

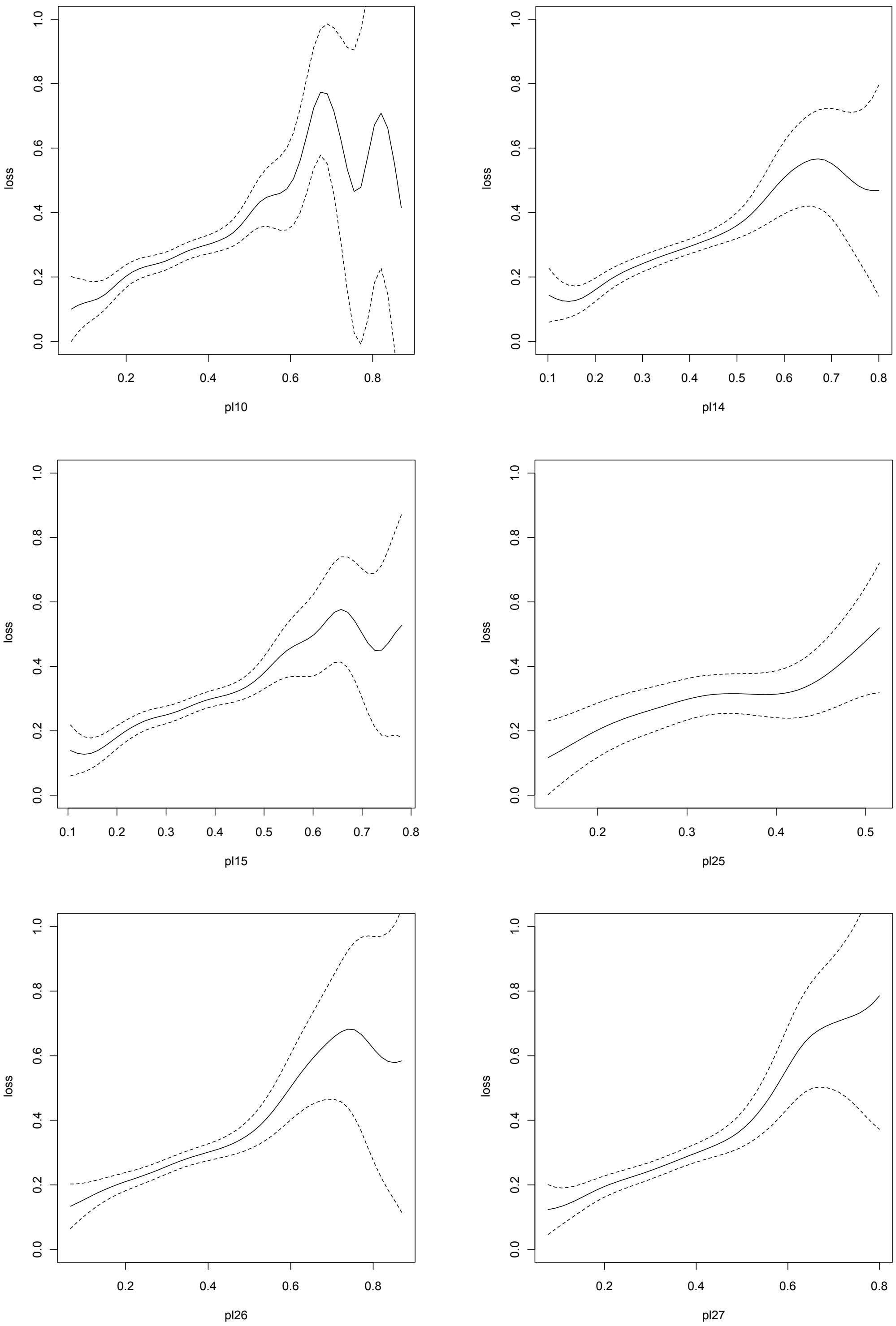

Figure A.8: nonparametric regression for loss - offices 10,14,15,25,26,27

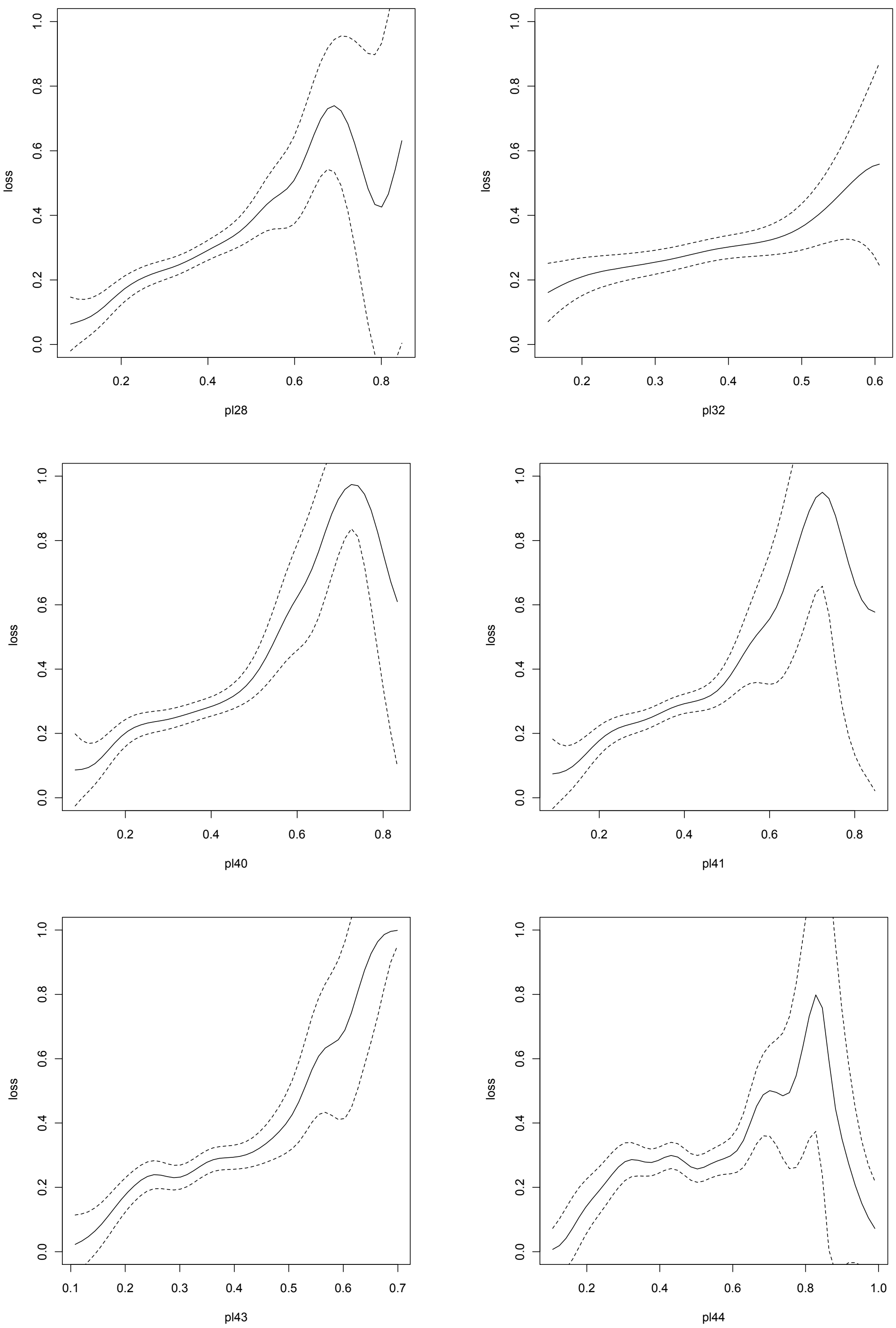

Figure A.9: nonparametric regression for loss - offices 28,32,40,41,43,44

Appendix B

Tables

Table B.1: OLS for prediction of goals - all variables

Model 15: OLS, using observations 1–2190 ($n = 1909$)
Missing or incomplete observations dropped: 281
Dependent variable: difference
Heteroskedasticity-robust standard errors, variant HC1

	Coefficient	Std. Error	t-ratio	p-value
const	−0.387083	0.611438	−0.6331	0.5268
avg_goals_score	0.351311	0.128953	2.7243	0.0065
avg_received	−0.365239	0.124285	−2.9387	0.0033
a_avg_scored	−0.423543	0.117660	−3.5997	0.0003
a_avg_received	0.347921	0.103103	3.3745	0.0008
points2	0.0208802	0.0371243	0.5624	0.5739
median_odds_0	0.150991	0.0637145	2.3698	0.0179
median_odds_1	0.109442	0.0604025	1.8119	0.0702
points4	−0.0563388	0.0389965	−1.4447	0.1487
points6	0.0655106	0.0286506	2.2865	0.0223
distance	0.000174574	0.000423546	0.4122	0.6803
apoints6	−0.0147358	0.0307071	−0.4799	0.6314
apoints4	0.0394723	0.0413965	0.9535	0.3404
apoints2	−0.0126615	0.0416341	−0.3041	0.7611
diffpoints	−0.0193842	0.0973611	−0.1991	0.8422

Mean dependent var	0.686223	S.D. dependent var	2.410565
Sum squared resid	10471.65	S.E. of regression	2.351352
R^2	0.055506	Adjusted R^2	0.048524
$F(14, 1894)$	8.083497	P-value(F)	4.97e–17
Log-likelihood	−4333.401	Akaike criterion	8696.802
Schwarz criterion	8780.117	Hannan–Quinn	8727.465

Table B.2: OLS for prediction of goals - reduced variables

Model 2: OLS, using observations 1–2190 (n = 1914)
Prediction of difference in goals - reduced variables
Missing or incomplete observations dropped: 276
Dependent variable: difference

	Coefficient	Std. Error	t-ratio	p-value
const	0.132480	0.522116	0.2537	0.7997
H_scored	0.372519	0.122450	3.0422	0.0024
H_received	−0.350521	0.118537	−2.9571	0.0031
points6	0.0331315	0.0179695	1.8438	0.0654
A_scored	−0.423576	0.0934574	−4.5323	0.0000
A_received	0.358581	0.0853472	4.2014	0.0000
median_odds_0	0.119395	0.0580236	2.0577	0.0398

Mean dependent var	0.690178	S.D. dependent var	2.410041
Sum squared resid	10547.28	S.E. of regression	2.351770
R^2	0.050759	Adjusted R^2	0.047773
$F(6, 1907)$	16.99562	P-value(F)	3.28e–19
Log-likelihood	−4349.134	Akaike criterion	8712.268
Schwarz criterion	8751.166	Hannan–Quinn	8726.582

Model 17: OLS, using observations 1–2190 (n = 1909)
Missing or incomplete observations dropped: 281
Dependent variable: win
Heteroskedasticity-robust standard errors, variant HC1

	Coefficient	Std. Error	t-ratio	p-value
const	0.492851	0.124255	3.9665	0.0001
avg_goals_score	0.0219772	0.0258093	0.8515	0.3946
avg_received	−0.0786067	0.0247136	−3.1807	0.0015
a_avg_scored	−0.0853706	0.0245797	−3.4732	0.0005
a_avg_received	0.0617935	0.0221840	2.7855	0.0054
median_odds_1	0.0224741	0.0115697	1.9425	0.0522
median_odds_0	0.0178422	0.0117261	1.5216	0.1283
points2	0.0119059	0.00795797	1.4961	0.1348
points4	−0.0125145	0.00821563	−1.5233	0.1279
points6	0.00947796	0.00614268	1.5430	0.1230
apoints2	−0.00676182	0.00870080	−0.7771	0.4372
apoints4	0.0104801	0.00873117	1.2003	0.2302
apoints6	0.000340248	0.00643172	0.0529	0.9578
distance	1.29343e–005	8.61027e–005	0.1502	0.8806
diffpoints	0.0135135	0.0133871	1.0094	0.3129

Mean dependent var	0.527501	S.D. dependent var	0.499374
Sum squared resid	457.1595	S.E. of regression	0.491297
R^2	0.039190	Adjusted R^2	0.032088
$F(14, 1894)$	7.128139	P-value(F)	1.39e–14
Log-likelihood	−1344.484	Akaike criterion	2718.969
Schwarz criterion	2802.284	Hannan–Quinn	2749.633

Table B.3: LPM for win - reduced variables

Model 4: OLS, using observations 1–2190 ($n = 2086$)
Linear probability model - WIN - reduced variables
Missing or incomplete observations dropped: 104
Dependent variable: win

	Coefficient	Std. Error	t-ratio	p-value
const	0.666389	0.0811321	8.2136	0.0000
points2	0.0117119	0.00610380	1.9188	0.0551
A_received	0.0588737	0.0129359	4.5512	0.0000
A_scored	−0.0730301	0.0139211	−5.2460	0.0000
H_received	−0.0793363	0.0200022	−3.9664	0.0001
H_scored	0.0315878	0.0209618	1.5069	0.1320

Mean dependent var	0.528284	S.D. dependent var	0.499319
Sum squared resid	505.4592	S.E. of regression	0.492960
R^2	0.027647	Adjusted R^2	0.025310
$F(5, 2080)$	11.82838	P-value(F)	2.65e–11
Log-likelihood	−1481.415	Akaike criterion	2974.831
Schwarz criterion	3008.689	Hannan–Quinn	2987.236

Model 24: OLS, using observations 1–2190 ($n = 1909$)
Missing or incomplete observations dropped: 281
Dependent variable: tie
Heteroskedasticity-robust standard errors, variant HC1

	Coefficient	Std. Error	t-ratio	p-value
const	0.0732857	0.0955149	0.7673	0.4430
avg_goals_score	0.0400990	0.0208482	1.9234	0.0546
avg_received	0.0124481	0.0192233	0.6476	0.5174
a_avg_scored	0.00921859	0.0198272	0.4649	0.6420
a_avg_received	0.00547931	0.0182237	0.3007	0.7637
median_odds_1	−0.00534259	0.00902416	−0.5920	0.5539
median_odds_0	−0.0168116	0.00909705	−1.8480	0.0648
points2	−0.00720230	0.00629362	−1.1444	0.2526
points4	0.00571512	0.00663458	0.8614	0.3891
points6	0.00295783	0.00516269	0.5729	0.5668
apoints2	0.00369207	0.00689326	0.5356	0.5923
apoints4	−0.0115780	0.00679786	−1.7032	0.0887
apoints6	0.00486640	0.00491117	0.9909	0.3219
distance	−1.25705e–005	6.96889e–005	−0.1804	0.8569
diffpoints	−0.00823505	0.00892436	−0.9228	0.3562

Mean dependent var	0.192247	S.D. dependent var	0.394170
Sum squared resid	293.2937	S.E. of regression	0.393515
R^2	0.010631	Adjusted R^2	0.003318
$F(14, 1894)$	1.644603	P-value(F)	0.060957
Log-likelihood	−920.8221	Akaike criterion	1871.644
Schwarz criterion	1954.959	Hannan–Quinn	1902.308

Table B.4: LPM for tie - reduced variables

Model 6: OLS, using observations 1–2190 (n = 2111)
Linear probability model - TIE - reduced variables
Missing or incomplete observations dropped: 79
Dependent variable: tie

	Coefficient	Std. Error	t-ratio	p-value
const	0.0817992	0.0412488	1.9831	0.0475
H_scored	0.0476001	0.0152639	3.1185	0.0018
diff_points	−0.0174773	0.0108587	−1.6095	0.1077

Mean dependent var	0.192326	S.D. dependent var	0.394221
Sum squared resid	326.2970	S.E. of regression	0.393433
R^2	0.004936	Adjusted R^2	0.003992
$F(2, 2108)$	5.228715	P-value(F)	0.005430
Log-likelihood	−1024.646	Akaike criterion	2055.291
Schwarz criterion	2072.256	Hannan–Quinn	2061.503

Table B.5: LPM for loss - all variables

Model 7: OLS, using observations 1–2190 ($n = 1909$)
Linear probability model - LOSS - all variables
Missing or incomplete observations dropped: 281
Dependent variable: loss

	Coefficient	Std. Error	t-ratio	p-value
const	0.365575	0.143262	2.5518	0.0108
H_scored	−0.0591817	0.0244885	−2.4167	0.0158
H_received	0.0644411	0.0235175	2.7401	0.0062
points6	−0.0123303	0.00559127	−2.2053	0.0276
A_scored	0.0756331	0.0230419	3.2824	0.0010
A_received	−0.0668168	0.0205969	−3.2440	0.0012
median_odds_0	−0.000273550	0.0114164	−0.0240	0.9809
points2	−0.00463816	0.00723057	−0.6415	0.5213
points4	0.00678962	0.00747114	0.9088	0.3636
apoints2	0.00335184	0.00783016	0.4281	0.6687
apoints4	0.00118961	0.00783968	0.1517	0.8794
apoints6	−0.00533725	0.00572366	−0.9325	0.3512
median_odds_2	0.0137011	0.0163607	0.8374	0.4025
median_odds_1	−0.00905588	0.0147134	−0.6155	0.5383
diff_points	−0.00621611	0.0176063	−0.3531	0.7241
distance	1.79662e–006	7.71373e–005	0.0233	0.9814

Mean dependent var	0.280251	S.D. dependent var	0.449240
Sum squared resid	368.5739	S.E. of regression	0.441252
R^2	0.042828	Adjusted R^2	0.035243
$F(15, 1893)$	5.646738	P-value(F)	1.85e–11
Log-likelihood	−1138.894	Akaike criterion	2309.787
Schwarz criterion	2398.657	Hannan–Quinn	2342.495

White's test for heteroskedasticity –
Null hypothesis: heteroskedasticity not present
Test statistic: LM = 164.759
with p-value = $P(\chi^2(133) > 164.759) = 0.0320885$

Table B.6: LPM for loss - reduced variables

Model 8: OLS, using observations 1–2190 ($n = 1916$)
Linear probability model - LOSS - reduced variables
Missing or incomplete observations dropped: 274
Dependent variable: loss

	Coefficient	Std. Error	t-ratio	p-value
const	0.337883	0.0879340	3.8425	0.0001
H_scored	−0.0679625	0.0224174	−3.0317	0.0025
H_received	0.0678527	0.0214769	3.1593	0.0016
points6	−0.00925389	0.00336197	−2.7525	0.0060
A_scored	0.0725170	0.0151305	4.7928	0.0000
A_received	−0.0649603	0.0142823	−4.5483	0.0000

Mean dependent var	0.279228	S.D. dependent var	0.448737
Sum squared resid	370.1340	S.E. of regression	0.440213
R^2	0.040142	Adjusted R^2	0.037629
$F(5, 1910)$	15.97555	P-value(F)	1.89e–15
Log-likelihood	−1143.610	Akaike criterion	2299.219
Schwarz criterion	2332.567	Hannan–Quinn	2311.491

Table B.7: Logit model for win - all variables

Model 9: Logit, using observations 1–2189 (n = 1909)
Logit model - WIN - all variables
Missing or incomplete observations dropped: 280
Dependent variable: win

	Coefficient	Std. Error	z	p-value
const	0.177359	0.688146	0.2577	0.7966
H_scored	0.0644576	0.116485	0.5534	0.5800
H_received	−0.307353	0.112868	−2.7231	0.0065
points6	0.0377761	0.0259709	1.4546	0.1458
A_scored	−0.331098	0.111910	−2.9586	0.0031
A_received	0.236789	0.0989830	2.3922	0.0167
median_odds_0	0.0782763	0.0563436	1.3893	0.1648
points2	0.0502069	0.0336216	1.4933	0.1354
points4	−0.0526269	0.0346469	−1.5189	0.1288
apoints2	−0.0296617	0.0362465	−0.8183	0.4132
apoints4	0.0430933	0.0362205	1.1897	0.2341
apoints6	0.00247352	0.0264460	0.0935	0.9255
median_odds_2	−0.0608725	0.0769693	−0.7909	0.4290
median_odds_1	0.0657330	0.0715120	0.9192	0.3580
diff_points	0.102557	0.108950	0.9413	0.3465
distance	4.25041e–005	0.000356928	0.1191	0.9052

Mean dependent var	0.527501	S.D. dependent var	0.249131
McFadden R^2	0.029534	Adjusted R^2	0.017416
Log-likelihood	−1281.334	Akaike criterion	2594.668
Schwarz criterion	2683.537	Hannan–Quinn	2627.375

Number of cases 'correctly predicted' = 1095 (57.4 percent)
Likelihood ratio test: $\chi^2(15) = 77.990$ [0.0000]

Table B.8: Logit model for win - reduced variables

Model 10: Logit, using observations 1–2189 ($n = 1916$)
Logit model - WIN - reduced variables
Missing or incomplete observations dropped: 273
Dependent variable: win

	Coefficient	Std. Error	z	p-value
const	1.12326	0.365301	3.0749	0.0021
points4	−0.0562348	0.0344291	−1.6334	0.1024
H_received	−0.359788	0.100407	−3.5833	0.0003
points2	0.0562493	0.0333322	1.6875	0.0915
points6	0.0484373	0.0249371	1.9424	0.0521
A_scored	−0.398560	0.0720765	−5.5297	0.0000
A_received	0.301454	0.0677578	4.4490	0.0000

Mean dependent var	0.528706	S.D. dependent var	0.249112
McFadden R^2	0.024401	Adjusted R^2	0.019118
Log-likelihood	−1292.581	Akaike criterion	2599.162
Schwarz criterion	2638.068	Hannan–Quinn	2613.478

Number of cases 'correctly predicted' = 1086 (56.7 percent)
Likelihood ratio test: $\chi^2(6) = 64.659$ [0.0000]

Table B.9: Logit model for tie - all variables

Model 11: Logit, using observations 1–2189 (n = 1909)
Logit model - TIE - all variables
Missing or incomplete observations dropped: 280
Dependent variable: tie

	Coefficient	Std. Error	z	p-value
const	−2.13544	0.877671	−2.4331	0.0150
points4	0.0384603	0.0432762	0.8887	0.3742
H_received	0.0751737	0.140760	0.5341	0.5933
points2	−0.0478648	0.0420761	−1.1376	0.2553
points6	0.0198936	0.0324488	0.6131	0.5398
A_scored	0.0346886	0.136933	0.2533	0.8000
A_received	0.0396817	0.121073	0.3278	0.7431
distance	−8.36041e–005	0.000445176	−0.1878	0.8510
H_scored	0.276890	0.145862	1.8983	0.0577
apoints2	0.0238676	0.0449958	0.5304	0.5958
apoints4	−0.0738604	0.0449010	−1.6450	0.1000
apoints6	0.0307832	0.0329016	0.9356	0.3495
median_odds_2	0.00798757	0.100606	0.0794	0.9367
median_odds_1	−0.0401858	0.0929855	−0.4322	0.6656
median_odds_0	−0.121458	0.0731643	−1.6601	0.0969
diff_points	−0.0802864	0.132565	−0.6056	0.5448

Mean dependent var	0.192247	S.D. dependent var	0.153180
McFadden R^2	0.011034	Adjusted R^2	-0.006090
Log-likelihood	−924.0790	Akaike criterion	1880.158
Schwarz criterion	1969.027	Hannan–Quinn	1912.866

Number of cases ‘correctly predicted’ = 1542 (80.8 percent)
Likelihood ratio test: $\chi^2(15)$ = 20.620 [0.1494]

Table B.10: Logit model for tie - all variables

Model 12: Logit, using observations 1–2189 ($n = 2008$)
Logit model - TIE - reduced variables
Missing or incomplete observations dropped: 181
Dependent variable: tie

	Coefficient	Std. Error	z	p-value
const	−1.87152	0.335999	−5.5700	0.0000
H_scored	0.242842	0.121628	1.9966	0.0459
points4	0.0348018	0.0216091	1.6105	0.1073
median_odds_0	−0.137054	0.0555430	−2.4675	0.0136

Mean dependent var	0.190239	S.D. dependent var	0.152674
McFadden R^2	0.007002	Adjusted R^2	0.002908
Log-likelihood	−970.1906	Akaike criterion	1948.381
Schwarz criterion	1970.801	Hannan–Quinn	1956.612

Number of cases ‘correctly predicted’ = 1626 (81.0 percent)
Likelihood ratio test: $\chi^2(3) = 13.681$ [0.0034]

Table B.11: Logit model for loss - all variables

Model 13: Logit, using observations 1–2189 ($n = 1909$)
Logit model - LOSS - all variables
Missing or incomplete observations dropped: 280
Dependent variable: loss

	Coefficient	Std. Error	z	p-value
const	−0.264171	0.777764	−0.3397	0.7341
H_scored	−0.257917	0.132749	−1.9429	0.0520
H_received	0.285958	0.127629	2.2405	0.0251
points6	−0.0605689	0.0291430	−2.0783	0.0377
A_scored	0.333897	0.125066	2.6698	0.0076
A_received	−0.297825	0.109743	−2.7138	0.0067
median_odds_0	−0.0215658	0.0665100	−0.3242	0.7457
points2	−0.0252877	0.0376783	−0.6711	0.5021
points4	0.0357136	0.0387920	0.9206	0.3572
apoints2	0.0159315	0.0402960	0.3954	0.6926
apoints4	0.00790179	0.0404193	0.1955	0.8450
apoints6	−0.0289608	0.0293908	−0.9854	0.3244
median_odds_2	0.0595437	0.0827013	0.7200	0.4715
median_odds_1	−0.0676075	0.0802616	−0.8423	0.3996
diff_points	−0.135415	0.152189	−0.8898	0.3736
distance	3.07315e–005	0.000396876	0.0774	0.9383

Mean dependent var	0.280251	S.D. dependent var	0.197157
McFadden R^2	0.037004	Adjusted R^2	0.022874
Log-likelihood	−1090.498	Akaike criterion	2212.996
Schwarz criterion	2301.865	Hannan–Quinn	2245.704

Number of cases 'correctly predicted' = 1389 (72.8 percent)
Likelihood ratio test: $\chi^2(15) = 83.806$ [0.0000]

Table B.12: Logit model for loss - reduced variables

Model 14: Logit, using observations 1–2189 ($n = 1916$)
Logit model - LOSS - reduced variables
Missing or incomplete observations dropped: 273
Dependent variable: loss

	Coefficient	Std. Error	z	p-value
const	−0.717234	0.458073	−1.5658	0.1174
H_scored	−0.346877	0.117347	−2.9560	0.0031
H_received	0.347935	0.112779	3.0851	0.0020
points6	−0.0461840	0.0176801	−2.6122	0.0090
A_scored	0.384642	0.0805602	4.7746	0.0000
A_received	−0.342453	0.0754741	−4.5374	0.0000

Mean dependent var	0.279228	S.D. dependent var	0.197266
McFadden R^2	0.034423	Adjusted R^2	0.029135
Log-likelihood	−1095.639	Akaike criterion	2203.277
Schwarz criterion	2236.625	Hannan–Quinn	2215.548

Number of cases 'correctly predicted' = 1389 (72.5 percent)
Likelihood ratio test: $\chi^2(5) = 78.119$ [0.0000]

Table B.13: Ordered Logit model - all variables

Model 15: Ordered Logit, using observations 1–2190 ($n = 1909$)
Missing or incomplete observations dropped: 281
Dependent variable: result

	Coefficient	Std. Error	z	p-value
H_scored	0.337722	0.106819	3.1616	0.0016
H_received	0.319383	0.103934	3.0730	0.0021
points6	0.0171564	0.0237828	0.7214	0.4707
A_scored	−0.258907	0.102939	−2.5152	0.0119
A_received	0.335079	0.0912758	3.6711	0.0002
median_odds_0	−0.173842	0.0515488	−3.3724	0.0007
points2	−0.0254198	0.0311623	−0.8157	0.4147
points4	0.0412660	0.0320615	1.2871	0.1981
apoints2	0.0255347	0.0337233	0.7572	0.4489
apoints4	−0.0571869	0.0337737	−1.6932	0.0904
apoints6	0.0256506	0.0247305	1.0372	0.2996
median_odds_2	0.0864569	0.0739032	1.1699	0.2421
median_odds_1	−0.250678	0.0659514	−3.8010	0.0001
diff_points	0.131882	0.0909300	1.4504	0.1470
distance	−2.74864e–005	0.000330171	−0.0832	0.9337
cut1	0.554562	0.640276	0.8661	0.3864
cut2	1.36543	0.640969	2.1303	0.0331

Mean dependent var	1.042954	S.D. dependent var	0.897959
Log-likelihood	−1957.197	Akaike criterion	3948.394
Schwarz criterion	4042.818	Hannan–Quinn	3983.146

Number of cases 'correctly predicted' = 911 (47.7 percent)
Likelihood ratio test: $\chi^2(15) = 167.183$ [0.0000]

Table B.14: Ordered Logit model - reduced variables

Model 16: Ordered Logit, using observations 1–2190 ($n = 1997$)
Missing or incomplete observations dropped: 193
Dependent variable: result

	Coefficient	Std. Error	z	p-value
H_received	0.315627	0.0960601	3.2857	0.0010
A_scored	−0.258111	0.0878873	−2.9368	0.0033
A_received	0.300522	0.0774453	3.8804	0.0001
points4	0.0504374	0.0172942	2.9164	0.0035
H_scored	0.322587	0.0973322	3.3143	0.0009
median_odds_1	−0.290902	0.0488081	−5.9601	0.0000
median_odds_0	−0.170730	0.0494390	−3.4534	0.0006
diff_points	0.147604	0.0886799	1.6645	0.0960
cut1	0.123500	0.462998	0.2667	0.7897
cut2	0.923534	0.463507	1.9925	0.0463

Mean dependent var	1.037556	S.D. dependent var	0.899283
Log-likelihood	−2047.887	Akaike criterion	4115.773
Schwarz criterion	4171.767	Hannan–Quinn	4136.335

Number of cases 'correctly predicted' = 945 (47.3 percent)
Likelihood ratio test: $\chi^2(8) = 168.094$ [0.0000]

Table B.15: Betting offices and their average margin for 1 - home win

num	company	average odds1	st. deviation	probability	margin1
4	Maxitip	1.897	0.306	0.501	0.026
3	Chance	1.926	0.338	0.501	0.018
1	Tipsport	1.945	0.317	0.501	0.013
2	Fortuna	1.945	0.33	0.501	0.013
44	Betfair	1.96	0.38	0.501	0.01
32	William Hill	1.972	0.443	0.501	0.006
14	Nike	2	0.313	0.501	-0.001
40	Bet-at-Home	2.003	0.364	0.501	-0.002
38	Startip	2.01	0.323	0.501	-0.003
8	STS	2.01	0.316	0.501	-0.003
15	Synot Tip	2.02	0.374	0.501	-0.006
41	Betway	2.022	0.333	0.501	-0.007
5	Sazka	2.023	0.312	0.501	-0.007
28	Eurobet	2.03	0.347	0.501	-0.009
43	Unibet	2.03	0.364	0.501	-0.009
27	Sporting Bet	2.03	0.35	0.501	-0.01
7	Gamebookers	2.041	0.342	0.501	-0.011
26	Bwin	2.04	0.42	0.501	-0.011
10	Expect	2.085	0.372	0.501	-0.022
25	Pinnacle Sports	2.107	0.400	0.501	-0.027

Table B.16: Betting offices and their average margin for 0 - tie

num	company	average odds1	st. deviation	probability	margin0
44	Betfair	3.86	0.704	0.204	0.055
2	Fortuna	4.088	0.158	0.204	0.04
14	Nike	4.11	0.162	0.204	0.04
1	Tipsport	4.095	0.158	0.204	0.04
10	Expect	4.121	0.202	0.204	0.039
4	Maxitip	4.176	0.151	0.204	0.035
32	William Hill	4.174	0.64	0.20	0.035
38	Startip	4.18	0.16	0.204	0.035
40	Bet-at-Home	4.214	0.212	0.204	0.033
41	Betway	4.21	0.192	0.204	0.033
26	Bwin	4.21	0.26	0.204	0.033
5	Sazka	4.208	0.164	0.204	0.033
15	Synot Tip	4.226	0.18	0.204	0.033
28	Eurobet	4.22	0.237	0.204	0.033
7	Gamebookers	4.238	0.182	0.204	0.032
27	Sporting Bet	4.26	0.218	0.204	0.031
43	Unibet	4.23	0.224	0.204	0.03
8	STS	4.309	0.283	0.204	0.028
3	Chance	4.028	0.338	0.204	0.018
25	Pinnacle Sports	4.53	0.434	0.204	0.017

Table B.17: Betting offices and their average margin for 2 - away win

num	company	average odds2	st. deviation	probability	margin2
44	Betfair	2.796	0.853	0.294	0.063
4	Maxitip	2.79	0.74	0.294	0.063
3	Chance	2.85	0.72	0.294	0.056
14	Nike	2.883	0.703	0.294	0.052
8	STS	2.879	0.652	0.294	0.052
32	William Hill	2.885	0.865	0.294	0.052
1	Tipsport	2.913	0.753	0.294	0.049
2	Fortuna	2.93	0.767	0.294	0.047
5	Sazka	2.946	0.734	0.294	0.045
41	Betway	2.973	0.771	0.294	0.042
38	Startip	2.996	0.792	0.294	0.039
28	Eurobet	3.032	0.777	0.294	0.035
7	Gamebookers	3.05	0.79	0.294	0.033
43	Unibet	3.074	0.837	0.294	0.031
15	Synot Tip	3.104	0.881	0.294	0.028
27	Sporting Bet	3.126	0.835	0.294	0.025
10	Expect	3.157	0.862	0.294	0.022
26	Bwin	3.213	0.946	0.294	0.017
25	Pinnacle Sports	3.276	0.918	0.294	0.011
40	Bet-at-Home	2.003	0.364	0.294	-0.002

Table B.18: efficiency of quoted odds - win

company	const	s.d. of const	beta	s.d. of beta	rejection
Tipsport	0.138	0.05	0.74	0.09	YES
Fortuna	0.136	0.048	0.737	0.09	YES
Chance	0.139	0.05	0.734	0.016	YES
Maxitip	0.139	0.05	0.736	0.092	YES
Sazka	0.157	0.049	0.706	0.093	YES
Gamebookers	0.119	0.049	0.770	0.091	YES
STS	0.124	0.052	0.761	0.097	YES
Expect	0.153	0.048	0.72	0.09	YES
Nike	0.128	0.05	0.754	0.09	YES
Synot Tip	0.138	0.049	0.74	0.092	YES
Pinnacle Sports	-0.022	0.143	1.065	0.286	NO(0.91)
Bwin	0.127	0.049	0.768	0.091	YES
Sporting Bet	0.114	0.056	0.800	0.104	NO(0.113)
Eurobet	0.035	0.062	0.96	0.116	NO(0.512)
William Hill	0.096	0.0966	0.766	0.182	NO(0.12)
Startip	0.05	0.065	0.867	0.121	NO(0.137)
Bet-at-Home	0.058	0.063	0.837	0.117	NO(0.022)
Betway	0.053	0.07	0.846	0.131	NO(0.055)
Unibet	0.07	0.07	0.81	0.141	NO(0.074)
Betfair	0.311	0.069	0.29	0.109	YES

Table B.19: efficiency of quoted odds - tie

company	const	s.d. of const	beta	s.d. of beta	rejection
Tipsport	-0.057	0.122	1.084	0.528	YES
Fortuna	-0.087	0.118	1.214	0.505	YES
Chance	-0.077	0.136	1.181	0.59	YES
Maxitip	-0.51	0.133	1.07	0.581	YES
Sazka	0.099	0.107	0.414	0.465	YES
Gamebookers	-0.091	0.118	1.25	0.51	YES
STS	0.12	0.141	0.319	0.602	YES
Expect	0.29	0.11	0.706	0.465	YES
Nike	0.004	0.1299	0.779	0.53	YES
Synot Tip	-0.1	0.13	1.28	0.54	YES
Pinnacle Sports	-0.113	0.261	1.418	1.17	NO(0.61)
Bwin	-0.102	0.087	1.305	0.38	YES
Sporting Bet	-0.082	0.121	1.22	0.53	YES
Eurobet	-0.059	0.14	1.097	0.605	YES
William Hill	-0.12	0.31	1.48	1.35	NO(0.65)
Startip	0.09	0.18	0.56	0.777	NO(0.414)
Bet-at-Home	0.09	0.157	0.596	0.67	NO(0.66)
Betway	0.07	0.17	0.702	0.727	NO(0.92)
Unibet	0.19	0.186	0.17	0.797	NO(0.5777)
Betfair	0.195	0.047	0.095	0.141	YES

Table B.20: efficiency of quoted odds - loss

company	const	s.d. of const	beta	s.d. of beta	rejection
Tipsport	0.01	0.034	0.757	0.091	YES
Fortuna	0.019	0.033	0.738	0.088	YES
Chance	0.005	0.034	0.771	0.091	YES
Maxitip	0.016	0.033	0.74	0.09	YES
Sazka	0.022	0.033	0.73	0.091	YES
Gamebookers	0.016	0.033	0.75	0.089	YES
STS	0.003	0.037	0.748	0.096	YES
Expect	0.03	0.031	0.73	0.090	YES
Nike	-0.016	0.036	0.804	0.094	YES
Synot Tip	0.02	0.033	0.737	0.09	YES
Pinnacle Sports	-0.016	0.106	0.96	0.315	NO (0.48)
Bwin	0.03	0.032	0.74	0.09	YES
Sporting Bet	-0.014	0.036	0.85	0.103	YES
Eurobet	-0.064	0.041	0.94	0.113	YES
William Hill	0.03	0.065	0.723	0.18	YES
Startip	-0.031	0.043	0.858	0.118	YES
Bet-at-Home	-0.014	0.041	0.827	0.115	YES
Betway	-0.025	0.047	0.84	0.13	YES
Unibet	-0.025	0.05	0.797	0.121	YES
Betfair	0.168	0.048	0.251	0.103	YES

Table B.21: efficiency of quoted odds - SUR estimation - results for loss

company	const	s.d. of const	beta	s.d. of beta
Tipsport	0.127	0.045	0.844	0.09
Fortuna	0.017	0.03	0.83	0.089
Chance	0.008	0.031	0.851	0.09
Maxitip	0.013	0.031	0.83	0.092
Sazka	0.027	0.031	0.79	0.091
Gamebookers	0.016	0.033	0.75	0.089
STS	0.0012	0.034	0.855	0.099
Expect	0.03	0.028	0.80	0.089
Nike	-0.01	0.03	0.9	0.096
Synot Tip	0.016	0.03	0.83	0.09
Pinnacle Sports	-0.038	0.09	1.066	0.28
Bwin	0.02	0.029	0.833	0.089
Sporting Bet	-0.01	0.032	0.915	0.101
Eurobet	-0.01	0.032	0.915	0.101
William Hill	0.016	0.059	0.83	0.18
Startip	-0.031	0.039	0.959	0.118
Bet-at-Home	-0.013	0.037	0.919	0.114
Betway	-0.027	0.042	0.94	0.12
Unibet	-0.014	0.045	0.911	0.14
Betfair	0.050	0.054	0.718	0.163

Table B.22: efficiency of quoted odds - SUR estimation - results for tie

company	const	s.d. of const	beta	s.d. of beta
Tipsport	0.863	0.074	-0.845	0.092
Fortuna	0.853	0.072	-0.832	0.09
Chance	0.87	0.075	-0.852	0.093
Maxitip	0.853	0.074	-0.83	0.092
Sazka	0.821	0.073	-0.79	0.091
Gamebookers	0.877	0.073	-0.86	0.092
STS	0.87	0.079	-0.855	0.1
Expect	0.82	0.071	-0.80	0.089
Nike	0.9	0.077	-0.898	0.097
Synot Tip	0.853	0.074	-0.83	0.092
Pinnacle Sports	1.04	0.221	-1.07	0.28
Bwin	0.855	0.07	-0.835	0.0899
Sporting Bet	0.92	0.083	-0.916	0.104
Eurobet	1.04	0.09	-1.06	0.115
William Hill	0.88	0.145	-0.83	0.18
Startip	0.979	0.094	-0.96	0.12
Bet-at-Home	0.95	0.09	-0.92	0.114
Betway	0.97	0.102	-0.94	0.13
Unibet	0.95	0.11	-0.91	0.139
Betfair	0.778	0.126	-0.72	0.163

Table B.23: efficiency of quoted odds - SUR estimation - results for win

company	const	s.d. of const	beta	s.d. of beta
Tipsport	0.010	0.031	0.844	0.092
Fortuna	0.13	0.04	0.83	0.09
Chance	0.12	0.045	0.851	0.092
Maxitip	0.13	0.044	0.83	0.092
Sazka	0.15	0.044	0.79	0.091
Gamebookers	0.116	0.044	0.858	0.090
STS	0.127	0.0475	0.855	0.0998
Expect	0.146	0.043	0.8	0.09
Nike	0.11	0.046	0.898	0.096
Synot Tip	0.13	0.044	0.83	0.091
Pinnacle Sports	-0.003	0.134	1.066	0.28
Bwin	0.125	0.044	0.833	0.088
Sporting Bet	0.09	0.05	0.914	0.101
Eurobet	0.03	0.055	1.06	0.114
William Hill	0.103	0.086	0.83	0.18
Startip	0.052	0.057	0.96	0.12
Bet-at-Home	0.06	0.056	0.92	0.11
Betway	0.05	0.06	0.936	0.13
Unibet	0.062	0.067	0.91	0.138
Betfair	0.171	0.0743	0.717	0.163