Regression tests and the efficiency of fixed odds betting markets

(work in progress)

Ruud H. Koning

IASE 2011 Prague

Abstract

The informational content of odds posted in sports betting market has been an ongoing topic of research. In this paper, I test whether fixed odds betting markets both in soccer and tennis are informationally efficient. The contributions of the paper are threefold: first, I propose a simple yet flexible statistical test to assess efficiency. Second, this test is applied to fixed odds betting markets in ten different countries, to professional tennis, and multiple seasons. Thereby, the empirical scope of the paper is much wider than that of research published so far. Finally, I examine significance of one variable that has been ignored int he literature so far: returns on earlier bets of the contestants.

Outline

Terminology

A bookmaker offers a payout on a certain event, that can be taken by a punter by staking some amount on that bet. The payout is fixed a few days before the contest takes place, hence the name ‘fixed odds betting market’.

The quotation of odds is in decimal notation, where the odds offered are the total payout per unit wagered. Hence, decimal odds exceed one. If the decimal odds of a home win are denoted by $\tilde{O}^{HW}$, the ‘implied probability’ of a home win is $1/\tilde{O}^{HW}$.

Terminology

A bookmaker offers a payout on a certain event, that can be taken by a punter by staking some amount on that bet. The payout is fixed a few days before the contest takes place, hence the name 'fixed odds betting market'.
The quotation of odds is in decimal notation, where the odds offered are the total payout per unit wagered. Hence, decimal odds exceed one. If the decimal odds of a home win are denoted by $\tilde{O}^{HW}$, the 'implied probability' of a home win is $1/\tilde{O}^{HW}$.

Terminology

Similarly, ‘implied probabilities’ for draws and away wins can be calculated, and invariably the sum of these ‘probabilities’ exceeds 1:

$$\theta^{HW} + \theta^{D} + \theta^{AW} = \frac{1}{\tilde{O}^{HW}} + \frac{1}{\tilde{O}^{D}} + \frac{1}{\tilde{O}^{AW}} \equiv 1 + \lambda, \tag{1}$$

The quotation of odds is in decimal notation, where the odds offered are the total payout per unit wagered. Hence, decimal odds exceed one. If the decimal odds of a home win are denoted by $\tilde{O}^{HW}$, the ‘implied probability’ of a home win is $1/\tilde{O}^{HW}$.

Probabilities that do sum up to 1 are now obtained by scaling (Pope and Peel,1989, Goddard and Asimakopoulos, 2004):

$$\pi^{HW} = \frac{\theta^{HW}}{1+\lambda} = \frac{1}{(1+\lambda)\tilde{O}^{HW}}. \tag{2}$$

Terminology

Similarly, 'implied probabilities' for draws and away wins can be calculated, and invariably the sum of these 'probabilities' exceeds 1:

$$\theta^{HW} + \theta^{D} + \theta^{AW} = \frac{1}{\tilde{O}^{HW}} + \frac{1}{\tilde{O}^{D}} + \frac{1}{\tilde{O}^{AW}} \equiv 1 + \lambda, \tag{1}$$

The quotation of odds is in decimal notation, where the odds offered are the total payout per unit wagered. Hence, decimal odds exceed one. If the decimal odds of a home win are denoted by $\tilde{O}^{HW}$, the 'implied probability' of a home win is $1/\tilde{O}^{HW}$.

Probabilities that do sum up to 1 are now obtained by scaling (Pope and Peel,1989, Goddard and Asimakopoulos, 2004):

$$\pi^{HW} = \frac{\theta^{HW}}{1+\lambda} = \frac{1}{(1+\lambda)\tilde{O}^{HW}}. \tag{2}$$

Terminology

Similarly, ‘implied probabilities’ for draws and away wins can be calculated, and invariably the sum of these ‘probabilities’ exceeds 1:

$$\theta^{HW} + \theta^{D} + \theta^{AW} = \frac{1}{\tilde{O}^{HW}} + \frac{1}{\tilde{O}^{D}} + \frac{1}{\tilde{O}^{AW}} \equiv 1 + \lambda, \tag{1}$$

The quotation of odds is in decimal notation, where the odds offered are the total payout per unit wagered. Hence, decimal odds exceed one. If the decimal odds of a home win are denoted by $\tilde{O}^{HW}$, the ‘implied probability’ of a home win is $1/\tilde{O}^{HW}$.

Probabilities that do sum up to 1 are now obtained by scaling (Pope and Peel,1989, Goddard and Asimakopoulos, 2004):

$$\pi^{HW} = \frac{\theta^{HW}}{1+\lambda} = \frac{1}{(1+\lambda)\tilde{O}^{HW}}. \tag{2}$$

Terminology

A capital market is said to be efficient if it fully and correctly reflects all relevant information in determining prices. Formally, the market is said to be efficient with respect to some information set . . . if security prices would be unaffected by revealing that information to all participants. Moreover, efficiency with respect to an information set . . . implies that it is impossible to make economic profits by trading on the basis [of that information set]. (Malkiel,1992).

Literature

- Dowie (1976) insider trading horse racing.
- Zuber et al. (1985) NFL

$$PS_{it} = \beta_0 + \beta_1 VL_{it} + \epsilon_{it}. \tag{3}$$

- Sauer et al (1988), Golec and Tamarkin (1991), Dare and MacDonald (1996) important refinements improve power and consistency of test (if team S wins by 5 points, B loses by 5 points).
- Pope and Peel (1989), Goddard and Asimakopoulos (2004) on soccer, and many others.

Literature

- Dowie (1976) insider trading horse racing.
- Zuber et al. (1985) NFL

$$PS_{it} = \beta_0 + \beta_1 VL_{it} + \epsilon_{it}. \tag{3}$$

- Sauer et al (1988), Golec and Tamarkin (1991), Dare and MacDonald (1996) important refinements improve power and consistency of test (if team S wins by 5 points, B loses by 5 points).
- Pope and Peel (1989), Goddard and Asimakopoulos (2004) on soccer, and many others.

Literature

- Dowie (1976) insider trading horse racing.
- Zuber et al. (1985) NFL

$$PS_{it} = \beta_0 + \beta_1 VL_{it} + \epsilon_{it}. \tag{3}$$

- Sauer et al (1988), Golec and Tamarkin (1991), Dare and MacDonald (1996) important refinements improve power and consistency of test (if team S wins by 5 points, B loses by 5 points).
- Pope and Peel (1989), Goddard and Asimakopoulos (2004) on soccer, and many others.

Literature

- Dowie (1976) insider trading horse racing.
- Zuber et al. (1985) NFL

$$PS_{it} = \beta_0 + \beta_1 VL_{it} + \epsilon_{it}. \tag{3}$$

- Sauer et al (1988), Golec and Tamarkin (1991), Dare and MacDonald (1996) important refinements improve power and consistency of test (if team S wins by 5 points, B loses by 5 points).
- Pope and Peel (1989), Goddard and Asimakopoulos (2004) on soccer, and many others.

Regression based tests

Since we take odds to be in decimal form, the expected payout (including stake) on a 1 unit bet is

$$O_{ijs}^{HW} \times \Pr(Y_{ijs}^{HW} = 1) + 0 \times \Pr(Y_{ijs}^{HW} = 0). \tag{4}$$

We assume that punters are collectively rational, that is, they do not engage in bets that have a negative expected return. Hence, if odds reflect all information, we have

$$O_{ijs}^{HW} \times \Pr(Y_{ijs}^{HW} = 1) + 0 \times \Pr(Y_{ijs}^{HW} = 0) = 1. \tag{5}$$

Now the small trick...

Regression based tests

Since we take odds to be in decimal form, the expected payout (including stake) on a 1 unit bet is

$$O_{ijs}^{HW} \times \Pr(Y_{ijs}^{HW} = 1) + 0 \times \Pr(Y_{ijs}^{HW} = 0). \tag{4}$$

We assume that punters are collectively rational, that is, they do not engage in bets that have a negative expected return. Hence, if odds reflect all information, we have

$$O_{ijs}^{HW} \times \Pr(Y_{ijs}^{HW} = 1) + 0 \times \Pr(Y_{ijs}^{HW} = 0) = 1. \tag{5}$$

Now the small trick...

Regression based tests

Since we take odds to be in decimal form, the expected payout (including stake) on a 1 unit bet is

$$O_{ijs}^{HW} \times \Pr(Y_{ijs}^{HW} = 1) + 0 \times \Pr(Y_{ijs}^{HW} = 0). \tag{4}$$

We assume that punters are collectively rational, that is, they do not engage in bets that have a negative expected return. Hence, if odds reflect all information, we have

$$O_{ijs}^{HW} \times \Pr(Y_{ijs}^{HW} = 1) + 0 \times \Pr(Y_{ijs}^{HW} = 0) = 1. \tag{5}$$

Now the small trick…

Regression based tests

In other words, informationally efficient odds imply $\Pr(Y_{ijs}^{HW} = 1) = 1/O_{ijs}^{HW}$. So estimate the logit model

$$\Pr(Y_{ijs}^{HW} = 1) = \frac{1}{1 + \exp(-\beta_0 - \beta_1 \log(O_{ijs}^{HW} - 1))}, \tag{6}$$

and test whether or not $\beta_0 = 0$ and $\beta_1 = -1$.

A more powerful test (Golec and Tamarkin, 1991) may be obtained by extending the logit model with variables z_{ijs} as in

$$\Pr(Y_{ijs}^{HW} = 1) = \frac{1}{1 + \exp(-\beta_0 - \beta_1 \log(O_{ijs}^{HW} - 1) - \gamma' z_{ijs})}, \tag{7}$$

and test whether $\beta_0 = 0$, $\beta_1 = -1$, and $\gamma = 0$.

Regression based tests

In other words, informationally efficient odds imply $\Pr(Y_{ijs}^{HW} = 1) = 1/O_{ijs}^{HW}$. So estimate the logit model

$$\Pr(Y_{ijs}^{HW} = 1) = \frac{1}{1 + \exp(-\beta_0 - \beta_1 \log(O_{ijs}^{HW} - 1))}, \qquad (6)$$

and test whether or not $\beta_0 = 0$ and $\beta_1 = -1$.
A more powerful test (Golec and Tamarkin, 1991) may be obtained by extending the logit model with variables z_{ijs} as in

$$\Pr(Y_{ijs}^{HW} = 1) = \frac{1}{1 + \exp(-\beta_0 - \beta_1 \log(O_{ijs}^{HW} - 1) - \gamma' z_{ijs})}, \qquad (7)$$

and test whether $\beta_0 = 0$, $\beta_1 = -1$, and $\gamma = 0$.

Economic tests

An alternative to these regression-based tests, are tests of efficiency that look for trading rules that provide the basis for a profitable betting strategy. Examples of such trading rules are: always bet on the home team, bet on the favorite, but they can also be more complex, such as, bet on the home team if the expected return of the bet is positive, the expectation being taken with respect to some statistical model. Efficiency is then tested by calculating the average return of the bets following the trading rule, and comparing that average to 0.

Are these results statistically significant?

Economic tests

An alternative to these regression-based tests, are tests of efficiency that look for trading rules that provide the basis for a profitable betting strategy. Examples of such trading rules are: always bet on the home team, bet on the favorite, but they can also be more complex, such as, bet on the home team if the expected return of the bet is positive, the expectation being taken with respect to some statistical model. Efficiency is then tested by calculating the average return of the bets following the trading rule, and comparing that average to 0.

Are these results statistically significant?

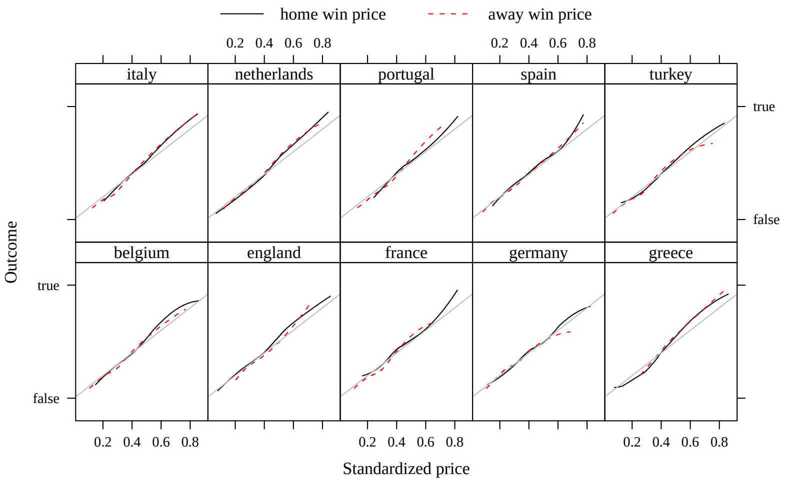
Data

The dataset consists of games from ten different highest level European leagues: Belgium, England, France, Germany, Greece, Italy, Netherlands, Portugal, Spain, and Turkey. Seasons covered are 2002/03 to 2009/10. The dataset consists of 25744 different soccer games. For each game, multiple odds offered by different bookmakers are available. The number of bookmakers vary by season and country. Bookmakers that appear in the dataset are Bet365, Blue Square, Bet&Win, Gamebookers, Interwetten, Ladbrokes, Sporting Odds, Sportingbet, Stan James, Stanleybet, Victor Chandler and William Hill.

Data

As a first check, we examine whether there are combinations of odds that offer a sure profit. This could be possible because odds of the same event vary between bookmakers. We find this is the case for 47 games, only 0.2% of all games. 34 of these games are from the 2004/05 season or earlier, when arbitrage by trading through the internet was perhaps less common, so we do not consider this to be pervasive evidence against full informational efficiency of betting odds.

Soccer efficiency



Soccer efficiency

Table: Estimates of single logit model for soccer bets, by result.

	home win		**draw**		**away win**	
	est.	std.err.	est.	std.err.	est.	std.err
β_0	0.121	0.014	0.550	0.081	0.121	0.02
β_1	-1.213	0.023	-1.600	0.081	-1.213	0.02
p-value	0.000		0.000		0.000	

True price is higher than observed price if $p > 0.5$, so true odds are lower than observed odds: favorite-longshot bias.

$$\Pr(Y^{HW} = 1) = \frac{1}{1 + e^{-\beta_0} e^{-(\beta_1 + 1)\log(O-1)} e^{-\gamma' z} e^{\log(O-1)}}$$

Increase in z, γ positive, decrease denominator, increase $\Pr(Y^{HW} = 1)$, (which justifies the parametrization).

Soccer efficiency

Table: Estimates of single logit model for soccer bets, by result.

	home win		draw		away win	
	est.	std.err.	est.	std.err.	est.	std.err
β_0	0.121	0.014	0.550	0.081	0.121	0.02
β_1	-1.213	0.023	-1.600	0.081	-1.213	0.02
p-value	0.000		0.000		0.000	

True price is higher than observed price if $p > 0.5$, so true odds are lower than observed odds: favorite-longshot bias.

$$\Pr(Y^{HW} = 1) = \frac{1}{1 + e^{-\beta_0} e^{-(\beta_1 + 1)\log(O-1)} e^{-\gamma' z} e^{\log(O-1)}}$$

Increase in z, γ positive, decrease denominator, increase $\Pr(Y^{HW} = 1)$, (which justifies the parametrization).

Additional variables

return	average return on one unit bets on last three games, for home team and away team
return H/A	average return on three one unit bets during last three home games for home team, and similar for away team
spread	best odds offered minus worst odds offered
time	dummy when game is played (season divided into ten deciles)

Additional variables

points	average points obtained during last three games, for home team and away team
position	position of home team and away team
goals	average number of goals scored during last three games, for home team and away team
goals H/A	average number of goals scored during last three home games for home team, and similar for away team

Additional variables

Table: *p*-values of test of efficiency with additional variables.

	home win		draw		away wir
	β, γ	γ	β, γ	γ	β, γ
base model	0.000		0.000		0.000
return	0.000	0.007	0.000	0.234	0.000
return H/A	0.000	0.000	0.000	0.004	0.000
spread	0.000	0.678	0.000	0.997	0.000
time	0.000	0.026	0.000	0.009	0.000
points	0.000	0.000	0.000	0.027	0.000
position	0.000	0.223	0.000	0.150	0.000
goals	0.000	0.817	0.000	0.918	0.000
goals H/A	0.000	0.321	0.000	0.253	0.000
all	0.000	0.010	0.000	0.375	0.000

Additional variables

- Mainly *HW* is affected, no variable significant in *AW* specification.
- Effects of away team performance not incorporated that well.
- Past return on bets is significant (away team only, $+$).
- Moving average past points obtained is significant (away team only, $+$).
- Instead, estimate

$$\Pr(Y_{ijs}^{HW} = 1) = \frac{1}{1 + \exp(-\beta_0 - g(\log(O_{ijs}^{HW} - 1)) - \gamma' z_{ijs})}, \tag{8}$$

  and test again $\gamma = 0$.

However, specification is not consistent (Dare and MacDonald, 1996).

Additional variables

- Mainly *HW* is affected, no variable significant in *AW* specification.
- Effects of away team performance not incorporated that well.
- Past return on bets is significant (away team only, $+$).
- Moving average past points obtained is significant (away team only, $+$).
- Instead, estimate

$$\Pr(Y_{ijs}^{HW} = 1) = \frac{1}{1 + \exp(-\beta_0 - g(\log(O_{ijs}^{HW} - 1)) - \gamma' z_{ijs})}, \tag{8}$$

  and test again $\gamma = 0$.

However, specification is not consistent (Dare and MacDonald, 1996).

Additional variables

- Mainly *HW* is affected, no variable significant in *AW* specification.
- Effects of away team performance not incorporated that well.
- Past return on bets is significant (away team only, $+$).
- Moving average past points obtained is significant (away team only, $+$).
- Instead, estimate

$$\Pr(Y_{ijs}^{HW} = 1) = \frac{1}{1 + \exp(-\beta_0 - g(\log(O_{ijs}^{HW} - 1)) - \gamma' z_{ijs})}, \tag{8}$$

and test again $\gamma = 0$.

However, specification is not consistent (Dare and MacDonald, 1996).

Additional variables

- Mainly *HW* is affected, no variable significant in *AW* specification.
- Effects of away team performance not incorporated that well.
- Past return on bets is significant (away team only, $+$).
- Moving average past points obtained is significant (away team only, $+$).
- Instead, estimate

$$\Pr(Y_{ijs}^{HW} = 1) = \frac{1}{1 + \exp(-\beta_0 - g(\log(O_{ijs}^{HW} - 1)) - \gamma' z_{ijs})}, \tag{8}$$

and test again $\gamma = 0$.

However, specification is not consistent (Dare and MacDonald, 1996).

Additional variables

- Mainly *HW* is affected, no variable significant in *AW* specification.
- Effects of away team performance not incorporated that well.
- Past return on bets is significant (away team only, $+$).
- Moving average past points obtained is significant (away team only, $+$).
- Instead, estimate

$$\Pr(Y_{ijs}^{HW} = 1) = \frac{1}{1 + \exp(-\beta_0 - g(\log(O_{ijs}^{HW} - 1)) - \gamma' z_{ijs})}, \tag{8}$$

and test again $\gamma = 0$.

However, specification is not consistent (Dare and MacDonald, 1996).

Additional variables

- Mainly *HW* is affected, no variable significant in *AW* specification.
- Effects of away team performance not incorporated that well.
- Past return on bets is significant (away team only, $+$).
- Moving average past points obtained is significant (away team only, $+$).
- Instead, estimate

$$\Pr(Y_{ijs}^{HW} = 1) = \frac{1}{1 + \exp(-\beta_0 - g(\log(O_{ijs}^{HW} - 1)) - \gamma' z_{ijs})}, \tag{8}$$

  and test again $\gamma = 0$.

However, specification is not consistent (Dare and MacDonald, 1996).

Multinomial logit

$$\Pr(Y_{ijs}^{HW} = 1) = \frac{\exp(-\beta_0^{HW} - \beta_1^{HW} LO_{ijs}^{HW})}{1 + \exp(-\beta_0^{HW} - \beta_1^{HW} LO_{ijs}^{HW}) + \exp(-\beta_0^{AW} - \beta_1^{AW} LO_{ijs}^{AW})}$$

$$\Pr(Y_{ijs}^{AW} = 1) = \frac{\exp(-\beta_0^{AW} - \beta_1^{AW} LO_{ijs}^{AW})}{1 + \exp(-\beta_0^{HW} - \beta_1^{HW} LO_{ijs}^{HW}) + \exp(-\beta_0^{AW} - \beta_1^{AW} LO_{ijs}^{AW})}$$

$$\Pr(Y_{ijs}^{D} = 1) = 1 - \Pr(Y_{ijs}^{HW} = 1) - \Pr(Y_{ijs}^{AW} = 1).$$

$$LO = \log(O - 1)$$

Multinomial logit

Table: Estimates of multinomial logit model for soccer bets (draw is the reference category).

	home win		**away win**	
	est.	std.err.	est.	std.err.
β_0	-0.058	0.024	-0.045	0.018
β_1	-1.243	0.036	-1.198	0.042
p-value	0.000			

Additional variables: to follow, but p=0 anyway...

Conclusions

- Simple, widely applicable test.
- Odds are not fully informationally efficient.
- Partly through nonlinearities.
- Past returns, past points away team significant as well.

To do:

- Extend multinomial logit model with additional variables.
- Similar test for tennis (spread between odds).
- Extend to online betting market (dependence, continuous time).

Conclusions

- Simple, widely applicable test.
- Odds are not fully informationally efficient.
- Partly through nonlinearities.
- Past returns, past points away team significant as well.

To do:

- Extend multinomial logit model with additional variables.
- Similar test for tennis (spread between odds).
- Extend to online betting market (dependence, continuous time).

Conclusions

- Simple, widely applicable test.
- Odds are not fully informationally efficient.
- Partly through nonlinearities.
- Past returns, past points away team significant as well.

To do:

- Extend multinomial logit model with additional variables.
- Similar test for tennis (spread between odds).
- Extend to online betting market (dependence, continuous time).

Conclusions

- Simple, widely applicable test.
- Odds are not fully informationally efficient.
- Partly through nonlinearities.
- Past returns, past points away team significant as well.

To do:

- Extend multinomial logit model with additional variables.
- Similar test for tennis (spread between odds).
- Extend to online betting market (dependence, continuous time).