Optimal Determination of Bookmakers' Betting Odds: Theory and Tests

Trinity Economic Paper Series
Technical Paper No. 96/9 (Revised: December 6, 1999)
JEL Classification: D82, G13

John Fingleton
Dept. of Economics
Trinity College
Dublin 2
Ireland
John.Fingleton@tcd.ie

Patrick Waldron
Dept. of Economics
Trinity College
Dublin 2
Ireland
Patrick.Waldron@tcd.ie

Abstract

This paper develops a theoretical model of how bookmakers' odds are determined, given varying levels of inside information on the part of punters. Bookmakers' attitudes towards risk and the degree of competition between them will influence bookmaker behaviour. Using a data set of 1,696 races in Ireland in 1993, we find that bookmakers are extremely risk-averse, and estimate that operating costs and monopoly rents combined account for up to 4 per cent of turnover and that between 3.1 and 3.7 per cent of betting is by punters with inside information.

Acknowledgements

We wish to thank Hyun Shin and Liam Mason (Irish Racing Services) for data and also participants at various seminars for useful suggestions. We remain responsible for any errors.

Non-Technical Summary

The odds or prices set by bookmakers on horse races exhibit four interesting properties which are acknowledged by practitioners and/or observed in empirical studies, the causes of which had been poorly understood by economic theoreticians until recently.

First, empirical odds invariably exhibit a favourite-longshot bias, whereby the prices of the favourites are relatively better value than those of the longshots. This bias is also observed in pool betting, with the interesting exception of data from Hong Kong. Second, the margin implicit in bookmakers' odds increases with the number of runners in the race. Third, this theoretical margin, calculated by summing the probabilities quoted by the bookmakers, overstates their realised operating profits, suggesting that punters can identify horses underpriced by bookmakers and exploit this 'inside information'. Fourth, margins vary greatly from country to country, even when market structure does not vary. In particular, bookmakers' prices are significantly higher (i.e. odds lower) in the Irish market than in the British market, although these markets differ only in size.

Shin (1991; 1992; 1993) provided a theoretical explanation for the first two phenomena, assuming inside information on the part of the punters. Since £1 bet on a horse at 3:1 exposes the bookmaker to a smaller potential loss than £1 bet on a horse at 30:1, the bookmaker will require a greater risk premium to insure himself against the possibility of inside information on a longshot. This is achieved by reducing the odds in respect of the longshot. The more horses in the race, the longer the odds on each, and thus the bigger the bookmaker's overall margin. Shin's empirical analysis estimates the extent of inside information without using the outcome of races to confirm the accuracy of that information.

Motivated by the differences between market outcomes in Ireland and Britain, this paper develops a more general model of determination of bookmaker betting odds, based on Shin's model, incorporating (a) infinite risk-aversion on the part of the bookmakers, (b) the possibility of anti-competitive behaviour among the bookmakers and (c) (as Shin does) different degrees of inside information on the part of the punters.

Shin's theoretical analysis is based on perfectly competitive, risk-neutral bookmakers. It ignores operating costs and assumes that any profits are competed away. In practice, bookmakers are often seen balancing their books so as to have identical liabilities across all horses. This would represent infinitely risk-averse behaviour and would not be profit maximising: a bookmaker can increase profits by setting slightly longer odds. While the level of risk-aversion among bookmakers alters the optimal prices, it does not affect the existence of either the favourite-longshot bias or the relationship between margins and the number of runners. Because optimal prices differ, we are able to test whether bookmakers are risk-neutral or infinitely risk-averse. We are interested in whether the higher bookmaker margins in the Irish market can be explained by greater risk aversion or greater levels of anti-competitive behaviour by bookmakers or by greater levels of inside information on the part of punters.

We use the results of 1,696 races in Ireland in 1993 to estimate jointly the extent of inside information, the operating profits earned and the degree of risk-aversion exhibited by bookmakers. Our methodology, using the closed form solution of Jullien & Salanié (1994) rather than the approximations of Shin (1993) and Vaughan Williams & Paton (1997), permits the estimation of "true" probabilities, which we find, for suitable parameter values, accurately reflect the subsequent outcomes.

Our conclusion suggests that bookmakers in Ireland are infinitely risk-averse and balance their books. We cannot distinguish between inside information and operating costs, merely concluding that combined they account for up to 3.7 per cent of turnover.

Section I: Introduction

A recent series of papers by Shin (1991; 1992; 1993) addresses, both theoretically and empirically, the determination of the prices of state contingent claims in a market (the horserace betting market) in which the presence of insider traders with superior information prevents the frictionless, symmetric-information, competitive equilibrium outcome of Arrow (1964) and Debreu (1959) from being attained. Shin, however, retains the traditional assumptions that there is perfect competition between risk-neutral bookmakers in this market and that transactions costs are negligible.

This paper, on the other hand, analyses the optimal determination of the prices (*i.e.* betting odds) set by bookmakers and faced by punters in such a market if instead bookmakers are risk-averse, engage in anti-competitive behaviour and/or face significant transactions costs. In practice, bookmakers are commonly viewed not as setting the risk-neutral odds of Shin's model, but as being preoccupied with guaranteeing a risk-free return by balancing their books, the equivalent of infinitely risk-averse behaviour. Such behaviour leads to an outcome not dissimilar to that attained by definition under a pool (*pari-mutuel*) betting system, so one could refer to such bookmakers as setting *pari-mutuel odds*. It will be argued that the least risk-averse bookmaker in the market will set the longest odds and do the most business, but even that bookmaker may be far from risk-neutral. It could also be the case that bookmakers earn monopoly rent, *i.e.* profits over and above those justified by the extent of informed trading and risk-aversion, or that the operating costs which they must recoup are substantial.

While there is an extensive literature on the efficiency of racetrack betting markets (see, for example, the collection of papers edited by Hausch, Lo & Ziemba (1994)), it is concerned mostly with pool betting. Until recently those few articles that dealt with bookmaker betting suffered from the joint-hypothesis problem (Fama, 1991) — market efficiency *per se* is not testable and must be tested jointly with some model of equilibrium, a model which does not seem to have existed prior to Shin's work.

The extensions of the theory presented in this paper are motivated by the very different market outcomes in the Irish and British horserace betting markets, which, while they could be explained by a more significant presence of insider traders in Ireland, suggest that Shin's model is not capturing all the factors at work in these markets.

Bookmakers in both markets quote state contingent claim prices in the form of an odds ratio, which can be viewed as representing the 'market probability' that the horse in question will win. The standard measure of market outcome used in the industry is the sum of these market probabilities, known as the *S.P. total percent*, described by the daily *Racing Post* as shown in Table 1.

The S.P. total percent will be denoted throughout this paper by β and the number of runners in the field by n. It will be seen that the positive relationship between β and n referred to in Table 1 is both predicted by our theory (if there are insider traders) and confirmed by our data. This empirical regularity motivates a second, less volatile, measure of the market outcome for a particular race, the *over-round per runner*, which is just $(\beta - 1)/n$. It is well-documented (*e.g.* Fingleton & Waldron (1995)) that the S.P. total percent is substantially higher in Ireland than elsewhere: for horseracing in Ireland in 1993, the over-round per runner averaged 3.33% over 1,696 races; for horseracing in Britain during one week in 1991,[1] it averaged only 1.86% over 136 races.

[1]This was the week used by Shin (1993) and Jullien & Salanié (1994) in their empirical work.

It will be argued that these substantial differences can be explained by a combination of a greater proportion of insider trading and higher gross margins in Ireland than in Britain and a greater degree of risk aversion exhibited by Irish bookmakers than by their British counterparts. As can be seen from Table 2, Ireland has far more races per capita than Britain, suggesting that insiders might account for a higher proportion of the Irish population than the British. Furthermore, betting per capita and, especially, per race, are much lower for Irish racing than British, exacerbating the potential significance of insiders in Ireland.

The extent of insider trading in horserace betting markets has previously been analysed empirically by Shin (1993), Jullien & Salanié (1994) and Schnytzer & Weiss (1996), all of whom have based their estimates solely on the betting odds observed before races, and none of whom have used the *results* of races either to confirm that those they identify as insiders are better informed than outsiders or to distinguish between competing models of the market. This paper uses race results to estimate jointly the extent of insider trading, the gross margins earned and the degree of risk aversion exhibited by bookmakers.

In Section II, we describe in more detail the structure of the horserace betting market. Section III presents detailed motivation for our explanations of the variations in S.P. total percent, which are then formalised in Section IV. In Section V, we compare different versions of our model, using data from races in Ireland in 1993. Our empirical results can be summarised as follows:

1. We reject the hypothesis that bookmakers behave in a risk-neutral manner;
2. We cannot reject the hypothesis that they are infinitely risk-averse;
3. We estimate gross margins to be up to 4 per cent of total on-course turnover; and
4. We estimate that 3.1 to 3.7 per cent (by value) of all bets are placed by punters with inside information.[2]

[2] Inside information of x per cent may be thought of as x per cent of wealth being held by punters with perfect foresight, or more wealth being held by less informed punters (see p.18).

Section II: The Bookmaking Market

The setting of odds by bookmakers is equivalent to pricing risky assets. A 'price' or odds ratio of five to two (5/2) represents an implicit probability of $\frac{2}{2+5} = \frac{2}{7}$, which is the price of a risky asset which pays out 1 in the event that the horse wins and 0 otherwise. Shorter odds (*e.g.*, two to one (2/1)) imply a higher probability and hence a higher price of the asset.

At racecourses in Ireland and Britain, a large number of independent bookmakers in the 'ring' (usually between 40 and 80) quote odds for each race, ostensibly in open competition with each other. The starting prices (S.P.) for a particular race are measured as those at which a substantial bet could have been placed with a leading bookmaker at the start or 'off' of the race. They are determined by sports news agencies and reported to the off-course market, where most bets are settled at S.P. In the on-course market, the price the punter pays for the asset is that which prevails when the bet is made, not the S.P. However, most on-course betting occurs close to the 'off' of the race and therefore at prices close to S.P.

Section III: Explanations for the Overround

In a perfectly competitive, zero-cost market, with risk-neutral bookmakers, no taxes or levies, and no inside information on the part of punters, total percent would equal to one. Anything less than one would enable a punter to earn risk-free arbitrage profits by betting on horses in proportion to their quoted probabilities so as to be sure of a net gain on the race. A value of total percent in excess of one would give monopoly rent that would be bid away.

There are four reasons in general why total percent might exceed one. These are:

1. **Positive operating costs.**

   These include licence fees as well as labour and equipment costs.

2. **Positive monopoly rent.**

   This arises if bookmakers collude or otherwise avoid competing directly.

3. **Risk aversion on the part of bookmakers.**

   If bookmakers are risk-averse, then equilibrium prices will incorporate a risk premium and exceed the competitive level. Infinitely risk-averse bookmakers would strive to equalise liabilities across all horses, thus ensuring themselves a risk-free race. Their profits would necessarily be lower than those of bookmakers who maximised expected profits. The latter, however, would be exposed to infinite risks with non-zero probability.

4. **Inside information on the part of punters.**
   If punters have inside information which is not available to bookmakers, then (even risk-neutral) bookmakers will price horses above the competitive level to insure themselves.

Section IV: The Model

Any beliefs concerning the outcome of a three-horse race can be represented by a point in the triangle OAB in Figure 1, with the point (p_1, p_2) assigning probabilities p_1 to the event that horse 1 wins, p_2 to the event that horse 2 wins and $1 - p_1 - p_2$ to the event that horse 3 wins.

The centroid, C, for example, where all probabilities are equal, can be thought of as representing either the beliefs of a completely agnostic (or perfectly uninformed) punter about any race, or the beliefs of any punter about a wide open race (i.e. equal true probabilities for all horses). Similarly, a vertex such as A can be thought of as representing either the beliefs of a perfectly informed punter about any race, or the beliefs of any punter about a race which is a foregone conclusion.

If horse 1, say, is destined to win, then the closer a punter's beliefs are to the vertex A, the better informed she is. In this case, punters with beliefs closer to A have positive inside information about horse 1; punters with beliefs further from A can be viewed as either uninformed or misinformed. On the other hand, if horse 1 is destined to lose, then the further a punter's beliefs are from A, the better informed she is. In this case, punters with beliefs further from A from have negative inside information about horse 1; punters with beliefs closer to A can now be viewed as either uninformed or misinformed.

The model assumes that there are two classes of punters, who have differential information about the chances of one horse, say horse 3, but agree on the relative probabilities of the others. The two classes of punters can thus be represented by points such as E and F on a ray such as OD in Figure 1. If there is positive inside information about horse 3, then the point E represents the informed beliefs and the point F represents the uninformed beliefs. On the other hand, if there is negative inside information about horse 3, then the identification is reversed. The extreme case of positive inside information is represented by the point O and the extreme case of negative inside information by the point D. If E represents uninformed beliefs, then any convex combination of O and E can represent informed beliefs, where the information is positive; and any convex combination of E and D can represent informed beliefs, where the information is negative. Note that misinformation in this framework merely represents a mislabelling of the informed and uninformed, so this possibility can safely be ignored.

More generally, when there are n runners in a race, if uninformed beliefs are represented by the probability vector

$$\mathbf{p} = (p_1, \ldots, p_n) \tag{1}$$

(where $0 \leq p_i \leq 1$ for $i = 1, \ldots, n$ and $\Sigma_{i=1}^{n} p_i = 1$), then

- the beliefs of a punter with positive inside information about horse i are represented by the vector

$$(1-\alpha)\,\mathbf{p}+\alpha\mathbf{e}_i, \tag{2}$$

where $\mathbf{e}_i = (0,\ldots,0,1,0,\ldots,0)$ is the degenerate distribution assigning all probability mass to the ith horse and $0 < \alpha \leq 1$ and

- the beliefs of a punter with negative inside information about horse i are represented by the vector

$$(1-\gamma)\,\mathbf{p}+\frac{\gamma}{1-p_i}\left(\mathbf{p}-p_i\mathbf{e}_i\right), \tag{3}$$

where $\mathbf{p}-p_i\mathbf{e}_i = (p_1,\ldots,p_{i-1},0,p_{i+1},\ldots,p_n)$ is the distribution assigning zero probability mass to the ith horse, $0 < \gamma \leq 1$ and $\frac{1}{1-p_i}$ normalises.

(As our empirical tests use the results of races, they implicitly assume that uninformed expectations are rational, which implies, barring misinformation, that informed expectations are also rational.)

The structure of the model is as follows. In **Stage I**, Bookmakers compete by quoting a vector of probabilities,

$$\boldsymbol{\pi} = (\pi_1, ..., \pi_n), \tag{4}$$

chosen subject to the constraint that the implicit total percent $\Sigma_{j=1}^{n} \pi_j$ does not exceed some perceived target value β^*, interpreted as the total percent which the most effective competitor is expected to set. The determination of β^* permits collusion between bookmakers, and will show below that, for a given attitude towards risk, the bookmaker setting the lowest total percent will set the longest odds about every horse in the race (since beliefs are homogenous across bookmakers). In **Stage II**, A bookmaker is approached by a representative punter with one unit of wealth. With probability $1 - z^*$ she is uninformed (like the bookmaker) and with probability z^* she is informed. She draws a horse at random from the distribution representing her beliefs and bets on that horse.[3]

We deal here with positive inside information (negative inside information is dealt with in Appendix A below). From the

[3]Equivalently, it could be assumed that there is a continuum of punters; that the fraction z^* of all punters are informed and the fraction $1 - z^*$ are uninformed; and that each subset bets on the n horses in proportion to their beliefs.

bookmaker's point of view, the representative punter encountered has beliefs

$$\begin{cases} \mathbf{p} & \text{with probability } 1 - z^* \\ (1-\alpha)\,\mathbf{p} + \alpha \mathbf{e}_i & \text{with probability } z^* \end{cases}. \tag{5}$$

The turnover on a race can be normalised to 1, enabling the expected liabilities to be denoted by $1 - R$ where $R \geq 0$ is the gross margin of the bookmaker. It is not possible to separate out the two components of the gross margin, operating costs and monopoly rent.

Conditional on horse i winning, the uninformed punter backs horse i with probability p_i and the informed punter backs horse i with probability $(1-\alpha)\,p_i + \alpha$. Thus, conditional on horse i winning, the bookmaker expects bets on it of

$$p_i\,(1 - z^*) + ((1-\alpha)\,p_i + \alpha)\,z^* = p_i\,(1 - \alpha z^*) + \alpha z^*. \tag{6}$$

As it is not possible to identify the parameters z^* and α separately, we can write $\alpha z^* = z$.[4] If the bookmaker quotes a probability of

[4]This makes precise the descripton of inside information in footnote 2 on page 10 and enables a more general interpretation of z than in the existing literature, which has always assumed $\alpha = 1$ (Shin, 1993; Jullien & Salanié, 1994).

π_i about horse i, then his (expected) liabilities conditional on its winning are

$$\frac{p_i\,(1-z)+z}{\pi_i}. \tag{7}$$

Since the bookmaker shares the beliefs of the uninformed, unconditional expected liabilities are

$$\sum_{i=1}^{n} p_i \frac{p_i\,(1-z)+z}{\pi_i}. \tag{8}$$

We next consider the behaviour first of the risk-neutral bookmaker who maximises expected profits and then of the infinitely risk-averse bookmaker who equalises expected liabilities on all horses. The model of Shin (1993) is a special case of the former in which $R = 0$ and $\alpha = 1$.

A : Expected Profit Maximisation

Following Shin (1993) and using (8), the expected profit of the bookmaker quoting $\boldsymbol{\pi}$ is

$$V\,(\boldsymbol{\pi}) = 1 - \sum_{j} \frac{zp_j + (1-z)\,p_j^2}{\pi_j}, \tag{9}$$

which is maximised subject to each π_j lying in $(0,1)$ and

$$\sum_j \pi_j \leq \beta^*. \tag{10}$$

There is an interior solution except when the perceived outcome of the race is close to a foregone conclusion. The total percent constraint binds and using the n first order conditions

$$zp_i + (1-z)\,p_i^2 = \lambda \pi_i^2 \tag{11}$$

gives the solution

$$\pi_i = \frac{y_i}{\Sigma_j\, y_j}\beta^*, \tag{12}$$

where

$$y_i = \sqrt{zp_i + (1-z)\,p_i^2}\,. \tag{13}$$

This confirms that the bookmaker setting the lowest total percent sets the longest odds for every horse.

Setting $V(\boldsymbol{\pi})$, evaluated at optimal $\boldsymbol{\pi}$, equal to the gross margin R gives

$$\pi_i = \frac{y_i\,\Sigma_j\, y_j}{1-R}. \tag{14}$$

Jullien & Salanié (1994) showed (for $R = 0$) that this solution can be inverted to recover the *ex ante* probability beliefs from the *ex post* quoted probabilities:

$$p_j = \frac{-z + \sqrt{z^2 + 4(1-z)\frac{1-R}{\beta}(\pi_j)^2}}{2(1-z)}. \qquad (15)$$

Since $\Sigma_{i=1}^{n} p_i = 1$, it is easily shown that z solves:

$$z = \frac{\Sigma_j \sqrt{z^2 + 4(1-z)\frac{1-R}{\beta}(\pi_j)^2} - 2}{n-2}. \qquad (16)$$

It can be shown that (16) has exactly two solutions (including $z = 1$) (see Appendix B below). From (16), given R, n and the S.P. of each horse (*i.e.* $\boldsymbol{\pi}$), we can find z iteratively using the recurrence relation:

$$z[0] := 0 \qquad (17)$$

$$z[i+1] := \frac{\Sigma_j \sqrt{z[i]^2 + 4(1-z[i])\frac{1-R}{\beta}(\pi_j)^2} - 2}{n-2}. \qquad (18)$$

Substituting in (15) allows us to recover an implicit $\mathbf{p}$ from $\boldsymbol{\pi}$ for each R. The optimal total percent (β) is

$$\frac{(\Sigma\, y_j)^2}{1-R}, \qquad (19)$$

which depends on the number of runners and on the *evenness of match* and which ranges from a minimum of $\frac{1}{1-R}$ for a foregone conclusion to a maximum of $\frac{(n-1)z+1}{1-R}$ for a wide open race. In our empirical work below, we calculate $\mathbf{p}(R)$ and $z(R)$ from $\boldsymbol{\pi}$ for several values of R for each race in our data set.

Note that the expected liability on horse i is

$$\frac{p_i(1-z)+z}{\pi_i} = \frac{y_i^2}{p_i\pi_i} = \frac{(1-R)}{\Sigma_j\, y_j}\frac{y_i}{p_i} = \sqrt{\frac{(1-R)}{\beta}\left(\frac{z}{p_i}+(1-z)\right)} \tag{20}$$

which is decreasing in p_i (given β) but approaches infinity as $p_i \to 0$. In races where some runners have virtually no chance, the possibility of bankruptcy therefore makes risk-neutral pricing impractical because of the absence of infinite credit lines.

B : Balanced Books

For the balanced books case, the bookmaker will equalise his liabilities across horses. Setting the expected liability on horse i to $1-R$ yields:[5]

$$\pi_i = \frac{p_i(1-z)+z}{1-R}. \tag{21}$$

[5]Expected liabilities exclude gross margin.

Inverting this we find

$$p_i = \frac{(1-R)\,\pi_i - z}{1-z} \tag{22}$$

and, using $\Sigma_{i=1}^{n} p_i = 1$,

$$z = \frac{(1-R)\,\beta - 1}{n-1}. \tag{23}$$

Substituting in (22) again allows us to recover an implicit $\mathbf{p}$ from $\boldsymbol{\pi}$ for each R. The optimal total percent (β) is

$$\frac{(n-1)\,z+1}{1-R}, \tag{24}$$

and thus

$$\pi_i = \frac{p_i\,(1-z)+z}{(n-1)\,z+1}\beta, \tag{25}$$

confirming again that the bookmaker setting the lowest total percent sets the longest odds for every horse. The optimal total percent in this case is independent of any evenness of match measure (and in fact coincides with the extreme case of risk-neutral pricing where all horses have equal chances). This confirms, as noted in the introduction, that risk-neutral odds always offer the punter better value, in terms of S.P. Total Percent, than do pari-mutuel odds. It follows that the least risk-averse bookmaker will do the most business.

In our empirical work using the balanced books version of the model, we again calculate $\mathbf{p}(R)$ and $z(R)$ for a range of values of R from data on $\boldsymbol{\pi}$.

C : Comparison

Both versions of the model predict both the familiar empirical regularity known as the *favourite-longshot bias* and the strong positive relationship between total percent and number of runners which is evident in our data.

In the case of the favourite-longshot bias, the proof given by Shin (1993, p.1147) for the risk-neutral case with $R = 0$ carries through straightforwardly to $R > 0.$ For the balanced books case, the expected return to an uninformed bet on horse i is

$$\frac{p_i}{\pi_i} = \frac{(1-R)\,p_i}{p_i\,(1-z)+z} \tag{26}$$

which is increasing in p_i, showing that the favourite-longshot bias is preserved when bookmakers balance books.

The starting price data for a given race yield two different invertible relationships between R and z:

1. derived from profit maximising behaviour (a non-linear relationship given by (16))
2. derived from balanced books behaviour (a linear relationship given by (23))

The two coincide in the case of perfectly even match.[6]

An example of the possible relationships between z and R is presented in Figure 2 which is based on a 22-runner race at Fairyhouse on January 1, 1993. The total percent is $\beta = 1.80713$ which is inconsistent with $R > 0.44664\%$. From (23), the (absolute) slope of the balanced books line is $\frac{\beta}{22-1} = 0.08605$ and its $z-$intercept is $\frac{\beta-1}{22-1} = 3.843\%$. The example confirms that the difference between the gross margin (R) of the bookmakers in both versions increases as the level of inside information increases.

Section V: Estimation and Results

We have two sets of data on starting prices, that used by Shin on 136 races run in Britain in July 1991 and the data on all

[6] Shin (1993, p.1146) points out that as the race becomes a foregone conclusion the solutions presented here no longer represent the optimal strategy, which instead involves some π_i exceeding 1, or 'No S.P. returned' as very occasionally happens in practice. This problem is greater, the bigger are z and R, but not great enough to have occured empirically in our data.

1696 races run in Ireland (32 counties) in 1993. Given two data sets and two versions of the model, our results are presented in four tables. The maximised expected profit tables (Tables 4 and 6) show the results of using (17) and (18) to calculate z for different values of R and those for balanced books (Tables 3 and 5) show the corresponding results using (23). Since both versions of the model predict that $\beta < \frac{1}{1-R}$, races violating this condition are dropped from the sample for the relevant values of R.

Table 3 shows the relationship for balanced books for the Irish data in 1993. Thus for a gross margin of 8 per cent, the average level of inside information is 2.5 per cent. By contrast, Table 4 for the same data set indicates that the average level of inside information would be 2.7 per cent if expected profits were being maximised.

The results for the British races in Shin's data set are presented in Table 5 for balanced books and Table 6 for expected profit maximisation. These tables illustrate the fact that the British market either has lower gross margins or lower inside information. For a similar 8 per cent gross margin, the level of inside information is just 1.1 per cent. Jullien & Salanié (1994), improving on the method of (Shin, 1993), cite levels of inside information of 2.3 per cent for this group of races. Our investigations reveal that this figure would be slightly lower at 2.2 per cent if bookmakers balanced books and considerably lower if account is taken of operating costs.

Any test we can carry out using the results of races will be a joint test of the model and of the hypothesis that punters have rational expectations. In particular, given rational expectations we expect that the favourite-longshot bias which is present in $\boldsymbol{\pi}$ should not be present in $\mathbf{p}\,(R)$. We investigate this hypothesis using the average pay out ratio (*APOR*), defined for any bet or betting strategy as the ratio of gamblers' winnings to stakes (which include all explicit taxes, levies or duties). For example, consider the strategy of backing a horse, chosen at random in proportion to the S.P. probabilities, to return a fixed amount. The total percent (β) is just the cost of backing all horses to return one unit. Thus, from a gross revenue of β, the bookmaker pays 95 pence to the punter (*i.e.* £1 less a 5 per cent levy[7]), so that the APOR for this strategy is $\frac{0.95}{\beta}$.

Tables 7 and 8 use the APORs for the strategies of back-

[7] On-course bookmaking in the Republic of Ireland incurs a 5 per cent levy to the Irish Horseracing Authority,

ing all horses at each S.P. to investigate the existence of a favourite-longshot bias in the quoted odds $\boldsymbol{\pi}$ and the underlying probabilities $\mathbf{p}(R)$ respectively. In both tables, the first column indicates the probability implicit in each of the 57 odds in the second column, which are the only S.P. odds returned in Ireland during 1993.[8] The third column in each table indicates the theoretical APOR for a punter who invested the same amount on each horse starting at the given odds, assuming that the 5 per cent levy is deducted from winnings. The figure in brackets is the number of horses starting at the given odds. In Table 8 the figure in brackets is the number of horses *estimated to have been* at the given odds, assuming a balanced books approach and a gross margin (R) of 2 per cent; in this table, $\mathbf{p}(R)$ has been rounded up for each horse to the next highest probability actually observed in the sample.

The favourite-longshot bias would be manifest as a positive relationship between the APOR and the probability, indicating that outsiders are priced relatively less favourably than favourites. With rational expectations on the part of the punters, we would expect:

[8] In practice, only a limited number of rational numbers are encountered as S.P. odds.

1. A positive relationship between APOR and $\boldsymbol{\pi}$;
2. No significant relationship between APOR and the implicit true probabilities $\mathbf{p}(R)$.

A crude estimator for R is the value which yields the weakest relationship between APOR and probability.

Table 9 reports the results of the regressions of the observed APOR for each cohort on the probability, for various versions of the model including that on which Table 8 is based. The first row of this table indicates a strongly significant positive relationship for the actual probabilities $\boldsymbol{\pi}$. For the regressions based on $\mathbf{p}(R)$, statistically significant positive relationships are still found except for the balanced books version of the model with values of R between 0 and 4 per cent. We thus conclude that we cannot accept the hypothesis that bookmakers maximise expected profits, but are unable to reject the balanced books version of the model for this range of values of R.

Conclusion

This paper has presented a model, of which Shin's is a special case, showing how three factors (informed trading, risk-aversion and the bookmaker margin) combine to determine optimal odds. The level of informed trading in Shin's model can be inferred exactly from the odds quoted on any given race, as can the true *ex ante* (uninformed) probabilities that each horse will win (Jullien & Salanié, 1994). In our model, however, we can make these inferences only by both assuming a bookmaker margin and fixing the (binary) choice between profit-maximising and book-balancing behaviour. While one can easily suggest reasons why the level of informed trading might vary from race to race, the other parameters of the model should be more or less constant across races, provided that the same population of bookmakers are participating on each race, so this seems a valid empirical approach. However, these latter parameters might be very different in Ireland from what they are in Britain.

In future work, we intend to develop more sophisticated measures of the closeness of $\mathbf{p}(R)$ to the outcome of the races and use these to obtain a better empirical estimate of R. We will also investigate the relationships between both R and z and various proxies for the level of operating costs and the extent of insider trading which can be easily observed.

Table 1: The *Racing Post*'s description of the S.P. total percent.

Betting Percentages

- THE S.P. total percent figure in our results shows by how much the betting was over-round.
- To calculate the percentage each price is worth, add a point and divide into 100. So 3–1 becomes four into 100, equals 25.
- A perfectly round book, with each price representing the true chance, would total 100.
- A book totalling 125 would be 25 percent over-round and give bookmakers a theoretical profit of 20 per cent (25 divided by 125).
- In general, the bigger the field, the more the percentage favours bookmakers.

Table 2: Betting turnover in the Republic of Ireland and the United Kingdom.

	Republic of Ireland	**United Kingdom**
Population	3.5m.	58.1m.
Races	1,612	7,200
per million pop.	455	124
Total betting	86.8m.	4561.0m.
per race	0.05m.	0.63m.
per capita	24.5	78.5

Sources: **?**, p.30), **?**, pp.16,19), Horserace Betting Levy Board accounts, United Nations Population Division.

Notes: The Republic of Ireland betting figure is in IEP and is for on-course turnover in the 1994 calendar year. Off-course turnover in the Republic of Ireland is mostly on races run in Britain and the proportion accounted for by domestic racing is much smaller than the on-course turnover.

The United Kingdom betting figure is in STG and is for off-course horseracing turnover in the 1994/5 financial year. While figures for on-course turnover in the United Kingdom are not readily available (as it is not subject to any levies or duties), it is believed to be small relative to off-course turnover.

The average exchange rate in 1994 was IEP1=STG0.9778.

Table 3: Summary Statistics for Total Percent and the implicit proportion of inside information in Irish racing, assuming risk-averse bookmakers.

		R					
	β	**0**	**0.02**	**0.04**	**0.06**	**0.08**	**0.1**
Means	1.426	0.037	0.034	0.031	0.028	0.025	0.023
St. Devs.	0.225	0.009	0.009	0.009	0.009	0.009	0.009
Maxima	2.426	0.097	0.085	0.073	0.061	0.054	0.052
Minima	1.048	0.012	0.008	0.003	0.001	0.001	0.000
Count	1696	1696	1696	1696	1692	1686	1671

Source: Irish data for 1993.

Notes: This table reports the total percent (β column) and the level of inside information (*i.e.* z-values) corresponding to six different profit levels (R) for 1696 races in Ireland in 1993 for the case where bookmakers balance their books. As R increases, the number of races with a total percent consistent with non-negative inside information declines and these are excluded.

Table 4: Summary statistics for total percent and the implicit proportion of inside information in Irish racing, assuming risk-neutral bookmakers.

		R					
	β	**0**	**0.02**	**0.04**	**0.06**	**0.08**	**0.1**
Means	1.426	0.040	0.037	0.034	0.030	0.027	0.024
St. Devs.	0.225	0.010	0.010	0.010	0.010	0.010	0.010
Maxima	2.426	0.107	0.093	0.079	0.066	0.062	0.059
Minima	1.048	0.012	0.008	0.003	0.001	0.001	0.000
Count	1696	1696	1696	1696	1692	1686	1671

Source: Irish data for 1993.
Notes: This table reports the total percent (β column) and the level of inside information (*i.e.* z-values) corresponding to six different profit levels (R) for 1696 races in Ireland in 1993 for the case where bookmakers maximise expected profits. As R increases, the number of races with a total percent consistent with non-negative inside information declines and these are excluded.

Table 5: Summary statistics for total percent and the implicit proportion of inside information in British racing, assuming risk-averse bookmakers.

		R					
	β	**0**	**0.02**	**0.04**	**0.06**	**0.08**	**0.1**
Mean	1.188	0.022	0.019	0.016	0.013	0.011	0.009
St. Dev	0.107	0.007	0.007	0.007	0.007	0.007	0.006
Max	1.584	0.049	0.046	0.042	0.039	0.035	0.032
Min	0.986	0.005	0.000	0.001	0.001	0.000	0.000
Count	136	135	135	133	130	120	106

Source: Shin data for July 1991.
Notes: This table reports the total percent (β column) and the level of inside information (*i.e.* z-values) corresponding to six different profit levels (R) for 136 races in the Shin sample for the case where bookmakers balance their books. One race in the sample had a total percent less than 1, giving 135 observations. As R increases, the number of races with a total percent consistent with non-negative inside information declines and these are excluded.

Table 6: Summary statistics for total percent and the implicit proportion of inside information in British racing, assuming risk-neutral bookmakers.

		R					
	β	**0**	**0.02**	**0.04**	**0.06**	**0.08**	**0.1**
Mean	1.188	0.023	0.020	0.016	0.013	0.011	0.009
St. Dev	0.107	0.008	0.008	0.007	0.007	0.007	0.007
Max	1.584	0.056	0.052	0.047	0.043	0.039	0.035
Min	0.986	0.005	0.000	0.001	0.001	0.000	0.000
Count	136	135	135	133	130	120	106

Source: Shin data for July 1991.

Notes: This table reports the total percent (β column) and the level of inside information (*i.e.* z-values) corresponding to six different profit levels (R) for 136 races in the Shin sample for the case where bookmakers maximise expected profits. One race in the sample had a total percent less than 1, giving 135 observations. As R increases, the number of races with a total percent consistent with non-negative inside information declines and these are excluded.

Table 7: The favourite-longshot bias in π.

Prob.	Odds	APOR	(No.)	Prob.	Odds	APOR	(No.)
0.004	250/1	0.0%	(1)	0.267	11/4	82.8%	(86)
0.005	200/1	0.0%	(1)	0.286	5/2	81.5%	(310)
0.007	150/1	0.0%	(4)	0.308	9/4	64.4%	(224)
0.010	100/1	0.0%	(84)	0.320	85/40	0.0%	(3)
0.012	80/1	0.0%	(1)	0.333	2/1	88.0%	(204)
0.015	66/1	0.0%	(157)	0.348	15/8	86.3%	(19)
0.020	50/1	15.1%	(640)	0.364	7/4	88.4%	(164)
0.024	40/1	0.0%	(123)	0.381	13/8	112.6%	(31)
0.029	33/1	9.8%	(1324)	0.400	6/4	90.1%	(174)
0.038	25/1	16.1%	(1224)	0.421	11/8	76.8%	(47)
0.048	20/1	16.9%	(2244)	0.444	5/4	74.8%	(135)
0.059	16/1	24.9%	(1317)	0.476	11/10	70.8%	(62)
0.067	14/1	33.7%	(1981)	0.500	1/1	89.1%	(96)
0.077	12/1	41.2%	(1828)	0.526	9/10	105.8%	(29)
0.083	11/1	43.0%	(53)	0.556	4/5	72.0%	(76)
0.091	10/1	52.0%	(2070)	0.579	8/11	93.8%	(21)
0.100	9/1	76.1%	(256)	0.600	4/6	96.0%	(33)
0.111	8/1	65.2%	(1442)	0.619	8/13	127.9%	(6)
0.118	15/2	115.4%	(84)	0.636	4/7	96.0%	(28)
0.125	7/1	62.7%	(946)	0.652	8/15	0.0%	(1)
0.133	13/2	78.3%	(173)	0.667	1/2	109.0%	(17)
0.143	6/1	72.9%	(782)	0.680	40/85	139.7%	(1)
0.154	11/2	90.3%	(229)	0.692	4/9	91.5%	(6)
0.167	5/1	71.8%	(683)	0.714	2/5	72.5%	(11)
0.182	9/2	83.4%	(357)	0.750	1/3	84.4%	(6)
0.200	4/1	55.8%	(511)	0.800	1/4	79.2%	(3)
0.222	7/2	92.8%	(373)	0.818	2/9	116.1%	(1)
0.231	100/30	89.1%	(134)	0.833	1/5	114.0%	(1)
0.250	3/1	81.1%	(342)				

Source: Irish data for 1993.

Notes: The first column gives the probability implicit in each of the 57 odds in the second column. The third column indicates the theoretical average pay out ratio (APOR) for a punter who invested the same amount on each horse starting at the given odds (with a 5% levy). The figure in brackets is the number of horses that started at the given odds.

Table 8: The favourite-longshot bias in $\mathrm{p}(R)$ assuming a balanced books approach and $R = 2\%$.

Prob.	Odds	APOR	(No.)	Prob.	Odds	APOR	(No.)
0.004	250/1	51.6%	(3237)	0.267	11/4	92.7%	(242)
0.005	200/1	122.4%	(156)	0.286	5/2	94.5%	(204)
0.007	150/1	160.0%	(269)	0.308	9/4	83.9%	(173)
0.010	100/1	116.8%	(575)	0.320	85/40	127.2%	(56)
0.012	80/1	106.6%	(433)	0.333	2/1	98.4%	(84)
0.015	66/1	37.7%	(507)	0.348	15/8	97.1%	(97)
0.020	50/1	53.4%	(817)	0.364	7/4	97.2%	(43)
0.024	40/1	62.5%	(810)	0.381	13/8	103.4%	(152)
0.029	33/1	61.1%	(899)	0.400	6/4	85.5%	(50)
0.038	25/1	56.4%	(1686)	0.421	11/8	84.8%	(125)
0.048	20/1	88.0%	(1451)	0.444	5/4	62.3%	(24)
0.059	16/1	70.8%	(1528)	0.476	11/10	85.9%	(137)
0.067	14/1	96.7%	(825)	0.500	1/1	109.4%	(33)
0.077	12/1	81.7%	(915)	0.526	9/10	73.2%	(37)
0.083	11/1	98.2%	(592)	0.556	4/5	81.4%	(63)
0.091	10/1	92.4%	(577)	0.579	8/11	98.5%	(30)
0.100	9/1	86.7%	(515)	0.600	4/6	126.7%	(10)
0.111	8/1	98.2%	(601)	0.619	8/13	98.7%	(28)
0.118	15/2	99.3%	(309)	0.636	4/7	0.0%	(2)
0.125	7/1	86.4%	(233)	0.652	8/15	118.4%	(16)
0.133	13/2	76.0%	(328)	0.667	1/2	71.3%	(2)
0.143	6/1	89.6%	(349)	0.680	40/85	111.8%	(5)
0.154	11/2	96.8%	(287)	0.692	4/9	85.8%	(8)
0.167	5/1	106.3%	(327)	0.714	2/5	33.3%	(4)
0.182	9/2	97.2%	(328)	0.750	1/3	101.3%	(5)
0.200	4/1	95.3%	(359)	0.800	1/4	89.1%	(4)
0.222	7/2	92.3%	(338)	0.818	2/9	116.1%	(1)
0.231	100/30	101.1%	(114)	0.833	1/5	—	(0)
0.250	3/1	81.3%	(159)				

Source: Irish data for 1993.

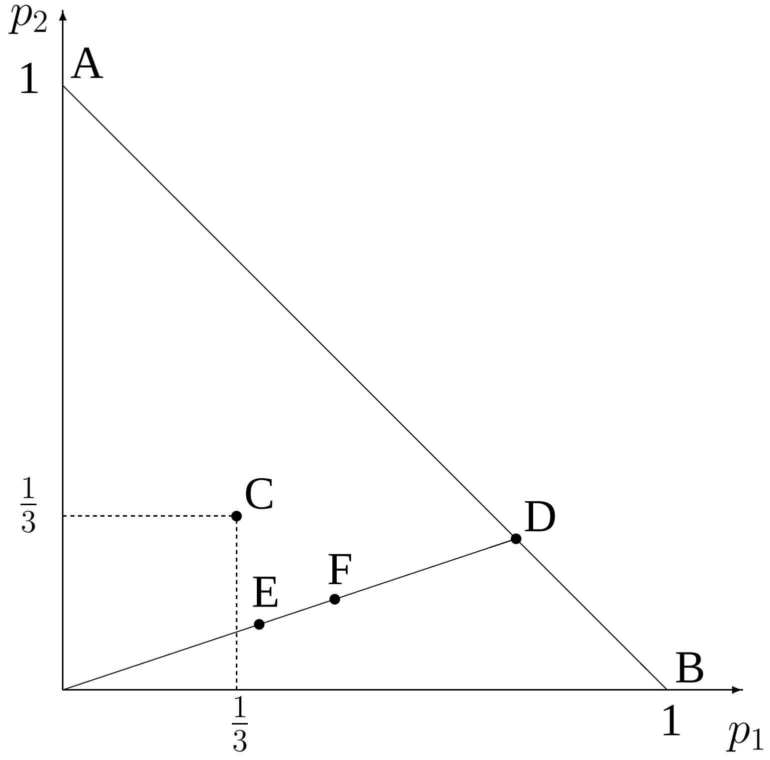
Notes: The first column gives the probability implicit in each of the 57 odds in the second column. The third column indicates the theoretical average pay out ratio (APOR) for a punter who invested the same amount on each horse starting at the given odds (with a 5% levy). The figure in brackets is the number of horses *estimated to have been* at the given odds.

Table 9: Results of regressing APOR on probability for different Rs and for both strategies.

		Intercept		**Prob**	
Profit	R^2	coeff	P-value	coeff	P-value
Actual ($\boldsymbol{\pi}$)					
—	0.423	34.2	0.000	98.4	0.000
Balanced Books ($\mathbf{p}$)					
0%	0.001	97.1	0.000	-5.4	0.820
2%	0.000	89.0	0.000	0.5	0.972
4%	0.123	78.1	0.000	32.9	0.009
6%	0.112	76.2	0.000	35.6	0.013
8%	0.086	77.7	0.000	33.6	0.031
10%	0.319	70.8	0.000	62.7	0.000
Maximise Expected Profits ($\mathbf{p}$)					
0%	0.164	69.8	0.000	58.7	0.002
2%	0.254	64.2	0.000	77.7	0.000
4%	0.237	64.4	0.000	76.1	0.000
6%	0.267	63.1	0.000	86.7	0.000
8%	0.234	63.7	0.000	83.9	0.000
10%	0.236	63.3	0.000	86.8	0.000

Source: Irish data for 1993.

Fig. 1: Different probability beliefs for a three-horse race.



A and B represent foregone conclusions or perfect information; C a wide open race or agnostic beliefs. Relative to uninformed beliefs at F, E represents positive information about horse 3, and D represents (extremely) negative information about horse 3.

Fig. 2: Relationship between z and R for both expected profit maximisation and balanced books strategies, based on the first race run in Ireland in 1993.

References

Arrow, K. J. (1964), 'The role of securities in the optimal allocation of risk-bearing', *Review of Economic Studies* **31**, 91–96.

Debreu, G. (1959), *Theory of Value: An Axiomatic Analysis of Economic Equilibrium*, Yale University Press, New Haven.

Fama, E. F. (1991), 'Efficient capital markets: II', *Journal of Finance* **46**, 1575–1617.

Fingleton, J. & Waldron, P. J. M. (1995), Margins and risk premia in bookmaking, *in* D. Neven & L.-H. Röller, eds, 'The Empirical Analysis of Industrial Organisation', Center for Economic Policy Research/Wissenschaftszentrum Berlin für Sozialforschung.

Hausch, D. B., Lo, V. S. Y. & Ziemba, W. T., eds (1994), *Efficiency of Racetrack Betting Markets*, Academic Press, San Diego.

Jullien, B. & Salanié, B. (1994), 'Measuring the incidence of insider trading: A comment on Shin', *Economic Journal* **104**, 1418–1419.

Schnytzer, A. & Weiss, A. (1996), Insider trading and state-contingent claims, working paper, Department of Economics, Bar-Ilan University.

Shin, H. S. (1991), ‘Optimal betting odds against insider traders’, *Economic Journal* **101**, 1179–1185.

Shin, H. S. (1992), ‘Prices of state contingent claims with insider traders, and the favourite-longshot bias’, *Economic Journal* **102**, 426–435.

Shin, H. S. (1993), ‘Measuring the incidence of insider trading in a market for state-contingent claims’, *Economic Journal* **103**, 1141–1153.

Vaughan Williams, L. & Paton, D. (1997), ‘Why is there a favourite-longshot bias in British racetrack betting markets?’, *Economic Journal* **107**, 150–158.

Appendix A: Negative Inside Information

In the case of negative inside information (about horse j), the representative punter encountered, from the bookmaker's point of view, has beliefs

$$\begin{cases} \mathbf{p} & \text{with probability } \; 1-z^* \\ (1-\gamma)\,\mathbf{p}+\frac{\gamma}{1-p_j}\left(\mathbf{p}-p_j\mathbf{e}_j\right), & \text{with probability } \; z^* \end{cases}. \tag{27}$$

Conditional on horse i winning, the uninformed punter backs horse i with probability p_i and the informed punter backs horse i with probability

$$(1-\gamma)\,p_i+\frac{\gamma}{1-p_j}p_i=\left(\frac{1-p_j\,(1-\gamma)}{1-p_j}\right)p_i. \tag{28}$$

Thus, conditional on horse i winning, the bookmaker expects bets on it of

$$p_i\,(1-z^*)+\left(\frac{1-p_j\,(1-\gamma)}{1-p_j}\right)p_i z^*$$

As it is not possible to identify the parameters p_j, z^* and γ separately, we can write[9]

$$\left(1-\left(\frac{1-p_j\,(1-\gamma)}{1-p_j}\right)\right)z^*=z. \tag{29}$$

[9]Note that in this case, z is negative, and the magnitude of z is positively related to z^*, γ and p_j.

If the bookmaker quotes a probability of π_i about horse i, then his (expected) liabilities conditional on its winning are

$$\frac{p_i\left(1-z\right)}{\pi_i}. \tag{30}$$

Since the bookmaker shares the beliefs of the uninformed, unconditional expected liabilities are

$$\sum_{i=1}^{n} p_i \frac{p_i\left(1-z\right)}{\pi_i}. \tag{31}$$

The expected profit of the bookmaker quoting $\boldsymbol{\pi}$ is now

$$V\left(\boldsymbol{\pi}\right) = 1 - \sum_{i=1}^{n} \frac{p_i^2\left(1-z\right)}{\pi_i}, \tag{32}$$

which is again maximised by the risk-neutral bookmaker subject to each π_j lying in $(0,1)$ and

$$\sum_j \pi_j \leq \beta^*. \tag{33}$$

The total percent constraint binds and using the n first order conditions

$$\left(1-z\right)p_i^2 = \lambda \pi_i^2 \tag{34}$$

gives the solution

$$\pi_i = p_i \beta^*. \tag{35}$$

In other words, the risk-neutral bookmaker facing negative inside information will set probabilities in proportion to uninformed beliefs. Using (30), his optimal expected liabilities on horse i are $\frac{(1-z)}{\beta^*}$, which is the same for all horses. In other words, the risk-neutral and risk-averse strategies coincide when inside information is negative in character.

Setting $V(\boldsymbol{\pi})$, evaluated at optimal $\boldsymbol{\pi}$, equal to the gross margin R now gives

$$R = 1 - \frac{(1-z)}{\beta}, \tag{36}$$

so that optimal total percent (β) is

$$\frac{1-z}{1-R}, \tag{37}$$

which is independent of the number of runners and the evenness of match. Since the quoted probabilities are proportional to the (rational) *ex ante* uninformed expectations, there is no favourite-longshot bias in this case.

The clear presence both of the positive relationship between the number of runners and the total percent and of the favourite-longshot bias in our empirical data suggests that the version of the model with positive inside information is much closer to reality than the version with negative inside information. One might even conclude that doping to win is much more likely than doping to lose.

Appendix B: Solutions to (16)

The proof that (16) has exactly two solutions (including $z = 1$) will follow from the well-known fact that a sum of strictly convex functions is strictly convex if it can be shown that $\sqrt{z^2 + 4(1-z)\frac{1-R}{\beta}(\pi_j)^2}$ is strictly convex. The first derivative of this expression (with respect to z) is

$$\frac{2z - 4\frac{1-R}{\beta}(\pi_j)^2}{2\sqrt{z^2 + 4(1-z)\frac{1-R}{\beta}(\pi_j)^2}}$$

Therefore the second derivative is positive if and only if

$$\begin{aligned} 4\left(\sqrt{\left(z^2 + 4(1-z)\frac{1-R}{\beta}\pi_j^2\right)}\right)^2 \; &> \; \left(2z - 4\frac{1-R}{\beta}\pi_j^2\right)^2 \\ \Leftrightarrow 4\left(z^2 + 4(1-z)\frac{1-R}{\beta}\pi_j^2\right) \; &> \\ & 4z^2 - 16z\frac{1-R}{\beta}\pi_j^2 + 16\left(\frac{1-R}{\beta}\right)^2\pi_j^4 \\ \Leftrightarrow 1 - z \; &> \; -z + \frac{1-R}{\beta}\pi_j^2 \end{aligned}$$

$$\Leftrightarrow \pi_j^2 \; < \; \frac{\beta}{1-R}$$

Given $\beta > 1$, $R > 0$ and $\pi_j < 1$, this inequality always holds.