Measuring the Incidence of Insider Trading in a Market for State-Contingent Claims

Hyun Song Shin


Stable URL:
http://links.jstor.org/sici?sici=0013-0133%28199309%29103%3A420%3C1141%3AMTIOIT%3E2.0.CO%3B2-Q

Your use of the JSTOR archive indicates your acceptance of JSTOR’s Terms and Conditions of Use, available at http://www.jstor.org/about/terms.html. JSTOR’s Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

The Economic Journal is published by Royal Economic Society. Please contact the publisher for further permissions regarding the use of this work. Publisher contact information may be obtained at http://www.jstor.org/journals/res.html.

The Economic Journal
©1993 Royal Economic Society

JSTOR and the JSTOR logo are trademarks of JSTOR, and are Registered in the U.S. Patent and Trademark Office. For more information on JSTOR contact jstor-info@umich.edu.

©2002 JSTOR
MEASURING THE INCIDENCE OF INSIDER TRADING IN A MARKET FOR STATE-CONTINGENT CLAIMS*

Hyun Song Shin

The classical view of markets is as a place which matches buyers and sellers at a common price – the price at which supply equals demand. In practice, however, most financial transactions take place through an intermediary, such as the market maker, who holds inventories of assets, and who sets prices for buyers and sellers. Invariably, the price facing buyers differs from that facing sellers. One of the main reasons for this divergence is that the market maker faces an adverse selection problem in which a customer may be trading on the basis of superior information. In this case the optimal bid-ask spread is determined in a trade-off between setting a large spread so as to minimise the loss to the insider traders, and setting the optimal spread against the ‘outsiders’. Bagehot (1971) hints at this source of the bid-ask spread, and subsequent papers by Copeland and Galai (1983), Glosten and Milgrom (1985) and Kyle (1985) have provided formal treatments of the problem.

Given that the spread is increasing with the incidence of insider trading, the size of the observed spread provides some indication of the severity of market distortion due to insider trading. This is particularly so in financial markets where market makers are engaged in active price competition, since we would expect that excessive spreads will be competed away. Given an appropriate model of asset price formation in which market makers engage in price competition, we may hope to infer the incidence of insider trading from the spreads quoted by market makers.

This paper is an attempt at such measurement through a model of price formation in a market for state-contingent claims. The specific setting for our investigation is the market for bets in a horse race, in which the role of the market makers are taken by the bookmakers, and the traders are played by the potential bettors (or ‘punters’). Thus, our description of betting is based on the system in the United Kingdom in which bookmakers set odds rather

* I have gained much from the comments of Stephen Bond, James Mirrlees and John Vickers. I also thank the editor and two referees for detailed comments and guidance. Earlier versions of this paper were presented at seminars in Cambridge, Southampton, QMWC, and Oxford, and at the European Summer Meeting of the Econometric Society in Cambridge, 1991.
than the system in North America in which odds are determined by the parimutuel method.

The betting market is an ideal vehicle for the implementation of our project for several reasons. First, the betting market provides an unambiguous and readily accessible measure of the market spread. In its simplest formulation, the market for bets in an \( n \)-horse race corresponds to a market for contingent claims with \( n \) states in which the \( i \)th state corresponds to the outcome in which the \( i \)th horse wins the race. In this market, the basic securities (or Arrow–Debreu securities) which pay a pound if a particular state obtains and nothing otherwise, have their prices determined by the betting odds. Then, the sum of the prices on all horses gives the price of a portfolio which pays a pound for sure at the end of the race. The divergence of this sum of prices from one represents the bookmakers' margin, and this margin provides an unambiguous measure of the size of the market spread. This is in marked contrast to more sophisticated markets, such as the stock market, in which the spread varies across assets and also across volumes traded.

More generally, the betting market is a particularly simple example of a financial market. The market convenes for about half an hour, at the end of which there is a definite and commonly acknowledged outcome. Again, this is in contrast to the complex decisions faced by traders and market makers in more sophisticated financial markets in which considerations of the distant future play an integral part in current decisions.

The betting market in the United Kingdom is well suited for our investigation. Bookmaking in the United Kingdom is a large and prominent industry, in which both the small independent bookmakers who operate at the track and the large chains of betting shops across the country engage in price competition. Also, there is a sizeable body of evidence, both systematic and anecdotal, which points to the prevalence of insider trading in the market for bets (see, for example, Crafts (1985)). The Committee of Inquiry of the Jockey Club acknowledges that 'trainers and their staff are insufficiently paid for their services and the majority have to resort to betting to make ends meet' (1968, p. 91). It would seem, therefore, that the activity of insider traders influences the outcome in the betting market in a significant way.

Our modelling of the betting market has been guided by two principles, both arising from the requirements of our empirical investigation. The first is that the predicted prices should be uniquely determined by the parameters of the problem, and moreover, that the predicted prices should be expressible in an explicit and reasonably simple way in terms of these parameters.

The second principle is that the predicted prices should conform to certain well-known stylised facts about prices in betting markets. Chief among these is the fact that the percentage mark-ups in the prices over the true probabilities are not uniform. In general, prices exhibit favourite-longshot bias in which, the normalised prices on the favourites of the race understate the winning chances of these horses, while the normalised prices on the longshots exaggerate their winning chances. Dowie (1976) confirms this bias, and Shin (1991, 1992) discusses some of the theoretical issues underlying this bias.

© Royal Economic Society 1993
The following table summarises Dowie’s (1976) investigation into the British flat racing season of 1973, as presented by Crafts (1985, p. 300). It shows the ratio of winnings to wager for subsets of the runners ordered by their betting odds. There were approximately 3,000 races in the 1973 flat season with almost 30,000 runners.

If the prices were proportional to true probabilities, the winnings/wager ratio would be constant over all subsets. However, the odds were biased in such a way that contingent claims on the favourites were cheaper (relative to true probabilities) than the claims on the longshots. This is the favourite-longshot bias.

In contrast to the betting system in the United Kingdom in which bookmakers set odds, the betting odds at North American race tracks are determined by the parimutuel method in which prices are proportional to amounts wagered. This distinction should be borne in mind when examining North American data (see Ali (1977), Asch et al. (1982), Quandt (1986), Ziemba and Hausch (1986) and Thaler and Ziemba (1988)).

The overview of our paper is as follows. In the next section, we present our model of the betting market in an n-horse race. The solution of the model follows in Section II. In Section III, we provide the transition from the theoretical model to the empirical part of the paper by deriving an empirical model. Section IV reports on the empirical findings of this paper.

## I. THEORETICAL MODEL

Our model of the betting market is a simplified version of that in Shin (1992). It describes a market for bets in a horse race with n horses. The market is organised as an extensive form game in which two bookmakers compete in setting odds in anticipation of the betting behaviour to follow. There are three players in the game – the incumbent bookmaker, a potential bookmaker, and the bettor with insider information (the ‘Insider’). There is also a set of n individuals called ‘outsiders’, labelled by the set \( \{1, 2, \ldots, n\} \). The outsiders are not modelled formally as players in the game. Rather, they are mechanical traders who do not act strategically. In particular, the ith outsider attaches probability 1 to the ith horse winning the race. However, although each outsider (taken individually) may be quite irrational, the game is designed so that the market prices which follow from outsiders’ demands for bets are fully revealing, in the sense that the equilibrium prices coincide with the true probabilities.
More elaborate modelling of the outsiders would have been possible here. The extreme assumption on beliefs could be relaxed somewhat to yield more realistic behaviour of the outsiders at an individual level. The qualitative features of the model would remain unchanged. The motivation in formulating the model as below has been to focus attention on the price-setting by the bookie. For this purpose, we have chosen to model the outsiders in the simplest way possible while still giving rise to two key features: firstly, that beliefs diverge among the outsiders, and yet (secondly) their aggregate behaviour reveals the true probabilities.

There are \( n \) types of tickets which are sold in the market. The \( i \)th type of ticket pays one pound if the \( i \)th horse wins and zero otherwise. The price of this ticket is denoted by \( \pi_i \). These prices correspond to betting odds in the usual way. Odds of \( k \) to \( l \) correspond to the price of \( l/(k+l) \). Negative odds are ruled out in our model, so that \( 0 \leq \pi_i \leq 1 \) for all \( i \). Fractional quantities of tickets may be sold.

The game is in extensive form and describes the encounter between one bookie and one punter. It has an alternative interpretation in which the bookie meets many punters, as described below. The game has three stages.

*Stage 1 (Bidding Stage).* The two bookies – the incumbent and the potential entrant – bid for monopoly rights to the betting market. Each submits a sealed bid of a positive real number. The bookie who submits the lower number wins the bid. If the bids are identical, the incumbent wins. The bookie who loses the bid gets a payoff of zero.

*Stage 2 (Price-setting Stage).* The bookie who wins the bid at stage 1 sets prices \( \pi_1, \pi_2, \ldots, \pi_n \) for the \( n \) types of tickets subject to the constraint that the sum \( \Sigma_i \pi_i \) does not exceed the bid submitted in stage 1 and that \( 0 \leq \pi_i \leq 1 \), for all \( i \).

*Stage 3 (Betting Stage).* Nature then performs two independent experiments. The bookie cannot observe the outcome in either experiment. In the first, Nature chooses the winner of the race according to given probabilities \( p_1, p_2, \ldots, p_n \). The \( i \)th horse is chosen with probability \( p_i \). We assume that \( p_i > 0 \) for all \( i \). In the second experiment, Nature chooses precisely one punter to bet against the price-setting bookie. With probability \( z \), the Insider is chosen to be the punter, and with probability \( (1-z) p_i \), the \( i \)th outsider is chosen to be the punter. If the Insider is chosen, she is permitted to observe the identity of the winning horse, and is free to buy tickets from the bookie at the posted prices up to the value of one pound. On the other hand, the \( i \)th outsider is an expected payoff maximiser with the belief that the \( i \)th horse wins with probability \( 1 \). Thus, the \( i \)th outsider always bets a pound on the \( i \)th horse. Once the betting has been completed, the race is run and the horse chosen by Nature is seen to win the race. The bookie then settles with the punter in accordance with the odds offered at stage 2.

The assumption that two bookies compete away their profits by bidding for monopoly rights is motivated by our main interest, which is the determination of prices given a zero profit condition. For us, the essential features are that (i) the bookmaker offers the full menu of bets, and (ii) expected profit is zero. Both these features are captured in our model. An alternative modelling strategy
would have been to allow fully-fledged Bertrand competition with an explicit market-sharing rule. Provided that the bookies offer the full menu of bets, the outcome of price competition coincides with that of our model.

Although our model is couched in terms of the encounter between one bookie and one punter, we could interpret the model as one in which the bookie meets a large number of punters, where the probabilities $p_i$ denote the proportion of each type of punter in the population. The problem for the bookie is identical.

II. EQUILIBRIUM PRICES

We now proceed to the solution of the game. We begin by solving for the bookie's optimal prices given the winning bid $\beta$, the parameter $z$ and the probabilities $p_1, \ldots, p_n$. By substituting these expressions into the bookie's profit function, we obtain an expression for the price-setting bookie's profit in terms of the bid $\beta$. By solving for the equilibrium bid $\beta$ and substituting into the expressions for $\pi_1, \ldots, \pi_n$, we obtain explicit solutions for the equilibrium prices in terms of the parameters of the model — namely, the incidence of insider trading, $z$, and the vector of winning probabilities, $p_1, \ldots, p_n$.

We denote by $\mathbf{p}$ the vector of winning probabilities $(p_1, \ldots, p_n)$ and by $\pi$ the vector of prices $(\pi_1, \ldots, \pi_n)$. We can represent the uncertainty facing the bookie at the price-setting stage in terms of the following diagram.

![Fig. 1.](image)

Fig. 1 shows the joint distribution over the sample space arising from the pair of independent experiments performed by Nature at stage 3 of the game. The $i$th row of the matrix represents the event that horse $i$ is chosen to win the race, while the $j$th column represents the event that the $j$th outsider is chosen to be the punter. The $(n + 1)$th column represents the event that the Insider is chosen to play. The shaded cells indicate those events in which the bookie pays out to the punter. Notice that the bookie always pays out to the Insider.

Any punter bets precisely one pound on a particular horse. If the punter bets on the $i$th horse, $1/\pi_i$ units of the $i$th ticket are sold to the punter. Thus, at stage

© Royal Economic Society 1993
2 of the game, the bookie expects to pay out \( z \sum_i (p_i / \pi_i) \) to the Insider, and \((1 - z) \sum_i (\beta^i / \pi_i)\) to the set of outsiders. Since the bookie’s revenue is constant at one pound, the bookie’s problem at stage 2 is to;

\[
\begin{align*}
\text{maximise} \quad & 1 - \sum_i \frac{z p_i + (1 - z) \beta^i}{\pi_i}, \\
\text{subject to} \quad & \sum_i \pi_i \leq \beta \quad \text{and} \quad 0 \leq \pi_i \leq 1 \quad \text{for all } i,
\end{align*}
\]

where \( \beta \) is the winning bid made at stage 1. The feasible set is given by the intersection of the set \( \{\pi | \sum_i \pi_i \leq \beta\} \) and the unit cube \( \{\pi | 0 \leq \pi_i \leq 1, \forall i\} \). Since both sets are convex, the feasible set is itself convex. The objective function is strictly concave in \( \pi \) as can be verified from the Hessian, which is a diagonal matrix with negative entries. Thus, any solution to the bookie’s maximisation problem at stage 2 is unique.

The solution of the game falls under two cases — namely, the case in which the optimal \( \pi \) lies in the interior of the unit cube and the case in which it lies on the boundary of the unit cube. Since profit is increasing in prices, to say that the optimal \( \pi \) lies on the boundary of the unit cube is to say that \( \pi_i = 1 \) for some \( i \). But since the \( i \)th ticket pays at most one pound, this is tantamount to saying that the bookie refuses to accept bets on the \( i \)th horse.

The case in which the equilibrium \( \pi \) lies on the boundary of the unit cube is a theoretical possibility which must be taken seriously, especially when \( z \) is large. In Shin (1991, 1992), corner solutions are examined under a variety of assumptions. For the purpose of this paper, however, we shall not dwell on this case. This entails no loss of generality for our empirical work, since it happens in practice that bets are accepted on all horses at finite odds. For the rest of this paper, then, we shall assume that the solution to the bookmaker’s optimisation problem at stage three lies in the interior of the unit cube. However, since expected profit is strictly increasing in prices, the constraint \( \sum_i \pi_i \leq \beta \) binds, and the optimal \( \pi \) satisfies the first-order conditions. Solving for \( \pi_i \) in terms of \( \beta, z \) and \( p \),

\[
\pi_i = \frac{\beta \sqrt{[zp_s + (1 - z) \beta^s]}}{\sum_s \sqrt{[zp_s + (1 - z) \beta^s]}}.
\]

Substituting (2) into (1), we have the following expression for expected profit in stage 1 in terms of \( \beta \).

\[
1 - \frac{1}{\beta} \left\{ \sum_s \sqrt{[zp_s + (1 - z) \beta^s]} \right\}^2.
\]

The payoff of both bookmakers must be zero in any equilibrium. It is at least zero since a bookie gets a payoff of zero by bidding a larger number than its rival. It is at most zero since the rival can undercut. Thus, in any equilibrium, the winning bid \( \beta \) is the number which sets (3) equal to zero. That is,

\[
\beta = \left\{ \sum_s \sqrt{[zp_s + (1 - z) \beta^s]} \right\}^2.
\]
Substituting (4) into (2), we can solve for $\pi_i$ in terms of the parameters $z$ and $p$.

$$\pi_i = \sqrt{zp_i + (1-z)p_t^2} \left\{ \sum_s \sqrt{zp_s + (1-z)p_t^2} \right\}. \quad (5)$$

This completes the solution of the game. We note the following properties of equilibrium prices.

**Property 1.** $p \leq \pi$ in any equilibrium. $p = \pi$ if and only if $z = 0$.

**Property 2.** If $z > 0$, equilibrium prices exhibit favourite-longshot bias. That is,

$$\pi_i/\pi_j < p_i/p_j \iff p_i > p_j. \quad (6)$$

To see that property 1 holds, note that

$$\sqrt{zp_i + (1-z)p_t^2} \geq \sqrt{zp_i + (1-z)p_t^2} = p_t$$

so that $\pi_i \geq p_t(\Sigma_s p_s) = p_t$. The inequality is strict whenever $z > 0$, and $\pi_i = p_i$ if and only if $z = 0$. In other words, in the absence of insider trading, the equilibrium prices coincide with the true probabilities, and any deviation of prices from the true probabilities is due to a positive incidence of insider trading.

To verify property 2, we note from (5) and the fact that all components of $p$ are non-zero, that:

$$\frac{\pi_i}{\pi_j} = \frac{\sqrt{zp_i + (1-z)p_t^2}}{\sqrt{zp_j + (1-z)p_t^2}} = \frac{p_i}{p_j} \sqrt{\frac{1-z+(z/p_t)}{1-z+(z/p_t)}}. \quad (7)$$

Thus, $p_i > p_j \iff z/p_t < z/p_t \iff \pi_i/\pi_j < p_i/p_j$. This is the favourite-longshot bias.

As a footnote to property 1, we offer the observation that the aggregation of information in our game mimics the determination of competitive prices in an economy consisting of the $n$ outsiders, where the $i$th outsider is endowed with share $p_i$ of the portfolio $(1, 1, \ldots, 1)$. Then, the total demand for the $i$th type of ticket is $p_i(\Sigma_s p_s)/p_i$, so that market clearing ensures $\pi_i = p_i$, $\Sigma_s p_s$, and prices are proportional to the true probabilities. Notice also that the normalised price on the $i$th horse is identical to the proportion of the total wealth wagered on the $i$th horse. In order words, competitive prices are the parimutuel betting odds (see Eisenberg and Gale (1959)).

**III. EMPIRICAL MODEL**

Having verified that our model is consistent with the stylised facts, we turn our attention to the task of estimating the parameter $z$, which is our measure of the incidence of insider trading. The focus of our attention will be on the expression for the sum of prices (4). The first step in our empirical investigation is to disentangle the effect on the sum of prices of the two parameters of our model $-z$ and $p$. For this purpose, consider the function $F(p_i) = \sqrt{zp_i + (1-z)p_t^2}$ and its second order approximation by Taylor expansion around the point $1/n$, which is $F(1/n) + F'(1/n) (p_i - 1/n) + \frac{1}{2} F''(1/n) (p_i - 1/n)^2$. Summing over $i$,

$$\sum_{i=1}^{n} F(p_i) = nF(1/n) + F'(1/n) \sum_{i=1}^{n} (p_i - 1/n) + \frac{1}{2} F''(1/n) \sum_{i=1}^{n} (p_i - 1/n)^2. \quad (7)$$

© Royal Economic Society 1993
Since $\mathbf{p}$ is a probability vector, the second term disappears. The sum of squares $\sum_i (p_i - 1/n)^2$ is the square of the Euclidean distance between $\mathbf{p}$ and the centre of the unit simplex. It gives an indication of the ‘evenness of match’ of the horses. With some abuse of terminology, we shall refer to $(1/n) \sum_i (p_i - 1/n)^2$ as the ‘variance’ of $\mathbf{p}$ and denote it by $\text{Var}(\mathbf{p})$. Since $nF(1/n) = \sqrt{[1 + z(n-1)]} + \frac{3}{4} nF''(1/n) \text{Var}(\mathbf{p})$. (8)

From the solution of the equilibrium bid $\beta$ given by (4), the sum of prices $\sum_i n_i$ is the square of $\sum_i F(p_i)$. Thus, by squaring both sides of (8) and subtracting one from both sides, we obtain an expression for the deviation of the sum of prices from one. We denote this deviation by $D$ and note that;

$$D = z(n-1) + n \sqrt{[1 + z(n-1)]} F''(1/n) \text{Var}(\mathbf{p}) + \frac{3}{4} n^2 [F''(1/n) \text{Var}(\mathbf{p})]^2. \quad (9)$$

Let $A = n \sqrt{[1 + z(n-1)]} F''(1/n)$ and $B = n^2 [F''(1/n)]^2$. Since $z$ is a constant, both $A$ and $B$ are non-linear algebraic functions of $n$, which are continuous in $n$ for $n > 0$. So, by appeal to the Weierstrass Approximation Theorem, $A$ and $B$ can be approximated to any arbitrary degree of accuracy by polynomials in $n$. Thus, let

$$A = \sum_{k=0}^{K} a_k n^k, \quad B = \sum_{k=0}^{K} b_k n^k,$$

for some integer $K$ and constants $\{a_k\}$ and $\{b_k\}$, where $k = 0, 1, \ldots, K$.

Substituting these into (9), we obtain the following equation, which expresses $D$ as a linear combination of the variables $(n-1), \{n^k \text{Var}(\mathbf{p})\}$, and $\{n^k [\text{Var}(\mathbf{p})]^2\}$, where $k = 0, 1, \ldots, K$.

$$D = z(n-1) + \sum_{k=0}^{K} a_k n^k \text{Var}(\mathbf{p}) + \sum_{k=0}^{K} b_k n^k [\text{Var}(\mathbf{p})]^2. \quad (10)$$

This equation is the basis of our empirical measurement of the incidence of insider trading. It decomposes the market spread $D$ into additive components. The term $z(n-1)$ can be interpreted as the part of $D$ which can be attributed ‘directly’ to $z$. The other terms may be seen as adjustment terms which measure the indirect effect of $z$ through the vector $\mathbf{p}$. Therefore, we may regard these terms as a ‘filter’ in the sense that they filter out the effects on $D$ of the vector $\mathbf{p}$. The fact that the direct effect $z(n-1)$ is increasing in the number of runners has an intuitive explanation. For a fixed bet of one pound, the Insider ends up with $1/n_i$ pounds, where $i$ is the winning horse. Hence, the Insider’s payoff will be large when there are many runners (so that longer odds are offered on the horses). Since the bookies must break even in expected terms, they raise their margins in order to recoup their loss to the Insider.

Although (10) decomposes $D$ into additive components, $\text{Var}(\mathbf{p})$ is not directly observable, and we must find a procedure which renders equation (10) empirically operational. An immediate thought might be to calculate the normalised prices $\tilde{\pi}_k = \pi_k / (\sum_i \pi_i)$, and use the vector of normalised prices $\tilde{\pi}$ as

© Royal Economic Society 1993
a proxy for the true probabilities $\mathbf{p}$. Provided that $z$ is small, we may expect $\hat{\mathbf{p}}$ to be a good proxy for $\mathbf{p}$. However, the favourite-longshot bias will imply that $\hat{\mathbf{p}}$ diverges systematically from $\mathbf{p}$, even though this divergence may be small in practice.

This problem may be overcome by utilising our explicit solution for the equilibrium prices $\pi$. From (5), the sum of squares of the normalised prices $\bar{\pi}_k$ can be seen to be $[z + (1 - z) \sum_i \hat{\pi}_i^2] / \beta$. In turn, Var ($\hat{\mathbf{p}}$) can be written in terms of this sum of squares as $(\sum_i \bar{\pi}_i^2 / n) - (1 / n^2)$. In a similar way, Var ($\mathbf{p}$) can be written as $(\sum_i \pi_i^2 / n) - (1 / n^2)$. Substituting out $\sum_i \bar{\pi}_i^2$ and rearranging, we derive an expression for Var ($\mathbf{p}$) in terms of Var ($\hat{\mathbf{p}}$) as follows:

$$\text{Var} (\mathbf{p}) = \frac{\beta}{1 - z} \text{Var} (\hat{\mathbf{p}}) + \frac{\beta - 1 - z(n - 1)}{n^2 (1 - z)}. \quad (11)$$

Thus, given an estimate of $z$, we can calculate Var ($\mathbf{p}$) from Var ($\hat{\mathbf{p}}$) and other observables. This suggests an iterative procedure for the estimation of $z$. It has three steps.

Step 1. Using Var ($\hat{\mathbf{p}}$) as a proxy for Var ($\mathbf{p}$) in equation (10), obtain an initial estimate of $z$ from an ordinary least squares regression of (10).

Step 2. From the initial estimate of $z$ obtained above, calculate Var ($\mathbf{p}$) by using (11). Then, use these values of Var ($\mathbf{p}$) to adjust the values of the regressors in equation (10) and re-estimate equation (10) to obtain a revised estimate of $z$.

Step 3. Repeat the adjustment of regressors in (10) and re-estimate (10) to obtain revised estimates of $z$ until these revised estimates converge.

The estimate of $z$ obtained through this procedure will be one which is consistent with the adjustments of the regressors for the favourite-longshot bias.

IV. ESTIMATION RESULTS

The data for our investigation is the set of ‘starting prices’ reported by the daily racing newspaper *Sporting Life*. Starting prices are the odds compiled by the journalists of the two racing dailies, *Sporting Life* and *Sporting Chronicle* who observe the market at the track for the ten minutes or so prior to the race and note the prices at which bets are accepted just before the start. Our sample is the set of races run in the United Kingdom in the week of Monday, July 1st
1991 to Saturday, July 6th 1991 (there was no racing on Sundays in the United Kingdom in 1991), as reported in the Sporting Life of Tuesday, 2nd July to Monday 8th July. There are 136 races in our sample. A number of summary statistics of the sample are of interest. The distribution of the sum of prices is shown on the previous page.

It is surprising to find a race in our sample in which the sum of prices is actually less than one. This is the 2:15 at Yarmouth on the 3rd of July, at which the sum of prices was 0.987. This is clearly an anomaly in the pricing system. One possible explanation is that betting odds on the day diverged across bookies, and starting prices were compiled from the cheapest prices available. This anomaly is instructive in its illustration of the degree of price competition in the betting market. Note also that the distribution is heavily skewed to the left, with the bulk of the observations falling in the range of 1.0 to 1.3. This makes the explanation of the outliers particularly important.

When the sample is grouped in terms of the number of runners, it gives rise to the following frequency distribution. As with the distribution of sum of

![Bar Chart](image.png)

Fig. 2.

prices, this is also skewed to the left, with a sizeable tail. Equation (10) suggests that there is a positive correlation between the numbers of runners and the sum of prices, and gives rise to the suspicion that the tails of both distributions consist of the same observations. We shall see that this suspicion is well-founded.

The first practical problem we encounter in implementing the iterative estimation procedure is how to truncate the polynomials in (10). We shall take

© Royal Economic Society 1993
a pragmatic line on this question, and obtain estimates of \( z \) for a variety of specifications and compare the results. If the estimate of \( z \) is stable to changes in the specification of the estimated equation, we would have strong arguments for selecting a parsimonious specification, using a polynomial with a low degree.

In fact, as we shall see below, the estimate of \( z \) is virtually unchanged as we vary the degree of the polynomials in (10). For this reason, the quadratic case is an important benchmark. The results of the iterative estimation procedure for the quadratic case is presented in table 1.

**Table 1**

**OLS regression of the equation:**

\[
D = z(n-1) + \sum \hat{a}_s n^s \text{Var}(p) + \sum \hat{b}_s n^s [\text{Var}(p)]^2
\]

after \( N \) adjustments of the regressors according to equation (11).

<table>
<thead>
<tr>
<th>( n-1 )</th>
<th>( N = 0 )</th>
<th>( N = 1 )</th>
<th>( N = 2 )</th>
<th>( N = 3 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \hat{a}_3 )</td>
<td>0.0246**</td>
<td>0.02079**</td>
<td>0.02668**</td>
<td>0.02668**</td>
</tr>
<tr>
<td>(20.66)</td>
<td>(17.18)</td>
<td>(15.81)</td>
<td>(15.81)</td>
<td></td>
</tr>
<tr>
<td>( \text{Var}(p) )</td>
<td>2.043</td>
<td>5.903</td>
<td>5.656</td>
<td>5.648</td>
</tr>
<tr>
<td>(0.384)</td>
<td>(1.864)</td>
<td>(1.827)</td>
<td>(1.826)</td>
<td></td>
</tr>
<tr>
<td>( \hat{a}_2 )</td>
<td>-0.922</td>
<td>-1.909*</td>
<td>-1.719*</td>
<td>-1.716*</td>
</tr>
<tr>
<td>(0.093)</td>
<td>(-2.109)</td>
<td>(-2.078)</td>
<td>(-2.078)</td>
<td></td>
</tr>
<tr>
<td>( \hat{a}_1 )</td>
<td>0.074</td>
<td>0.131*</td>
<td>0.127*</td>
<td>0.127*</td>
</tr>
<tr>
<td>(-0.097)</td>
<td>(2.111)</td>
<td>(2.085)</td>
<td>(2.084)</td>
<td></td>
</tr>
<tr>
<td>( \text{Var}(p) )</td>
<td>22.52</td>
<td>-59.26</td>
<td>-57.70</td>
<td>-57.65</td>
</tr>
<tr>
<td>(0.514)</td>
<td>(-1.763)</td>
<td>(-1.735)</td>
<td>(-1.734)</td>
<td></td>
</tr>
<tr>
<td>( \hat{a}_2 )</td>
<td>-0.951</td>
<td>-2.647</td>
<td>-2.320</td>
<td>-2.217</td>
</tr>
<tr>
<td>(0.993)</td>
<td>(1.499)</td>
<td>(1.389)</td>
<td>(1.388)</td>
<td></td>
</tr>
<tr>
<td>( \hat{a}_1 )</td>
<td>-3.730</td>
<td>-2.21</td>
<td>-2.22</td>
<td>-2.22</td>
</tr>
<tr>
<td>(1.321)</td>
<td>(-1.09)</td>
<td>(-1.09)</td>
<td>(-1.09)</td>
<td></td>
</tr>
<tr>
<td>( R^2 = 0.71 )</td>
<td>( R^2 = 0.70 )</td>
<td>( R^2 = 0.70 )</td>
<td>( R^2 = 0.70 )</td>
<td></td>
</tr>
</tbody>
</table>

* indicates significance at 5% level.
** indicates significance at 1% level.

The initial estimate of \( z \) is from the ordinary least squares regression of (10) in which \( \text{Var}(\hat{t}) \) is used as a proxy for \( \text{Var}(p) \). It appears in the first column of table 1. This regression yields an initial estimate of \( z \) of around 2.5%. It is worthy of note that our estimate of \( z \) is significant at the 1% level throughout, lending support to the central prediction of our theoretical model that the sum of prices is positively correlated with the number of runners. Successive regressions of equation (10) following the adjustments of regressors appear in the successive columns of table 1. Our revised estimates of \( z \) converge to five decimal places after three iterations to a value of 2.9%.

One noticeable feature of table 1 is that the "filtering" variables — i.e., variables which involve \( \text{Var}(p) \) — gain in significance after the first adjustment. By comparing the first and second columns, we see that the t scores of these
variables increase markedly, with two of them becoming significant at the 5% level. This appears to be strong empirical vindication of the adjustment procedure adopted in the paper.

We shall now turn to the robustness of our estimate of $z$ as we change the specification of the estimated equation. A series of estimations was carried out in which the degree of the polynomial approximation in (10) was raised at each step. For each degree of the polynomial and for each round in the iteration, the estimate of $z$ was noted. These estimates of $z$ are presented in table 2. The

<table>
<thead>
<tr>
<th></th>
<th>$N = 0$</th>
<th>$N = 1$</th>
<th>$N = 2$</th>
<th>$N = 3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quadratic</td>
<td>0.02464</td>
<td>0.02079</td>
<td>0.02068</td>
<td>0.02068</td>
</tr>
<tr>
<td>Cubic</td>
<td>0.02513</td>
<td>0.02033</td>
<td>0.02029</td>
<td>0.02029</td>
</tr>
<tr>
<td>Quartic</td>
<td>0.02517</td>
<td>0.02013</td>
<td>0.02040</td>
<td>0.02039</td>
</tr>
<tr>
<td>Quintic</td>
<td>0.02497</td>
<td>0.02043</td>
<td>0.02069</td>
<td>0.02068</td>
</tr>
</tbody>
</table>

Estimates of $z$ given: (i) degree of polynomial approximation (rows), and (ii) number of iterations ($N$) in the adjustment of regressors (columns).

columns indicate the number of iterations and the rows indicate the degree of the polynomial in the estimated equation. The last column of this table is particularly noteworthy, since it contains the final estimates of $z$ for each specification. It is striking that these estimates are virtually identical. Indeed, the bottom right hand estimate of $z$ (for the quintic case) is identical up to 5 decimal places to the estimate in the quadratic case. In the light of this evidence, the quadratic specification examined above takes on particular significance as a benchmark case, and it would appear that we have good reason to place credence in the estimate of $z$ that it provides.

We conclude this section with a discussion of alternative estimation procedures. Polynomial approximations such as that used here have become less popular among practitioners as more powerful non-linear estimation procedures have become available. For our purposes, however, the linear specification has a number of advantages. Foremost among these is the fact that it delivers an additive decomposition of the market spread. By apportioning $D$ between the ‘direct’ effect due to insider trading and the ‘indirect effect’ due to the evenness of match, we are able to compare the effect of $z$ across two races with different numbers of runners. This apportionment ties in well with the central feature of the theoretical model – namely, that the sum of prices increases with the number of runners as the bookie tries to recoup greater losses to the insider by raising the prices faced by outsiders. Our linear estimate of $z$ provides a measure of the steepness of this relationship.

Nevertheless, if the estimate of $z$ had not been robust to alternative truncations of the polynomial, we would have had little choice but to adopt a direct, nonlinear estimation procedure. In the event, as shown in table 2, our estimate of $z$ turns out to be insensitive to the exact specification of the

© Royal Economic Society 1993
polynomial. In general, however, there is no guarantee that the procedure used in this paper will deliver satisfactory results, and future research should investigate nonlinear estimation methods.

V. CONCLUDING REMARKS

The conclusion of our empirical investigation is that the incidence of insider trading in the week of the 1st of July 1991 was around 2%. The key empirical feature at the heart of this paper has been the strong positive correlation between the sum of prices and the number of runners. Any alternative account of the betting market would have to provide a convincing alternative explanation of this feature.

University College, Oxford

Date of receipt of final typescript: January 1993

REFERENCES


© Royal Economic Society 1993