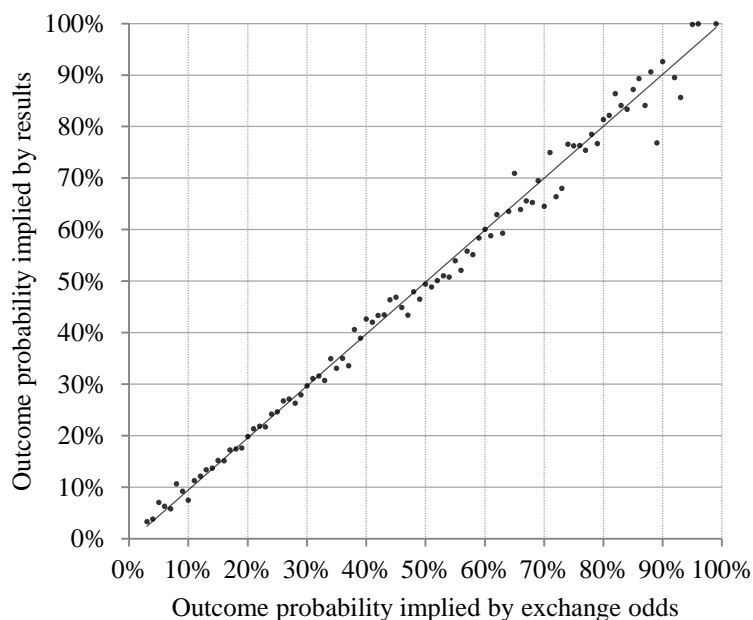


THE WISDOM OF THE CROWD

At my local fair recently there was a competition to guess the number of flower petals stuffed into a box. Rather than bother to take a close look, I simply asked the stall holder for the list of previous guess. So far there had been 40 guesses from which I calculated the average: 245. I submitted that figure as my guess. It turned out the correct answer was 295, meaning I was out by 17%, not too bad. Additionally, following my attempt there were a further 13 guesses, a total of 54, with an average of 272, just 8% lower than the correct answer. Seemingly, 54 completely unconnected people acting independently of one another collectively produced an answer that was a pretty close estimate of the true one. When you consider that the range of guess was from 53 to 866, perhaps you'll agree that something magical is going on here. Not only was the final average pretty close to the actual number, it was also more accurate than the vast majority of individual guesses. That magic is called the 'Wisdom of the Crowd.'

The wisdom of the crowd phenomenon was first observed in the early 20th century by the eminent anthropologist, Sir Francis Galton. At a 1906 country fair in Plymouth, 787 people participated in a contest to estimate the weight of butchered ox. Galton calculated that the median guess to be 1207 pounds, a figure accurate to within 1% of the true weight of 1198 pounds and again more accurate than the majority of individual estimates. The wisdom of the crowd is a very real and repeatedly observable phenomenon in life, not least in the world of betting markets that are dominated by player psychology. Remove the influence of the bookmaker's margin or betting exchange's commission and we find that betting markets actually do a phenomenal job of replicating the 'true' probabilities of outcomes, despite not knowing *a priori* what the results of sporting events will be. This is perhaps best observed at betting exchanges where the favourite-longshot bias is eliminated. The chart below, based on 52,411 Betfair odds from worldwide football league matches during the period 29th October 2004 to 31st October 2005, compares the probabilities implied *a priori* by volume weighted average betting prices with the probabilities implied *a posteriori* by the actual results. There is an almost perfect correlation ($r = 0.995$).

Comparison of outcome probabilities implied by Betfair odds versus actual results



What is it about a betting crowd that makes their collective opinion so accurate, where individually so many can be inaccurate? Provided individual errors are not systematic and in the same direction they will tend to cancel each other out. Each individual guess has two components: signal (information) and noise (error). Remove the noise and what's left behind is the signal, that is to say the collective wisdom. Four conditions are necessary for a

crowd to be wise in this manner: diversity, independence, decentralisation and aggregation. Arguably they are all present in betting markets..

Having a diverse set of opinions is a prerequisite for collective wisdom. Where everyone is thinking or doing exactly the same thing, the probability of systematic error or bias increases. Diversity is the basis for any competitive market; let different ideas or products compete against one another and that's usually a recipe for the best ones succeeding. Google inherently understand the significance of diversity; it invests a lot of time, money and effort into lots of little start-up ideas like Google Earth, Google Glass and a driverless car. Not all of them will succeed but by having lots of eggs in your basket, you increase the chances that at least some of them hatch. In prediction markets like betting, diversity is virtually a given because of the environment of uncertainty and the typically large number of people acting in them with difference opinions, risk preferences and approaches to forecasting. When attempting to forecast the outcome of a game, for example, there are potentially limitless ways to skin that cat. Some prediction methods try to determine the intrinsic probability of outcome (for example value betting). Others adopt a more psychological approach, believing a market to be more a reflection of opinions (and more importantly opinions about opinions) with all their biases, making use of methods such a technical analysis to study trends and directions in betting prices. Then there are different types of prediction models: linear or nonlinear, static or dynamic, deterministic, probabilistic, or dynamic. Most will be wrong but the pooling of diverse ideas encourages collective accuracy.

Diversity arises out of independence of thought. Arguably, this is the most important ingredient for crowd wisdom. If everyone thinks the same way and does the same thing we frequently end up with poor outcomes. Everyone was betting on the 2015 UK General Election to return a hung parliament, because that's what all the polls and pundits were saying was going to happen. Evidently they were all doing the same thing and failing to take account of a couple of very important influences: the shy Tory effect and the lazy Labour voter. (Of course, it's easy to say this with hindsight.) Perhaps if more polls and more pundits had demonstrated a greater independence of thought using what economists call 'private information', such ideas (and others) could have been used to collectively improve their predictions. Of course this isn't always the case. Teaching people to serve a tennis ball or to do differential calculus requires a narrow field of learning through repetition. But such activities are sufficiently predictable with clear relationships between cause and effect. In prediction markets under uncertainty, by contrast, learning through pattern recognition is limited because the patterns are largely random. What signal exists is deafened by noise, with good and bad luck dominating outcomes. In such environments having people acting independently helps to eliminate that noise, because it offers the best chance for keeping people's errors from becoming correlated. When mistakes are random they will cancel out.

The final two pieces of the jigsaw that make the wisdom of the crowd such a powerful mechanism are decentralisation and aggregation. A system is said to be decentralised if it's not acting under the influence of a top-down central authority. By definition, independence and diversity of thought and decision making will be encouraged where central regulation is not restricting outputs. Decentralisation ensures that a crowd of self-interested, independent people working without top-down interference will collectively find a better solution than anything else you could come up with. The process happens as if by magic. It's the mechanism behind bird flocking, fish shoaling and insect swarming, the emergence of complex and seemingly coordinated behaviour out of a few simple rules followed by the self-interested individuals. In the case of birds there are just four: stay close to the middle; keep sufficient distance between neighbours; avoid collisions; and flee predatory attack. For human interactions, the 18th century economist Adam Smith labelled this magic the '*invisible hand*,' describing the unintended social benefits resulting from individual actions.

Decentralisation, however, will only be of benefit if there exists a way of coordinating or aggregating together all the information. In a betting market that aggregation process is explicit: the conversion of private information and expression of opinions into a piece of public property – the price. The odds for a football team publically aggregate all the private information that exists. They represent the current balance of opinions about the likelihood of a team winning as expressed by the amounts of money wagered for and against it.

The magic lies in the emergence of wisdom without individuals having a complete understanding of what the market is doing and without anyone knowing what the 'true' answer, if there is one, will be. People with only partial knowledge and limited calculating abilities actually arrive collectively at the right answer. A wonderful demonstration of this magic was accomplished by Nobel Prize-winning economist Vernon Lomax Smith. In 1956 he set out to determine whether people with limited information would confirm to the hypothesis of market clearing, where prices of traded assets adjust up or down such that quantity supplied at the market-clearing price equals the quantity demanded at the market-clearing price. Such a price is also called the equilibrium price. Giving his 22 students cards with a dollar price tag, he made half of them buyers and half of them sellers. The sellers were instructed not to sell at less than this price, whilst the buyers were instructed not to buy at more than their card value. A difference achieved between card value and actual contract price could be regarded as profit for the player. Strict anonymity was applied such that no one knew the value of anyone else's card. The students were then asked to start trading, calling out bids and offers which may, or may not, be accepted. Sellers and buyers were free to accept a bid or offer. If they were refused, further price compromise or bartering would be required until they were accepted. The successful trades were recorded publically on the classroom blackboard. Economic theory was matched by reality. Traded prices quickly converged on one price, the equilibrium price or what we might also call the expectation price, despite players being completely unaware of their competitors' demands and despite none of them preferring this outcome (self-interested traders after all want more profit). Collectively the convergence on the market-clearing price yielded the best possible outcome, even if some of the players had been blessed with additional knowledge telling them how they should trade. The brilliance of this experiment was that it demonstrated that for markets under uncertainty, imperfect people could collectively produce near-perfect outcomes. What allowed it to happen was a decentralised independence of action and the aggregation of privately anonymous information via the publication of a price.

Essentially, price convergence is exactly what happens at a betting exchange like Betfair. This invisible hand is a kind of Bayesian process in which prices are continually and dynamically updated to reflect changes in supply and demand. At a betting exchange odds move simply in response to supply and demand. The market maker sits completely outside the contest, skimming his commission percentage from the action. This process is otherwise known as 'price discovery', a mechanism for determining the price of an asset in the marketplace through the interactions of buyers and sellers, or in this case backers and layers. Remarkable as it may seem, the betting public collectively 'knows' the 'true' probability of outcome of a sporting event through their betting actions. Odds shorten on the fancied competitors and lengthen on the least fancied, settling at values that reflect all the private information that has been consumed by the players. This price discovery is dynamic with the equilibrium price never completely stationary because there will always be new information arriving randomly on to the market.

For a bookmaker, things are a little different but only because they are part of the action; the fundamental process remains the same. Odds shorten because too much money has been bet on one outcome, giving the bookmaker a large liability in the event that it happens. Bookmakers are always looking to reduce their liability; in this case they can achieve this by shortening the odds to discourage further interest from customers. At the same time they lengthen the odds on the opposition to attract money. Through this Bayesian price clearing process they attempt to balance their book. If they get it right they won't care which team or player wins, and in effect they become more like an exchange. When betting at such a bookmaker, the punter should understand that he is not really betting against the bookmaker but against his customers who have taken a different opinion to his.

Some traditional bookmakers nevertheless still prefer to take some sort of position on an event, and they do this by offering attractive prices that possess positive value expectation relative to the collective market and by refusing to drop those prices when others around them are doing so. Frequently they are then exposed to some risk on the side of the book that has attracted a disproportionate level of action. For them there are other methods of managing liability. One option is to restrict customers' betting activity. Another is to lay off the risk at a betting exchange or another bookmaker with a smaller margin. One bookmaker that rarely takes positions on games is Pinnacle Sports, who instead rely on professional odds management algorithms allowing the market to make its

own mind up. With its small margins and laissez-faire exchange-model approach unmistakably Pinnacle Sports has become synonymous with high-volume action. Of course, there is one significant consequence of Pinnacle's market being wiser than all the others: it makes it much harder to beat.

We can, however, use Pinnacle's market wisdom to estimate what the 'true' chances of a result might be. To do this we simply need to remove the influence of the margin or overround that Pinnacle applies to its odds. This involves a two-step process: firstly calculate the overall margin of a book; secondly determine the relative margin weights applied to each outcome in the book. It is now well-established that bookmakers apply differential shortening to their odds, with more shortening taking place for longer odds. This is known as the favourite-longshot bias and arises because punters have a tendency to overbet long prices relative to short ones. Calculating the margin for a book is easy, and is performed as follows for a home-draw-away football betting market:

$$M = (1/H + 1/D + 1/A) - 1$$

where M is the margin expressed as a decimal, H is the betting price for the home win, D is the draw price and A is the away price. For example for a home-draw-away book with prices 1.44, 4.42 and 6.25 respectively, $M = 0.08$ (or 8% expressed as a percentage). For such a book, the overround would be said to be 108%.

Determining how a bookmaker weights this margin differentially across the home, draw and away prices is a little trickier. This is not something any bookmaker will reveal publically so we are forced to guess at how they might do it. One method might be to apply margin weights in proportion to the size of the odds. Hence, for a home-draw-away betting market with 3 possible outcomes:

$$M_H = \frac{MH_f}{3}$$

$$M_D = \frac{MD_f}{3}$$

$$M_A = \frac{MA_f}{3}$$

where H_f , D_f and A_f are the fair home, draw and away odds respectively.

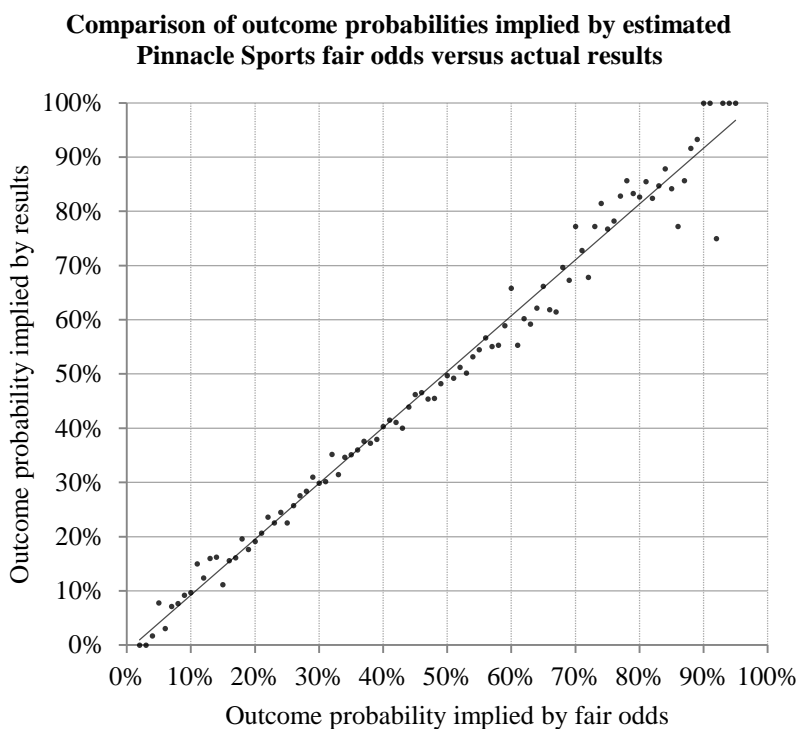
For example, a home-draw-away book with fair odds of 1.5, 5 and 7.5 and where the book margin to apply was 8%, the differential margins for home, draw and away would be 0.04, 0.133 and 0.200 respectively. To calculate the actual prices one then simply divides the fair price by the margin weight plus 1. For the home odds, for example, this is $1.5 \div 1.04 = 1.44$. Similarly, the draw and away prices are $5 \div 1.133 = 4.42$ and $7.5 \div 1.200 = 6.25$. If a margin weight of 8% had been applied equally to home, draw and away prices, we would have 1.39, 4.63 and 6.94 respectively. You can see from this exercise that a differential weighting of odds in this manner shortens longshots more significantly than favourites.

With a little bit of algebraic rearranging, we can reverse the process to calculate what fair odds the bookmaker will have estimated in the first place, given his book margin and applying this model of differential margin weighting. Hence, for any published odds, O, the fair odds from which they came in a 3-outcome market will be given by:

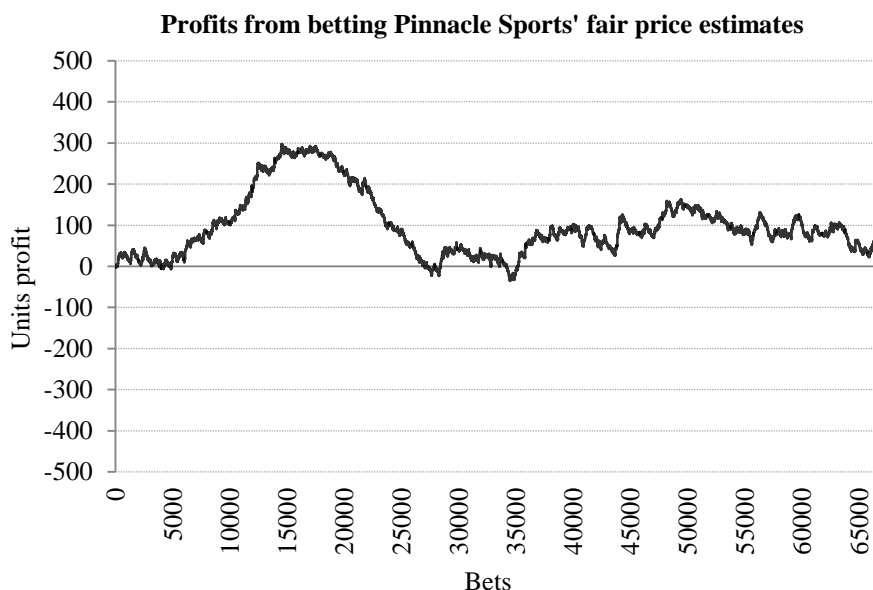
$$O_f = \frac{3O}{3 - MO}$$

So in our example above, a published price of 6.25 will have fair odds of $(3 \times 6.25) \div (3 - (0.08 \times 6.25)) = 7.5$. Whilst the basis for this simple odds model is just conjecture and probably an oversimplification, it does appear to closely reflect the betting prices for many of the major brands. We can also test how accurate the model's

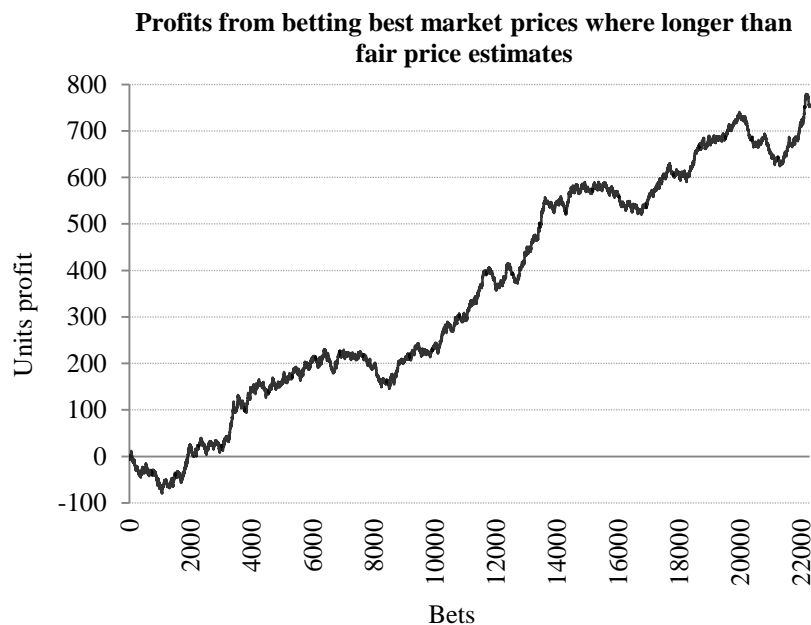
estimates of fair odds would have been retrospectively, and by implication how wise the Pinnacle Sports betting market really is. For this I have used 3 full seasons (2012/13 to 2014/15) of European football league home-draw-away Pinnacle Sports betting odds (a total of 22,318 games).



As for the Betfair exchange data I showed earlier, my model estimates for Pinnacle Sports' fair prices do a pretty good job of predicting actual outcome frequencies. Another way to test how wise these fair prices have been is to see whether they would have broken even if all of them had been bet blindly. Fair prices, by definition, should break even over the long term, allowing for shorter term periods of good and bad luck to even out. The next chart show the evolution of profits from level stakes betting all matches to these fair prices. In fact the closing yield from these 22,318 matches (66,954 home, draw and away bets in total) is 0.08%, as close to break even as we might reasonably expect.

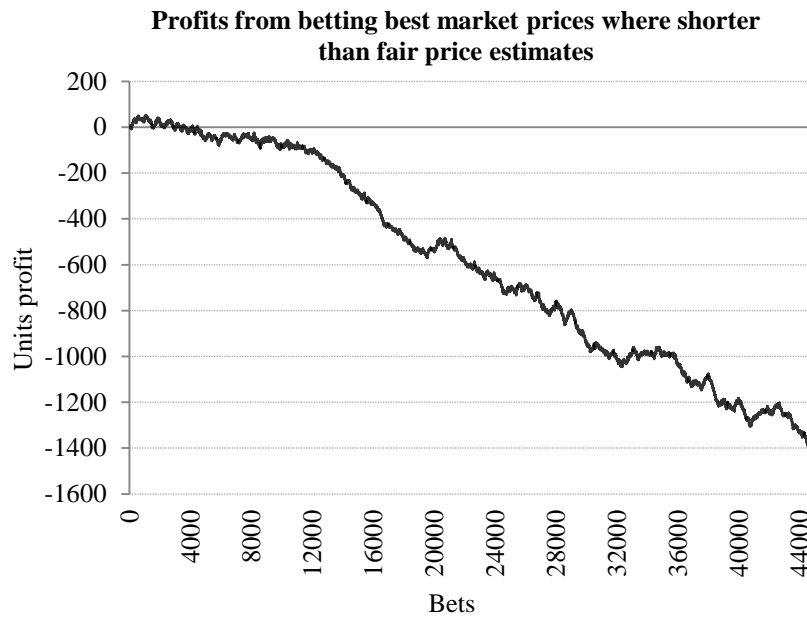


An obvious question now arises: if Pinnacle Sports' football home-draw-away betting market is so accurate, can we use that wisdom to identify mistakes elsewhere with a view to potentially making a profit? The answer, it would appear, is yes. Alongside the betting odds for Pinnacle Sports, I have also recorded the best market prices (as published by the odds comparison Betbrain.com). Betting every home, draw or away price (22,281 in total out of the possible 66,954) where the best market price was longer than the fair Pinnacle price (as estimated by my model) gave a yield of 3.4% and the following profit trend.



In fact this was a little better than one might expect given the prices that were bet. The average advantage over the modelled fair prices was 2.2% from which *a priori* one would expect to see a similarly sized yield. Those best market prices came from theoretical best books with an average overround of 100.36%. In other words, if we had blindly bet all possible outcomes with appropriate staking we would have lost about 0.4% on turnover. Evidently, our fair prices did a good job of finding which price amongst the home-draw-away market was the value one, something not possible with traditional arbitrage betting. This means we can find many more betting opportunities than arbitrage hunting will be able to achieve, since we don't always need an underround book to have a value price. Indeed, over two-thirds of these value opportunities were found in books that were still overround at best prices.

How would we have performed had we decided to back best market prices that were shorter than the model-estimated fair prices? This time our yield from the remaining 44,673 bets would have been -3.08% (with an average disadvantage against fair odds of 2.37%). The profit evolution is again shown below.



Of course, we could choose to be more selective with our betting criteria. We might for example just decide to bet when our advantage over the fair odds is greater than 1%, 2%, 3% or higher. Naturally, this will reduce the number of betting opportunities available, but in theory it should increase the yield we will achieve. Would this have happened for this 3-season sample? Yes; the table below shows how.

Advantage over fair odds greater than	Bets	Yield	Average advantage over fair odds
0%	22281	3.40%	2.23%
1%	14837	4.60%	3.07%
2%	9196	6.63%	4.04%
3%	5474	8.83%	5.11%
4%	3243	12.70%	6.26%
5%	1927	13.22%	7.51%

Evidently, for each sample performance was better than would be predicted from the theoretical advantage gained over the fair odds. Presumably this is simply a result of good fortune. The right hand column figures are probably more representative of what we should expect to achieve by way of yields. Nevertheless, it would appear that the wisdom of Pinnacle Sports' home-draw-away betting market, coupled with this rudimentary model at estimating fair prices from it, can provide profitable betting opportunities at bookmakers more prone to offering mistakes.

Naturally, there are a couple of caveats with this approach. Firstly, given the relatively small yields involved, one should reasonably expect to suffer fairly long periods of treading water, or worse still, losing, lasting hundreds and perhaps thousands of bets. Secondly, it is to be expected that the sort of bookmaker that will offer betting prices in excess of Pinnacle Sports' fair price estimates will also be the sort of bookmaker who won't like a customer consistently exploiting such generosity. This is usually offered to attract new customers or to advertise the impression that the brand offers good value. If customers repeatedly take advantage of those superior prices they can often expect to have their betting activity curtailed. Advising how a punter can avoid detection in this respect is beyond the scope of this article. However, it has at least identified that a 'wisdom-of-the-crowd' approach can identify where bookmakers have made mistakes, and that technically at least it should be possible to exploit them.