

Pundits: The confidence trick

Better confident than right?

Media pundits are the supreme example of self-belief and confidence in their own opinions. Through TV, newspapers and blogs they tell us with sublime certainty what will happen. But are they right? And does it matter if they are wrong? **Ben Smith** and **Jadrian Wooten** ask what we demand from pundits – accuracy or confidence?

We live in a world full of pundits, professionals hired to make predictions in various media outlets. Some pundits make predictions about finance, some about politics and some about sport. Often it appears they are not particularly accurate. In 2008, in the early stages of the financial crisis the well-known and popular Jim Cramer famously endorsed Bear Stearns on his television punditry show, *Mad Money*:

Bear Stearns is fine ... Bear Stearns is not in trouble. Don't be silly ... don't move your money.

One week later, Bear Sterns collapsed¹.

Cramer later attempted to justify his remarks by suggesting that being too negative would have resulted in a mass panic. In fact, a popular novice investment strategy is to invest in Cramer-endorsed stocks to see small overnight gains, and then short sell the stocks to capitalise on the surge in prices². It has even been reported that betting against Cramer's "buy" recommendations could yield 25% in the initial month³. Sites like PunditTracker suggest that the most famous pundits are also among some of the worst-performing pundits.

Indeed, whenever pundits do turn out to be extremely accurate, the public often considers them outliers. Nate Silver's accurate prediction of the 2012 presidential election is a case in point. He had done it before: in the 2008 presidential election, Silver by intelligent use of

statistics was able to correctly predict the winner for 49 of the 50 states, as well as the winner for all 35 senatorial races. While often considered mild-mannered and low-key compared to most pundits, even Nate Silver has the occasional moment of confident bravado. After being criticised – before the election – by an MSNBC pundit about his prediction that Barack Obama had a 75% chance of winning re-election, Silver responded with thousand-dollar confidence and a public challenge⁴:



Obama to win – Pundits' toss-up or \$1,000 certainty? © iStockphoto.com/EdStock

If you think it's a toss-up, let's bet. If Obama wins, you donate \$1,000 to the American Red Cross. If Romney wins, I do. Deal?

Ironically, Silver's confident wager was published publicly on his Twitter account (@fivethirtyeight) – the punditry medium that we analyse here.

Silver's accuracy was what made him unusual. If pundits are not held to high stands for accuracy, what is it that makes them so popular with the public? This was the original question we set out to answer. We were inspired by a basic idea from psychology: people do not like uncertainty. Uncertain situations make people uncomfortable, and people are willing to give something up to avoid uncertainty. If this holds true for media pundits, the public would allow some inaccuracy for an increased sense of certainty. This forms our hypothesis: when tested statistically, we should see increased popularity

from confident pundits, when controlling for all other factors – accuracy included.

The process

Sadly, financial stocks do not have a terminal date. Therefore it is hard to prove pundits wrong in their predictions. They can always claim we simply did not wait long enough for the stock to go up or down as they predicted. While Jim Cramer may have inspired our investigation, we turned to the world of sports pundits to test our hypothesis. When a pundit makes a prediction about a game, they are either right or wrong and we know exactly when we can record an outcome either way. This makes sport the ideal subject matter for testing the popularity of pundits.

Once the subject matter was resolved, we focused on an ideal data source. We could have watched a whole lot of television and read a lot of newspapers to collect every prediction we



Pope resigns! No pundit predicts it! Photo: Kancelaria Prezydenta RP

How accurate are they?

PunditTracker (www.pundittracker.com) is a recently launched website that aims "to bring accountability to the punditry industry". As they put it, pundits are incentivised to make confident and bold predictions; "we keep track of those predictions to keep them honest".

They point out that there is self-reporting bias in punditry: the pundits proudly report their hits (correct predictions) far more frequently than their misses (the things they got wrong). There is also a "positive reporting bias": just as in science very few published papers show negative results, so in punditry, predictions of negatives – things that will not happen – are rapidly forgotten, whether the things happen or not; predictions of things that do come to pass are remembered.

We are also psychologically wired to remember hits more than misses. Unusual information has an outsized grip on our memory. Bold calls are typically incorrect, so we quickly forget those – but bold calls that turn out right are unusual and therefore stick in our mind. And because we tend to confuse ease of recall with frequency, we develop a warped sense of the pundit's batting average. Hence we are we force-fed a skewed sample of prediction outcomes.

PunditTracker aims to fix what they call this "moral hazard" by playing the role of public scorekeeper. It keeps track of 120 well-known US pundits in the fields of politics, finance and sport. It grades them according to accuracy of their predictions (what proportion of them come true); but it also gives them credit for boldness (unlikely predictions that do actually come true). Likelihood is based on the odds you might get at a betting shop. Predicting successfully that the sun will rise tomorrow would gain little credit. If any pundit had predicted ahead of time that Pope Benedict was going to resign, he would have gained huge credit. (As far as we know, not one pundit in the world did predict it.)

Pundits are scored on a scale of A (best and boldest) to F (least accurate and least bold). In politics, among household names, George Will, outspoken syndicated columnist for the *Washington Post* and contributing analyst with ABC News, is graded F, the lowest for accuracy and boldness. Eleanor Clift, television pundit, contributing editor for *Newsweek* and blogger on the *Daily Beast* website, shares that dishonour.

In finance, Jim Cramer is also an F; but Nouriel Roubini, who blogs at www.economonitor.com/nouriel/, and Barron's Roundtable, an annual round-up from the financial newspaper *Barron's*, both scored A+, the best. Barron's Roundtable has picked successfully performing stocks. Roubini anticipated both the collapse of the US housing market and the world-wide recession.

see. However enjoyable such research would have been, it poses three problems:

1. Sites like PunditTracker already record pundit predictions on a regular basis – see the box.
2. Regardless of how much television we watch, we will miss a large number of predictions while we sleep and go to work, which means our data set will be relatively small.
3. This will only examine predictions made by *professional* pundits from a limited number of media outlets.

Specifically looking at the last issue, a reasonable person might argue that, in fact, the public does not want confident pundits. Perhaps, the news/entertainment industry simply *thinks* we want confidence and therefore gives us only the confident pundits, denying us the choice. What we need is a data source where both professionals and amateurs alike make predictions on games: this is one of the reasons we used Twitter.

Twitter has a number of interesting properties that make it ideal for this type of analysis. First, the cost of providing content to one more subscriber for the tweet producer is zero. Pundits wish to serve as many people as possible, but the number following each pundit is entirely up to the individual followers. This allows us to observe the preferences

Table 1. A sample of Twitter verified accounts during the 2012 Major League Baseball play-offs. These punditry programmes and individuals claim expertise, but accuracies of 15%, 28% and so on are not impressive

Account	Predictions	Accuracy	Confidence
ESPN	27	41%	52%
SportsCenter	32	31%	44%
Baseball Tonight	39	28%	44%
Mike Greenberg	13	15%	53%
Lou Holtz	9	33%	11%
Chris Rose	11	36%	36%

of the people following the pundit and not the preferences of the pundits themselves.

Twitter is also fairly representative of the general population⁵. While the network is more urban and younger than the US population, it matches educational attainment and income rather well. Because Twitter is a real market, and not an artificial environment, it avoids a set of observer bias effects common to surveys and experiments.

Further, both professionals and amateurs alike can make predictions, and Twitter conveniently separates those groups for all of their users. Because famous people commonly have the problem that they might be impersonated on the web, Twitter has created the “verified account”. Twitter checks to make sure that the claimed user is, in fact, the user. Anyone who is famous because he or she is on TV, on the radio or writes for a newspaper will have this special flag.

Additionally, users have a biography section where they have the opportunity to describe themselves. We can examine this information for claims of sports expertise. There are many ways to claim expertise; we use a technique called “regular expression” to capture as many of these as possible. Regular expression is discussed in more detail a little later. In the final analysis, anyone claiming to be a sports expert but is not verified we classified as an *amateur* pundit; anyone claiming to be a sports expert with a verified account we classified as a *professional*.

The next step was to collect the data from Twitter. Twitter allows users to programmatically “watch” a set of words. After you register a set of words, a tweet is sent in real time to your computer if the words occur in the tweet. Starting one week before the 2012 baseball play-offs, we started collecting every tweet containing any of the team names, nicknames, or city names. This resulted in over a billion items, most of which were not predictions, but

that was OK because we could filter those out, again using “regular expressions”.

If you think about sentence structure, only a few words in each sentence matter in terms of determining the meaning of the entire sentence. We only need to be able to identify these key elements of the sentence. One method of doing this is to build a large table of regular expressions; it is a technique where a number

It may be that media outlets give us overconfident pundits because they believe that that is what the public wants

of phrases can be generalized. A simplified example of a regular expression would be:

```
\b(Bears)(?:(!\b((not)|(won['']t))\b)).)*\b((destroy)|(annihilate))\b.+ \b(Dogs)\b
```

This would match any phrase that says that the Bears will destroy or annihilate the Dogs. But, unlike a normal search, this structure allows for variations on that theme and will still be picked up by the regular expression (e.g. “the Bears will totally destroy the Dogs”); however, it specifically disallows variations with the opposite meaning (“the Bears will not destroy the Dogs”). Making a large table of these expressions results in us picking up all the predictions. (To the best of our knowledge, at least: we continued to add expressions until our list matched all the forms of predictions we were seeing on Twitter.)

We determined the strength of the each of these predictions by using the work of Chklovski and Pantel⁶, who have ranked the strength of words. Thus the verb “startle” is stronger than “surprise”, and “shock” is stronger

than “startle”. This allowed us to mark a prediction like “the Bears will destroy the Dogs” as confident, while marking a prediction like “the Bears will beat the Dogs” as not confident. “Confident” here of course means “sounding confident” or “giving a confident impression.” We have no way of telling the strength of the pundits’ own inner convictions.

Using regular expression, along with the strength of phrases, we thinned our billion plus mass of collected tweets down to 1.6 million predictions, with information about followers, confidence, accuracy, verified status and many other aspects of the Twitter account. At this point the problem becomes a simple regression problem.

Using a technique called Box–Cox parameter transformation estimation, we found that a log model best fits the data. We then use ordinary least squares and iterated generalized methods of moments to estimate the importance of accuracy and of confidence while controlling for other observable factors, such as the age of the account, engagement in the Twitter community and the number of tweets per year.

The findings

The public appears to heavily value confidence and places a much smaller, although still positive, emphasis on accuracy. Among professional pundits, perfect accuracy – predicting every baseball play-off game correctly – would only result in a 3½% increase in popularity. Being consistently confident, by contrast, would result in an almost 17% increase in popularity. Both results are statistically significant. For the pundit who wants an audience, sounding confident is overwhelmingly more important.

Ideally, we would all appreciate confident pundits with perfect accuracy, but realistically it is difficult to be perfectly accurate. Pundits can, however, control how confident their predictions sound, which results in many pundits making strong predictions regardless of the probability of their statement being accurate. To judge by the resulting popularity, they are right (in career terms at least) to do so. By focusing on confidence, pundits on Twitter achieve higher follower counts, which is a rough gauge of consumer demand.

There is a parallel in the world of casinos. Imagine two roulette tables side by side with two equally competent gamblers placing bets on a red outcome. While each has approximately a 50% chance of being correct (if we



The Bears will annihilate the opposition... or will destroy them... or might win... or could perhaps manage it if they are lucky... © iStockphoto.com/Matt_Brown

ignore the zeros on the wheel), people will gravitate to watch (and to follow with their own bets) the more animated player. Gamblers trying to confidently predict the ball landing in red will usually gather a crowd, while the quieter gambler will play by himself.

Those were the results for the professionals. Amateur pundits experience a similar story. They see an increase in followers of 7% for perfect accuracy, compared to almost 20% for being consistently confident. The ability of amateur pundits, especially for those operating as aspiring bloggers, to gain followers is pivotal to their success. So here too there is a stronger incentive to appear confident than to be right.

This result is particularly important when compared to the professional pundits. Because the two results are not particularly dissimilar, we can reject the idea that it is the media that drives the confidence in punditry. Amateurs on Twitter have no intermediary (like a network or newspaper) to satisfy, so they directly serve the public's desires.

Regardless of the punditry status, one thing is certain: the most popular pundits may not be right all the time, but they tend to make their predictions more confidently than the other pundits.

Implications

Pundits are confident because that is what the public demands. One hypothesis was that pundits are confident because the networks are asking them to operate that way, but our results have suggested otherwise. The amateur pundits have no intermediary dictating their confidence level, and yet they attract a larger following when being confident than when expressing their opinions more moderately. So while the TV networks like ESPN or CNBC have ultimate discretion over which pundits host their shows, the demand for pundits is actually derived from the viewing audience.

When you watch a pundit on television, remember that their job is to maximize eyeballs, not accuracy. Their employer is in the business of selling advertising, which means the networks will choose those pundits who provide the most advertising revenue. That might not necessarily be the most accurate pundit. It could just be the one who is most outrageously overconfident.

References

1. Huffington Post (2008) Cramer called Bear Stearns "fine" just a week ago. *Huffington Post*,

March 28th. http://www.huffingtonpost.com/2008/03/17/cramer-called-bear-stearn_n_91878.html

2. Engelberg, J., Sasseville, C. and Williams, J. (2012) Market madness? The case of *Mad Money*. *Management Science*, 58(2), 351–364.

3. Alpert, B. (2009) Cramer's star outshines his stock picks. *Barron's*, February 9th. <http://online.barrons.com/article/SB123397107399659271.html>

4. Sullivan, M. (2012). Under attack, Nate Silver picks the wrong defense. Public Editor's Journal, *New York Times*, November 1st. <http://publiceditor.blogs.nytimes.com/2012/11/01/under-attack-nate-silver-picks-the-wrong-defense/>

5. Pew Internet (2012) Trend data (adults). Pew Research Center. Retrieved from [http://pewinternet.org/Static-Pages/Trend-Data-\(Adults\)/Usage-Over-Time.aspx](http://pewinternet.org/Static-Pages/Trend-Data-(Adults)/Usage-Over-Time.aspx) in August 2012.

6. Chklovski, T. and Pantel, P. (2004) VERBOCEAN: Mining the web for fine-grained semantic verb relations. In *Proceedings of Conference on Empirical Methods in Natural Language Processing*, Vol. 4, pp. 33–40. East Stroudsburg, PA: Association for Computational Linguistics.

Jadrian Wooten and Ben Smith are PhD candidates at the School of Economic Sciences at Washington State University.