“I’ll trade you some accuracy for that confidence”

The story of pundit accuracy and viewer preferences

Ben Smith (BensResearch.com)

Jadrian Wooten (JadrianWooten.com)

Pundits make predictions about everything: sports, stocks, politics, etc. The funny thing is that they’re terrible at it. Research over the last fifty years has shown that experts, but pundits in particular, perform only marginally better than chance (Meehl, 1954; Camerer, 1997; Spence, 1997) — but are really confident (use strong words) in their predictions (Krug, 2007; Swann, 1997; Tetlock, 2009). If they are so wrong, why are they on TV?

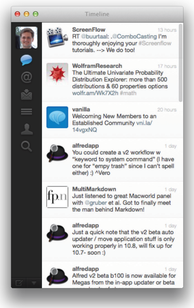
Consider the psychology of the audience — we want the confidence. When people must wait for a result, stress builds up due to the uncertainty (Osuna, 1985) — people wish to avoid this stress. In fact, a review of anxiety literature describe humans as ‘certainty maximizers’ (Bammer, 2008), especially for unimportant events.

> *We theorize that due to psychological costs of uncertainty, consumers are willing to trade accuracy for confidence when choosing pundits, even if the confidence is unwarranted.*

Using twitter data we show, that at least amongst sports pundits, this hypothesis holds true.

Twitter Subscriber Market

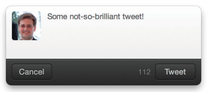
The Twitter Timeline/Stream (Demand)



- Every “tweet” from someone you follow shows up in the timeline
- Each “tweeter” has their own rate of creating tweets
- There is a limit to how many people you can follow which is governed by your attention and the tweet rate of the people you follow

Content Creators (Supply)

- As a content creator, there is zero cost to any additional subscriber
- There is obviously a cost to creating the tweet, but the cost is fixed and must be satisfied if you are tweeting
- The creation process (and distribution) is identical regardless of follower count



So…

What we are looking at is an environment where demanders choose who to follow and suppliers would prefer to have everyone follow them.

On twitter, you directly observe the demand curve

Method

During the 2012 Baseball Playoffs we collected nearly every prediction made by a pundit of member or the general pundit. How?

Collection (Twitter Streaming API)

Using twitter’s streaming API, a developer can register a set of words that they want to watch (for instance, the names and nicknames of the major league baseball teams). Then twitter sends tweets that contain these words through the API as they occur.

This results in a lot of tweets that aren’t predictions, but that’s OK because non-predictions don’t match any of the regular expressions and are thrown away.

Extraction (Regular Expression)

Using a table of regular expressions (a technique where a large number of phrases can be generalized), we extracted the key aspects of a particular tweet: the event they were predicting, their prediction and how strongly they felt about that prediction (made possible by Chklovski & Pantel’s word strength — confidence — rankings). A simplified example of a regular expression would be:

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

Which 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 (but exclude phrases that have the opposite meaning).

2012 Baseball Playoffs Results

Summary Statistics and Definitions

Professional Pundits:
Accounts where they’re both verified and the user claims sports expertise. Verified accounts are a mechanism for celebrities to claim their name on twitter. Those who have a famous identity external to twitter must verify that they are who their account claims to be.

Amateur Pundits:
Claims sports expertise (doesn’t include verified accounts)

Content Creator	Mean Accuracy	Mean Confidence
Professional Pundit	.472	.480
Amateur Pundit	.455	.313

Model and Regression Results

Given our reasonable satisfaction of the underlying assumptions as well as statistical confirmation of a semi-logarithmic relationship in our datasets, we are able to use simple OLS to estimate the demand model:

$$Log(F_i) = \alpha + \beta_1 A_i + \beta_2 C_i + \beta_3 M_i + \sum_{k=1}^{j} \gamma_k X_{ik} + \varepsilon_i$$

Where F_i is the number of followers (quantity demanded), C_i is the subjective confidence, A_i is accuracy, M_i is the number of tweets per year and X_{ik} is a set of control variables.

Controls include account age (to control for discovery), number of lists created by the content creator (a measure of engagement), and the amount of time between the prediction and the event

Content Creator	Accuracy ($ß_1$)	Confidence ($ß_2$)
Professional Pundit	.034	.169
Amateur Pundit	.073	.197

all significant at 99%, Dependant Variable: Log(F_i) (Followers)

We see similar results from the 2013 Super Bowl. Next up: NCAA Basketball and Stock predictions