METHOD

How To Be Right More Often Than The Experts

Calibration, base rates, and a benchmark that keeps its own score.

THE METHOD

01	Accuracy is calibration
02	The base rate first
03	Estimate before price
04	Update in proportion
05	Ask answerable questions
06	Score every forecast
07	Benchmark against the market

Forecasting without a score

Public forecasting is unscored. The commentators, analysts and panelists who state what will happen are not recorded against what did, and so the accuracy of the profession is unknown to its own practitioners. This is not a lapse. It is the condition under which the activity has always operated.

The consequence is that expertise in a subject and accuracy in forecasting it have never been separated, and the first is routinely mistaken for the second. Deep knowledge of a field is not the same as a well-calibrated sense of what the field will produce next, and studies that measure the two find them only loosely related.

Accuracy is measurable. It is a property of a forecaster's numbers over time, not of any single call, and the measurement is straightforward once numbers are recorded. What follows sets out what accuracy means, how it is scored, and where a benchmark exists that keeps its own score continuously and for free.

THE METHOD

01	Accuracy is calibration	05	Ask answerable questions
02	The base rate first	06	Score every forecast
03	Estimate before price	07	Benchmark against the market
04	Update in proportion		

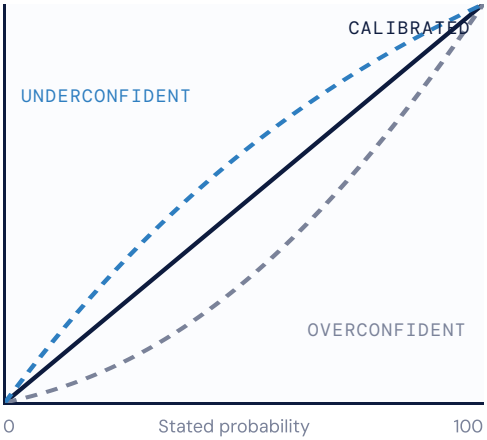
STANDING NOTE

Worth stating plainly.

The major US prediction markets are CFTC-designated contract markets, regulated under the same framework that covers futures and commodities. Readers new to the category frequently assume the opposite, and the assumption is the single most common factual error about how these venues operate.

Accuracy is calibration, not a hit rate

FIG. 01 · A CALIBRATION PLOT



Schematic, not data.

The intuitive measure of a forecaster is how often they are right. It is the wrong measure, and adopting the right one is the first step in this guide.

A forecaster who says "80 percent" and is right is not more accurate than one who says "60 percent" and is right, if the event was in fact a 60 percent proposition. The first was overconfident and happened to be lucky. The second was correct about the probability, which is the thing a forecast claims.

Calibration is the measure. A forecaster is calibrated when, across all the occasions on which they said 70 percent, the event occurred about 70 percent of the time; across all the occasions they said 90, about 90; and so on. The plot on this page shows what that looks like: stated probability on one axis, observed frequency on the other, and a perfectly calibrated forecaster on the diagonal.

Above the diagonal is underconfidence: events happened more often than stated. Below it is overconfidence, the more common failure and the more expensive one, since it produces positions sized for a certainty that did not exist.

Hit rate rewards boldness. Calibration rewards honesty about uncertainty. Only the second is a skill, and only the second can be improved by practice, because only the second produces a score that can be compared across time.

The base rate as an opening position

FIG. 02 · REFERENCE CLASS TO ADJUSTED

REFERENCE CLASS	Scheduled rate decisions where guidance was unchanged
BASE RATE	The share of those that matched guidance
ADJUSTED	Base rate, moved for what is genuinely different this time

Every forecast should begin with a question that most forecasters skip: how often does this kind of thing happen?

That frequency is the base rate, and it is the single most powerful input available, because it is the answer that is correct in the absence of any other information. A scheduled rate decision that matches the prior meeting's guidance; a sitting government retaining its majority; a measure that has passed one chamber passing the other. Each has a historical frequency, and that frequency is where the estimate should start.

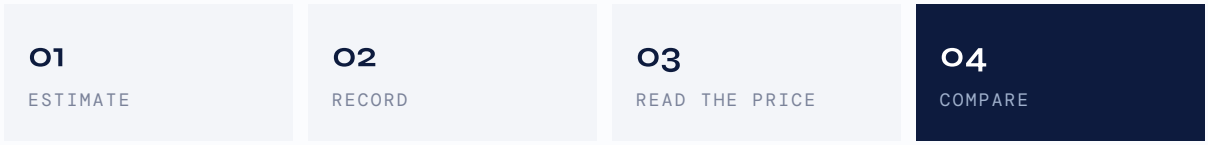
The error, made almost universally, is to start instead from the specifics of the case. The specifics are vivid, the base rate is dull, and vividness is mistaken for relevance. A forecaster who begins with what makes this case special will overweight it, because everything looks special when it is the thing being examined.

The discipline is reference class selection: choosing the set of past cases that this one most resembles, finding the frequency within that set, and only then adjusting for what is genuinely different. The adjustment is usually smaller than intuition suggests.

The base rate does not decide the forecast. It anchors it. But an anchor set from the specifics rather than the frequency will drift, and it will drift in the direction of whatever made the case seem interesting.

Estimate before price

FIG. 03 · THE ORDER IS THE METHOD



The order is the method.

The order in which a forecaster forms an estimate and reads the market price is not a detail. It determines whether the forecaster has an estimate at all.

A price seen before an estimate is formed becomes the estimate. The effect is anchoring, it is robust, it operates on experts as reliably as on novices, and it is not undone by awareness of it. The number seen first sets the range within which the subsequent number will fall, and the forecaster will then generate reasons why the price is approximately right, mistaking the exercise for analysis.

The procedure is simple and rarely followed. Form the estimate from the base rate and the specifics. Record it, in writing, as a number.

Then, and only then, read the price. The two can now be compared, and the comparison is informative: agreement suggests the estimate is defensible, and disagreement identifies a question worth pursuing, since one of the two numbers has incorporated something the other has not.

A forecaster who reads the price first has no second number and therefore no comparison. They have the market's view, dressed as their own.

The cost of skipping this step is invisible, which is why it is skipped. Nothing feels wrong. The forecaster simply never learns what they would have thought.

Updating

FIG. 04 · THE ASYMMETRY SCHEMATIC



Same weight of evidence. Different size of update.

New information should move an estimate by an amount proportional to how much it changes the probability. That principle is easy to state and, in practice, rarely followed, because the size of the update is governed less by the strength of the evidence than by whether the forecaster welcomed it.

The asymmetry is well documented. Evidence supporting a held position produces prompt and full revision. Evidence against it is discounted, attributed to noise, or held for confirmation that never quite arrives. The forecaster is not being dishonest. They are applying different standards of proof to the two directions without noticing that the standards differ.

The correction is not to eliminate the preference, which is not possible, but to counter it procedurally. Two questions serve. On welcome evidence: what would I need to see to move back? On unwelcome evidence: would I discount this if it pointed the other way? A forecaster who answers the second honestly will update more, and more evenly, than one who does not ask it.

There is a second failure in the opposite direction: over-updating, where a single data point swings the estimate far beyond what its weight supports. This is common in fast-moving situations and is checked by the base rate. An update that carries an estimate across the base rate should be examined before it is accepted.

Proportional updating is what distinguishes a forecaster from a commentator. The commentator has a view. The forecaster has an estimate with a revision history.

The answerable question

FIG. 05 · FORECASTABLE AGAINST INTERESTING

FORECASTABLE	INTERESTING
Resolves	Does not resolve
Has a written rule	Resolves by shifting consensus
Has a date	Has no date

Not every question that is interesting is one that can be forecast, and the failure to distinguish them costs more time than any other error in this guide.

A forecastable question has three properties. It resolves: at some point the answer becomes known. It has a rule: the criterion by which it resolves is stated precisely enough that two observers would agree on the outcome. It has a date: the resolution occurs within a horizon that permits the forecast to be scored.

Whether a policy is wise, whether a leader is competent, whether a company is well managed: these are interesting and unforecastable. They do not resolve, or they resolve only by consensus that shifts, or they have no date. A forecast about them cannot be scored, and a forecast that cannot be scored cannot improve.

Whether a rate decision matches guidance, whether a bill passes by a date, whether a release prints above a threshold: these are forecastable. They are also, frequently, less interesting than the unforecastable questions they sit inside, which is why forecasters drift toward the latter.

The discipline is to translate. Every interesting question contains forecastable ones, and the work is to find them. "Is the economy strong?" is not answerable. "Does the next release print above the consensus?" is. A forecaster who works only on answerable questions accumulates a score. One who does not accumulates opinions.

Scoring

ILLUSTRATIVE EXAMPLE

FIG. 06 · SIX FORECASTS, SCORED

FORECAST	OUTCOME	SCORE
0.70	1	0.09
0.60	0	0.36
0.85	1	0.02
0.40	0	0.16
0.75	1	0.06
0.55	1	0.20
Average		0.15

A forecaster who always says 0.50 scores 0.25. Lower is better.

A recorded forecast can be scored, and the standard score is simple enough to compute by hand.

The Brier score for a single forecast is the square of the difference between the stated probability and the outcome, with the outcome written as 1 if the event occurred and 0 if it did not. A forecast of 70 percent on an event that occurred scores 0.7 minus 1, squared, which is 0.09. The same forecast on an event that did not occur scores 0.7 minus 0, squared, which is 0.49. Lower is better. Zero is perfect.

Two properties make the score useful. A forecaster who always says 50 percent scores exactly 0.25 on every forecast, so 0.25 is the line below which a forecaster is adding information and above which they are subtracting it. And the score is proper: a forecaster minimizes it by stating their true probability, so there is no advantage in hedging or in exaggerating.

The table on this page shows six illustrative forecasts scored this way and averaged into a single figure. No arithmetic is demanded of the reader beyond seeing that the average sits on one side or the other of 0.25.

Over a few dozen forecasts, the average Brier score is a measure of the forecaster that no panel, column or commentary has ever produced about its author. It is the difference between claiming to be accurate and being able to show it.

The market as benchmark

ILLUSTRATIVE EXAMPLE

FIG. 07 · FORECASTER AGAINST PRICE

THE FORECASTER	THE PRICE
0.19	0.16

Same twenty questions, same period.

A prediction market price is a forecast. It can be scored by the same method, on the same questions, over the same period, and the result is a benchmark that keeps its own score continuously and for free.

Beating it means producing a lower Brier score than the price across a meaningful number of resolved questions. That happens. It happens rarely, it happens more often on thin and obscure contracts than on liquid ones, and it happens to forecasters who have done the work on the preceding six pages more often than to those who have not.

A persistent gap in either direction is informative. A forecaster who consistently scores worse than the price is systematically overconfident or systematically anchored, and the calibration plot from page three will show which. A forecaster who consistently scores better on a particular class of question has found something, and should examine what before assuming it will persist.

The benchmark also states the risk plainly. A price of 70 that resolves against its holders three times in ten is not a failed benchmark. It is a correct one. Capital committed against any price is capital that can be lost, and a well-calibrated forecaster loses it at exactly the rate their numbers predict. Beating the market does not mean avoiding loss. It means losing less often than the price says, over time, and being able to prove it.

A NECESSARY STATEMENT OF LIMITS

Calibration improves the accuracy of a forecaster's probabilities. It does not convert probabilities into certainties. A forecaster who has done everything in this guide correctly will still be on the wrong side of resolved outcomes at the rate their own numbers state, and capital committed to any market is capital exposed to that loss. The method is a way of knowing how often. It is not a way of avoiding it.

Where iPredicta fits

The preceding page treated the market price as a single benchmark. It is not one number. The same question is priced on more than one market, and the prices differ.

For a forecaster keeping score, that matters twice. The benchmark is not one Brier score but two, and a forecaster who beats one and not the other has learned something about which population they resemble. And the gap between the two prices is itself worth recording: two populations have examined one question and reached different probabilities.

ILLUSTRATIVE · NOT LIVE PRICES

FIG. 08 · ONE QUESTION, TWO BOOKS

TERMS DIFFER

Does the scheduled decision change the policy rate?

TERMS	Venue A	Venue B
Implied probability, Yes	48	53
Settlement source	Published decision	Published decision
Cutoff	Meeting date	Month end
Last trade	4 min ago	1 h 12 min ago

Seeing both requires both to be visible at once, with the settlement terms alongside so that the gap can be attributed rather than merely observed. By hand, that takes longer than either price is stable for.

iPredicta exists for that view. Every market, wherever it is priced, in one place, with the terms shown alongside the levels.

For a forecaster who keeps a scorecard, it is the column the scorecard was missing: the benchmark, in full, on the same question, at the same moment as the estimate it is being compared to.

The board is at ipredicta.co, and the benchmark described here is on it, priced on both venues.

What we are, and how we are paid

iPredicta is a media and data company. It is not a venue and not an operator. No deposits are taken, no client money is held, and nothing is executed here.

Revenue comes from affiliate commission when a reader opens an account at a venue. That is stated rather than buried because it determines the incentive, and the incentive is worth being able to see: a business paid on breadth has reason to present every venue and no reason to favor any one of them.

Editorial follows from the same position. The site publishes what the prices say, which includes publishing that a question is thinly traded, poorly specified, or receiving more attention than its book can support.

WHAT WE DO NOT DO

Take deposits or hold client money.

Execute anything.

Tell you what to buy.

One platform, every market.

IN SUMMARY

The method

01	Calibration, not hit rate.
02	Base rate first.
03	Estimate, then price.
04	Update in proportion.
05	Forecast only what resolves.
06	Score every one.
07	Benchmark against the market.

Accuracy is calibration, calibration is measurable, and the market is the benchmark that keeps its own score.

Over the coming days a short video follows, along with the single idea from this guide that carries the most weight of the seven, and a worked example run end to end on one question.

Score one forecast against the board today.

ipredicta.co