

What is reference class forecasting?

Reference class forecasting predicts a plan's outcome from how comparable past cases turned out, and corrects the plan's own estimate toward that record.

Author	Published	Reading time	Words
Sinan Isoglu, MBA (Quantic) Commercial growth leader, lecturer and doctoral researcher	18 October 2026	19 min	4,270

Management summary Reference class forecasting predicts the outcome of a planned action, such as its cost, duration or demand, from the actual outcomes of a class of comparable past actions rather than from the plan's own scenarios. This concept guide defines it and the inside and outside views behind it (Kahneman and Lovallo; Flyvbjerg), explains why forecasts built from a plan tend to be optimistic, sets out the three steps and the cost uplifts used for UK transport budgets, shows how small and favourable a class recalled from memory turned out, reports a film revenue test, and gives a one-page record for a revenue forecast. The sources are one conceptual essay, one account of the method in planning practice and one paper with two empirical studies.

Contents	§ 1	What is reference class forecasting?	2
	§ 2	Why do forecasts built from a plan tend to be optimistic?	3
	§ 3	Which steps turn a class of past cases into a forecast?	5
	Figure 1	Budget uplift for a 50 and an 80 per cent chance of staying within it	6
	§ 4	Which past cases belong in the class?	7
	§ 5	Did looking at the class change the forecasts?	9
	§ 6	Why do planners resist the outside view?	10
	§ 7	Which fields make a reference class forecast checkable?	11
	Figure 2	A reference class forecast record	11
	§ 8	What do the three papers leave untested for a firm's revenue forecast?	13
	References	14	

What is reference class forecasting?

§ 1 of 8

It is a way of predicting the outcome of a planned action from how comparable actions turned out. Flyvbjerg (2008)¹, who describes its first use in planning practice, defines it this way: “A reference class forecast of a given planned action is based on knowledge about actual performance in a reference class of comparable actions already carried out.” He calls it “a method for systematically taking an outside view on planned actions”.

¹ Flyvbjerg, B. (2008). Curbing optimism bias and strategic misrepresentation in planning: Reference class forecasting in practice. *European Planning Studies*, 16(1), 3–21.

The terms inside view and outside view come from Kahneman and Lovallo (1993)², who credit the comparison of a case with its class to earlier work by Kahneman and Tversky. The inside view is the forecast a planning team makes without being asked to: “An inside view forecast is generated by focusing on the case at hand, by considering the plan and the obstacles to its completion, by constructing scenarios of future progress, and by extrapolating current trends.” The outside view does the opposite. “It essentially ignores the details of the case at hand, and involves no attempt at detailed forecasting of the future history of the project. Instead, it focuses on the statistics of a class of cases chosen to be similar in relevant respects to the present one.” The case at hand is then compared with the other members of the class, to judge where it is likely to fall among their outcomes. The same question, the authors write, runs through forecasting and through choices under risk: “whether a particular problem of forecast or decision is treated as unique, or as an instance of an ensemble of similar problems.”

² Kahneman, D., & Lovallo, D. (1993). Timid choices and bold forecasts: A cognitive perspective on risk taking. *Management Science*, 39(1), 17–31.

In this guide, reference class forecasting means forecasting a quantity of a planned action, such as its cost, duration or demand, from the distribution of actual outcomes in a class of comparable completed actions, and then placing the case in that distribution, instead of building the number from the plan’s own steps and scenarios. Where the plan’s estimate is kept, it is corrected toward the class rather than trusted as it stands. It is the outside view done with data. It differs from three things it is easily confused with: a forecast that extends the case’s own history, as a time series of past sales does; a single analogy, one past case that looks most like this one; and a scenario,

which tells how this case might unfold. For a product with no sales history, the [new-product forecasting essay](#) argues for combining independent methods around a defined decision; a reference class adds an estimate built from outcomes rather than from the plan's scenarios, though in the actual-over-plan form used below, it still scales with the plan figure, so it corrects the plan rather than replacing it.

Why do forecasts built from a plan tend to be optimistic?

§ 2 of 8

Kahneman and Lovallo's³² thesis is that decision makers “have a strong tendency to consider problems as unique”, and that “Overly optimistic forecasts result from the adoption of an inside view of the problem, which anchors predictions on plans and scenarios.” Their reasoning is about probability. “In general, the future of a long and complex undertaking is simply not foreseeable in detail.” The plan's own scenario, flawless execution, may be much more probable than any one specific sequence of events that would make the project take four times as long as planned, yet “the less favorable outcome could be more likely overall, because there are so many different ways for things to go wrong”. The inside view does not always err toward optimism, they add: parents of a rebellious teenager can worry more than the statistics warrant. “To decision makers with a goal and a plan, the same way of thinking offers absurdly optimistic forecasts.”

³ Lovallo, D., Clarke, C., & Camerer, C. (2012). Robust analogizing and the outside view: Two empirical tests of case-based decision making. *Strategic Management Journal*, 33(5), 496–512.

Their illustration is an incident from 1976. A team of academics and teachers, about a year into developing a high-school curriculum on judgment and decision making under uncertainty in Israel, wrote down how many months they would need to complete a draft for the Ministry of Education: “The estimates, including my own, ranged from 18 to 30 months.” Kahneman then asked the team's curriculum expert how comparable teams had fared. “About 40% of them eventually gave up”, the expert said, and of the rest, “I cannot think of any that was completed in less than seven years, nor of any that took more than ten”. He could think of nothing that set this team apart for the better, and his own first estimate had been similar to the others'. The forecast was quietly dropped, and the project reached “eventual completion some eight years later”. Among the lessons the authors draw: “consensus on a forecast is not necessarily an indication of its

validity”, and “statistical knowledge that is known to the forecaster will not necessarily be used, or indeed retrieved, when a forecast is made by the inside approach”. The incident is one remembered case, offered as an illustration.

They also cite an interview study of new entrepreneurs by Cooper and colleagues, which they present as a striking example of the contrast between the two views. “Over 80% of entrepreneurs perceived their chances of success as 70% or better”, and “Fully one-third of them described their success as certain.” Asked about a business like theirs, the same entrepreneurs gave a mean chance of success of 59 per cent. Kahneman and Lovallo² set both answers against a five-year survival rate for new firms of about 33 per cent, a figure they take from a Dun and Bradstreet report of 1967.

Organizations, they argue, can discipline the forecasts they treat as recurrent: “We have no reason to doubt the ability of organizations to impose forecasting discipline and to reduce or eliminate obvious biases in recurrent problems.” Their example of a recurrent problem is “forecasts of the sales of existing product lines”. But “all significant forecasting problems have features that make them appear unique”, and where projects compete for resources, “The only projects that have a good chance of surviving in this competition are those for which highly favorable outcomes are forecast”. The winning project is therefore “more likely than others to be associated with optimistic errors”, and “A forecast readily becomes a target”. These are arguments built on the literature they review, not results of a study of their own.

Flyvbjerg (2008)¹ adds a second cause to the cognitive one. The cognitive cause is optimism bias, “a cognitive predisposition found with most people to judge future events in a more positive light than is warranted by actual experience”, and he describes the underlying error, which he credits to Kahneman and Tversky’s² work and calls, after Lovallo and Kahneman³, the planning fallacy: people underestimate the costs, completion times and risks of planned actions and overestimate their benefits. The second cause is strategic misrepresentation, in which “forecasters and planners deliberately and strategically overestimate benefits and underestimate costs” to win approval and funding. Both deceive; the difference is intent, since “optimism bias is self-deception”. In the database of transport projects from his earlier studies, actual construction costs exceeded the forecasts made at the decision to build by 44.7 per cent on average for rail, 33.8 per cent for bridges and tunnels and 20.4 per cent for roads,

in constant prices. The two explanations, he argues, complement each other, since “one is strong where the other is weak”: strategic misrepresentation has the stronger claim where political and organizational pressures are high, which he says was the situation for the projects in those studies.

Which steps turn a class of past cases into a forecast?

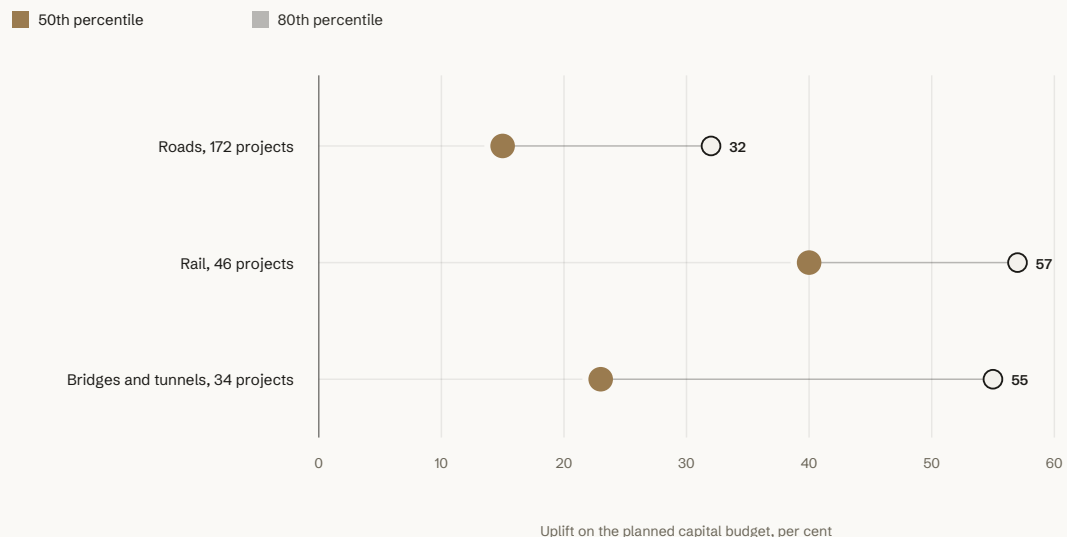
§ 3 of 8

Flyvbjerg sets out three. First, “Identification of a relevant reference class of past, similar projects.” The class “must be broad enough to be statistically meaningful but narrow enough to be truly comparable with the specific project.” Second, “Establishing a probability distribution for the selected reference class”, which “requires access to credible, empirical data for a sufficient number of projects within the reference class to make statistically meaningful conclusions.” Third, “Comparing the specific project with the reference class distribution, in order to establish the most likely outcome for the specific project.” The method does not try to foresee the events that will affect this project; it places the project among the outcomes of the class. In statisticians’ terms, he writes, following Kahneman and Tversky², that means “regressing forecasters’ best guess toward the average of the reference class” and “expanding their estimate of credible interval toward the corresponding interval for the class”. Lovallo et al. (2012)³ give the procedure in five steps that make the correction explicit: choose an unbiased class, assess the distribution of its outcomes, estimate the outcome intuitively, assess how predictable outcomes are, and correct the intuitive estimate statistically.

The UK applied the method to transport budgets. In a study for the Department for Transport, Flyvbjerg and Cowi¹ grouped the department’s transport schemes into categories in which statistical tests, benchmarking and other analyses showed that the risk of cost overrun could be treated as statistically similar. “For roads, for example, a class of 172 completed and comparable projects was used”. From each class’s distribution of cost overruns the study derived an optimism bias uplift on the planned capital budget, to be applied to the budget at the decision to build, in constant prices, and the uplift depends on the risk the investor will accept: “The lower the acceptable risk for cost overrun, the higher the uplift.” From August 2004, local authorities applying to the Department for Transport or HM Treasury for funding had to allow for optimism bias with these uplifts. Figure 1 shows them for the three classes for which a distribution was available.

Figure 1 Budget uplift for a 50 and an 80 per cent chance of staying within it

Moving from a 50 to an 80 per cent chance of staying within budget adds 17 points of uplift for roads and for rail, and 32 for bridges and tunnels.



Each dot is the percentage added to the planned capital budget. At the 80th percentile the risk that the project still costs more than the uplifted budget is 20 per cent; at the 50th it is 50 per cent. The values are 15 and 32 per cent for roads, 40 and 57 for rail, and 23 and 55 for bridges and tunnels, which Flyvbjerg's table calls fixed links.

Source Flyvbjerg (2008), p. 16, Table 4: optimism bias uplifts on capital expenditure at the 50% and 80% percentiles, constant prices, applied to the budget at the decision to build; class sizes from Figures 2 to 4, pp. 13 and 14. Building, IT and civil engineering have only ranges, from another study, and are not drawn.

The 80th percentile “is the level of risk that the UK Department for Transport is typically willing to accept for large investments in local transportation infrastructure”. For a new motorway estimated at GBP 100 million, planners who want an 80 per cent chance of staying within budget “would use an uplift of 32% on their estimated capital expenditure budget”, which gives GBP 132 million; at 50 per cent the budget would be GBP 115 million. The median uplift, Flyvbjerg writes, suits “the investor with a large project portfolio, where cost overruns on one project may be offset by cost savings on another.” A planner may go below the class uplift only with evidence: “Only if planners have historical evidence to demonstrate that they would be significantly better at estimating costs for the project at hand than their colleagues were for the projects in the reference class would the planners be justified in using lower uplifts”. The rule also runs the other

way: “Conversely, if there is evidence that the planners are worse at estimating costs than their colleagues, then higher uplifts should be used.”

The first use was for Edinburgh’s Tram Line 2, in October 2004. “The business case had estimated a base cost of GBP 255 million and an additional allowance for contingency and optimism bias of GBP 64 million”, about 25 per cent, for a total of about GBP 320 million. Reviewing the business case for the Scottish Parliament, Ove Arup applied the rail uplifts to the base cost: GBP 357 million at the 50th percentile and GBP 400 million at the 80th. Arup judged even these figures likely to be low, because “Tram Line 2 had not yet even reached the outline business case stage”, when risks are higher still. The paper reports the forecast; it does not report what the line cost.

Which past cases belong in the class?

§ 4 of 8

The class is the hard part. In the curriculum story it was easy to find. “Other cases are more ambiguous”, Kahneman and Lovallo² write. For a firm costing an investment in an unfamiliar technology, they ask whether the class is the firm’s own recent ventures in new technologies or the most similar developments carried out by other firms. Their answer: “Neither is perfect, and the recommendation would be to try both”. The class may also need a unit that makes cases of different size comparable: “The ratio of actual spending to planned expenditure is an example of a convenient unit that permits meaningful comparisons across diverse projects.”

Flyvbjerg’s¹ classes were built by testing. “Projects were placed in the same reference class if tests showed no significant differences between projects and if similar regulatory and construction regimes applied to projects.” On that rule, “rail projects in Europe and North America were placed in the same reference class, whereas rail projects in developing nations and Japan were excluded from the reference class”. He argues that projects are more alike in statistical terms than their planners assume, and that the hard part is the data: “The real challenge in doing a reference class forecast lies in assembling a valid dataset that will allow a reliable forecast.” His own classes were possible only because years of research on cost overruns in transport projects had produced such a dataset.

In the one study here that asked people to recall a class, it came out small and favourable. Lovallo et al. (2012)³ asked 38 employees of private equity firms about an investment opportunity they were reviewing or a recent investment in their portfolio. Each first described

the most probable path for the investment and stated its expected rate of return, then named two categories of similar investments, was asked to list the five most relevant examples in each, and estimated the average return of each category. Of the 27 who listed similar investments, 10 named only one category; only 14 of the 33 usable respondents listed five projects in each; and the great majority of the reference projects were described as successes, a label the questionnaire also allowed for an investment the respondent only expected to succeed: it asked whether each example “was, or you believe will be, a success or failure”. The authors call these results disturbing: they suggest, they write, that the method “is itself open to bias in the recollection of” reference projects. Sampling from memory, they warn, is likely to take too few cases and “analogies that are too favorable”.

How many cases, and weighted how? Lovallo et al.³’s second study forecast United States box office revenue for 19 films, with the similarity ratings collected before their summer release. Drawing on a database of 1,751 films released in the United States between 1993 and 2004, 358 moviegoers rated how similar each target was to past films they knew. The films judged most similar, between 1 and 15 per target (Table 1), formed each target’s class, and each was weighted by its similarity; the authors call this similarity-based forecasting. Its mean absolute relative error was 30.09 per cent, against 62.42 per cent for a regression on film features estimated on the 1,751 earlier films and 66.36 per cent for the same regression with interactions (Table 2). Using only the single most similar film did far worse: the text gives a mean error of 135 per cent, and the errors listed in Table 2 average about 151 per cent. Two points temper the result. The step that turned similarity into dollars was a regression fitted on the 19 target films’ own revenues, with opening-week cinemas and a sequel marker, and $R^2 = 0.55$ “for the sample”, whereas the benchmark regression was estimated on the earlier films. And by the errors in Table 2 the similarity forecast was closer for 15 of the 19 films, where the text says 17. On this guide’s reading, the study is consistent with the authors’ case for several weighted cases over one, but it is not an out-of-sample test of either comparison.

Did looking at the class change the forecasts?

§ 5 of 8

In the private equity study it did, in one direction. The mean expected return for the investments in question was 29.35 per cent, against 18.72 and 20.82 per cent for the two categories of similar investments ($p < 0.001$). Of the 17 respondents whose expected return lay above both of their category averages, 14 lowered it when asked whether they would revise it; the mean revision the authors report for this group was 5.96 percentage points, which they note is more than a fifth of the expected return. None revised upward, and those whose investment lay between or equal to their category averages changed nothing. The authors call the gap between the investments and their categories “strongly indicative” but “not conclusive evidence” that inside view forecasts are more optimistic than outside view forecasts; the reason they give is that the participants viewed their current projects as particularly attractive.

The design limits what the study can show. Every participant answered the inside view questions first and the outside view questions second; no group skipped the outside view, and the paper reports no returns that the investments actually earned. On this guide’s reading, the study shows that forecasts moved toward the class when people were asked to look at one; it cannot show that the moved forecasts were more accurate.

Flyvbjerg reports no accuracy test of his own either. With the first reference class forecasts made in 2004 and 2005, “it is too early to draw definitive conclusions about the impacts of adopting the practice of reference class forecasting”; he observes that in the UK the method raised awareness of optimism bias and strategic misrepresentation in transport planning. His statement that the method “has been demonstrated to be more accurate than conventional forecasting methods” rests on experimental work by Kahneman and others that he cites and this guide has not read. Kahneman and Lovallo’s³² own statement, that “when both methods are applied with equal intelligence and skill, the outside view is much more likely to yield a realistic estimate”, follows from the argument above rather than from a measured result, and so do their prescriptions: “We suggest that errors of intuitive prediction can sometimes be reduced by adopting an outside view, which forecasts the outcome without attempting to forecast its history”, and an outside view “will generally reduce the optimistic bias”.

Why do planners resist the outside view?

§ 6 of 8

Kahneman and Lovallo's³² main observation is psychological: “the inside view is overwhelmingly preferred in intuitive forecasting”. The inside view is valued as a serious attempt to deal with the unique case, and the outside view is rejected “for relying on crude analogy from superficially similar instances”. Physicians and lawyers, they note, often argue against statistical reasoning about particular cases, and there the preference for the inside view “almost bears a moral character”. Deliberately setting aside what makes a case unique “can appear irresponsible”, and the curriculum team met a further cost: “facing the facts can be intolerably demoralizing”. The authors also doubt that realism always repays its price. Drawing on work they cite, they note that optimism can help people persist, and that “a successful effort to improve the realism of assessments could do more harm than good in an organization that relies on unfounded optimism to ward off paralysis”. They call it “a genuine dilemma that will not yield to any simple rule”.

Flyvbjerg separates the situations by cause. Where optimism bias drives the error, forecasters can be assumed to make honest mistakes, and “the potential for using the outside view and reference class forecasting will be good”. Where strategic misrepresentation drives it, as when cities compete for national funds, “the potential for reference class forecasting is low”, because the inaccuracy is deliberate, and “incentives must be aligned to reward accurate forecasts and punish inaccurate ones”: grants capped at the estimate made at the decision to build, funding that depends on using the method, and due diligence by banks and independent bodies. He adds two cautions. A class “may fail to predict extreme outcomes, that is, those that lie outside all historical precedents”. And large uplifts create reserves that can invite spending: “Reserves will be spent simply because they are there, as many believe in the construction business”, so the uplifts need tight contracts and cost control alongside them.

Which fields make a reference class forecast checkable?

§ 7 of 8

Six, in one row per forecast: the decision the number will inform; the inside view estimate as the plan states it, and how it was built; the reference class, with where its cases come from and the rule for including them; the class distribution in a unit that compares cases of different size, such as actual over plan; the evidence for placing this case above or below the middle of the class; and the forecast at the certainty the decision needs. The fields follow Flyvbjerg's¹ three steps and his rule on lower and higher uplifts, Kahneman and Lovallo's³² advice on classes and units, and Lovallo et al.'s finding on classes recalled from memory.

Figure 2 A reference class forecast record

Write down the plan figure, the class, its distribution and the case's position before the adjusted number goes into a budget.

Decision	Inside view	Class	Distribution	Position	Forecast
What will this number decide?	Plan figure, built how?	Which past cases, chosen how?	Actual over plan, which percentiles?	Evidence this case differs?	Number at the certainty needed?

Read the percentile the decision needs. Shift the case only on evidence that this team plans better, or worse, than the class did.

One row per forecast. Keep the plan figure next to the adjusted one. If the class was recalled from memory rather than drawn from records, say so in the class field.

Source Author's decision worksheet. The inside view field follows Kahneman and Lovallo (1993), p. 25; the class and distribution fields follow Flyvbjerg (2008), p. 8, and Kahneman and Lovallo (1993), p. 25; the position field follows Flyvbjerg (2008), p. 16; the forecast field follows Flyvbjerg (2008), pp. 12 to 16; the warning on classes recalled from memory follows Lovallo et al. (2012), pp. 500 and 509. The decision field and the application to revenue are this guide's own. The fields are reader inputs; no company data is supplied.

An illustration, with no company data: a software firm plans an add-on module, and its product team forecasts first-year bookings from the plan, from the accounts it will target, the share it expects to buy and the price. The record keeps that figure as the inside view. The class is the firm's own earlier add-on launches and, if the data can be had, comparable launches at other firms; Kahneman and Lovallo's³² advice is to try both. The unit is actual first-year bookings divided by the figure in each launch's plan, which lets launches of different size be compared. For revenue the risk runs the other way from a cost overrun, so the stricter number is a low percentile of the class: if four in five past launches reached at least 0.6 of their plan, the record states 0.6 times the plan figure as the level that four in five reached. The median suits a portfolio of launches whose shortfalls and surprises can offset each other; a single launch whose shortfall the budget cannot absorb needs the stricter figure, as Flyvbjerg argues for costs. The certainty the decision needs sets which percentile is read. The position field shifts the case within the class, and only on evidence: toward the plan figure if the team's past launches have come closer to their plans than those in the class, further below it if they have fallen further short, as Flyvbjerg, for costs, allows lower uplifts and calls for higher ones. If the class holds only a handful of launches, the record says so, because the second step needs enough cases for statistically meaningful conclusions, and a class recalled from memory may come out small and favourable, as it did in Lovallo et al.'s questionnaire. Keeping the plan figure and the adjusted figure side by side is the record the forecast override guide sets out for any change from a baseline forecast to a final forecast, and the actual result later shows whether the change improved the forecast, the comparison the forecast value added essay makes. These are the author's examples and have not been tested.

What do the three papers leave untested for a firm's revenue forecast?

§ 8 of 8

Most of what a revenue team would want to know. Kahneman and Lovallo (1993)² is a conceptual essay: it argues for the two views from psychological research, an incident and studies by others, and does not test the outside view itself. Flyvbjerg (2008)¹ describes the method and its first uses in UK transport planning and reports no test of whether those forecasts were more accurate. Lovallo et al. (2012)³ ran two studies: in one, the forecasts of private equity employees moved toward a class they recalled themselves, with no outcomes observed; in the other, similarity-weighted classes forecast the revenue of 19 films with less error than a regression, with the fitting step noted above. None of the three papers forecasts a firm's sales pipeline, bookings or subscription revenue; the nearest is box office revenue for films.

On this guide's reading, one thing carries over with some confidence, and a second as a warning. A forecast built from a plan alone is prone to optimism when the undertaking is not routine: Kahneman and Lovallo² argue it, and Flyvbjerg's¹ database shows transport projects costing more than forecast on average, although for those projects he gives strategic misrepresentation the greater weight. And a reference class recalled from memory may come out small and tilted toward successes: it did in one questionnaire study of private equity employees, with 33 usable responses, where a success could also be one the respondent only expected, and the authors judge that sampling from memory is likely to do the same. Which class and unit fit a firm with few comparable launches, how much a class improves a revenue forecast, and how forecasters behave once a class is required, the sources do not say. Kahneman and Lovallo place forecasts of the sales of existing product lines among the recurrent problems that organizations can already discipline; on that reading, the outside view earns its place on the non-routine number.

 End of essay

References

- 1** Flyvbjerg, B. (2008). Curbing optimism bias and strategic misrepresentation in planning: Reference class forecasting in practice. *European Planning Studies*, 16(1), 3–21.
<https://doi.org/10.1080/09654310701747936>
- 2** Kahneman, D., & Lovallo, D. (1993). Timid choices and bold forecasts: A cognitive perspective on risk taking. *Management Science*, 39(1), 17–31.
<https://doi.org/10.1287/mnsc.39.1.17>
- 3** Lovallo, D., Clarke, C., & Camerer, C. (2012). Robust analogizing and the outside view: Two empirical tests of case-based decision making. *Strategic Management Journal*, 33(5), 496–512.
<https://doi.org/10.1002/smj.962>

How to cite this essay

Isoglu, S. (2026, October 18). What is reference class forecasting? Sinan Isoglu.
<https://isoglu.com/insights/journal/what-is-reference-class-forecasting/>
Online version:
<https://isoglu.com/insights/journal/what-is-reference-class-forecasting/>

Sinan Isoglu, MBA (Quantic)

Commercial growth leader, lecturer and doctoral researcher
© 2026 Sinan Isoglu · isoglu.com · All rights reserved