

Was ist eine Referenzklassenprognose?

Die Referenzklassenprognose leitet das Ergebnis eines Vorhabens aus dem Verlauf vergleichbarer früherer Fälle ab und rückt die Planzahl in deren Richtung.

Autor	Veröffentlicht	Lesezeit	Wörter
Sinan Isoglu, MBA (Quantic)	18. Oktober 2026	24 Min.	4.971
Führungskraft für Umsatzwachstum, Dozent und Doktorand			

Management Summary Die Referenzklassenprognose schätzt das Ergebnis eines geplanten Vorhabens, etwa seine Kosten, seine Dauer oder seine Nachfrage, aus den tatsächlichen Ergebnissen einer Klasse vergleichbarer früherer Vorhaben statt aus den Szenarien des Plans. Dieser Konzeptleitfaden definiert sie und die Innen- und Außensicht, auf denen sie beruht (Kahneman und Lovallo; Flyvbjerg), erklärt, warum aus dem Plan abgeleitete Prognosen zu Optimismus neigen, beschreibt die drei Schritte und die Kostenzuschläge für britische Verkehrsbudgets, zeigt, wie klein und erfolgslastig eine aus dem Gedächtnis gebildete Klasse ausfiel, berichtet einen Test mit Einspielergebnissen von Filmen und bietet ein Blatt für eine Umsatzprognose. Die Quellen sind ein konzeptioneller Aufsatz, ein Bericht über die Methode in der Planungspraxis und ein Aufsatz mit zwei empirischen Studien.

Inhalt	§ 1	Was ist eine Referenzklassenprognose?	3
	§ 2	Warum neigen Prognosen aus dem Plan zu Optimismus?	5
	§ 3	Welche Schritte machen aus einer Klasse früherer Fälle eine Prognose?	8
	Abbildung 1	Zuschläge auf das Investitionsbudget bei 50 und bei 80 Prozent Wahrscheinlichkeit, im Budget zu bleiben	9
	§ 4	Welche früheren Fälle gehören in die Klasse?	11
	§ 5	Hat der Blick auf die Klasse die Prognosen verändert?	13
	§ 6	Warum sträuben sich Planer gegen die Außensicht?	14
	§ 7	Welche Felder machen eine Referenzklassenprognose überprüfbar?	15
	Abbildung 2	Ein Blatt für eine Referenzklassenprognose	16

§ 8	Was lassen die drei Arbeiten für die Umsatzprognose eines Unternehmens ungeprüft?	18
	Referenzen	19

Was ist eine Referenzklassenprognose?

§ 1 von 8

Sie schätzt das Ergebnis eines geplanten Vorhabens aus dem, was vergleichbare Vorhaben tatsächlich erreicht haben. Flyvbjerg (2008)¹, der ihren ersten Einsatz in der Planungspraxis beschreibt, definiert sie so: Eine Referenzklassenprognose für ein geplantes Vorhaben beruht auf dem Wissen über die tatsächlichen Ergebnisse einer Referenzklasse vergleichbarer, bereits abgeschlossener Vorhaben („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.“). Er nennt sie eine Methode, um systematisch die Außensicht auf geplante Vorhaben einzunehmen („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.

Die Begriffe Innen- und Außensicht gehen auf Kahneman und Lovallo (1993)² zurück, die den Vergleich eines Falls mit seiner Klasse früheren Arbeiten von Kahneman und Tversky zuschreiben. Die Innensicht ist die Prognose, die ein Planungsteam von sich aus erstellt: Sie entsteht aus dem Blick auf den vorliegenden Fall, auf den Plan und die Hindernisse auf dem Weg zu seiner Umsetzung, aus Szenarien des weiteren Verlaufs und aus der Fortschreibung aktueller Trends („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.“). Die Außensicht geht umgekehrt vor. Sie sieht von den Einzelheiten des Falls weitgehend ab und versucht nicht, seinen künftigen Verlauf im Detail vorherzusagen („It essentially ignores the details of the case at hand, and involves no attempt at detailed forecasting of the future history of the project.“); stattdessen stützt sie sich auf die Statistik einer Klasse von Fällen, die dem vorliegenden in den maßgeblichen Merkmalen ähneln („Instead, it focuses on the statistics of a class of cases chosen to be similar in relevant respects to the present one.“). Anschließend wird der Fall mit den übrigen Mitgliedern der Klasse verglichen, um einzuschätzen, wo er unter ihren Ergebnissen voraussichtlich liegt. Dieselbe Frage, schreiben die Autoren, zieht sich durch Prognosen wie durch Entscheidungen unter Risiko: ob ein Prognose- oder Entscheidungsproblem als einzigartig behandelt wird oder als ein Fall unter vielen ähnlichen („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 diesem Leitfaden heißt Referenzklassenprognose: eine Größe eines geplanten Vorhabens, etwa seine Kosten, seine Dauer oder die Nachfrage, aus der Verteilung der tatsächlichen Ergebnisse in einer Klasse vergleichbarer abgeschlossener Vorhaben zu schätzen und den Fall in diese Verteilung einzuordnen, statt die Zahl aus den Schritten und Szenarien des Plans aufzubauen. Wird die Schätzung aus dem Plan beibehalten, rückt man sie in Richtung der Klasse, statt sie unverändert zu übernehmen. Es ist die Außensicht, mit Daten umgesetzt. Sie unterscheidet sich von dreierlei, womit sie leicht verwechselt wird: von einer Prognose, die die eigene Vergangenheit des Falls fortschreibt, wie eine Zeitreihe früherer Umsätze; von einer einzelnen Analogie, also dem einen früheren Fall, der diesem am meisten ähnelt; und von einem Szenario, das erzählt, wie sich dieser Fall entwickeln könnte. Für ein Produkt ohne eigene Absatzhistorie plädiert der Beitrag zur Prognose neuer Produkte dafür, unabhängige Methoden für eine klar umrissene Entscheidung zu kombinieren; eine Referenzklasse liefert dazu eine Schätzung aus Ergebnissen statt aus den Szenarien des Plans. In der Form Ist zu Plan, die weiter unten verwendet wird, wächst diese Schätzung allerdings mit der Planzahl und korrigiert den Plan, statt ihn zu ersetzen.

Warum neigen Prognosen aus dem Plan zu Optimismus?

§ 2 von 8

Die These von Kahneman und Lovallo² lautet, dass Entscheidungsträger stark dazu neigen, Probleme als einzigartig zu betrachten („have a strong tendency to consider problems as unique“), und dass zu optimistische Prognosen aus der Innensicht entstehen, die Vorhersagen an Plänen und Szenarien verankert („Overly optimistic forecasts result from the adoption of an inside view of the problem, which anchors predictions on plans and scenarios.“). Ihre Begründung stützt sich auf Wahrscheinlichkeiten. Die Zukunft eines langen und komplexen Vorhabens lässt sich im Einzelnen nicht vorhersehen („In general, the future of a long and complex undertaking is simply not foreseeable in detail.“). Das Szenario des Plans, die reibungslose Umsetzung, kann viel wahrscheinlicher sein als jede einzelne bestimmte Abfolge von Ereignissen, durch die das Vorhaben viermal so lange wie geplant dauern würde; dennoch kann das ungünstigere Ergebnis insgesamt wahrscheinlicher sein, weil so vieles auf so viele Arten schiefgehen kann („because there are so many different ways for things to go wrong“). Die Innensicht fällt nicht immer zu optimistisch aus, fügen sie hinzu: Eltern eines aufsässigen Teenagers sorgen sich mitunter mehr, als die Statistik rechtfertigt. Wer aber ein Ziel und einen Plan hat, dem liefert dieselbe Denkweise absurd optimistische Prognosen („To decision makers with a goal and a plan, the same way of thinking offers absurdly optimistic forecasts.“).

Ihr Beispiel ist ein Vorfall aus dem Jahr 1976. Ein Team aus Wissenschaftlern und Lehrkräften, das seit etwa einem Jahr einen Lehrplan zu Urteilen und Entscheidungen unter Unsicherheit für weiterführende Schulen in Israel entwickelte, schrieb auf, wie viele Monate es noch bis zu einem fertigen Entwurf für das Bildungsministerium brauchen würde. Die Schätzungen, Kahnemans eigene eingeschlossen, lagen zwischen 18 und 30 Monaten („The estimates, including my own, ranged from 18 to 30 months.“). Kahneman fragte daraufhin den Lehrplanexperten des Teams, wie es vergleichbaren Teams ergangen war. Etwa 40 Prozent hätten irgendwann aufgegeben („About 40% of them eventually gave up.“), antwortete dieser, und von den übrigen falle ihm keines ein, das in weniger als sieben Jahren fertig geworden sei, und keines, das mehr als zehn gebraucht habe. Er sah nichts, was das eigene Team positiv von den anderen abhob, und seine eigene erste Schätzung hatte denen der anderen geähnelte. Die Prognose verschwand stillschweigend aus der Diskussion, und das Projekt wurde rund acht Jahre später abgeschlossen („eventual completion some eight years later“). Zu den

Lehren, die die Autoren ziehen, gehört, dass Einigkeit über eine Prognose nicht unbedingt für ihre Gültigkeit spricht („consensus on a forecast is not necessarily an indication of its validity“), und dass statistisches Wissen, über das der Prognostiker verfügt, bei einer Prognose aus der Innensicht nicht unbedingt genutzt oder auch nur abgerufen wird („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“). Der Vorfall ist ein einzelner erinnelter Fall und dient als Veranschaulichung.

Sie zitieren außerdem eine Interviewstudie von Cooper und Kollegen mit Gründern neuer Unternehmen, die sie als eindrückliches Beispiel für den Gegensatz der beiden Sichtweisen anführen. Über 80 Prozent der Gründer schätzten ihre Erfolgchancen auf 70 Prozent oder mehr („Over 80% of entrepreneurs perceived their chances of success as 70% or better.“), und ein volles Drittel hielt den eigenen Erfolg für sicher („Fully one-third of them described their success as certain.“). Nach einem Unternehmen wie dem ihren gefragt, nannten dieselben Gründer im Mittel eine Erfolgchance von 59 Prozent. Kahneman und Lovallo² stellen beide Antworten einer Fünfjahres-Überlebensrate neuer Unternehmen von rund 33 Prozent gegenüber, die sie einem Bericht von Dun and Bradstreet aus dem Jahr 1967 entnehmen.

Organisationen können, so ihr Argument, bei Prognosen, die sie als wiederkehrend behandeln, für Disziplin sorgen („We have no reason to doubt the ability of organizations to impose forecasting discipline and to reduce or eliminate obvious biases in recurrent problems.“). Als Beispiel für ein wiederkehrendes Problem nennen sie Absatzprognosen für bestehende Produktlinien („forecasts of the sales of existing product lines“). Doch alle bedeutenden Prognoseprobleme haben Züge, die sie einzigartig erscheinen lassen („all significant forecasting problems have features that make them appear unique“), und wo Projekte um Ressourcen konkurrieren, haben nur jene eine gute Chance, für die sehr günstige Ergebnisse vorhergesagt werden. Das siegreiche Projekt ist daher mit größerer Wahrscheinlichkeit als andere mit optimistischen Fehlern behaftet („more likely than others to be associated with optimistic errors“), und eine Prognose wird leicht zur Zielvorgabe („A forecast readily becomes a target“). Das sind Argumente, die auf der von ihnen gesichteten Literatur aufbauen, keine Ergebnisse einer eigenen Studie.

Flyvbjerg (2008)¹ stellt der kognitiven Ursache eine zweite zur Seite. Die kognitive Ursache ist die optimistische Verzerrung (optimism bias), eine bei den meisten Menschen anzutreffende Neigung, künftige

Ereignisse positiver zu beurteilen, als es die tatsächliche Erfahrung rechtfertigt („a cognitive predisposition found with most people to judge future events in a more positive light than is warranted by actual experience“). Den zugrunde liegenden Fehler, den er auf die Arbeiten von Kahneman und Tversky² zurückführt, nennt er nach Lovallo und Kahneman³ den Planungsfehlschluss: Menschen unterschätzen Kosten, Dauer und Risiken geplanter Vorhaben und überschätzen ihren Nutzen. Die zweite Ursache ist die strategische Falschdarstellung, bei der Prognostiker und Planer den Nutzen bewusst und gezielt zu hoch und die Kosten zu niedrig ansetzen („forecasters and planners deliberately and strategically overestimate benefits and underestimate costs“), um Genehmigung und Finanzierung zu erhalten. Beides ist Täuschung; der Unterschied liegt in der Absicht, denn die optimistische Verzerrung ist Selbsttäuschung („optimism bias is self-deception“). In der Datenbank von Verkehrsprojekten aus seinen früheren Studien lagen die tatsächlichen Baukosten im Mittel um 44,7 Prozent bei Schienenprojekten, um 33,8 Prozent bei Brücken und Tunneln und um 20,4 Prozent bei Straßen über den Prognosen zum Zeitpunkt der Bauentscheidung, zu konstanten Preisen. Beide Erklärungen ergänzen sich, so Flyvbjerg, denn die eine ist stark, wo die andere schwach ist („one is strong where the other is weak“): Die strategische Falschdarstellung erklärt mehr, wo politischer und organisatorischer Druck hoch ist, und das war nach seiner Angabe bei den Projekten jener Studien der Fall.

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.

Welche Schritte machen aus einer Klasse früherer Fälle eine Prognose?

§ 3 von 8

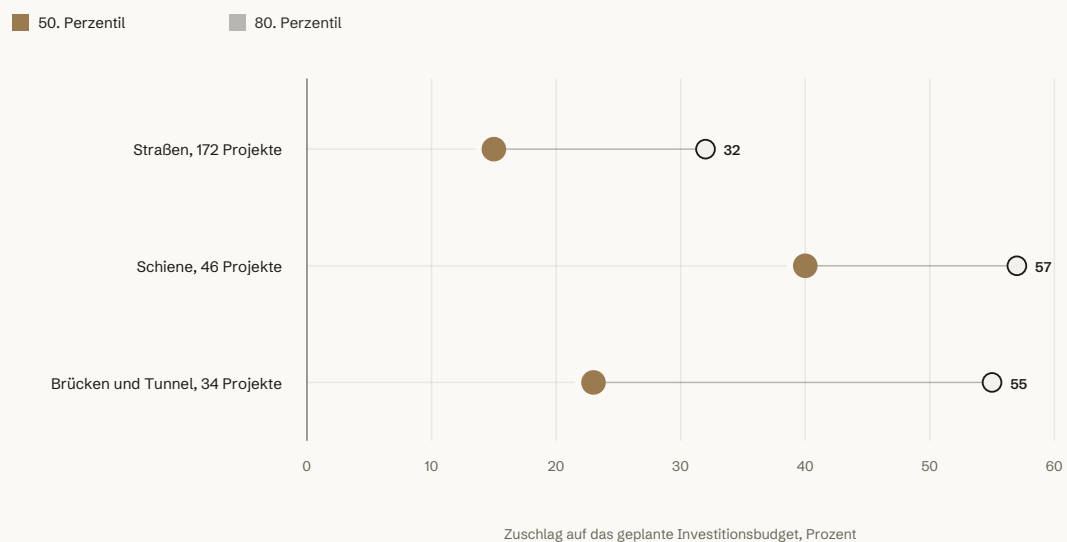
Flyvbjerg nennt drei. Erstens die Bestimmung einer relevanten Referenzklasse früherer, ähnlicher Projekte („Identification of a relevant reference class of past, similar projects.“). Die Klasse muss breit genug sein, um statistisch aussagekräftig zu sein, aber eng genug, um mit dem konkreten Projekt wirklich vergleichbar zu sein („The class must be broad enough to be statistically meaningful but narrow enough to be truly comparable with the specific project.“). Zweitens das Aufstellen einer Wahrscheinlichkeitsverteilung für die gewählte Referenzklasse („Establishing a probability distribution for the selected reference class“), wofür glaubwürdige empirische Daten zu genügend Projekten der Klasse nötig sind, um statistisch belastbare Schlüsse zu ziehen. Drittens der Vergleich des konkreten Projekts mit der Verteilung der Referenzklasse, um sein wahrscheinlichstes Ergebnis zu bestimmen („Comparing the specific project with the reference class distribution, in order to establish the most likely outcome for the specific project.“). Die Methode versucht nicht, die Ereignisse vorherzusehen, die dieses Projekt treffen werden; sie ordnet das Projekt unter den Ergebnissen der Klasse ein. Statistisch gesprochen, schreibt er im Anschluss an Kahneman und Tversky², wird die beste Schätzung der Prognostiker zum Mittelwert der Referenzklasse hin verschoben („regressing forecasters’ best guess toward the average of the reference class“) und die Spanne ihrer Schätzung auf die entsprechende Spanne der Klasse hin erweitert. Lovallo et al. (2012)³ beschreiben das Verfahren in fünf Schritten, die diese Korrektur ausdrücklich machen: eine unverzerrte Klasse wählen, die Verteilung ihrer Ergebnisse bestimmen, das Ergebnis intuitiv schätzen, beurteilen, wie gut sich Ergebnisse vorhersagen lassen, und die intuitive Schätzung statistisch korrigieren.

Großbritannien wandte die Methode auf Verkehrsbudgets an. In einer Studie für das Verkehrsministerium (Department for Transport) teilten Flyvbjerg und Cowi¹ die Verkehrsvorhaben des Ministeriums in Kategorien ein, in denen statistische Tests, Benchmarking und andere Analysen zeigten, dass das Risiko einer Kostenüberschreitung als statistisch ähnlich gelten konnte. Für Straßen etwa diente eine Klasse von 172 abgeschlossenen, vergleichbaren Projekten („For roads, for example, a class of 172 completed and comparable projects was used“). Aus der Verteilung der Kostenüberschreitungen jeder Klasse leitete die Studie einen Zuschlag für optimistische Verzerrung auf das geplante Investitionsbudget ab, anzuwenden auf das Budget zum Zeitpunkt der Bauentscheidung, zu konstanten Preisen, und der

Zuschlag hängt davon ab, welches Risiko der Investor hinnimmt: Je geringer das akzeptierte Risiko einer Kostenüberschreitung, desto höher der Zuschlag („The lower the acceptable risk for cost overrun, the higher the uplift.“). Seit August 2004 mussten Kommunen, die beim Verkehrsministerium oder beim Finanzministerium (HM Treasury) Mittel für Verkehrsprojekte beantragten, die optimistische Verzerrung mit diesen Zuschlägen berücksichtigen. Abbildung 1 zeigt sie für die drei Klassen, für die eine Verteilung vorlag.

Abbildung 1 Zuschläge auf das Investitionsbudget bei 50 und bei 80 Prozent Wahrscheinlichkeit, im Budget zu bleiben

Der Schritt von 50 auf 80 Prozent Wahrscheinlichkeit, im Budget zu bleiben, erhöht den Zuschlag bei Straßen und bei Schienenprojekten um 17 Punkte, bei Brücken und Tunneln um 32.



Jeder Punkt ist der Prozentsatz, der auf das geplante Investitionsbudget aufgeschlagen wird. Beim 80. Perzentil liegt das Risiko, dass das Projekt auch das erhöhte Budget überschreitet, bei 20 Prozent, beim 50. bei 50 Prozent. Die Werte sind 15 und 32 Prozent für Straßen, 40 und 57 für Schienenprojekte sowie 23 und 55 für Brücken und Tunnel, die Flyvbjergs Tabelle als feste Verbindungen (fixed links) führt.

Quelle Flyvbjerg (2008), S. 16, Tabelle 4: anwendbare Zuschläge für optimistische Verzerrung auf die Investitionsausgaben beim 50. und 80. Perzentil, zu konstanten Preisen, anzuwenden auf das Budget zum Zeitpunkt der Bauentscheidung; Klassengrößen aus den Abbildungen 2 bis 4, S. 13 und 14. Hochbauprojekte, IT-Projekte und Ingenieurbau tragen Spannen aus einer anderen Studie ohne Wahrscheinlichkeitsverteilung und sind nicht dargestellt.

Das 80. Perzentil entspricht dem Risiko, das das britische Verkehrsministerium bei großen Investitionen in die lokale

Verkehrsinfrastruktur typischerweise hinzunehmen bereit ist („is the level of risk that the UK Department for Transport is typically willing to accept for large investments in local transportation infrastructure“). Für eine neue Autobahn mit geschätzten Kosten von 100 Millionen Pfund würden Planer, die mit 80 Prozent Wahrscheinlichkeit im Budget bleiben wollen, einen Zuschlag von 32 Prozent ansetzen („would use an uplift of 32% on their estimated capital expenditure budget“), was 132 Millionen Pfund ergibt; bei 50 Prozent läge das Budget bei 115 Millionen Pfund. Der Median-Zuschlag, schreibt Flyvbjerg, passt zu einem Investor mit einem großen Projektportfolio, in dem Überschreitungen bei einem Projekt durch Einsparungen bei einem anderen ausgeglichen werden können. Unter den Zuschlag der Klasse darf ein Planer nur mit Belegen gehen: nur wenn er historisch nachweisen kann, dass er die Kosten des vorliegenden Projekts deutlich besser schätzen würde, als es seine Kollegen bei den Projekten der Referenzklasse taten („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“). Die Regel gilt auch umgekehrt: Zeigen Belege, dass die Planer die Kosten schlechter schätzen als ihre Kollegen, sind höhere Zuschläge anzusetzen („Conversely, if there is evidence that the planners are worse at estimating costs than their colleagues, then higher uplifts should be used.“).

Erster Anwendungsfall war im Oktober 2004 die Straßenbahnlinie 2 in Edinburgh. Der Business Case hatte Basiskosten von 255 Millionen Pfund und einen Aufschlag für Unvorhergesehenes und optimistische Verzerrung von 64 Millionen Pfund angesetzt („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“), rund 25 Prozent, zusammen etwa 320 Millionen Pfund. Ove Arup, beauftragt, den Business Case für das schottische Parlament zu prüfen, wandte die Zuschläge für Schienenprojekte auf die Basiskosten an: 357 Millionen Pfund beim 50. und 400 Millionen Pfund beim 80. Perzentil. Arup hielt selbst diese Werte für wahrscheinlich zu niedrig, weil die Linie noch nicht einmal das Stadium des Outline Business Case erreicht hatte („Tram Line 2 had not yet even reached the outline business case stage“), in dem die Risiken noch höher sind. Der Aufsatz berichtet die Prognose; was die Linie tatsächlich kostete, berichtet er nicht.

Welche früheren Fälle gehören in die Klasse?

§ 4 von 8

Die Klasse ist der schwierige Teil. In der Geschichte vom Lehrplan war sie leicht zu finden; andere Fälle sind weniger eindeutig („Other cases are more ambiguous.“), schreiben Kahneman und Lovallo². Für ein Unternehmen, das die Kosten einer Investition in eine ungewohnte Technologie schätzt, fragen sie, ob die Klasse aus den eigenen jüngeren Vorhaben mit neuen Technologien bestehen soll oder aus den ähnlichsten Entwicklungen anderer Unternehmen. Ihre Antwort: Keine der beiden ist vollkommen, und man sollte beide ausprobieren („Neither is perfect, and the recommendation would be to try both“). Unter Umständen braucht die Klasse außerdem eine Einheit, die Fälle unterschiedlicher Größe vergleichbar macht; als Beispiel nennen sie das Verhältnis der tatsächlichen zu den geplanten Ausgaben („The ratio of actual spending to planned expenditure is an example of a convenient unit that permits meaningful comparisons across diverse projects.“).

Flyvbjergs Klassen entstanden durch Tests. Projekte kamen in dieselbe Referenzklasse, wenn Tests keine signifikanten Unterschiede zeigten und ähnliche Regulierungs- und Bauregime galten („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.“). Nach dieser Regel landeten Schienenprojekte in Europa und Nordamerika in derselben Klasse, während Schienenprojekte in Entwicklungsländern und in Japan ausgeschlossen wurden. Er argumentiert, dass sich Projekte statistisch ähnlicher sind, als ihre Planer annehmen, und dass die eigentliche Schwierigkeit in den Daten liegt: in einem belastbaren Datensatz, der eine verlässliche Prognose erlaubt („The real challenge in doing a reference class forecast lies in assembling a valid dataset that will allow a reliable forecast.“). Seine eigenen Klassen waren nur möglich, weil jahrelange Forschung zu Kostenüberschreitungen bei Verkehrsprojekten einen solchen Datensatz hervorgebracht hatte.

In der einzigen Studie dieses Leitfadens, die eine Klasse aus dem Gedächtnis abfragte, fiel sie klein und erfolgslastig aus. Lovallo et al. (2012)³ befragten 38 Beschäftigte von Private-Equity-Gesellschaften zu einer Investitionsmöglichkeit, die sie gerade prüften, oder zu einer jüngeren Investition in ihrem Portfolio. Alle beschrieben zunächst den wahrscheinlichsten Verlauf der Investition und nannten die erwartete Rendite, benannten dann zwei Kategorien ähnlicher Investitionen, sollten in jeder die fünf relevantesten Beispiele aufzählen und schätzten die durchschnittliche Rendite jeder Kategorie. Von den 27,

die ähnliche Investitionen auflisteten, nannten 10 nur eine Kategorie; nur 14 der 33 verwertbaren Teilnehmenden führten in jeder Kategorie fünf Projekte auf; und die große Mehrheit der Referenzprojekte wurde als Erfolg beschrieben, was der Fragebogen auch für Investitionen zuließ, deren Erfolg die Befragten erst erwarteten: Er fragte, ob ein Beispiel ein Erfolg oder ein Misserfolg war oder nach Einschätzung der Befragten sein werde („was, or you believe will be, a success or failure“). Die Autoren nennen diese Ergebnisse beunruhigend: Sie deuteten darauf hin, dass die Methode selbst anfällig für Verzerrungen bei der Erinnerung an Referenzprojekte ist („is itself open to bias in the recollection of“). Wer aus dem Gedächtnis auswählt, warnen sie, greift wahrscheinlich zu wenige Fälle und zu günstige Analogien heraus („analogies that are too favorable“).

Wie viele Fälle, und wie gewichtet? In ihrer zweiten Studie prognostizierten Lovallo et al.³ das Einspielergebnis von 19 Filmen an den US-Kinokassen, wobei die Ähnlichkeitsurteile vor dem Sommerstart erhoben wurden. Auf Grundlage einer Datenbank von 1.751 Filmen, die zwischen 1993 und 2004 in den USA angelaufen waren, beurteilten 358 Kinogänger, wie ähnlich jeder Zielfilm früheren Filmen war, die sie kannten. Die als am ähnlichsten beurteilten Filme, je Zielfilm zwischen 1 und 15 (Tabelle 1), bildeten die Klasse jedes Zielfilms und wurden nach ihrer Ähnlichkeit gewichtet; die Autoren nennen das ähnlichkeitsbasierte Prognose (similarity-based forecasting). Ihr mittlerer absoluter relativer Fehler lag bei 30,09 Prozent, gegenüber 62,42 Prozent für eine Regression auf Filmmerkmale, die an den 1.751 früheren Filmen geschätzt war, und 66,36 Prozent für dieselbe Regression mit Interaktionen (Tabelle 2). Nur den ähnlichsten einzelnen Film heranzuziehen, schnitt weit schlechter ab: Der Text nennt einen mittleren Fehler von 135 Prozent, die in Tabelle 2 aufgeführten Fehler ergeben im Mittel rund 151 Prozent. Zwei Punkte relativieren das Ergebnis. Der Schritt, der Ähnlichkeit in Dollar übersetzte, war eine Regression, die an den Einspielergebnissen der 19 Zielfilme selbst geschätzt wurde, mit der Zahl der Kinos in der Startwoche und einem Merkmal für Fortsetzungen, und $R^2 = 0,55$ für diese Stichprobe („for the sample“), während die Vergleichsregression an den früheren Filmen geschätzt wurde. Und nach den Fehlern in Tabelle 2 lag die ähnlichkeitsbasierte Prognose bei 15 der 19 Filme näher am Ergebnis, wo der Text 17 nennt. Nach der Lesart dieses Leitfadens passt die Studie zum Argument der Autoren für mehrere gewichtete Fälle statt eines einzigen, ist aber für keinen der beiden Vergleiche ein Test außerhalb der Schätzstichprobe.

Hat der Blick auf die Klasse die Prognosen verändert?

§ 5 von 8

In der Private-Equity-Studie ja, und zwar in eine Richtung. Die mittlere erwartete Rendite der geprüften Investitionen lag bei 29,35 Prozent, gegenüber 18,72 und 20,82 Prozent für die beiden Kategorien ähnlicher Investitionen ($p < 0,001$). Von den 17 Teilnehmenden, deren erwartete Rendite über beiden Kategoriemitteln lag, senkten 14 sie, als sie gefragt wurden, ob sie sie anpassen würden; die mittlere Korrektur, die die Autoren für diese Gruppe angeben, betrug 5,96 Prozentpunkte und nach ihrer Angabe mehr als ein Fünftel der erwarteten Rendite. Niemand korrigierte nach oben, und wer mit seiner Investition zwischen oder auf den Kategoriemitteln lag, änderte nichts. Die Autoren nennen den Abstand zwischen den Investitionen und ihren Kategorien einen starken Hinweis („strongly indicative“), aber keinen schlüssigen Beleg („not conclusive evidence“) dafür, dass Prognosen aus der Innensicht optimistischer sind als solche aus der Außensicht; als Grund nennen sie, dass die Teilnehmenden ihre aktuellen Projekte als besonders attraktiv ansahen.

Das Design begrenzt, was die Studie zeigen kann. Alle Teilnehmenden beantworteten zuerst die Fragen zur Innensicht und danach die zur Außensicht; keine Gruppe ließ die Außensicht aus, und der Aufsatz berichtet keine Renditen, die die Investitionen tatsächlich erzielten. Nach der Lesart dieses Leitfadens zeigt die Studie, dass sich Prognosen in Richtung der Klasse bewegten, als die Befragten gebeten wurden, eine solche zu betrachten; ob die korrigierten Prognosen genauer waren, kann sie nicht zeigen.

Auch Flyvbjerg berichtet keinen eigenen Genauigkeitstest. Da die ersten Referenzklassenprognosen 2004 und 2005 erstellt wurden, sei es zu früh für endgültige Schlüsse über die Wirkung dieser Praxis („it is too early to draw definitive conclusions about the impacts of adopting the practice of reference class forecasting“); er beobachtet, dass die Methode in Großbritannien das Bewusstsein für optimistische Verzerrung und strategische Falschdarstellung in der Verkehrsplanung geschärft hat. Seine Aussage, die Methode sei nachweislich genauer als herkömmliche Prognoseverfahren („has been demonstrated to be more accurate than conventional forecasting methods“), stützt sich auf experimentelle Arbeiten von Kahneman und anderen, die er zitiert und die dieser Leitfaden nicht gelesen hat. Die eigene Aussage von Kahneman und Lovallo², die Außensicht liefere bei gleich klugem und geschicktem Einsatz beider Methoden weit eher eine realistische Schätzung („when both methods are applied with equal intelligence and

skill, the outside view is much more likely to yield a realistic estimate“), folgt aus der oben dargestellten Argumentation, nicht aus einem gemessenen Ergebnis, und dasselbe gilt für ihre Empfehlungen: Fehler intuitiver Vorhersagen ließen sich manchmal verringern, indem man die Außensicht einnimmt, die das Ergebnis vorhersagt, ohne seinen Verlauf vorherzusagen („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“), und die Außensicht werde die optimistische Verzerrung im Allgemeinen verringern („will generally reduce the optimistic bias“).

Warum sträuben sich Planer gegen die Außensicht?

§ 6 von 8

Die wichtigste Beobachtung von Kahneman und Lovallo² ist psychologisch: Beim intuitiven Prognostizieren wird die Innensicht weit überwiegend bevorzugt („the inside view is overwhelmingly preferred in intuitive forecasting“). Die Innensicht gilt als ernsthafter Versuch, dem einzigartigen Fall gerecht zu werden, und die Außensicht wird abgelehnt, weil sie sich auf grobe Analogien zu oberflächlich ähnlichen Fällen stütze („for relying on crude analogy from superficially similar instances“). Ärzte und Juristen, so die Autoren, wenden sich oft gegen statistisches Denken über einzelne Fälle, und dort hat die Vorliebe für die Innensicht beinahe moralischen Charakter („almost bears a moral character“). Bewusst auszublenden, was einen Fall einzigartig macht, kann verantwortungslos wirken („can appear irresponsible“), und das Lehrplanteam stieß auf eine weitere Hürde: Sich den Tatsachen zu stellen, kann unerträglich entmutigend sein („facing the facts can be intolerably demoralizing“). Die Autoren bezweifeln zudem, dass sich Realismus immer lohnt. Gestützt auf Arbeiten, die sie zitieren, halten sie fest, dass Optimismus beim Durchhalten helfen kann und dass ein erfolgreiches Bemühen um realistischere Einschätzungen in einer Organisation, die sich mit unbegründetem Optimismus vor Lähmung schützt, mehr schaden als nützen könnte („could do more harm than good in an organization that relies on unfounded optimism to ward off paralysis“). Sie sprechen von einem echten Dilemma, das sich keiner einfachen Regel fügt („a genuine dilemma that will not yield to any simple rule“).

Flyvbjerg unterscheidet die Situationen nach ihrer Ursache. Wo die optimistische Verzerrung den Fehler treibt, lässt sich annehmen, dass Prognostiker ehrliche Fehler machen, und die Aussichten für die Außensicht und die Referenzklassenprognose sind gut („the potential for using the outside view and reference class forecasting will be

good“). Wo die strategische Falschdarstellung ihn treibt, etwa wenn Städte um nationale Fördermittel konkurrieren, ist das Potenzial der Methode gering („the potential for reference class forecasting is low“), weil die Ungenauigkeit gewollt ist, und Anreize müssen so gesetzt werden, dass sie genaue Prognosen belohnen und ungenaue bestrafen („incentives must be aligned to reward accurate forecasts and punish inaccurate ones“): Zuschüsse, die bei der Kostenschätzung zum Zeitpunkt der Bauentscheidung gedeckelt sind, eine Finanzierung, die an die Anwendung der Methode geknüpft ist, und eine Prüfung durch Banken und unabhängige Stellen. Zwei Einschränkungen fügt er hinzu. Eine Klasse kann Extremergebnisse verfehlen, die außerhalb aller historischen Vorläufer liegen („may fail to predict extreme outcomes, that is, those that lie outside all historical precedents“). Und hohe Zuschläge schaffen Reserven, die zum Ausgeben verleiten können („Reserves will be spent simply because they are there, as many believe in the construction business.“); die Zuschläge brauchen daher straffe Verträge und Kostenkontrolle an ihrer Seite.

Welche Felder machen eine Referenzklassenprognose überprüfbar?

§ 7 von 8

Sechs, in einer Zeile je Prognose: die Entscheidung, die die Zahl stützen soll; die Schätzung aus der Innensicht, wie der Plan sie nennt, und wie sie zustande kam; die Referenzklasse, mit der Herkunft ihrer Fälle und der Regel, nach der sie aufgenommen wurden; die Verteilung der Klasse in einer Einheit, die Fälle unterschiedlicher Größe vergleichbar macht, etwa Ist zu Plan; die Belege dafür, diesen Fall über oder unter der Mitte der Klasse einzuordnen; und die Prognose mit der Sicherheit, die die Entscheidung braucht. Die Felder folgen Flyvbjergs drei Schritten und seiner Regel zu niedrigeren und höheren Zuschlägen, den Hinweisen von Kahneman und Lovallo² zu Klassen und Einheiten und dem Befund von Lovallo et al.³ zu Klassen aus dem Gedächtnis.

Abbildung 2 Ein Blatt für eine Referenzklassenprognose

Planzahl, Klasse, Verteilung und Position festhalten, bevor die korrigierte Zahl in ein Budget eingeht.

Entscheidung	Innensicht	Klasse	Verteilung	Position	Prognose
Was soll diese Zahl entscheiden?	Planzahl, wie entstanden?	Welche früheren Fälle, wie gewählt?	Ist zu Plan, welche Perzentile?	Belege, dass der Fall abweicht?	Zahl bei nötiger Sicherheit?

Das Perzentil ablesen, das die Entscheidung braucht. Den Fall nur mit Belegen verschieben, dass dieses Team besser oder schlechter plant als die Klasse.

Eine Zeile je Prognose. Die Planzahl neben der korrigierten Zahl stehen lassen. Wurde die Klasse aus dem Gedächtnis statt aus Aufzeichnungen gebildet, das im Feld Klasse vermerken.

Quelle Entscheidungsblatt des Autors. Das Feld zur Innensicht folgt Kahneman und Lovallo (1993), S. 25; die Felder zu Klasse und Verteilung folgen Flyvbjerg (2008), S. 8, und Kahneman und Lovallo (1993), S. 25; das Feld zur Position folgt Flyvbjerg (2008), S. 16; das Feld zur Prognose folgt Flyvbjerg (2008), S. 12 bis 16; der Hinweis auf Klassen aus dem Gedächtnis folgt Lovallo et al. (2012), S. 500 und 509. Das Feld zur Entscheidung und die Übertragung auf Umsätze stammen von diesem Leitfaden. Die Felder sind Eingaben der Lesenden; es werden keine Unternehmensdaten geliefert.

Ein Beispiel ohne Unternehmensdaten: Ein Softwareanbieter plant ein Zusatzmodul, und sein Produktteam prognostiziert den Auftragseingang im ersten Jahr aus dem Plan, aus den Zielkunden, dem erwarteten Anteil der Käufer und dem Preis. Der Eintrag hält diese Zahl als Innensicht fest. Die Klasse bilden die früheren Zusatzmodule des Anbieters und, soweit die Daten zu bekommen sind, vergleichbare Einführungen anderer Unternehmen; Kahneman und Lovallo² raten, beides zu versuchen. Die Einheit ist der tatsächliche Auftragseingang im ersten Jahr geteilt durch die Zahl im Plan jeder Einführung, sodass sich Einführungen unterschiedlicher Größe vergleichen lassen. Beim Umsatz liegt das Risiko auf der anderen Seite als bei einer Kostenüberschreitung, daher ist die strengere Zahl ein niedriges Perzentil der Klasse: Erreichten vier von fünf früheren Einführungen

mindestens das 0,6-Fache ihres Plans, nennt der Eintrag das 0,6-Fache der Planzahl als den Wert, den vier von fünf erreichten. Der Median passt zu einem Portfolio von Einführungen, deren Fehlbeträge und Überraschungen sich ausgleichen können; eine einzelne Einführung, deren Fehlbetrag das Budget nicht auffangen kann, braucht die strengere Zahl, wie Flyvbjerg für Kosten argumentiert. Die Sicherheit, die die Entscheidung braucht, bestimmt, welches Perzentil abgelesen wird. Das Feld Position verschiebt den Fall innerhalb der Klasse, und zwar nur mit Belegen: näher an die Planzahl, wenn die früheren Einführungen des Teams näher an ihren Plänen lagen als die in der Klasse, weiter darunter, wenn sie ihre Pläne deutlicher verfehlten, so wie Flyvbjerg für Kosten niedrigere Zuschläge zulässt und höhere verlangt. Umfasst die Klasse nur eine Handvoll Einführungen, vermerkt der Eintrag das, denn der zweite Schritt braucht genug Fälle für statistisch belastbare Schlüsse, und eine aus dem Gedächtnis gebildete Klasse kann klein und erfolgslastig ausfallen, wie in der Fragebogenstudie von Lovallo et al.³ Planzahl und korrigierte Zahl nebeneinander festzuhalten, ist der Eintrag, den der Leitfaden zum Forecast-Override für jede Änderung von einer Basisprognose zu einer Endprognose beschreibt, und der spätere Istwert zeigt, ob die Änderung die Prognose verbessert hat, der Vergleich, den der Beitrag zu Forecast Value Added anstellt. Das sind Beispiele des Autors, sie sind nicht geprüft.

Was lassen die drei Arbeiten für die Umsatzprognose eines Unternehmens ungeprüft?

§ 8 von 8

Das meiste, was ein Umsatzteam wissen möchte. Kahneman und Lovallo (1993)² ist ein konzeptioneller Aufsatz: Er begründet die beiden Sichtweisen mit psychologischer Forschung, einem Vorfall und Studien anderer und testet die Außensicht selbst nicht. Flyvbjerg (2008)¹ beschreibt die Methode und ihre ersten Anwendungen in der britischen Verkehrsplanung und berichtet keinen Test, ob diese Prognosen genauer waren. Lovallo et al. (2012)³ führten zwei Studien durch: In der einen bewegten sich die Prognosen von Private-Equity-Beschäftigten auf eine selbst erinnerte Klasse zu, ohne dass Ergebnisse beobachtet wurden; in der anderen prognostizierten Ähnlichkeitsgewichtete Klassen das Einspielergebnis von 19 Filmen mit geringerem Fehler als eine Regression, mit dem oben genannten Vorbehalt zum Schätzschrift. Keine der drei Arbeiten prognostiziert die Vertriebspipeline, den Auftragseingang oder die Abonnementumsätze eines Unternehmens; am nächsten kommen die Einspielergebnisse von Filmen.

Nach der Lesart dieses Leitfadens lässt sich eines mit einiger Zuversicht übertragen, ein Zweites als Warnung. Eine Prognose, die allein aus dem Plan entsteht, neigt zu Optimismus, wenn das Vorhaben nicht zur Routine gehört: Kahneman und Lovallo² begründen es, und Flyvbjergs Datenbank zeigt, dass Verkehrsprojekte im Mittel teurer wurden als prognostiziert, wobei er für diese Projekte der strategischen Falschdarstellung das größere Gewicht gibt. Und eine aus dem Gedächtnis gebildete Referenzklasse kann klein und erfolgslastig ausfallen: So war es in einer Fragebogenstudie mit Private-Equity-Beschäftigten (33 verwertbare Antworten), wobei als Erfolg auch gelten konnte, was die Befragten erst erwarteten, und die Autoren halten es für wahrscheinlich, dass eine Auswahl aus dem Gedächtnis ebenso ausfällt. Welche Klasse und welche Einheit zu einem Unternehmen mit wenigen vergleichbaren Einführungen passen, wie stark eine Klasse eine Umsatzprognose verbessert und wie sich Prognostiker verhalten, sobald eine Klasse vorgeschrieben ist, sagen die Quellen nicht. Kahneman und Lovallo zählen Absatzprognosen für bestehende Produktlinien zu den wiederkehrenden Problemen, bei denen Organisationen bereits für Disziplin sorgen können; nach dieser Lesart hat die Außensicht ihren Platz bei Zahlen für Vorhaben, die keine Routine sind.

 Ende des Beitrags

Referenzen

Zitate stehen in der Sprache des Originals; die deutschen Wiedergaben im Fließtext sind sinngemäß und stammen vom Autor.

- 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>

Zitierweise

Isoglu, S. (2026, 18. Oktober). Was ist eine Referenzklassenprognose? Sinan Isoglu.

<https://isoglu.com/de/insights/journal/was-ist-referenzklassenprognose/>

Online-Fassung:

<https://isoglu.com/de/insights/journal/was-ist-referenzklassenprognose/>

Sinan Isoglu, MBA (Quantic)

Führungskraft für Umsatzwachstum, Dozent und Doktorand

© 2026 Sinan Isoglu · isoglu.com · Alle Rechte vorbehalten