REVENUE OPERATIONS & AI

Forecast accuracy can hide offsetting errors

A forecast can be unbiased on average while missing every period: keep signed error, absolute error, distribution, horizon, and decision cost separate.

Sinan Isoglu, MBA (Quantic)
Commercial growth leader, lecturer and doctoral researcher

PUBLISHED
1 September 2026

READING TIME
10 min

WORDS
2,237

https://isoglu.com/insights/journal/forecast-accuracy-can-hide-offsetting-errors/

CONTENTS

MANAGEMENT SUMMARY

A forecast can have zero average signed error and still miss the actual outcome in every period. That is not a paradox; it is cancellation. This data note defines signed error, absolute error, bias, dispersion, distribution, horizon, and denominator as separate review objects. A four-period synthetic example shows how opposite errors disappear inside a mean. Fildes, Goodwin and De Baets report a bounded demand-planning corpus in which accuracy and bias are distinct and results vary across datasets and units. Fildes, Goodwin, Lawrence and Nikolopoulos show why adjustment direction and size also deserve separate inspection. The article turns those boundaries into a row-level audit. It does not offer a universal forecast formula, a B2B benchmark, or private operating diagnosis.

Keywords: Forecast accuracy · Forecast bias · Forecast error · Forecast Value Added · Sales forecasting

A forecast can be right on average and wrong in every period.

That sentence sounds contradictory because forecast accuracy is often treated as one score. But a score is the end of an aggregation process. If positive and negative errors are added together before anyone looks at the rows, they can cancel. The resulting zero can be mathematically correct and operationally misleading.

The short answer is precise: a zero average signed error shows no net direction in the selected rows; it does not show that the forecast was close to the actual outcome. To understand a forecast, keep the signed error, its absolute magnitude, its distribution, its horizon, its denominator, and the decision it supports as separate objects.

Why can aggregate forecast accuracy hide offsetting stage errors?

Begin with a convention. In this article, the signed forecast error is:

forecast error = forecast - actual outcome

An underforecast is therefore negative and an overforecast is positive. Another convention can be valid, but the sign must be declared before the result is interpreted. The following four periods are synthetic. They are designed to expose the aggregation problem, not to represent a company or a market.

Table 1 The forecast-error cancellation trap

Period	Actual outcome	Forecast	Signed error	Absolute error
1	100	80	-20	20
2	100	120	+20	20
3	100	80	-20	20
4	100	120	+20	20
Mean	100	100	0	20

Author's synthetic example. Values are illustrative and do not describe an observed forecast set.

The signed errors sum to zero: -20 + 20 - 20 + 20 = 0. The mean signed error is therefore zero as well. If a team calls that result unbiased, the statement is limited and technically correct. The forecast had no net tendency to be too high or too low across these rows.

The absolute errors tell a different story. Every period is wrong by 20 units. The mean absolute error is 20, not zero. The table has no correct period to hide behind. The positive and negative signs describe direction. Removing the signs reveals magnitude.

That distinction is the first release gate for a forecast review. A directional summary can answer whether the process tends to overshoot or undershoot. It cannot answer how close each forecast was. Those are different questions.

One word called accuracy hides several measures

Forecast teams often use accuracy as a convenient label for a selected error measure. The label becomes dangerous when the measure is not named. The following objects should stay visible:

Table 2 One word called accuracy hides several measures

Object	What it answers	What it can hide
Mean signed error	Is the selected set directionally high or low?	Opposite errors and the size of each miss
Mean absolute error	How large is the typical miss in the selected unit?	Whether errors are high or low
Root mean squared error	How strongly do large misses affect the summary?	The reason for an extreme miss and the ordinary case
Median absolute error	What is the middle miss after magnitude is taken?	A small number of severe tail errors
Quantiles or tail shares	How is error distributed across cases?	The causal reason for the distribution
Error by horizon or segment	Where does the process behave differently?	The aggregate average across unlike objects

Table from this essay. Sources and interpretation are given in the article.

None of these is the single true measure. Each is a lens. The correct lens depends on the decision. A capacity decision may care about the upper tail. A staffing decision may care about persistent bias. A cash decision may care about the timing and size of misses. A forecast used to allocate scarce inventory may value an underforecast differently from an overforecast.

The word accuracy should therefore be followed by a definition: accuracy according to which error, over which horizon, for which object, with which denominator, and for which decision?

Bias is not the same as accuracy

Bias is directional. Under the convention above, a positive mean signed error indicates that forecasts are high on average; a negative mean indicates that they are low on average. An unbiased mean can be created by alternating errors, as the synthetic table shows.

Accuracy is about closeness under a chosen measure. A forecast can have low bias and poor accuracy if its positive and negative errors are large. It can have modest accuracy but noticeable bias if most errors are small and point in one direction. The two measures can move together, but they do not contain the same information.

This is not only a mathematical warning. It changes how a process should be reviewed. If bias is the only reported metric, a team may celebrate cancellation. If absolute error is the only reported metric, a team may miss a systematic direction that creates inventory, capacity, or cash exposure. The review needs both the sign and the magnitude before it chooses a response.

The denominator can move the result

Percentage error adds another choice. A common form divides the error by the actual outcome. That may be reasonable when the actual is a stable positive quantity. It becomes unstable when the actual approaches zero, equals zero, or differs in scale across items. A 20-unit miss on an actual of 100 is not the same percentage as a 20-unit miss on an actual of 1,000.

Other reviews divide by the forecast, an average of forecast and actual, a cohort total, or a revenue base. Each denominator creates a different comparison. The result may be useful, but it must be named.

The same problem appears when a forecast moves from units to revenue, from a weekly horizon to a quarter, or from a product-level object to an account-level aggregate. Addition is not neutral when the rows carry different prices, exposure, or timing. A total revenue forecast can look stable while a set of important items carries large offsetting errors underneath.

Before comparing two accuracy summaries, record:

1. the forecast object and unit;
2. the issue date and outcome horizon;
3. the sign convention;
4. the denominator, including zero handling;
5. the inclusion and exclusion rules;
6. the treatment of cancellations, substitutions, and missing actual outcomes.

Without that metadata, a change in the number may be a change in the measurement frame rather than a change in forecasting performance.

Distribution matters before the mean

The synthetic table has the same error magnitude in every period. Real error sets are usually less tidy. Consider two four-period sets under the same sign convention:

Table 3 Distribution matters before the mean

Set	Signed errors	Mean signed error	Mean absolute error	Shape
A	-20, +20, -20, +20	0	20	Repeated moderate miss
B	-40, 0, 0, +40	0	20	Two exact periods and two larger misses

Table from this essay. Sources and interpretation are given in the article.

Both sets have the same mean signed error and the same mean absolute error. They do not carry the same operating risk. Set A misses continuously. Set B is quiet twice and then misses more severely twice. A review that only reports the two means cannot distinguish the service, capacity, or escalation pattern.

The next check is therefore distribution. Inspect the median, upper quantiles, tail share, and the rows that generate the tails. Then split by horizon, product, segment, geography, or process stage when those categories are part of the forecast object. The split should be decided before looking for a preferred story. Otherwise segmentation becomes another way to select the number that sounds best.

What the empirical studies actually show

Fildes et al. (2025) analyze 147,131 forecasts and actuals from 10 organizations and 22 business units. Their evidence is organized as six datasets of one-step-ahead system forecasts, final forecasts, and actual outcomes. The reported overall medians show 51.5% of SKUs with improved Forecast Value Added and 55.6% with improved bias. The result is useful for this article because the study treats accuracy and bias as distinct dimensions and reports variation across datasets and organizations.

It does not mean that 51.5% is a target for every forecasting process. It does not mean that bias is an adequate substitute for accuracy. It does not mean that a commercial team can transfer the result to a pipeline without testing the object, horizon, information set, and outcome definition.

The earlier Fildes et al. (2009) field study analyzes 68,984 complete forecast triples across four companies. It reports that upward adjustments were wrong-signed 66% of the time in the manufacturers and 83% at the retailer, compared with 46% for downward adjustments. Those figures make adjustment direction worth recording. They do not turn a supply-chain result into a rule that every upward sales override is wrong or every downward override is useful.

Together, the studies support a disciplined measurement habit: preserve the baseline or prior forecast, the intervention if one exists, the actual outcome, and the error fields before a summary is used. The studies do not eliminate the need to declare the decision and the measurement boundary.

Do not let FVA answer a different question

The existing forecast value added process audit owns a different but adjacent object. It asks whether a defined intervention added value compared with a preserved baseline, using an actual outcome, a selected metric, and the cost of the review process. The existing forecast override article owns the judgmental forecast call and the accountability risk around moving a number.

This article begins before either conclusion. It asks whether the error rows were still visible when the process calculated its average. A final forecast can beat a baseline on mean absolute error and still hide a severe tail. A mean signed error can be zero for both baseline and final forecast while their error distributions differ materially. The FVA comparison is stronger when those underlying fields remain available.

That is the boundary between the routes. The FVA article is a process comparison. The override article is a judgment and governance problem. This article is an aggregation-order and error-distribution problem. They should link to each other, not compete to own the same phrase.

The row-level forecast-error audit

Before publishing a score, build a small audit table. It can be implemented in a spreadsheet, warehouse, or forecasting system, but the fields should remain inspectable.

Table 4 The row-level forecast-error audit

Audit field	Required question	Failure if omitted
Forecast object	What exactly was forecast: units, revenue, opportunities, capacity, or something else?	Unlike rows are silently combined
Baseline and final forecast	Was an intervention applied, and what was the untouched value?	FVA and override effects cannot be separated
Actual outcome	What event closed the forecast window?	Error is calculated against a moving target
Signed error	Which direction convention is in force?	Bias changes sign or becomes uninterpretable
Absolute or squared error	How large was the miss regardless of direction?	Cancellation is mistaken for accuracy
Horizon	How far ahead was the forecast issued?	Short and long forecasts are treated as identical
Denominator	What base creates the percentage comparison?	Ratios change without a visible measurement change
Segment and tail flag	Which rows create persistent or severe errors?	Aggregate averages erase the operational problem
Decision cost	What does an over- or underforecast make the organization do?	Metric improvement is mistaken for decision value

Table from this essay. Sources and interpretation are given in the article.

Run the review in this order:

1. freeze the row-level forecast and actual definitions;
2. declare the sign convention and denominator;
3. calculate signed, absolute, and at least one distributional view;
4. inspect horizons and segments before the aggregate is presented;
5. retain excluded or missing rows with a reason code;
6. compare a baseline with a final forecast when judgment or automation changes the number;
7. connect the metric to the decision cost rather than to a generic accuracy target.

This order makes a tidy average harder to manufacture accidentally. It also makes disagreement more useful. Two people can disagree about the right decision while still agreeing on which rows, definitions, and errors the decision rests on.

Three conclusions to resist

First, do not infer intent from bias. A positive average may result from a changed mix, a new segment, a different issue horizon, or missing low outcomes. The number flags a pattern. It does not identify the motivation.

Second, do not infer operational safety from low mean absolute error. A small average can hide a tail that matters to a capacity constraint or a major account. Tail analysis is not optional when the downside is concentrated.

Third, do not infer decision quality from metric improvement alone. A more accurate forecast may arrive too late, use a definition that does not match the decision, or improve a measure while worsening a service, cash, or resource outcome. The metric needs a declared use.

Boundary

The claim here is narrow: forecast accuracy cannot be interpreted responsibly until signed error, absolute error, distribution, horizon, denominator, and decision cost remain visible. The two cited studies provide bounded demand-planning and supply-chain evidence for separating accuracy, bias, and adjustment direction. They do not validate a universal B2B sales-forecast rule. A commercial transfer hypothesis would need its own baseline, event definition, repeated outcomes, and test design.

Fildes, R., Goodwin, P., & De Baets, S. (2025). Forecast value added in demand planning. International Journal of Forecasting, 41, 649–669. https://doi.org/10.1016/j.ijforecast.2024.07.006

Fildes, R., Goodwin, P., Lawrence, M., & Nikolopoulos, K. (2009). Effective forecasting and judgmental adjustments: An empirical evaluation and strategies for improvement in supply-chain planning. International Journal of Forecasting, 25(1), 3–23. https://doi.org/10.1016/j.ijforecast.2008.11.010

END OF ESSAY

HOW TO CITE THIS ESSAY

Isoglu, S. (2026, September 1). Forecast accuracy can hide offsetting errors. Sinan Isoglu. https://isoglu.com/insights/journal/forecast-accuracy-can-hide-offsetting-errors/

KEEP READING

Revenue operations & AI

What is a forecast override? Judgment is an intervention in the data

https://isoglu.com/insights/journal/what-is-forecast-override/

Revenue operations & AI

Forecast value added is a process audit

https://isoglu.com/insights/journal/forecast-value-added-is-a-process-audit/

Revenue operations & AI

What are forecast categories? Names need stable decision rules

https://isoglu.com/insights/journal/what-are-forecast-categories/

Sinan Isoglu, MBA (Quantic)
Commercial growth leader, lecturer and doctoral researcher

Online version: https://isoglu.com/insights/journal/forecast-accuracy-can-hide-offsetting-errors/