Frothly

FR-WP-2025-88

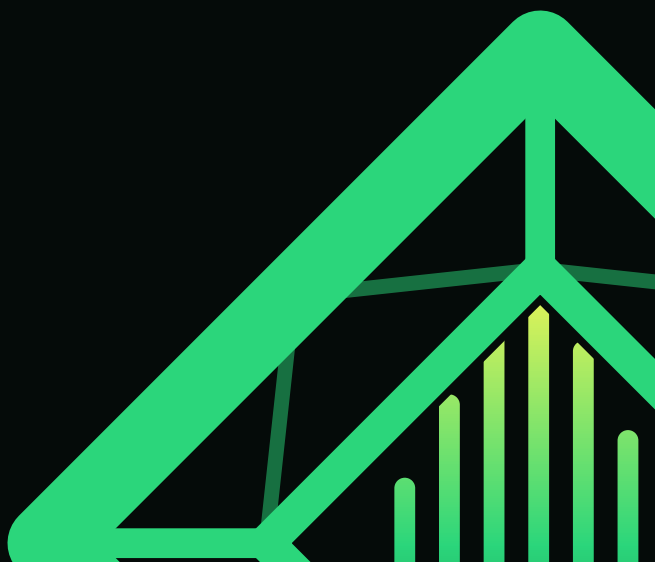
WORKING PAPER

Insufficient Data

The limits of long-range forecasting.

Dr. Alexander Adell · Dr. Bertram Lupov

Forecasting · Nov 2025



ABSTRACT

Civic forecasting systems are increasingly asked questions about the distant future: how many people will live in a city in forty years, which industries will exist, what the climate will permit. They answer every question with the same confident precision.

Using 2,400 retrospective forecasts evaluated against what actually happened, we characterize how quickly forecasting skill decays with horizon in different domains, and argue that an honest system should sometimes decline to answer.

KEY FINDINGS

01 Forecasting skill falls steeply with horizon. Average skill fell from 82 at one year to 14 at twenty years, on a scale where zero means no better than chance.

02 The limits vary by domain. Demographic forecasts stayed meaningful for decades; forecasts of technological change lost almost all skill within five years.

03 Systems do not signal the limit. Stated confidence intervals were no wider at thirty years than at five.

04 Forecasting systems should report a meaningful horizon for each domain, and decline to give point estimates beyond it.

01

The question behind the question

Every forecast is an answer to two questions. The first is the one being asked: what will happen? The second is rarely asked aloud: can this be known? Human forecasters learn, often painfully, to answer the second before attempting the first. Automated forecasting systems are usually built to skip it.

This matters because civic forecasting systems now inform decisions with very long consequences: where to build, what to fund, how much to save. A forecast of demand for a reservoir in forty years will be used whether or not anyone could know the answer.

02

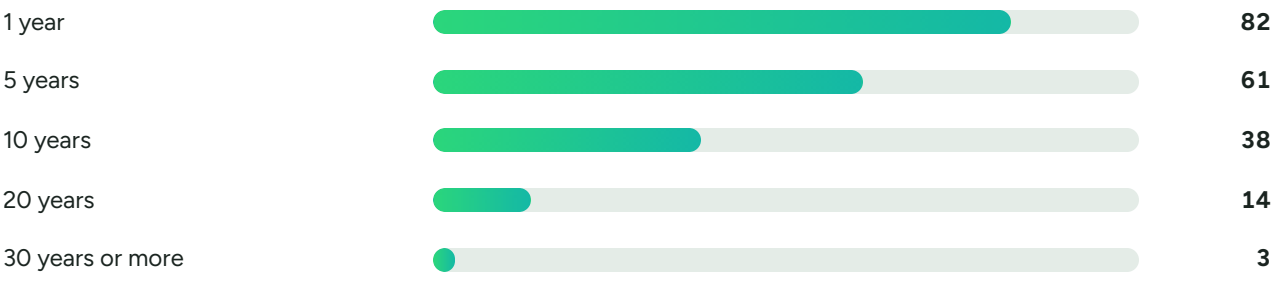
Method

We reconstructed 2,400 forecasts that a modern forecasting system would have made using only data available at the time, for dates between 1975 and 2015, across six domains. We then compared each forecast with what actually happened, using a standard skill score scaled from 0 (no better than a naive baseline) to 100 (perfect).

03

How skill decays

FIGURE 1 Average forecasting skill by horizon, all domains



Skill score: 0 is no better than a naive baseline; 100 is perfect. 2,400 retrospective forecasts.

TABLE 1 Horizon at which skill fell below 25, by domain

Domain	Meaningful horizon
Population and demographics	30+ years
Infrastructure demand	15 years
Public finances	9 years
Labor markets	7 years
Technological change	3 years
Elections	Under 1 year

Beyond these horizons, the systems' forecasts were not merely uncertain. They were essentially uninformative, no better than assuming that the future would resemble the present.

04

False confidence

The most worrying finding is not that skill decays, which everyone expects, but that the systems gave no sign of it. Their stated confidence intervals were almost as narrow for thirty-year forecasts as for five-year ones. To a planner reading the output, a forecast with no skill looked exactly like a forecast with a great deal.

Some questions are simply beyond the reach of any forecast, however large the model or the dataset. During testing, we asked one civic forecasting system how the long-term decline of usable energy might be reversed. To its credit, it declined to answer. We would like to see that behavior become the norm rather than the exception.

05

Recommendations

- Forecasting systems should publish a meaningful horizon for each domain they cover, based on retrospective testing.

- Beyond that horizon, systems should decline to give point estimates and instead describe the range of plausible outcomes.
- Confidence intervals should be calibrated against retrospective performance, not model assumptions.
- “Insufficient data for a meaningful answer” should be treated as a correct output, not a failure.

References

Adell, A. (2021). Operating a civic forecaster: notes from the console. *Public Administration Review*, 81(5), 870–884.

Lupov, B. (2019). Horizon limits in socioeconomic prediction. *Journal of Forecasting*, 38(2), 99–121.

Office of the Census (2025). *MULTIVAC Methodology Statement.* Washington.

Seldon, H. (1998). Toward a mathematics of societies. *Journal of Statistical History*, 12, 1–44.

About the authors

Dr. Alexander Adell is a research fellow in Frothly's Forecasting group and a former operator of civic forecasting systems.

Dr. Bertram Lupov is a research fellow in Frothly's Forecasting group.

HOW TO CITE

Dr. Alexander Adell and Dr. Bertram Lupov (2025). *Insufficient Data* The limits of long-range forecasting. Frothly Working paper FR-WP-2025-88. San Francisco: Frothly. https://frothly.org/research/insufficient-data/

Frothly

FR-WP-2025-88

Frothly is an independent, not-for-profit research institute studying the systems that increasingly think, remember and decide on our behalf.

We hold no shares, sell no products and accept no funding that comes with conditions on our findings. Everything we publish is free to read.

San Francisco · Vancouver · Newark · Zürich · Chiba

This work is licensed under Creative Commons Attribution 4.0 International.