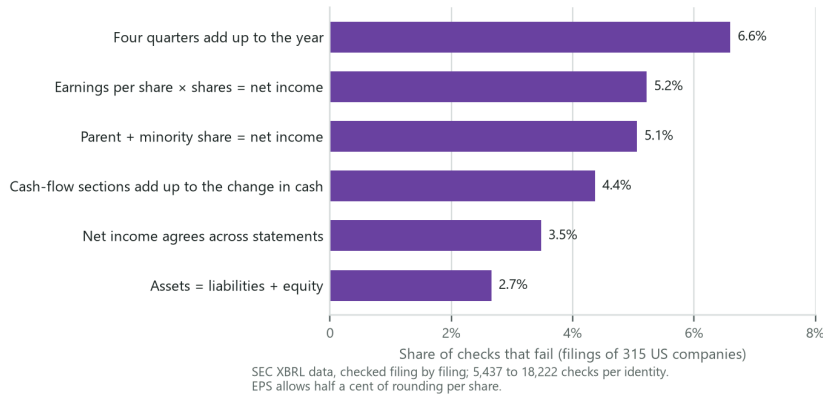


The Time Machine Problem

How much do AI forecasters remember instead of predict? A point-in-time SEC database, and a memory test for models with known training cutoffs.

How often SEC filings break basic accounting identities



THE QUESTION

Every AI investing product is tested by asking a model about the past. If the model read the answer in training, the backtest flatters it and live results disappoint. Measuring that takes data that knows exactly what was public, and when.

4.9M

reported numbers from 315 companies, each stamped with when it became public

7.6%

of those numbers later re-reported with a different value

6,868

company-quarters in the memory test, on eight models

WHAT I BUILT

- A point-in-time database of SEC XBRL filings for 315 US companies, sampled as of 2019 so later-delisted firms stay in. A number becomes visible when its filing was accepted; quarterly earnings, at the press-release date.
- Fourth quarters are never filed on their own. They're derived as the full year minus nine months, and match Apple's and Microsoft's press releases exactly.
- An audit of every filing against six accounting identities.
- A memory test: does a model give the true reported EPS more probability than nine plausible alternatives? Run on eight models with training cutoffs from 2016 to 2024.

WHAT I FOUND

1. SEC machine-readable financials are mostly consistent but not clean: 2.7% of balance sheets and 4.4% of cash-flow statements break their identities, and quarters fail to add up to the year 6.6% of the time.
2. 7.6% of numbers are later re-reported differently, so a backtest on today's data uses information nobody had at the time.
3. Small date-limited models (124 million to 1.55 billion parameters) don't memorize reported earnings at all: they score at chance before and after their cutoffs, while a control shows the test does catch well-known facts. The memory question has to be put to large commercial models.

HOW IT HOLDS UP

No look-ahead by construction: every query is answered as of a date, with the 4:00 pm filing rule. Quarter derivations are tested against hand-checked filings, and the estimators are tested on simulated data with planted answers.

NEXT

Forecasts from a commercial model with the company named versus masked, before and after its training cutoff, to separate memory from the effect of seeing a name; and a one-page checklist for backtesting with AI.