

Alternative Data and AI in Finance and Economics



Insights from a practitioner

Akhilesh Tilotia, CFA

XKDR Emerging Market Dialogues
March 30, 2026

What is alternative data?

Data beyond traditional financial statements and market prices that provides unique, often real-time, insights into economic activity

Traditional data

- Quarterly earnings reports
- Annual filings (10-K, AR)
- GDP, CPI, IIP (monthly/quarterly)

Alternative data

- Daily prices: mandi, cement, hotels, airfare...
- RTO-level vehicle registrations
- E-commerce pricing and reviews
- Conference call transcripts

Why it matters

- Faster signals (days, not months)
- Granular (district, RTO, product)
- Cross-sector insight linkages
- NowCasting capabilities

Alternate data today tends to become traditional data tomorrow

Get data faster

Why wait weeks for official releases when daily and weekly signals exist?

Electricity generation, coal inventory

Daily power generation data from Central Electricity Authority, coal inventory, power traded,...

Prices and quantities/SKUs

Daily commodity arrivals from eNAM and AgMarkNet mandis across India; deep-dive on e-commerce

Pollution statistics

Get data in 15-minute intervals to track how the plants are performing

"Bestsellers"

Daily insights across categories on what is gaining or losing market share

On Thurro: 2,000+ tables refreshed regularly...and gigabytes of text data

🔍 Get more granular data

From national aggregates to district, RTO, port, and molecule-level detail

Vehicle registrations by RTO

7.7M+ rows

Vahan data: maker, category, fuel type, EV vs ICE vs CNG adoption, tractor registrations, 3-wheeler logistics

Export data by port and molecule

24.7M+ rows

DGCIS: commodity exports by country, port, and HS code (2/4/6/8-digit granularity)

Customs at port level

600K+ daily records

ICEGATE: daily customs revenue and drawback pendency by EDI location

Pharma drug level data

190K+ rows

US CMS Medicare and Medicaid spending by drug - which molecules go to the US?

It is easier to deep down to find insights – but you need to know where and how to look



Cross-sector linkages for insights

Using data from one domain to generate insight for a completely different sector

**Power generation
(CEA, POSOCO daily)**



**Cement and steel
capacity utilization**

Higher electricity consumed by industrial states signals rising capacity utilization in heavy industry

**Temperature and rainfall data
(IMD district-level)**



**FMCG and agriculture;
consumer demand**

District rainfall patterns predict crop yields, rural income, and consumer spending

**Amazon and e-commerce
pricing and reviews**



**FMCG brand
market share**

Track 400M+ product data points to gauge brand popularity, pricing power, and competition

The value of alternative data multiplies when you connect datasets across domains

NowCasting: from data to prediction

Building models that predict current-quarter results before official numbers are released

How Thurro builds NowCasts

- 1 Identify early indicators** Select alternative data that historically correlates with the target (e.g., store locators to revenues)
↓
- 2 Build regression models** Train models on [5-10] years of historical data to establish relationships and coefficients
↓
- 3 Combine multi-source signals** Blend signals from different datasets with appropriate weights to create ensemble forecasts
↓
- 4 Validate and backtest** Test models on held-out data; measure accuracy against actual quarterly results from prior years
↓
- 5 Update and publish** Refresh nowcasts monthly as new data arrives; track prediction error to validate model fit

More than two-thirds of NowCast are within +/-2% of the actual final reading

The data landscape

Thurro covers 2,000+ datasets across sectors, regions, and time frequencies

Automobiles 10M plus Vahan (RTO), OEMs, EV, tractors	Energy and power 35M plus CEA, IEX, POSOCO, HPX daily	Agriculture 75M plus eNAM, AgMarkNet, IMD rainfall
Trade and ports 40M plus DGCIS, ICEGATE, MoCI exports	E-commerce 450M plus Amazon, Flipkart, Zomato, Swiggy	Real estate 2M plus Anarock, MagicBricks, JLL, hotels
Financial markets 20M plus BSE, NSE, MF portfolios, FPI	Macro economy 500K plus RBI, MOSPI, GST, inflation	Healthcare 300K plus US FDA shortages, CMS drug data
Credit markets 5M plus RBI bank credit, industry-wise exposure	Logistics and transport 50M plus NHAI toll data, FastTag transactions, freight	Commodities 10M plus Metals, fertilizers, energy, spot prices

....and many types of reports: annual reports, investor presentations, credit ratings, industry associations, etc.

Case study: alternative data for credit risk

A three-layer early warning system using Thurro datasets across corporate and retail credit

Real-time signals

Daily / Weekly

Credit rating actions (1M+ records)
MCA charge filings (new borrowing)
Commodity prices (cement, steel, oil)
Toll collections (FASTag daily)

Monthly leading indicators

Monthly

NACH bounce rates (EMI stress)
Vahan hypothecation data
GST e-way bill volumes
EPFO payroll by industry

Quarterly validation

Quarterly

Bank NPA by industry (30+ sectors)
NBFC product-wise sanctions
Capacity utilization surveys
Anarock unsold inventory

Corporate credit risk

Industry-wise bank NPA data (infra, textiles, metals, chemicals)

cross-referenced with sector proxies:

- toll traffic for roads; freight data
- power generation, imports, iridescence for energy
- production and dispatches for manufacturing; pollution
- railway freight for logistics, etc.

Retail credit risk

- NACH EMI bounce rates (return rates)
- Vehicle registration trends as proxy for location/consumption
- Vahan hypothecation, additions and terminations
- Credit card spend patterns
- UPI merchant category shifts

Case study: credit risk signals in detail

Four distinct alternative data sources that provide early warning before NPAs surface in quarterly filings

Vahan hypothecation data

25.7 lakh HP additions in 2025

Tracks new vehicle loans (HP addition), repayments (HP termination), and repossessions (fresh RC to financier). The termination-to-addition ratio fell from 4.3x in 2020 to 2.3x in 2026, showing credit cycle recovery. Fresh RC to financier rose 41% in 2025 — a repossession stress signal

NACH EMI bounce rates

184.9M mandate presentations (Feb 2026)

NACH debit data can reveal real-time retail EMI stress. Of 184.9M mandates presented in February 2026, 45.8M were returned — a 24.7% bounce rate. This is the single most powerful system-wide retail credit health indicator, available monthly

Credit rating actions

1M+ rating records across CRISIL, CARE, ICRA

Daily rating actions with upgrade/downgrade tracking, outlook migrations (stable to negative), and issuer-not-cooperating flags. The upgrade-to-downgrade ratio by industry serves as a real-time sector credit sentiment index

MCA charge filings

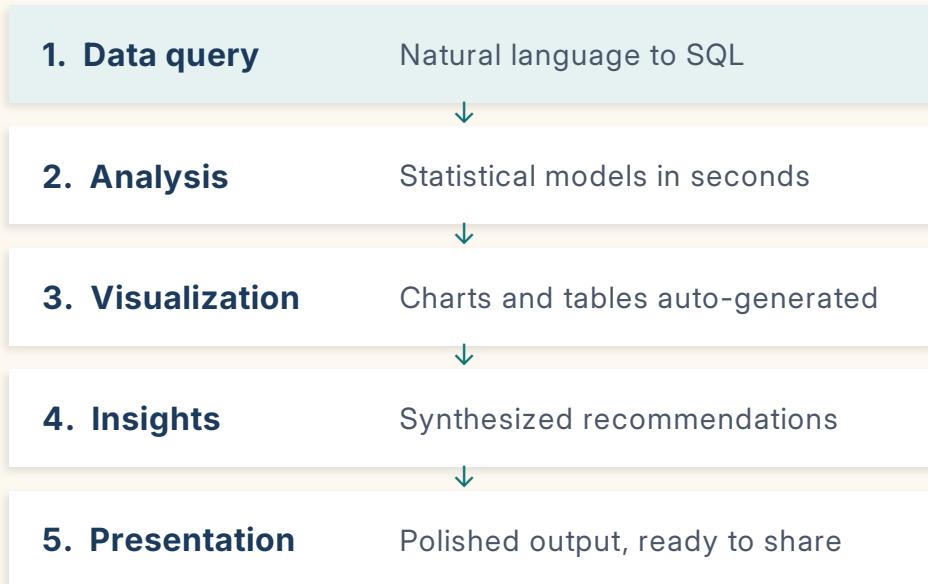
45,000+ charge records with lender details

Every bank loan creates a charge filing at MCA. Track new charge creation versus satisfaction by sector to build a real-time leverage thermometer. Repeated modifications to the same charge signal restructuring activity — an early stress indicator

AI: lowering the barrier to analysis

From data to analysis to presentation in minutes, but with critical guardrails.

The promise



The guardrails

Curated data quality

AI must access only verified, structured datasets, not the open internet

Step-by-step verification

Every calculation must be checked. AI can hallucinate numbers convincingly

Restricted data access

Limit AI to approved databases via MCP connectors, not free-form browsing

Human-in-the-loop

Misses can occur at every step. Domain expertise remains essential

The changing role of the analyst

When data feeds are live via MCP connectors, analysis that took days now takes minutes.

Yesterday's analyst







- Days to gather data from PDFs, websites
- Hours building Excel models manually
- Limited to 5-10 familiar datasets
- Linear workflow: collect, model, present
- Value equals execution speed

Tomorrow's analyst

- Knows hundreds of alternative datasets
- Understands cross-domain linkages
- Develops intuition for non-obvious signals
- Designs the question, AI executes
- Value equals insight quality and creativity

Fund managers can now do on-the-fly analysis via MCP-connected data feeds from MOSPI, Thurro, and others.

Key takeaways

-  Speed wins. Alternative data arrives daily or weekly, not months later. This is the era of nowcasting
-  Granularity reveals. RTO-level, port-level, molecule-level insights give actionable insights
-  Linkages multiply. Power data predicts cement, rainfall predicts spending, e-commerce data predicts brand wins
-  Credit risk gets a new lens. NACH bounce rates, vehicle hypothecation trends, rating actions, and MCA charge filings provide months of lead time over quarterly NPA data
-  AI accelerates but guardrails are non-negotiable. Locked data sources, step-by-step verification, and domain expertise beat open-ended AI every time
-  The analyst evolves. Tomorrow belongs to those who know 100s of datasets, understand interlinkages, and ask questions AI cannot

Thank you

thurro.com

326, DBS Center, 31 A Cathedral Garden Road, Chennai, 600034

BHive, 3rd floor, Inspire, BKC, Bandra East, Mumbai, 400051