Discussion of "Geoeconomic Pressure"

Clayton, Coppola, Maggiori, Schreger

FRBSF Macroeconomics and Monetary Policy Conference

Matteo Iacoviello
Federal Reserve Board

March 28, 2026

What the Paper Does

1. **Measurement:** Uses ≈300,000 earnings calls and analyst reports from firms globally to systematically measure geoeconomic pressures at the firm level
   - Covers tariffs, sanctions, export controls, and supply chain disruptions
   - Uses open-source LLMs (Llama 3) for scalable, reproducible inference

2. **Facts:** Uses the resulting data to uncover a set of novel stylized facts
   - Hegemons disproportionately target strategic chokepoints — sectors where the adversary has concentrated leverage
   - Sanctions and tariffs trigger systematic supply chain adjustments at the firm level — reshoring, supplier diversification, inventory builds

Overview of My Discussion

1. **Illustration:** What LLMs can do for this type of analysis — a simple example using US earnings calls and tariff sentiment in 2024-2025.

2. **From sentiment to investment:** What a follow-up prompt can reveal about firm responses

3. **Four suggestions** for the authors

Simple Illustration: Tariff Sentiment from Earnings Calls

Data: Earnings calls of US firms (S&P universe), 2024–2025

Step 1: Extract large snippets of text around tariff-related keywords

Step 2: Send each snippet to Claude Sonnet 4.5 with a simple prompt

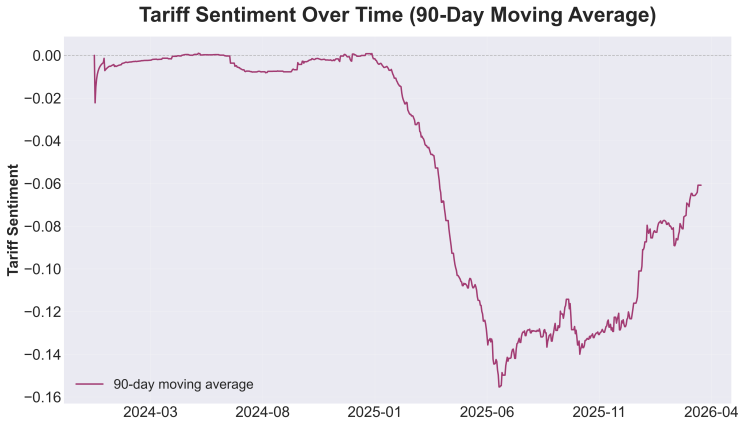
The prompt

"Companies in their earnings calls may talk about trade policy, import and export tariffs and barriers, customs and duties. You will be analyzing text from earnings calls that may mention topics related to trade tariffs, not complete transcripts.
Analyze this company's sentiment toward trade tariffs using a -1 to $+1$ scale, with values ranging all the way from -1.0 to 1.0 in 0.1 increments [...] Return 0 if content refers to non-trade tariffs (utility rates, electricity tariffs, etc.)."

Model: Claude Sonnet 4.5 — a "dense" model (uses all parameters at every inference step), not open-source, estimated to be 3–5$\times$ more complex than Llama 3.3

Cost: $\approx$ \$24 for 8K documents mentioning tariffs; $\approx$6 hours on a laptop via API.

Tariff Sentiment of US Corporations



90-day moving average of tariff sentiment scores across S&P earnings calls.
Closely tracks the aggregate time series in Clayton et al.

From Sentiment to Investment: A Follow-Up Prompt

Step 3: For calls with non-zero tariff sentiment, send text back with a second prompt asking about investment intentions

Investment Intentions

"Analyze planned or actual changes in capital expenditure, capacity expansion, new facilities, R&D spending **because of tariffs** *using a -1.0 to 1.0 scale (0.1 increments)*

-1.0*: Major cut or pause in investment relative to before*

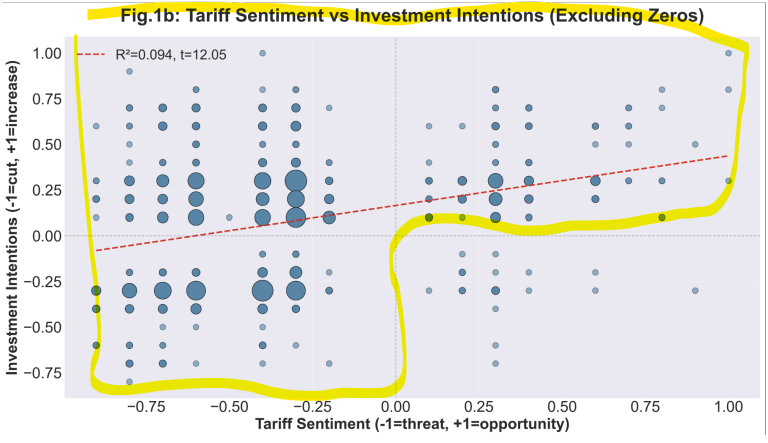
0.0*: Neutral/unclear/no specific investment plans mentioned*

$+1.0$*: Strong increase in investment because of tariffs*

Provide score and reasoning. Only score specific response of investment."

A single API call extracts both investment intentions, plus the reasoning behind each score.

Tariff Sentiment and Investment Intentions: Firm-Level



Each dot is a firm–quarter. Dots proportional to number of obs. Tariff optimism is positively correlated with investment intentions.

What the LLM Says About Why Firms Adjust Investment

Negative Tariff / Positive Intention

n = 182

- Supply chain diversification investments offset by cost discipline and efficiency focus
- Defensive repositioning through inventory builds and existing facility optimization, not expansion
- Maintaining pre-planned investments despite tariffs, not accelerating because of them

Positive Tariff / Positive Intention

n = 56

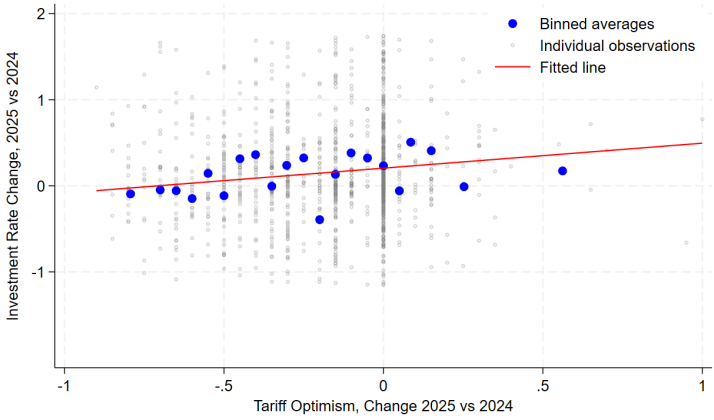
- Tariffs create reshoring opportunities; firms invest in domestic capacity expansion
- Strategic supply chain repositioning drives moderate infrastructure and inventory investments
- Competitive positioning through tariff advantages enables selective capacity and technology investments

Negative Tariff / Negative Intention

n = 111

- Tariff uncertainty triggers investment pauses and project delays pending policy clarity
- Cost pressures drive defensive actions: expense cuts, facility closures, inventory reductions
- Capital reallocation from growth to mitigation: supply chain shifts without capacity expansion

Validation: Text vs. Actual Investment Data

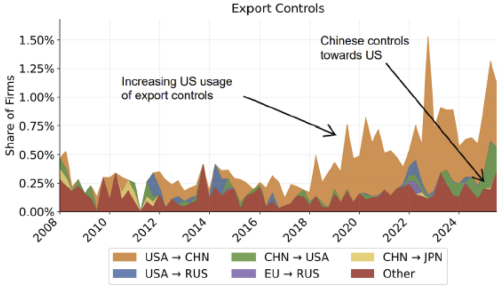


Source: Caldara, Ferrante, Iacoviello, Schott (in progress).
Text-based sentiment changes between 2024 and 2025 are strongly associated with changes in *actual* investment rates across firms.

Additional Validation: Sanctions and Export Controls

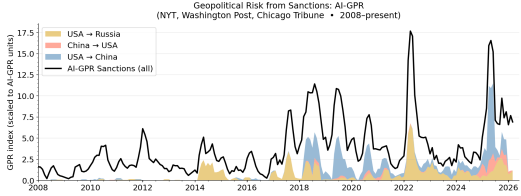
Independent text-based indices tell a consistent story

Export Controls Exposure
(Clayton et al., this paper)



Share of firms in earnings calls reporting exposure to export controls.

AI-GPR Sanctions Index
(Iacoviello–Tong)



Newspaper-based index of sanctions-related geopolitical risk, constructed from NYT, WaPo, and Chicago Tribune using GPT-4o-mini.

Both series spike around the same events — strong cross-validation across very different text sources and methodologies.

Suggestion 1: Fewer Questions

The authors' prompt asks 55 binary (0/1) questions per document.

- Is the goal to provide a comprehensive, real-time downloadable dataset? If so, this is very valuable, but is it all needed?
- For the paper: many 0/1 answers may not communicate as much as fewer, well-targeted questions scored on a richer scale
- Binary indicators lose information: "investment up" says nothing about why, how much, or with what confidence

Suggestion

Consider small set of continuous scores for the main outcomes, supplemented by narrative summaries
Simplify the 55-field schema
Feed the full transcript rather than the LLM-generated summary.

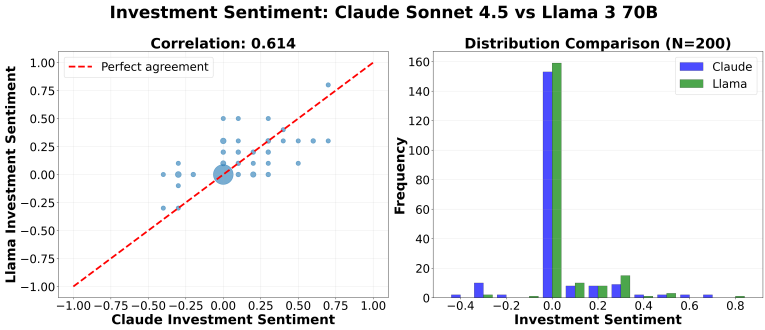
Suggestion 2: Make Data Accessibility Central

The citation count of the paper may depend in large part on making data available in an accessible form.

Is the goal to offer (1) aggregate time series, or (2) firm-level data too?

I think (2) is far more valuable, with the caveat that measurement error concerns are more salient when working with firm-level observations (and corporations may go after you).

Suggestion 3: Keep Track of Model Vintage



Simpler/smaller models may default to 0 more readily — they may lack the sophistication to recognize that "moving a facility" implies investment. Will newer, more capable models more likely assign non-zero scores?

However, when given the same prompt, models are in broad agreement, though there are puzzling differences (e.g. Llama seems not to see declines in investment).

Suggestion 4: Provide Cost Estimates, if Cost an Issue

- The authors use the Stanford Sherlock high performance computing cluster (8 NVIDIA A100-80GB GPUs + 4 H100-80GB GPUs) — a major infrastructure commitment
- But inference costs are declining rapidly: an equivalent analysis via commercial API would cost $\approx$ \$700 for the full corpus
- Not clear which infrastructure is cheaper / more accessible for replication purposes (hardware vs software)
- Clearer accounting would help researchers assess the trade-off between open-weight local inference (reproducible, confidential, perhaps large hardware cost) vs proprietary API-based inference (cheaper, accessible, proprietary, maybe gets better over time)

Suggestion 4 (cont'd): The Sherlock Facility

Sherlock High Performance Computing Cluster | Stanford Research Computing

Visit >