Geoeconomic Pressure

April 2026

Christopher Clayton
Yale & NBER

Antonio Coppola
Stanford

Matteo Maggiori
Stanford & NBER

Jesse Schreger
Columbia & NBER

THE GLOBAL CAPITAL ALLOCATION PROJECT

Geoeconomics and Economic Statecraft

- Hegemonic governments use their economic strength from existing financial and trade relationships to achieve geopolitical and economic goals
- Many prominent examples:
  - US government imposition of export controls in semiconductors
  - China threatening to cut off supply of rare earths
  - Trump administration using tariffs as negotiation tools
- How can we systematically measure pressure and responses at the firm level?

This Paper

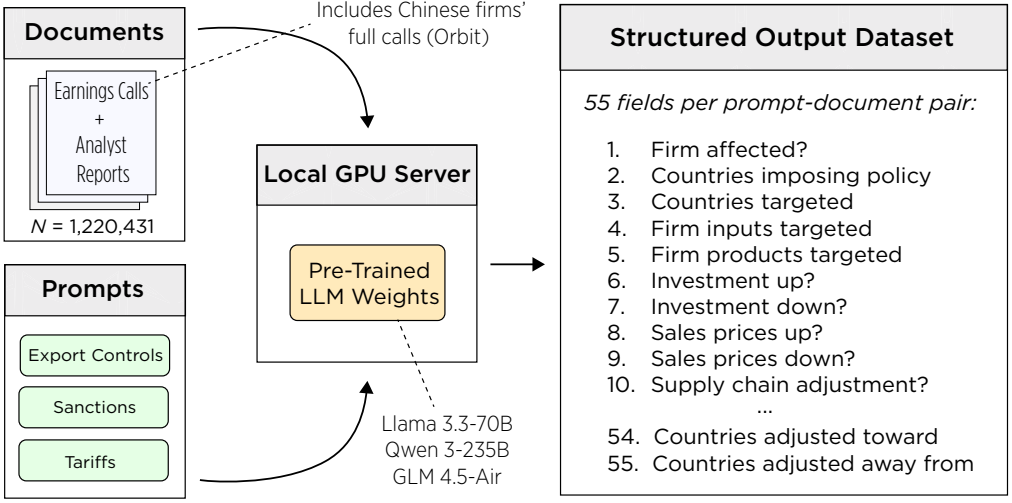
- Characterize the application of geoeconomic pressure around the world
  - Firm-level corporate text combined with large-scale, replicable LLM inference
  - Approach: approximate firm survey with LLMs and corporate text
  - Subjective Causal Effect: measure of firm's reported behavior change in response to policy
- How does geoeconomic pressure change firm behavior?:
  1. **Export controls** drive R&D by targeted firms
  2. **Sanctions** drive increasing supply chain adjustment in recent episodes
  3. Winners in **trade wars** rely relatively little on foreign inputs but still raise their prices
- US targets **strategic chokepoints** when applying export controls and sanctions
- Multi-pronged approach to assessing measurement uncertainty and reliability:
  - Prompt and model perturbations; human validation; debiasing and error correction; comparison with standard NLP; alternate text sources; computational reproducibility

Related Literature

- **Text-Based Measurement for Macroeconomics and Finance:** Baker, Bloom and Davis (2016), Hassan et al. (2019, 2024a,b), Caldara and Iacoviello (2022), Ahir, Bloom, and Furceri (2022), Gentzkow and Shapiro (2010), Gentzkow, Kelly, and Taddy (2019), Handlan (2002), Juhász et al. (2022), Goldberg et al. (2024), Flynn and Sastry (2022), Flynn et al. (2025), Chen, Kelly, and Xiu (2022), Ottonello, Song, and Sotelo (2024), Bybee (2023), Egami et al. (2023), Battaglia et al. (2024), Dell (2024), Sarkar (2025), Lagakos, Michalopoulos, and Voth (2025), Ludwig et al. (2025), Bisbee and Spirling (2025)

- **Geoeconomics and Economic Statecraft:** Hirschman (1945), Baldwin (1985), Kindleberger (1973), Keohane and Nye (1977), Blackwill and Harris (2016), Drezner (2003, 2024), Drezner, Farrell, and Newman (2021), Farrell and Newman (2019, 2023), Parks et al. (2022), Clayton, Maggiori, Schreger (2023, 2024, 2025), Thoenig (2023), Kleinman, Liu, and Redding (2024), Alekseev and Lin (2024), Becko and O'Connor (2024), Broner et al. (2024, 2025), Liu and Yang (2024), Kooi (2024), Mattoo, Ruta, and Staiger (2024), Pflueger and Yared (2024), Fernández-Villaverde, Mineyama, and Song (2024), Gopinath et al. (2024), Aiyar, Malacrino, and Presbitero (2024), Hakobyan, Meleshchuk, and Zymek (2023), Bonadio et al. (2024), Crosignani et al. (2024), Mohr and Trebesch (2024), Egorov, Korovkin, Makarin, Nigmatulina (2025)

Using LLMs to Extract Large-Scale Structured Data From Text



Estimating the Subjective Causal Effect of Geoeconomic Pressure

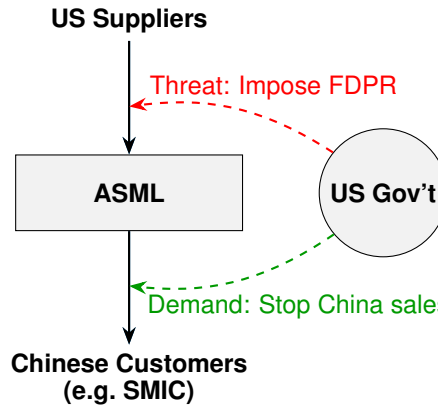
- The **Subjective Causal Effect (SCE)** asks whether the firm itself says the policy *caused* a response, not just whether the response happened.

- Consider two potential statements of an American firm
  1. "President Trump's Liberation Day tariffs put significant stress on our global supply chains. To overcome the challenges raised by these tariffs, we have decided to expand our production facility in the United States."
  2. "The second quarter of 2025 featured a changing global environment, with our supply chains under significant stress from the Liberation Day tariffs. Against this challenging global backdrop, we were pleased to see continued strong demand for our products and have decided to expand our production facility in the United States."

- Humans and LLMs recognize both firms affected by tariffs and undertook domestic investment. Both report that tariffs ***caused*** investment only for the first firm.

Example of LLM Inference From Textual Data

- ASML flagged as affected by export controls in multiple earnings calls and reports (2021-25)
- Structured field outputs from LLM:
  1. Countries imposing controls: US, Netherlands
  2. Country receiving export controls: China
  3. Firm's products targeted: EUV and DUV systems, lithography tools
  4. Overall impact on firm: negative
  5. Firm's responses: lower sales
  6. Country of lower sales: China



- Additionally, we also capture data from Chinese customers such as SMIC

Understanding the Prompts: Export Controls

Prompt Excerpt

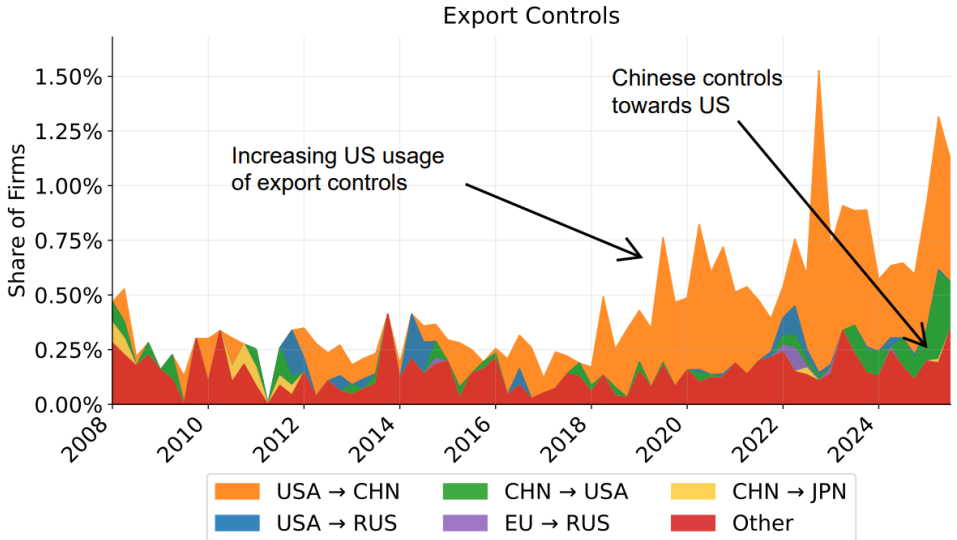
- **Export controls** are defined as restrictions on the sale of certain goods, technologies, or services. These must be imposed by the exporting country in order to be classified as export controls. Export controls are not tariffs, financial sanctions, or boycotts.

[. . .]

- A boolean flag called ***effect_any***, which should be 1 if the firm discusses export controls at any point in the call, and 0 otherwise. Even if the term “export controls” is not explicitly used throughout the call, return a 1 if the firm discusses impacts on its business that clearly relate to export controls.

- A field called ***countries_imposing***, listing the countries whose export control policy the firm discusses, if any. For example, if the firm reports current or future export controls imposed by the US government on goods or services exported to China, this field should say “USA”.

The Geography and Trends of Export Controls



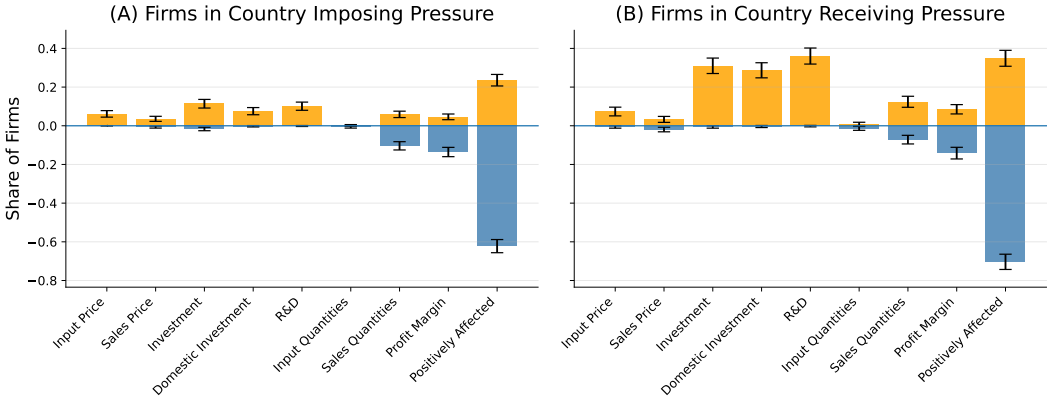
Identifying Firms' Responses to Export Controls

Prompt Excerpt

- A boolean flag called ***investment_up***, which should be 1 if the firm says it is increasing or planning to increase its investment as a result of current or future export controls and 0 otherwise. Only set the flag to 1 if the company explicitly attributes a change in its investment decisions to export controls.

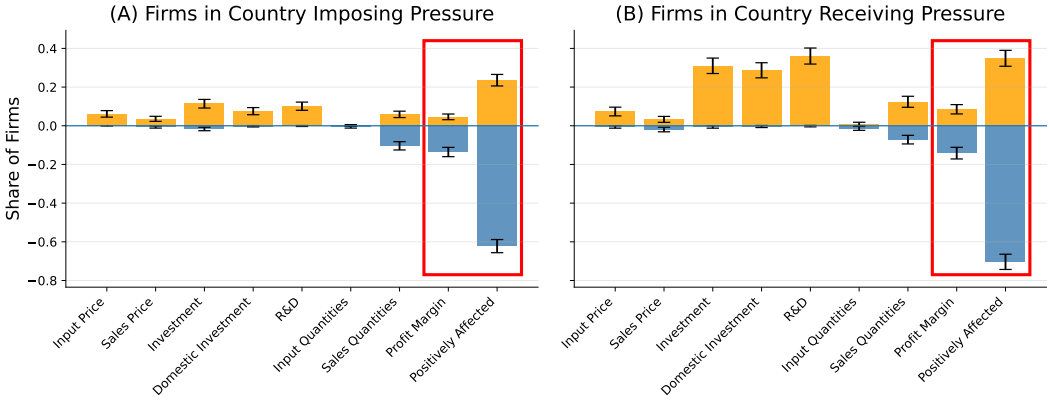
- A boolean flag called ***rd_up***, which should be 1 if the firm says it is increasing or expecting to increase the amount of research and development it undertakes as a result of current or future export controls and 0 otherwise. Only set the flag to 1 if the company explicitly attributes a change in its research and development decisions to export controls.

The Response to Export Controls, By Source and Destination



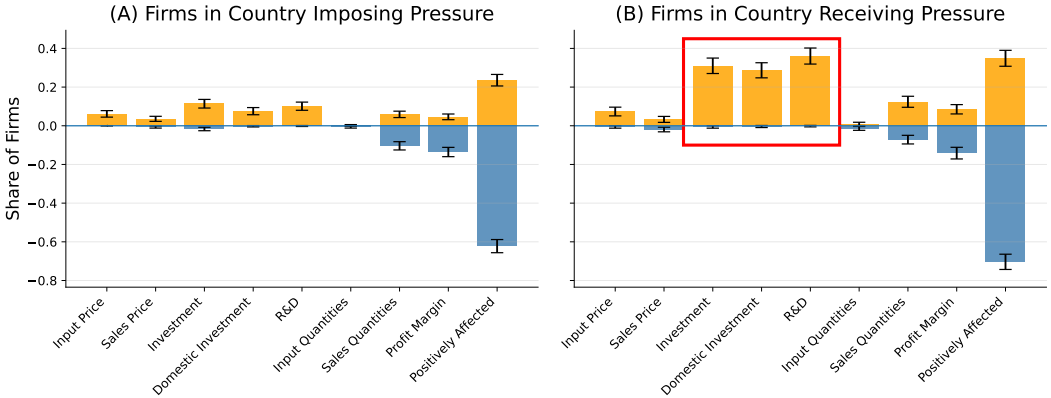
Examples Third Parties

The Response to Export Controls, By Source and Destination



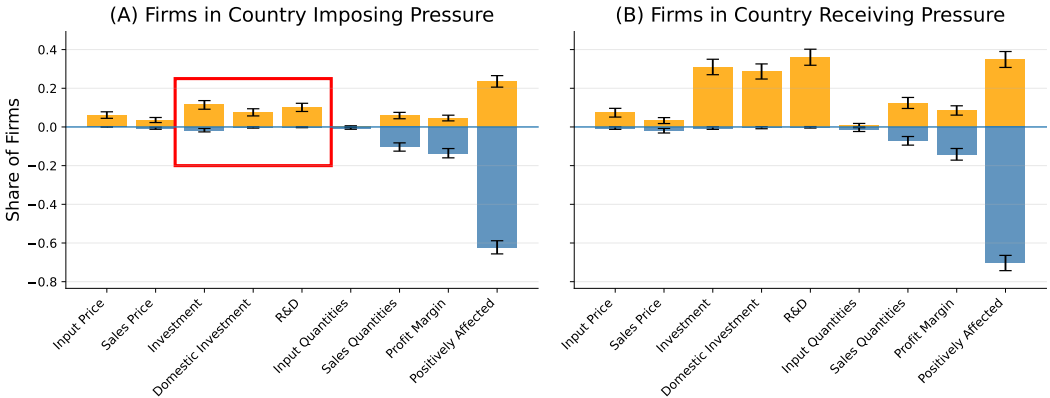
Examples Third Parties

The Response to Export Controls, By Source and Destination



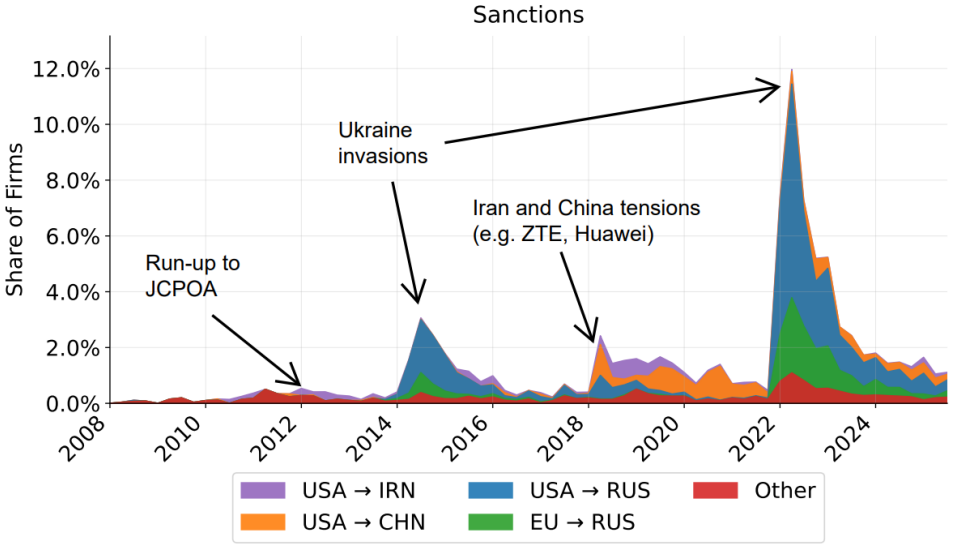
Examples Third Parties

The Response to Export Controls, By Source and Destination

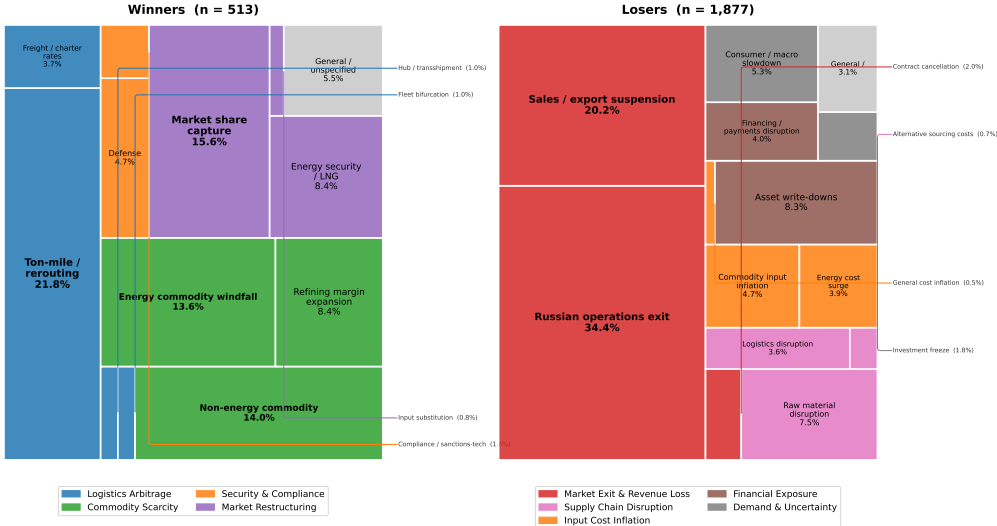


Examples Third Parties

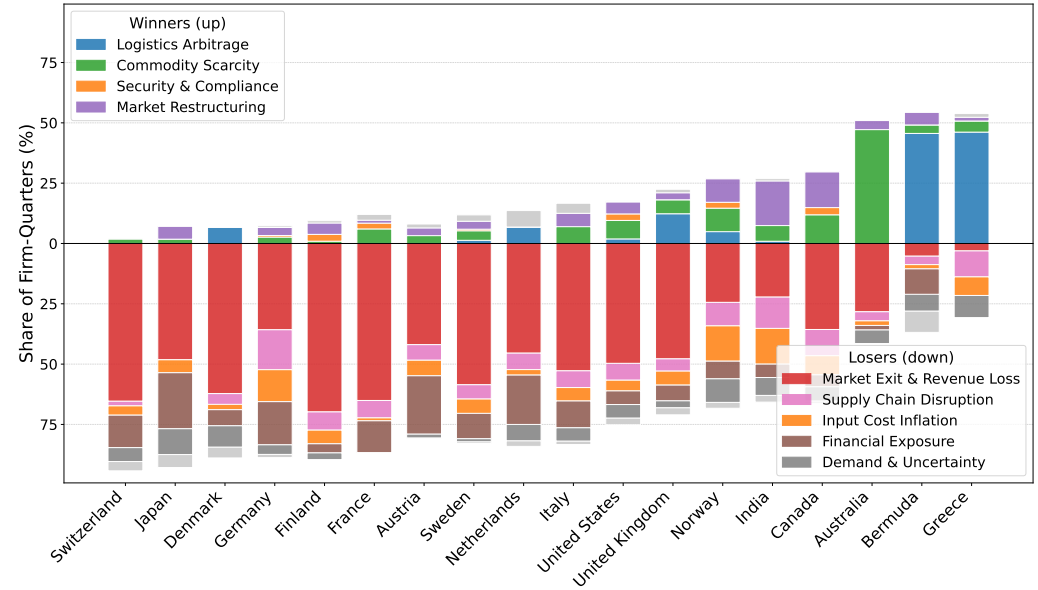
Aggregate Patterns in the Impact of Sanctions on Global Firms



Why do firms report winning or losing from 2022 Russia Sanctions?



Why do firms report winning or losing from 2022 Russia Sanctions



Analogous Prompts for Other Instruments: Tariffs

Prompt Excerpt

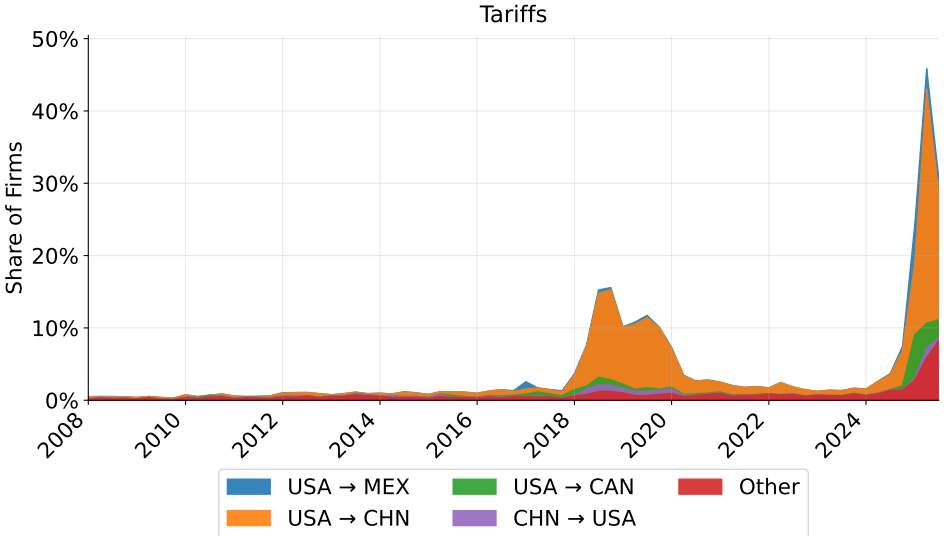
- Tariffs are defined as taxes imposed on imported foreign goods. These must be imposed by the importing country in order to be classified as tariffs. Tariffs are not export restrictions, quotas, embargoes, financial sanctions, boycotts, or non-tariff barriers.

  [. . .]

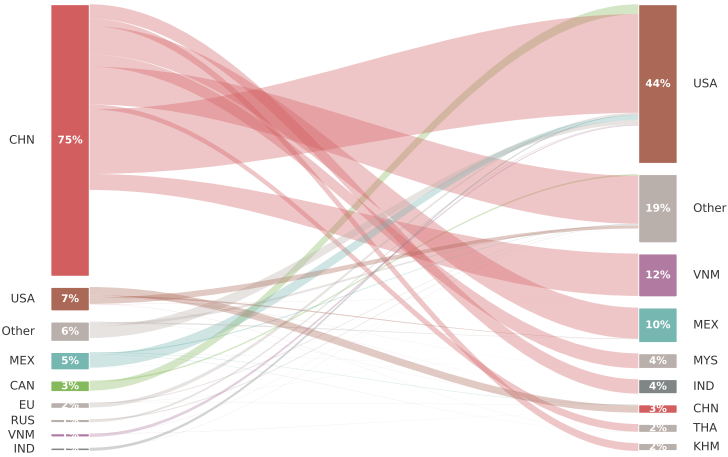
- A boolean flag called ***effect_any***, which should be 1 if the firm discusses tariffs at any point in the call, and 0 otherwise. Even if the term “tariffs” is not explicitly used throughout the call, you should return a 1 if the firm discusses impacts on its business that clearly relate to tariffs.

- A field called ***countries_imposing***, listing the countries whose tariffs policy the firm discusses, if any. For example, if the firm reports concerns about tariffs imposed by the US government on goods imported from China, this field should say “USA”.

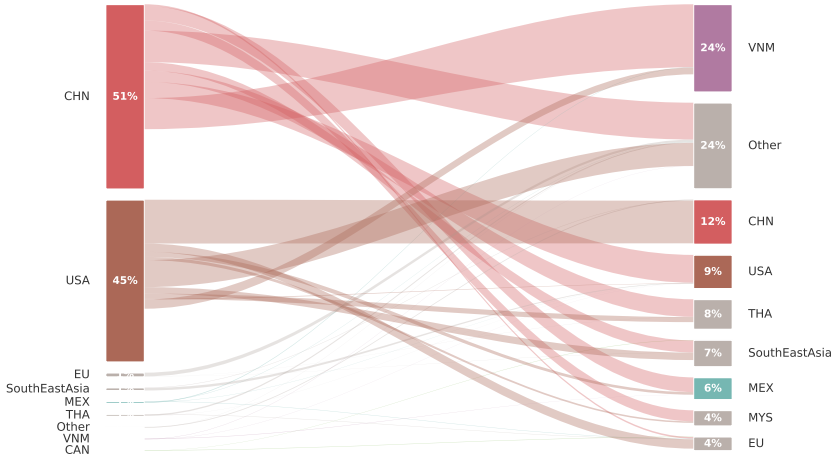
Aggregate Patterns in the Impact of Tariffs on Global Firms



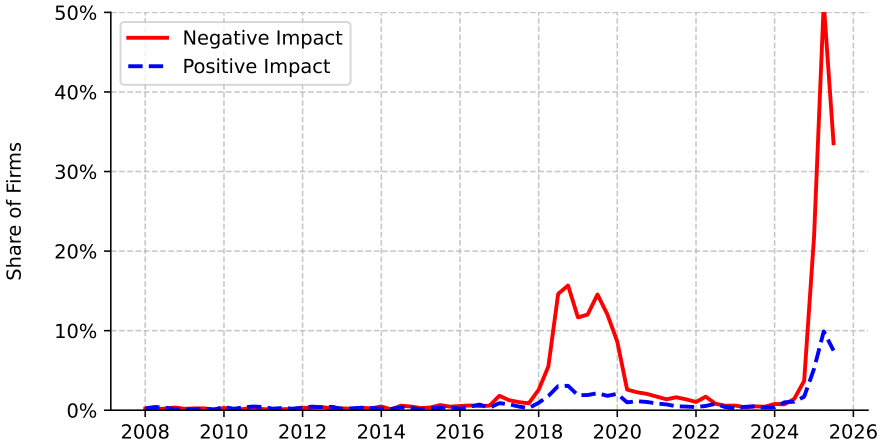
Supply Chain Reshuffling: U.S. Firms



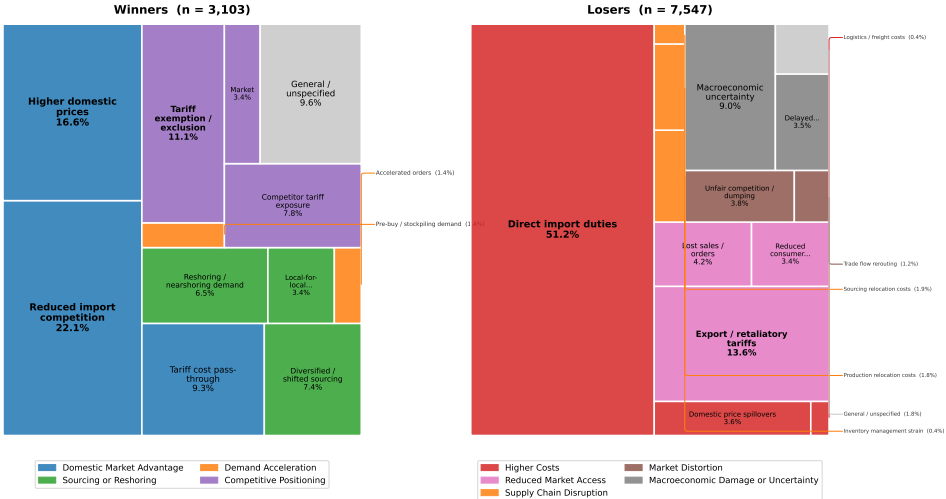
Supply Chain Reshuffling: Chinese Firms



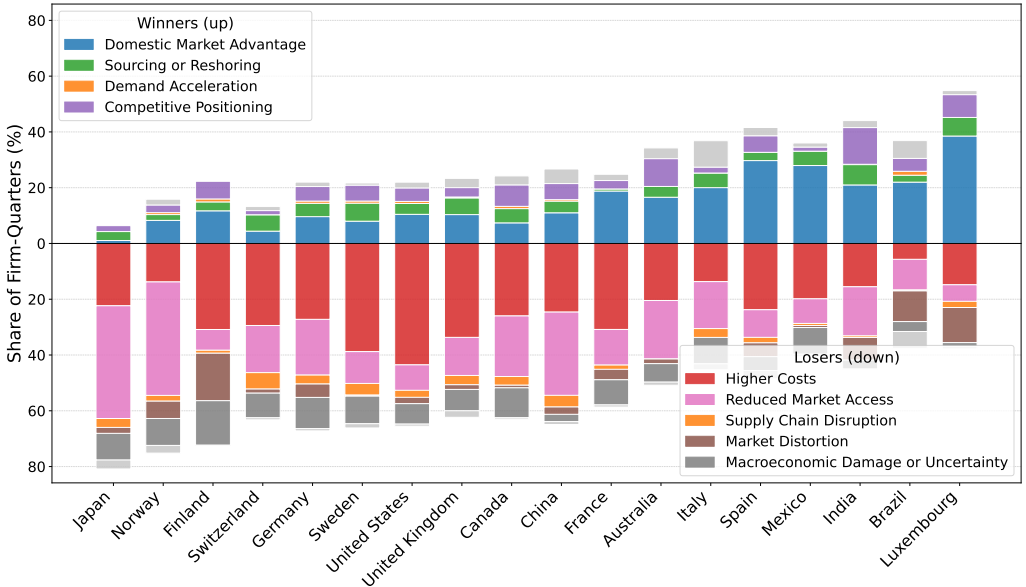
Winning and Losing from Tariffs



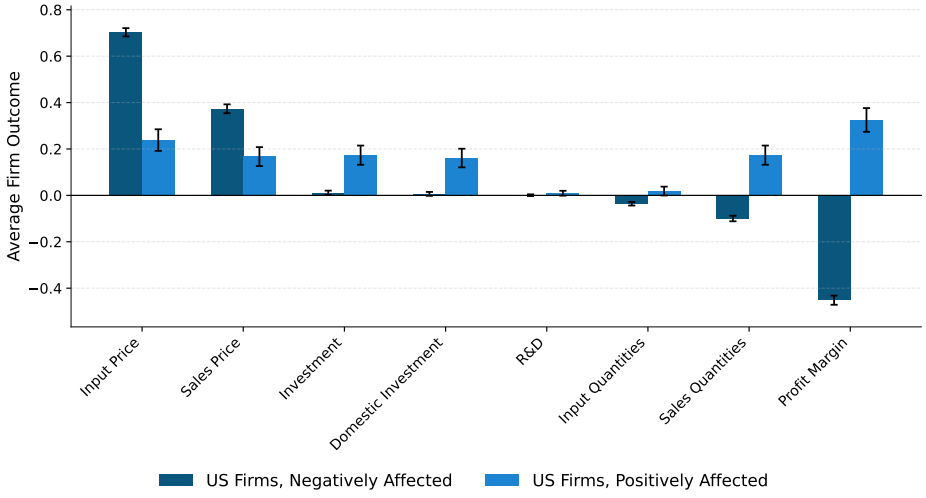
Why do firms report winning or losing from tariffs?



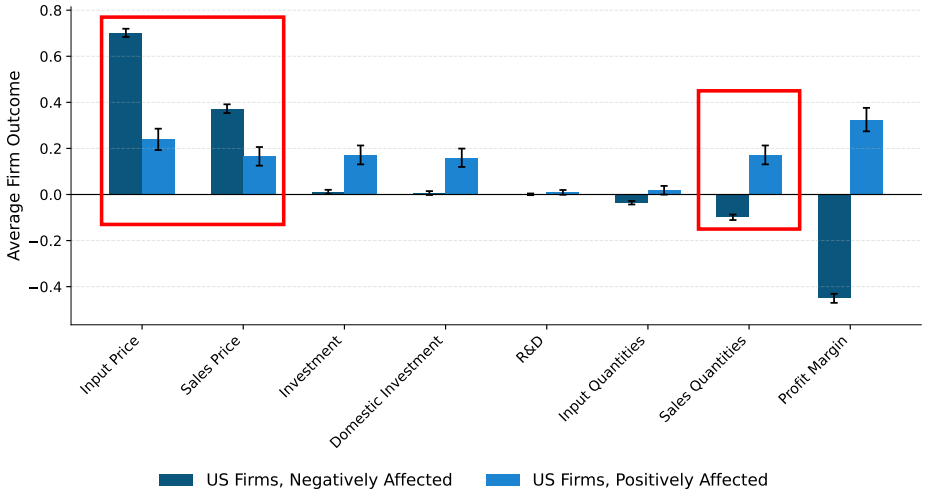
Why do firms report winning or losing from tariffs?



An Ongoing Look at the Trade War of 2025: US Firms' Adjustment



An Ongoing Look at the Trade War of 2025: US Firms' Adjustment



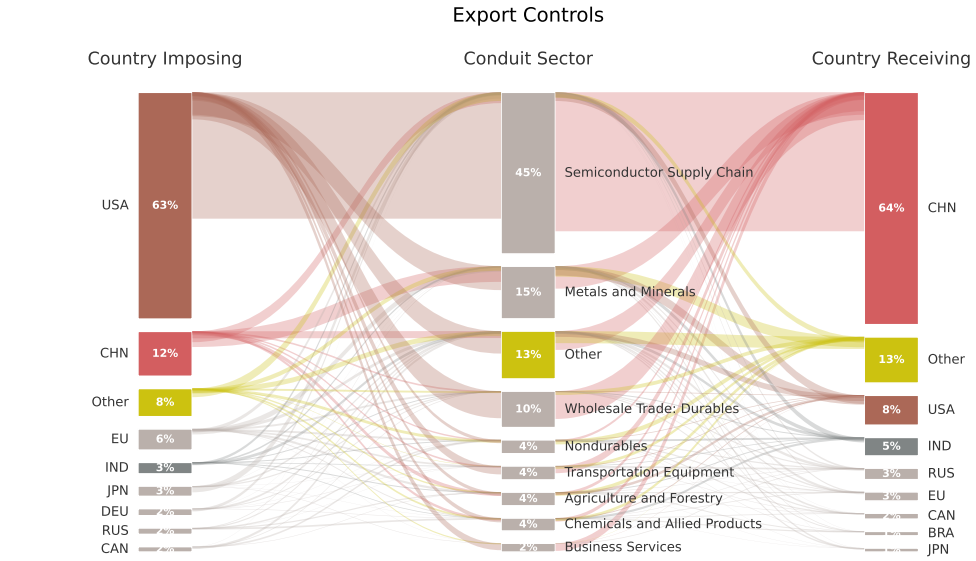
Identifying the Products Involved in Export Controls

Prompt Excerpt

- A field called ***product_receiving***, listing the goods or services that the firm sells that are targeted by the export controls, if any. For example, if the firm reports current or future export controls imposed by the US government on semiconductors that it exports to China, this field should say “semiconductors”.

- A field called ***input_receiving***, listing the goods or services that the firm buys that are targeted by the export controls, if any. For example, if the firm reports current or future export controls imposed by the US government on semiconductors that it buys from the US, this field should say “semiconductors”.

Which Sectors Are Used as the Means of Pressure?

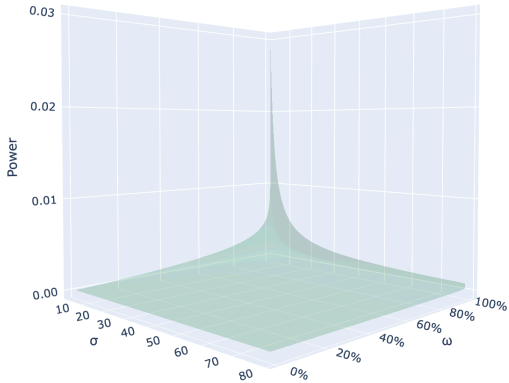


Geoeconomic Pressure: Do Hegemons Target Strategic "Chokepoints"?

- Percentage loss to target n from losing access to hegemon's input in sector j:

$$\text{Power}_{nj} \propto -\underbrace{\Omega_{nj}}_{\substack{\Omega_{nj}:\text{ Expenditure}\\ \text{share on sector } j}} \times \underbrace{(\sigma_j - 1)^{-1}}_{\substack{\sigma_j\text{: Elasticity of}\\ \text{substitution in sector } j}} \times \underbrace{\log(1 - \omega_{nj})}_{\substack{\omega_{nj}\text{: Share of expenditures in}\\ \text{sector } j \text{ controlled by hegemon}}}$$

- Simplified calibration, full formula in Clayton et al. (24); analogy to Arkolakis et al. (12)



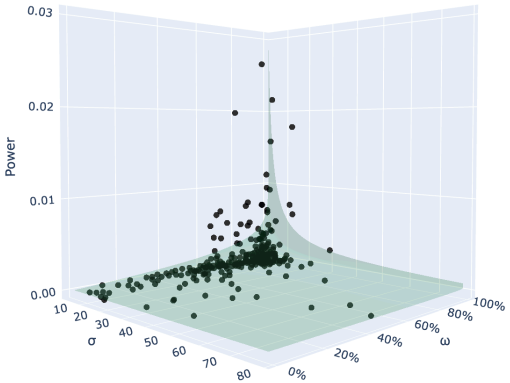
▸ General Formula

Geoeconomic Pressure: Do Hegemons Target Strategic "Chokepoints"?

- Percentage loss to target n from losing access to hegemon's input in sector j:

$$\text{Power}_{nj} \propto -\underbrace{\Omega_{nj}}_{\substack{\Omega_{nj}\text{ : Expenditure}\\ \text{share on sector } j}} \times \underbrace{(\sigma_j - 1)^{-1}}_{\substack{\sigma_j\text{: Elasticity of}\\ \text{substitution in sector } j}} \times \underbrace{\log(1-\omega_{nj})}_{\substack{\omega_{nj}\text{: Share of expenditures in}\\ \text{sector } j \text{ controlled by hegemon}}}$$

- Simplified calibration, full formula in Clayton et al. (24); analogy to Arkolakis et al. (12)
- Measure Power_{nj} at SIC4 sectoral level using hard data (BACI, Fontagne et al. 22)



▸ General Formula

Geoeconomic Pressure: Do Hegemons Target Strategic "Chokepoints"?

- Percentage loss to target n from losing access to hegemon's input in sector j:

$$\text{Power}_{nj} \propto -\underbrace{\Omega_{nj}}_{\substack{\Omega_{nj}:\text{ Expenditure}\\ \text{share on sector } j}} \times \underbrace{(\sigma_j - 1)^{-1}}_{\substack{\sigma_j\text{: Elasticity of}\\ \text{substitution in sector } j}} \times \underbrace{\log(1-\omega_{nj})}_{\substack{\omega_{nj}\text{: Share of expenditures in}\\ \text{sector } j \text{ controlled by hegemon}}}$$

- Simplified calibration, full formula in Clayton et al. (24); analogy to Arkolakis et al. (12)
- Measure Power_{nj} at SIC4 sectoral level using hard data (BACI, Fontagne et al. 22)
- Use LLMs to measure intensity of sector usage by the US in export controls and sanctions towards China



▸ General Formula

Addressing Measurement Uncertainty and Reliability

1. **Reproducibility:** deterministic local inference
2. **Conditional vs. unconditional prompts:** SCE robust to causal framing
3. **Policy stripping & identity masking:** robust to removing identifying info
4. **Inter-model agreement:** high agreement across 3 LLMs
5. **Reliability:** prompt/model perturbations yield high reliability
6. **Strategic disclosure:** analyst reports replicate earnings call results
7. **Simple NLP baseline:** LLM recovers bag-of-words for simplest fields
8. **Human validation:** high agreement with human labelers
9. **Frontier Model Comparison**: High overlap but more conservative

Conclusion

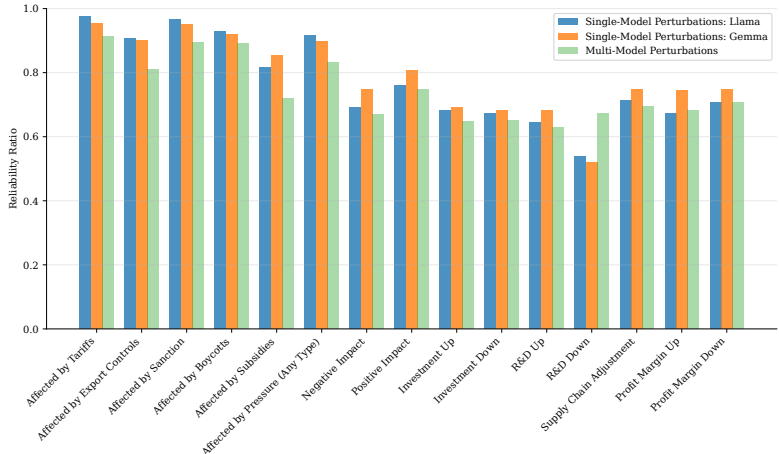
- Novel, systematic approach to measuring **geoeconomic pressure** using firm-level text
- Use firm-level measurement to characterize firms' responses to geoeconomic pressure:
  - **Export controls** drive R&D by targeted firms
  - **Sanctions** drive increasing supply chain adjustment in recent episodes
  - Winners in **trade wars** rely relatively little on foreign inputs but still raise their prices
- US targets **strategic chokepoints** when applying export controls
- Reproducible LLM inference pipelines; analysis of measurement uncertainty and reliability
- **Geoeconomic Monitor** at globalcapitalallocation.com

Extra Slides

Measurement Reliability Is Comparable to High-Quality Survey Data

We vary models and prompts; model outcome Y_{ik} for document i and perturbation k as:

$$Y_{ik} = \bar{\eta} + \eta_i + \varepsilon_{ik} \quad \Longrightarrow \quad \text{Reliability} = \frac{\text{Var}(\eta_i)}{\text{Var}(\eta_i) + \text{Var}(\varepsilon_{ik})}$$



Details

Human Validation: The Gold Standard?

- We ask humans (RAs) to perform the same labeling task:
  - Expensive and time-consuming; open question if better ground truth (Bisbee Spirling 25)
  - Also use human validation sample for formal debiasing (Carlson Dell 25, Ludwig et al. 25)
- Sample at random 500 documents in 2025; test multiple models (Llama, Gemma, GLM) and human annotators on identical prompts; focus here on binary "affected by tariffs"
  - When all 3 models agree (94% of cases) $\Longrightarrow$ High agreement with humans

(a) When All Models Agree

	LLMs = 0	LLMs = 1	Total
Human = 0	254	5	259
Human = 1	23	187	210
Total	277	192	469

Human Validation: The Gold Standard?

- We ask humans (RAs) to perform the same labeling task:
  - Expensive and time-consuming; open question if better ground truth (Bisbee Spirling 25)
  - Also use human validation sample for formal debiasing (Carlson Dell 25, Ludwig et al. 25)
- Sample at random 500 documents in 2025; test multiple models (Lllama, Gemma, GLM) and human annotators on identical prompts; focus here on binary "affected by tariffs"
  - When models disagree $\Longrightarrow$ Type 1 and type 2 errors against humans

(a) LLaMA

	0	1	Total
Human = 0	8	1	9
Human = 1	18	3	21
Total	26	4	30

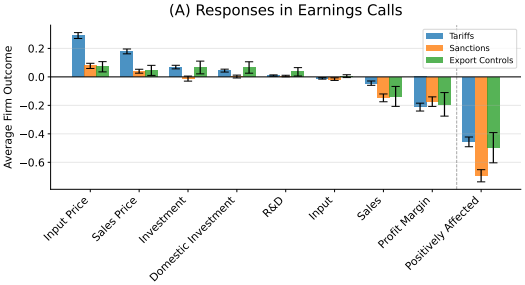
(b) Gemma

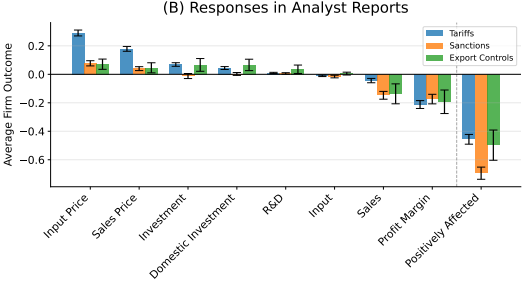
	0	1	Total
Human = 0	6	3	9
Human = 1	9	12	21
Total	15	15	30

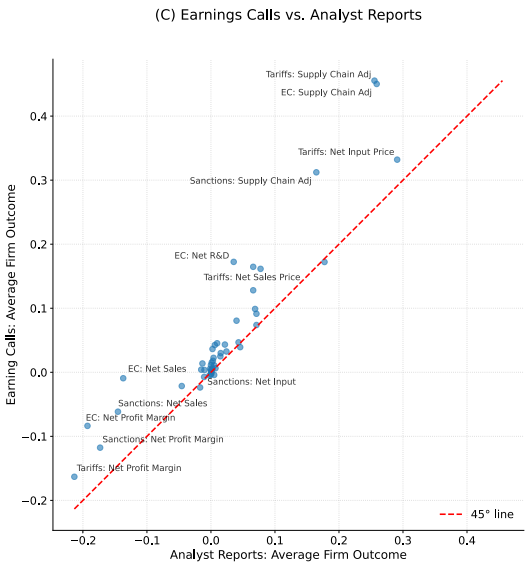
(c) GLM

	0	1	Total
Human = 0	3	6	9
Human = 1	6	15	21
Total	9	21	30

Strategic Disclosure? Responses in Earnings Calls vs. Analyst Reports

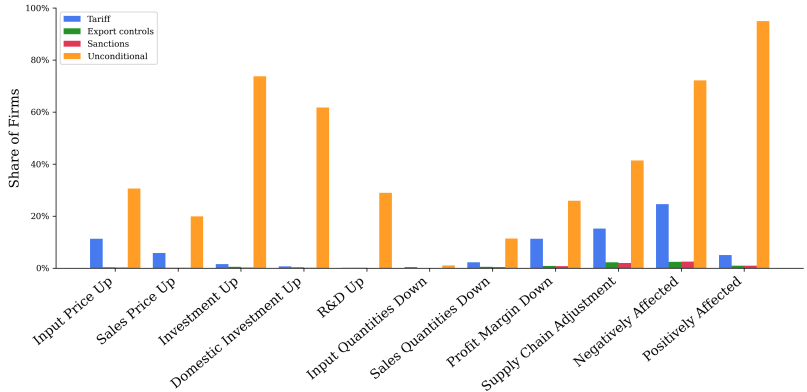






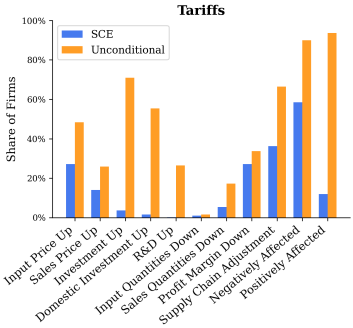
Validation of the SCE: Conditional vs. Unconditional Prompts

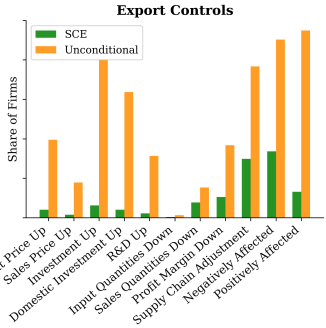
- Compare baseline SCE prompt to unconditional prompt without explicit causal conditioning.

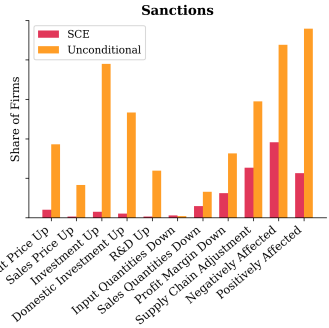


Validation of the SCE: Conditional vs. Unconditional Prompts

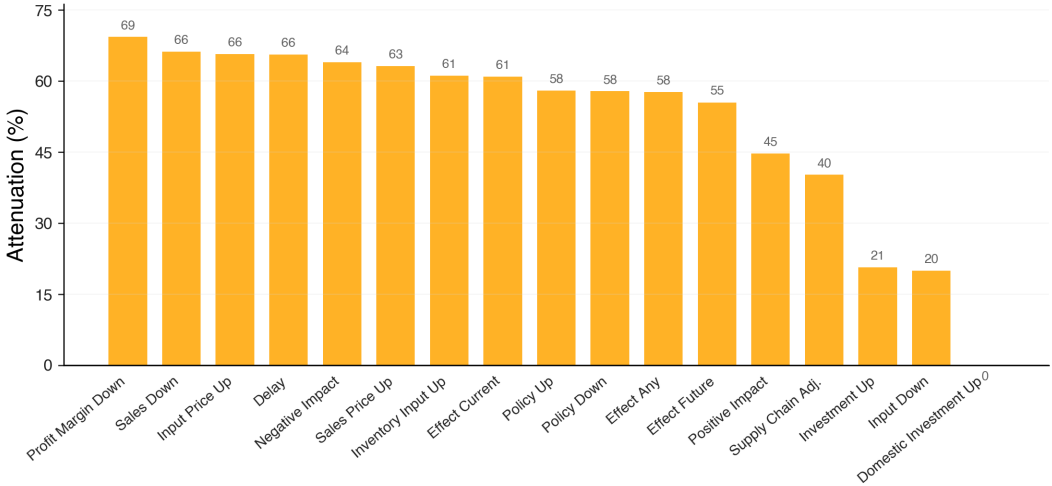
- Compare baseline SCE prompt to unconditional prompt without explicit causal conditioning.



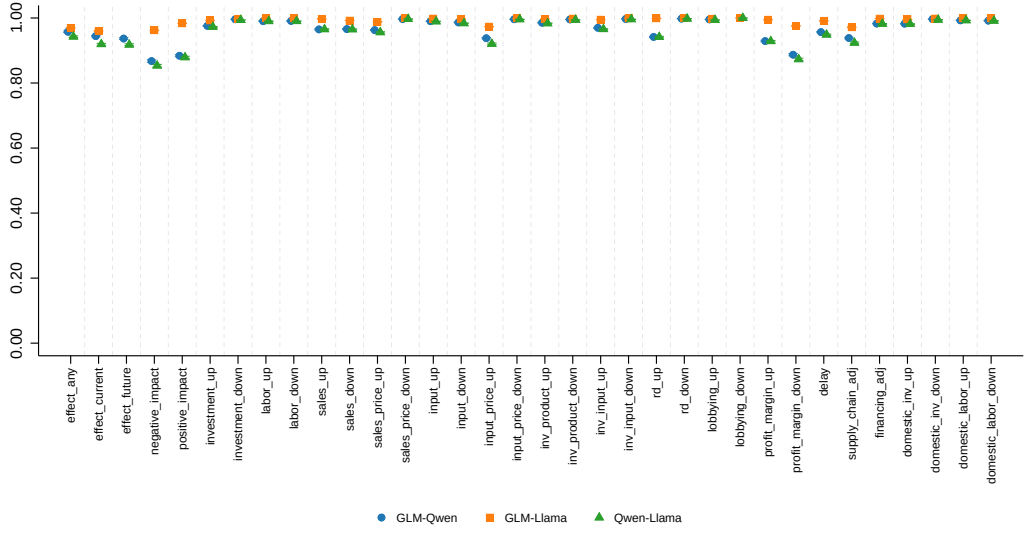




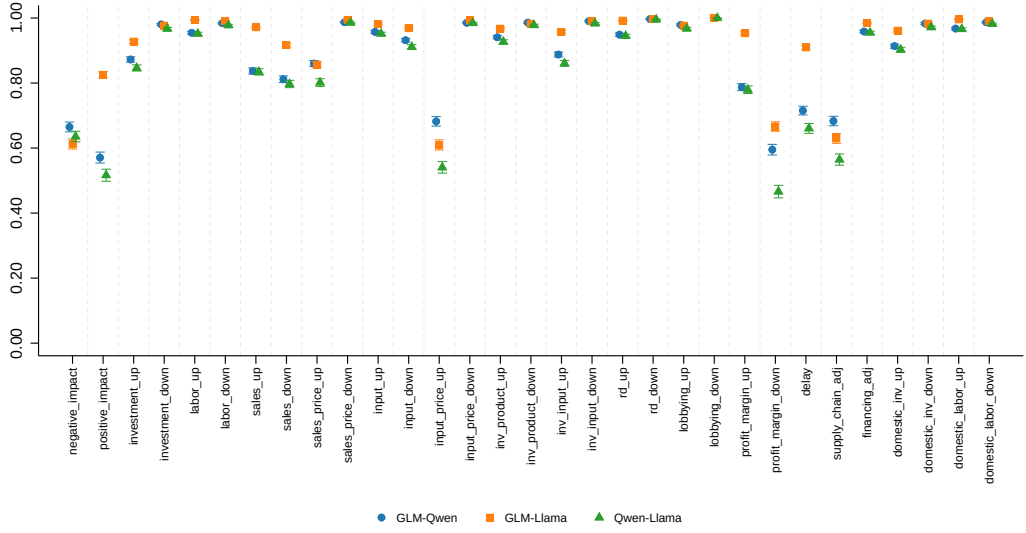
Validation of the SCE: Policy Stripping



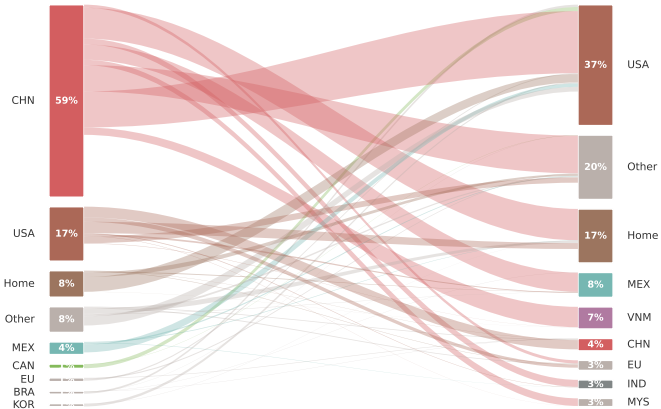
Inter-Model Agreement: Tariff Decomposition



Inter-Model Agreement: Tariff Effects Conditional on Policy Flag



Supply Chain Reshuffling Under Tariffs: Rest-of-World Firms

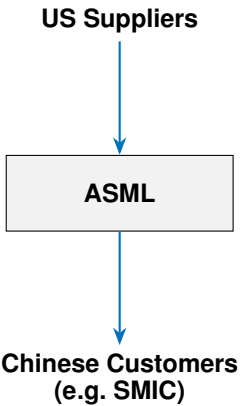


Rest-of-world firms also move away from China while expanding toward the U.S. and other Asian hubs.

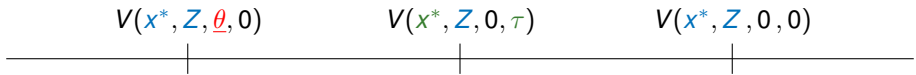
Organizing Framework: Imposing Hegemons, Receivers, and Pressure Means



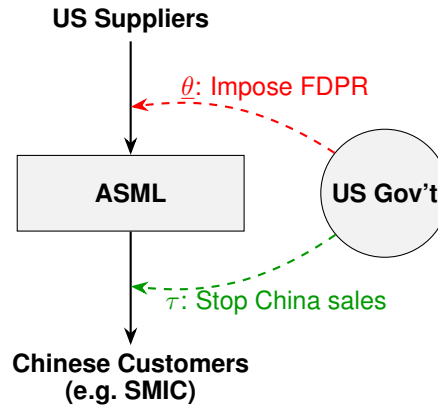
- Entity's privately optimal actions: $x^* = x(Z, \cdot, \cdot)$
- Takes as given a set of aggregate variables Z



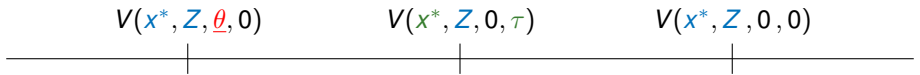
Organizing Framework: Imposing Hegemons, Receivers, and Pressure Means



- Entity's privately optimal actions: $x^* = x(Z, \theta, \tau)$
- Takes as given a set of aggregate variables Z
- Hegemon threatens $\underline{\theta}$ if demands not accepted
- Demands that entity undertake costly actions τ

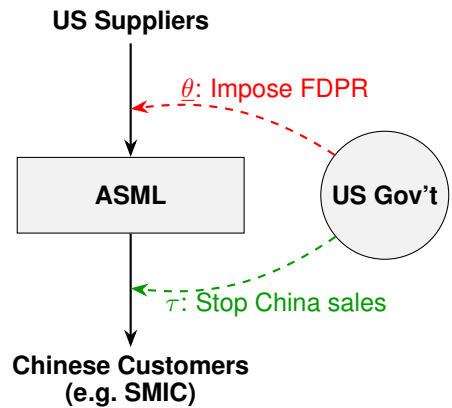


Organizing Framework: Imposing Hegemons, Receivers, and Pressure Means



- Entity's privately optimal actions: $x^* = x(Z, \theta, \tau)$
- Takes as given a set of aggregate variables Z
- Hegemon threatens $\underline{\theta}$ if demands not accepted
- Demands that entity undertake costly actions τ
- **Participation constraint**:

$$V(x^*, Z, 0, \tau) \geq V(x^*, Z, \underline{\theta}, 0)$$



Broader Tools of Economic Statecraft (θ)

Export or Import Restrictions
- Export and import taxes (tariffs)
- Export and import controls
- Boycotts
- Licensing and regulation

Financial Sanctions
- Asset freezes
- Asset expropriation
- Access and taxes on international payments
- Aid withdrawals
- FDI inbound and outbound screening
- Access and taxes on portfolio allocations
- Access and taxes on insurance
- Access and taxes on trade credit

Industrial Policy
- Producer and infrastructure subsidies
- Creation and allocation of sovereign wealth funds
- Long term procurement contracts and price floors
- Creation and use of strategic reserve funds

Macroeconomic Restrictions
- Ability to use lender of last resort facilities
- Ability to participate in int'l regulatory meetings
- Participation in sovereign renegotiations
- Participation in bilateral and multilateral treaties

Back

Understanding Why Particular Sectors Are Used: Quantifying Power

Clayton et al. (2024) use a nested CES structure to show that percentage loss from losing access to hegemon's inputs, $\Delta \log V_n$, is:

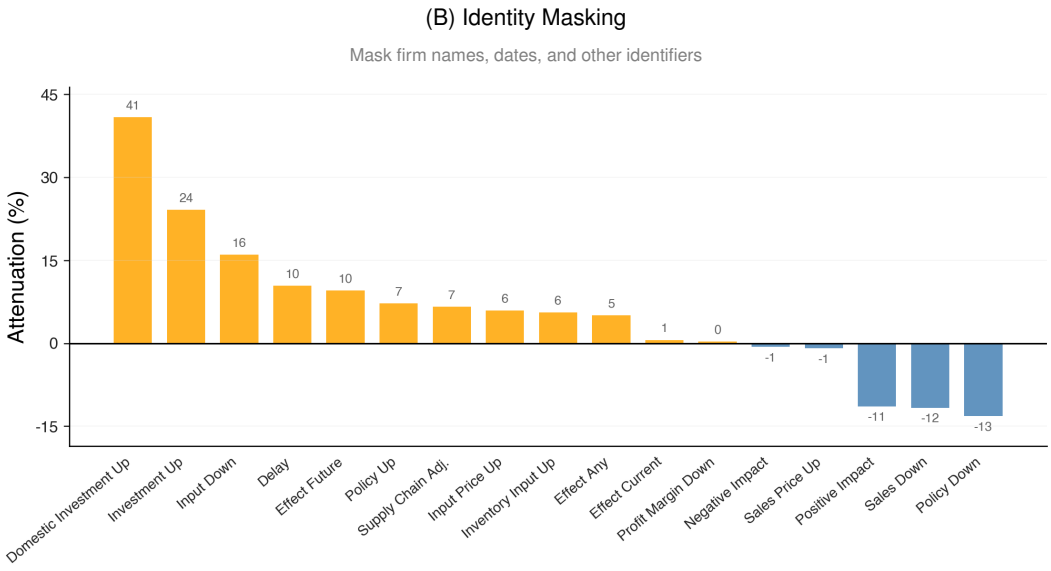
$$\frac{\beta}{1-\beta}\frac{1}{1-\varrho}\log\left(\sum_{G\in\{M,F\}}\Omega_{nG}\left(\sum_{J\in J_G}\Omega_{nGJ}\left(1-\Omega_{nJR}+\Omega_{nJR}\left(1-\omega_{nJR}\right)^{\frac{\zeta_J-1}{\sigma_J-1}}\right)^{\frac{\rho_G-1}{\zeta_J-1}}\right)^{\frac{\varrho-1}{\rho_G-1}}\right)$$

where:

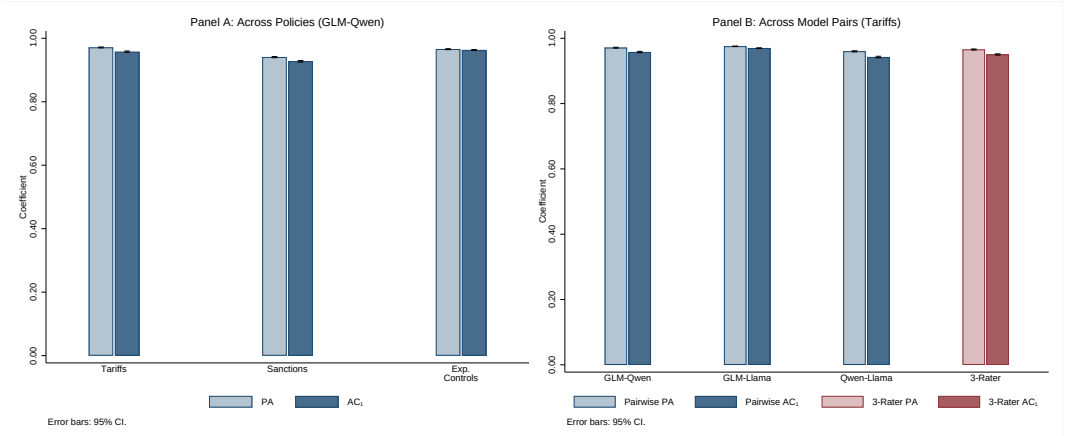
- Ω_{nG}: share on sector (financial services F, all other goods and services M)
- Ω_{nGJ}: share of sectoral spending on sub-sector J
- Ω_{nJR}: share of sub-sector J spending on foreign inputs
- ω_{nJR}: share of foreign inputs coming from hegemon
- $\zeta_J, \sigma_J, \rho_G$: elasticities of substitution at different nesting levels
- β, ϱ: parameters of aggregate production function, invariant by sector

Back

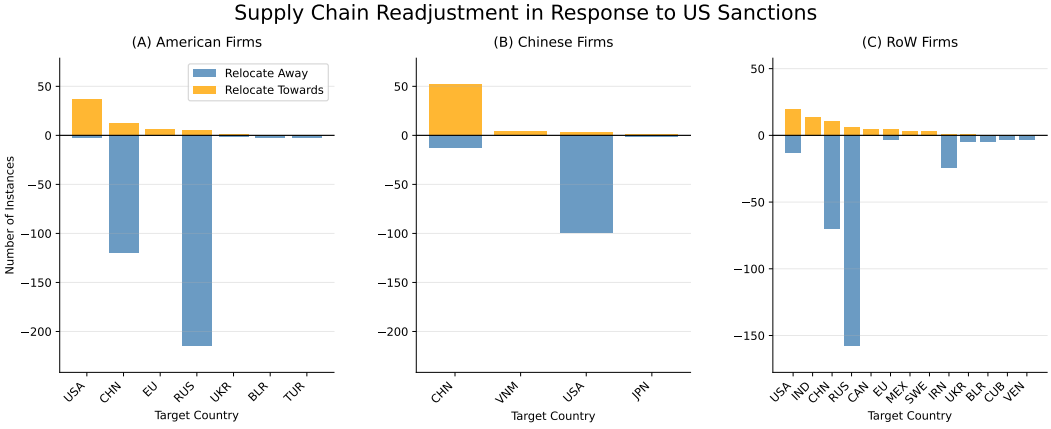
Validation of the SCE: Policy Stripping and Identity Masking



Inter-Model Agreement: Policy Classification

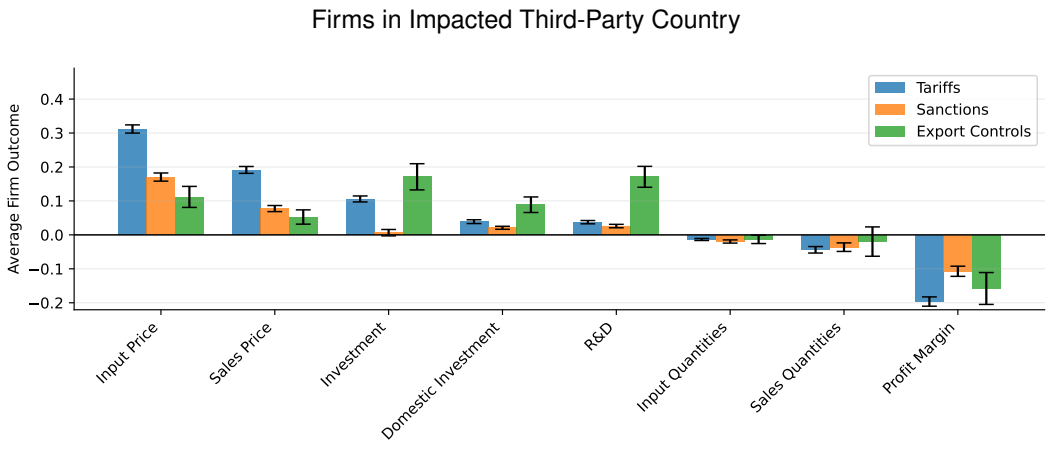


Firms' Responses: Assessing Supply Chain Reshuffling Systematically



Back

Responses to Pressure Are Heterogeneous by Firm's Country Role



Back

How Firms Respond to Export Controls, Examples: SMIC

Company	LLM's Automated Analysis	Verbatim Evidence in Transcripts
SMIC (Semiconductor Manufacturing Intl. Corp.) China's largest integrated circuit foundry. Added to the U.S. Entity List in December 2020.	"The firm discusses export controls explicitly... impacting the firm's business... affecting their ability to obtain certain equipment and parts..." "The firm is also investing in research and development to improve their products and services..."	"The current international situation is becoming more and more complicated... For the United States **export control**... in the short term, [it] has a certain impact on the company... [we have allocated] 7 billion yuan for advanced and mature technology R&D project reserve funds, [and] our second generation of advanced technology n+1... has entered a small trial production."

Back

How Firms Respond to Export Controls, Examples: Sugon

Company	LLM's Automated Analysis	Verbatim Evidence in Transcripts
Sugon (Dawning Information Industry Co.) A leading Chinese company in supercomputing, servers, and data storage. Added to the U.S. Entity List in June 2019.	"The firm discusses the impact of being on the entity list... risks such as shortages of raw materials from upstream parts manufacturers will affect the company's operating revenue." "The firm's behavior has been impacted... with the company mentioning that it has been... investing in R&D and guaranteeing the supply of key components."	"Since entering the **entity list**... risks such as shortage of raw materials from upstream parts manufacturers will affect the company's operating revenue... [so the company will] continue to... increase R&D investment, [and] can guarantee the supply of key components."

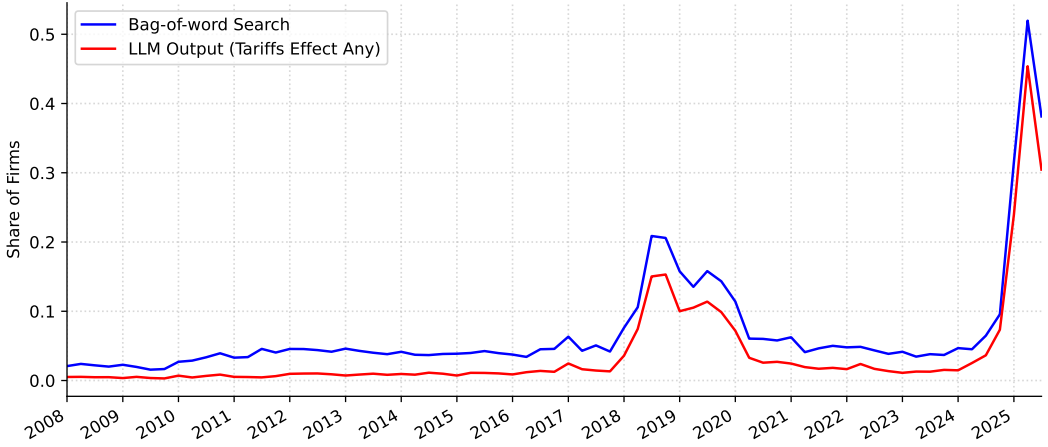
Back

How Firms Respond to Export Controls, Examples: iFLYTEK

Company	LLM's Automated Analysis	Verbatim Evidence in Transcripts
iFLYTEK (iFLYTEK CO.,LTD.) A leading Chinese AI and speech technology company. Added to the U.S. Entity List in October 2019.	"The firm discusses export controls explicitly, specifically the US government's restrictions on the export of high-end chips to China... iFlytek... is subject to sanctions." "The firm has also collaborated with Huawei to develop... and is building a domestic large model architecture on the basis of independently innovative software and hardware."	"The US government further tightens **export controls** on cutting-edge artificial intelligence chips... After iFlytek was listed as a US entity... iFlytek and Huawei... [are] collaborating... to develop a high-performance operator library, jointly Building a New Base for... independently innovative software and hardware."

Back

Comparing LLM Measures to Bag-of-Words Identification of Tariffs



Back

Reproducibility: Are the Results Computationally Replicable and Stable?

- Our LLM pipeline uses open-weights models on a stable computing infrastructure:
  - Stable weights stored locally: no uncontrolled model changes, no risk of deprecation
  - Other researchers can fully replicate the same inference pipeline
- Greedy sampling, deterministic inference engine (vLLM), single-GPU inference:
  - Two runs of same prompt on same data $\Rightarrow$ same output
- Different from running models over an API with closed-weight model
- Tradeoffs for researchers' best practices: cost, reproducibility, flexibility, leap-frogging

Back

Assessing Measurement Reliability Using Prompt and Model Variations

We examine size of LLM-induced noise using perturbations approach:

- We re-run measurement with different models and varying the prompts
- Prompt variations maintain same meaning, but alter structure and substantive language
- Range from minor to major; intended to span space of reasonable prompt alternatives
- Sample from models deterministically (zero-temperature: no arbitrary sampling variation)

Back

Assessing Measurement Reliability Using Prompt and Model Variations

- For a given field, model outcome for a document i and a given prompt perturbation k as:

$$Y_{ik} = \bar{\eta} + \eta_i + \varepsilon_{ik}, \quad \sigma_\eta^2 = \mathrm{Var}(\eta_i), \quad \sigma_\varepsilon^2 = \mathrm{Var}(\varepsilon_{ik})$$

- Reliability is defined as:

$$\boxed{\text{Reliability} = \frac{\sigma_\eta^2}{\sigma_\eta^2 + \sigma_\varepsilon^2}}$$

- We measure this ratio of variances empirically:

$$\hat{\sigma}_\varepsilon^2 = \frac{1}{N}\sum_{i=1}^{N} \widehat{\mathrm{Var}}(\{Y_{ik}\}_{k=1}^{K}), \qquad \hat{\sigma}_\eta^2 = \widehat{\mathrm{Var}}(\{\bar{Y}_i\}_{n=1}^{N})$$

- We use a total of $K = 60$ perturbations (20 prompt variations $\times$ 3 models)

Back

Formal Error Correction Methods With Human-Validated Sample

- Literature in CS on de-biasing LLM measure; more recent literature in economics
- We implement MAR-S methodology of Carlson and Dell (25) with human-validated sample
- Debiasing of first moments comes down to difference in means; example for tariffs in 2025:

$$\underbrace{\mu_{MARS}}_{\text{De-Biased Estimate}} = \underbrace{\hat{\mu}}_{\text{LLM Est.: Full Sample}} + \underbrace{\hat{M}}_{\text{Human Est.: Validation Sample}} - \underbrace{\hat{\mu}'}_{\text{LLM Est.: Validation Sample}}$$

Formal Error Correction Methods With Human-Validated Sample

- Literature in CS on de-biasing LLM measure; more recent literature in economics
- We implement MAR-S methodology of Carlson and Dell (25) with human-validated sample
- Debiasing of first moments comes down to difference in means; example for tariffs in 2025:

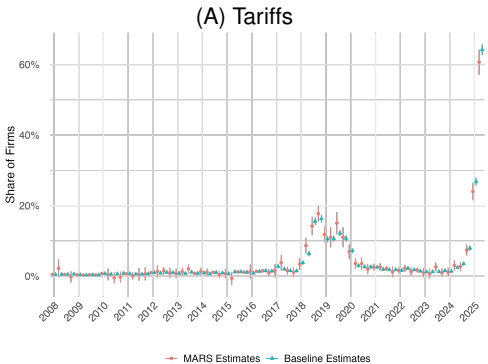
$$\underbrace{\mu_{MARS}}_{\text{De-Biased Estimate}} = \underbrace{\hat{\mu}}_{\text{LLM Est.: Full Sample}} + \underbrace{\hat{M}}_{\text{Human Est.: Validation Sample}} - \underbrace{\hat{\mu}'}_{\text{LLM Est.: Validation Sample}}$$

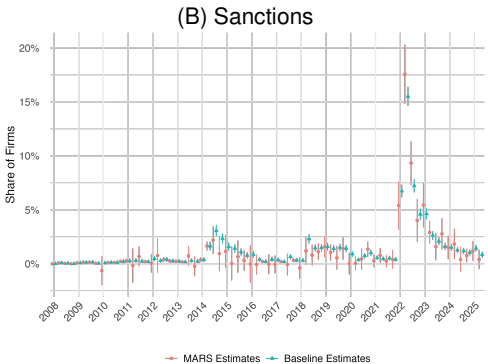
$$\underbrace{43.77\%}_{\text{De-Biased Estimate}} = \underbrace{37.26\%}_{\text{LLM Est.: Full Sample}} + \underbrace{46.57\%}_{\text{Human Est.: Validation Sample}} - \underbrace{40.04\%}_{\text{LLM Est.: Validation Sample}}$$

Back

Error Correction Methods: Further Illustration From LLM-as-Judge

- Expensive to extend human validation beyond 2025 to full time series, but we provide illustration using a complementary LLM-as-judge approach
- Implement MAR-S framework of Carlson-Dell (2025) to debias estimated moments taking "ground truth" to be 5% random sample from one model inference run

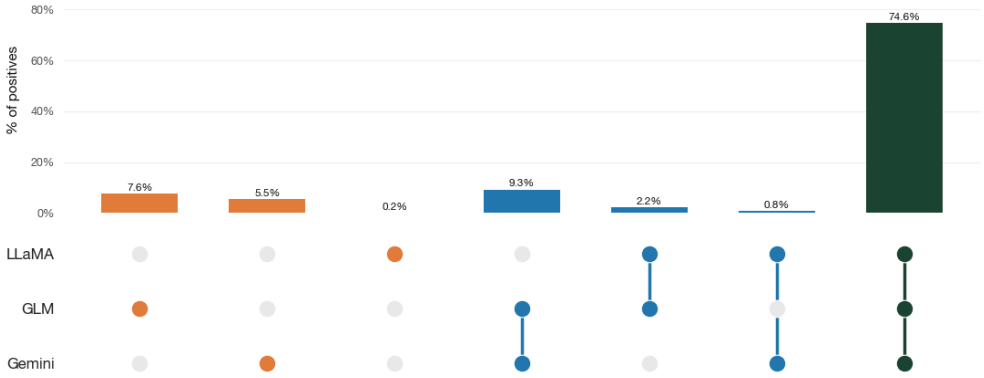




Back

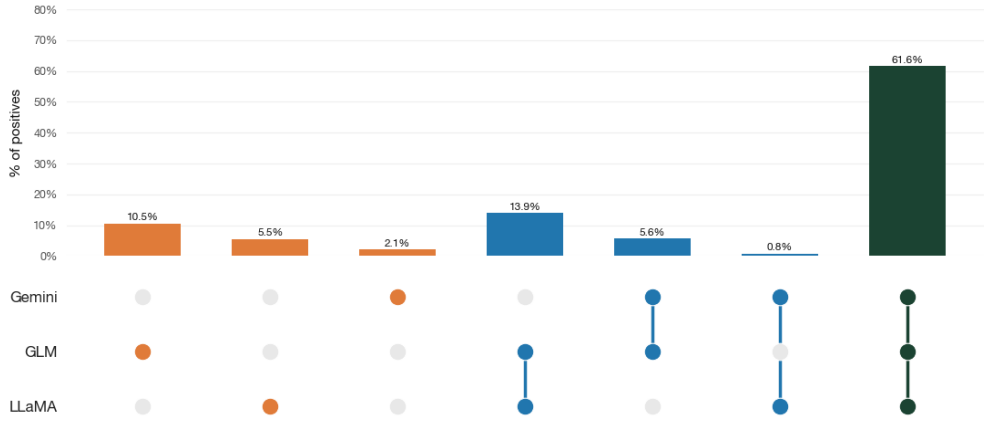
Frontier Model Comparison: Tariffs

Back



Frontier Model Comparison: Supply Chain Adjustment, Tariffs

Back



Frontier Model Comparison: Export Controls

Back

