

Policy documents across 185 countries predominantly rely on evidence from the Global North

Sebastian Ramirez-Ruiz (European University Institute)

Roman Senninger (Aarhus University)

Leibniz Open Science Day 2025: “Better Science for Better Policies”

October 27, 2025



In this project, we analyze data from **1.2 million government policy documents** from **185 countries** to find that:

In this project, we analyze data from **1.2 million government policy documents** from **185 countries** to find that:

- Governments in the global South more frequently cite **foreign materials**

In this project, we analyze data from **1.2 million government policy documents** from **185 countries** to find that:

- Governments in the global South more frequently cite **foreign** materials, while higher-income countries tend to draw from **domestic** expertise




In this project, we analyze data from **1.2 million government policy documents** from **185 countries** to find that:

- Governments in the global South more frequently cite **foreign** materials, while higher-income countries tend to draw from **domestic** expertise
- Countries in the global North (e.g., , , and ) lead in **production** of *policy-based & academic research* cited by governments worldwide.

In this project, we analyze data from **1.2 million government policy documents** from **185 countries** to find that:

- Governments in the global South more frequently cite **foreign** materials, while higher-income countries tend to draw from **domestic** expertise
- Countries in the global North (e.g., , , and ) lead in **production** of *policy-based & academic research* cited by governments worldwide.
- While some **policy domains** appear to be more *academic* (or *policy-based*) **evidence oriented**, the pattern holds

In this project, we analyze data from **1.2 million government policy documents** from **185 countries** to find that:

- Governments in the global South more frequently cite **foreign** materials, while higher-income countries tend to draw from **domestic** expertise
- Countries in the global North (e.g., , , and ) lead in **production** of *policy-based & academic research* cited by governments worldwide.
- While some **policy domains** appear to be more *academic* (or *policy-based*) **evidence oriented**, the pattern holds: evidence from economically powerful nations is consistently more prominent.

Background (aka 'Why do we care?')

Background (aka 'Why do we care?')

- Over the past decades, evidence-informed policymaking has gained widespread support in national and international policy circles (incorporating practices to better integrate **research**, **evaluation**, and **analysis** into decision-making).

Background (aka 'Why do we care?')

- Over the past decades, evidence-informed policymaking has gained widespread support in national and international policy circles (incorporating practices to better integrate **research**, **evaluation**, and **analysis** into decision-making).
- Yet, despite this normative shift, **we still know surprisingly little** about what types of evidence actually make it into policy

Background (aka ‘Why do we care?’)

- Over the past decades, evidence-informed policymaking has gained widespread support in national and international policy circles (incorporating practices to better integrate **research**, **evaluation**, and **analysis** into decision-making).
- Yet, despite this normative shift, **we still know surprisingly little** about what types of evidence actually make it into policy—and **where that evidence originates**.

Background (aka ‘Why do we care?’)

- Over the past decades, evidence-informed policymaking has gained widespread support in national and international policy circles (incorporating practices to better integrate **research**, **evaluation**, and **analysis** into decision-making).
- Yet, despite this normative shift, **we still know surprisingly little** about what types of evidence actually make it into policy—and **where that evidence originates**.
- Government agencies, IGOs, think tanks, and research institutions **all produce (relevant) knowledge**—but the distribution of their reach remains unclear.

Background (aka ‘Why do we care?’)

- Over the past decades, evidence-informed policymaking has gained widespread support in national and international policy circles (incorporating practices to better integrate **research**, **evaluation**, and **analysis** into decision-making).
- Yet, despite this normative shift, **we still know surprisingly little** about what types of evidence actually make it into policy—and **where that evidence originates**.
- Government agencies, IGOs, think tanks, and research institutions **all produce (relevant) knowledge**—but the distribution of their reach remains unclear.
- Understanding who informs policymaking matters: **information is central** to most major theories of the policy process.

Background (aka ‘Why should *we all* care?’)

Background (aka ‘Why should *we all* care?’)

Because evidence is central to decision-making—
and what counts as credible knowledge shapes

Background (aka ‘Why should *we all* care?’)

Because evidence is central to decision-making—
and what counts as credible knowledge shapes
which voices are **heard**, which problems are
prioritized, and which **solutions** get **attention**.

Background (aka ‘Why should *we all* care?’)

Because evidence is central to decision-making—
and what counts as credible knowledge shapes
which voices are **heard**, which problems are
prioritized, and which **solutions** get **attention**.

There is catch:

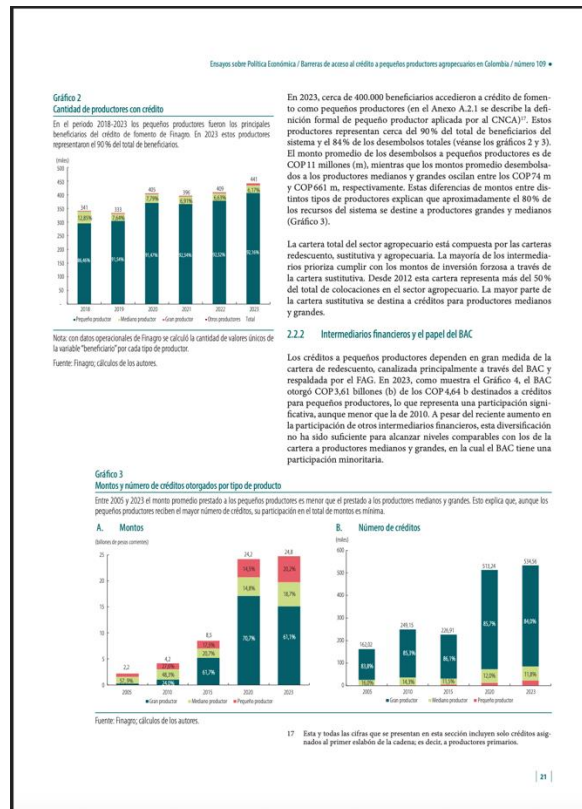
Access to, and **use** of, evidence are **not neutral**.

We asked ourselves: Is there a *plausible vehicle* to study this (at scale)?

We asked ourselves: Is there a ***plausible vehicle*** to study this (at scale)?

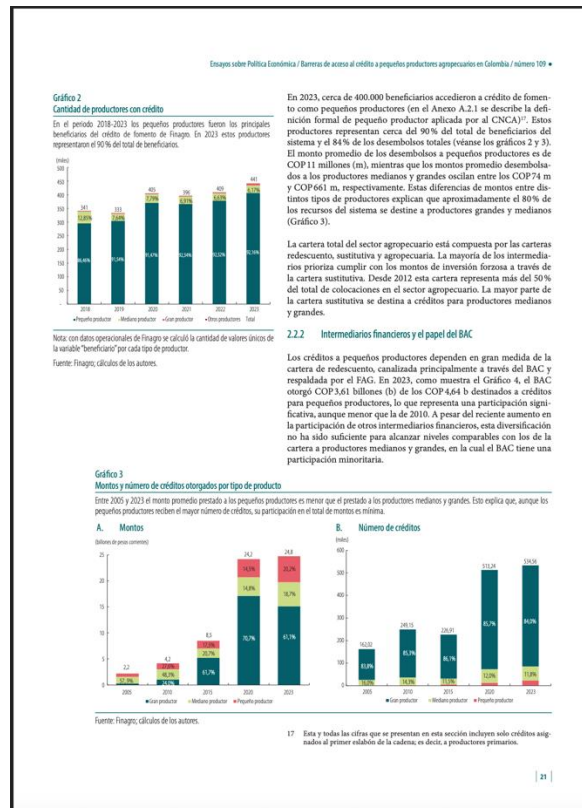
Our heads immediately went to one of the most **widespread** artifacts of modern **policy** and **knowledge dissemination** practice across the world:

We asked ourselves: Is there a *plausible vehicle* to study this (at scale)?



Our heads immediately went to one of the most **widespread** artifacts of modern **policy** and **knowledge dissemination** practice across the world: the humble *policy document*.

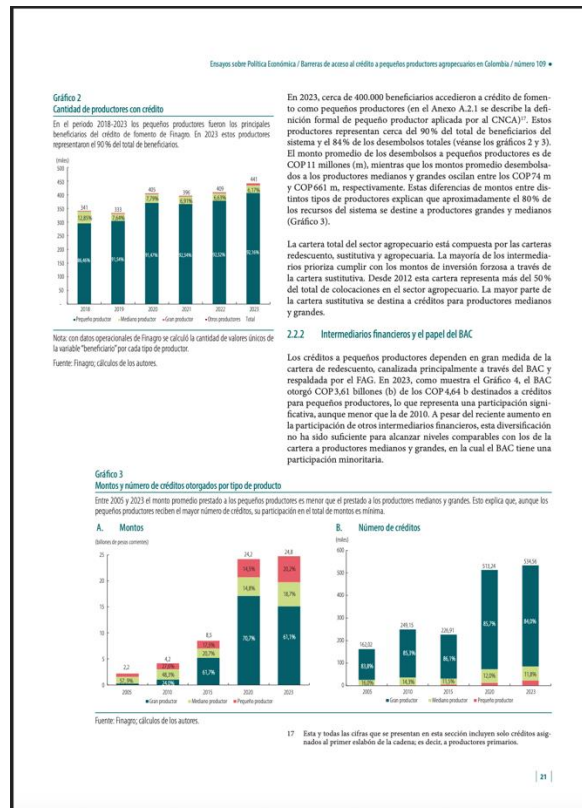
We asked ourselves: Is there a *plausible vehicle* to study this (at scale)?



Our heads immediately went to one of the most **widespread** artifacts of modern **policy** and **knowledge dissemination** practice across the world: the humble *policy document*.

We realized that **policy documents**—from white papers to regulatory assessments—offer a unique window into that relationship.

We asked ourselves: Is there a *plausible vehicle* to study this (at scale)?

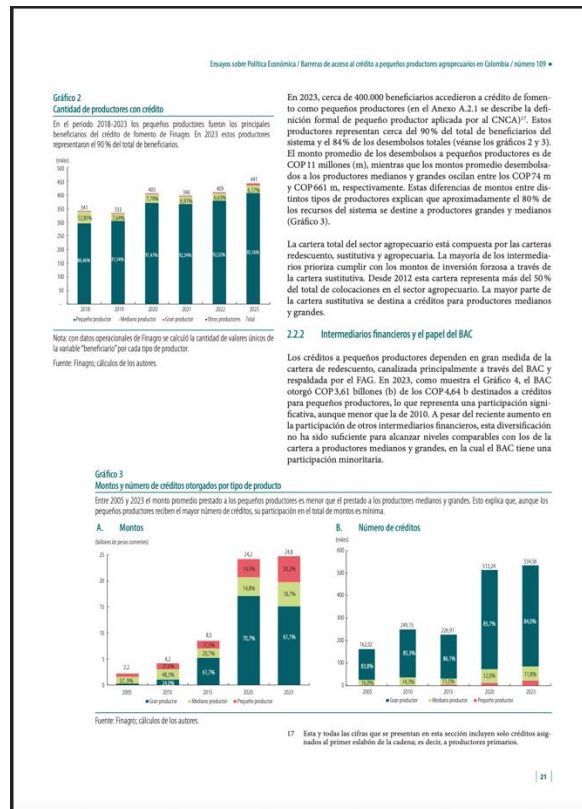


Our heads immediately went to one of the most **widespread** artifacts of modern **policy** and **knowledge dissemination** practice across the world: the humble **policy document**.

We realized that **policy documents**—from white papers to regulatory assessments—offer a unique window into that relationship.

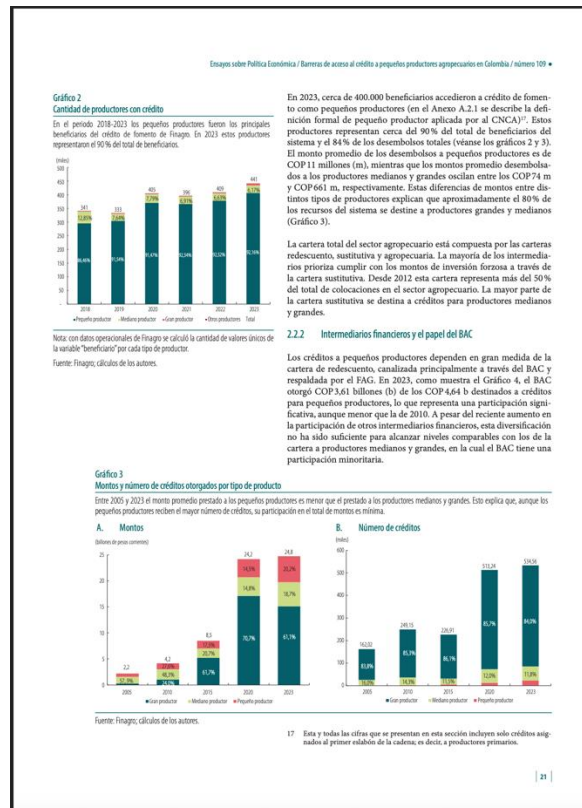
These aren't just administrative outputs. They **signal institutional priorities**, reflect **norms around credibility**, and show **what kinds of knowledge** get referenced.

We asked ourselves: Is there a *plausible vehicle* to study this (at scale)?



We found that existing research looking at **policy documents** is often:

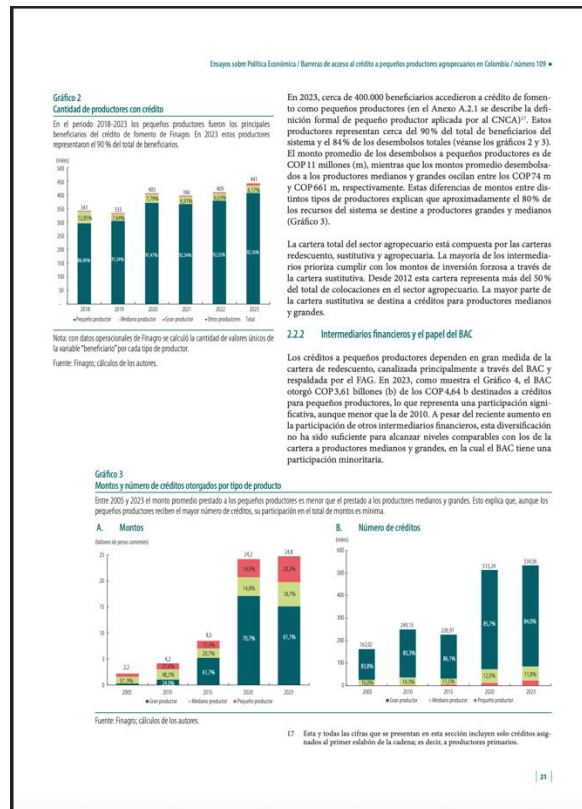
We asked ourselves: Is there a *plausible vehicle* to study this (at scale)?



We found that existing research looking at **policy documents** is often:

- Narrow in scope (e.g., focused on scientific references only),

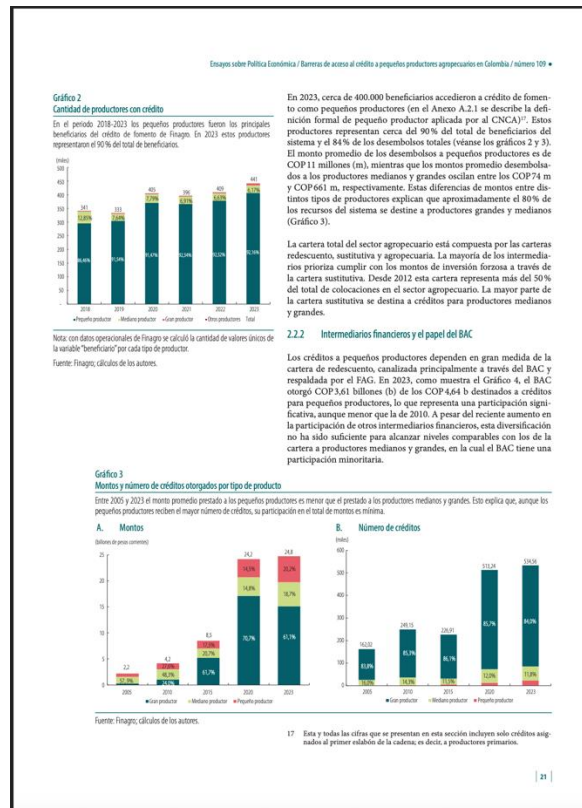
We asked ourselves: Is there a *plausible vehicle* to study this (at scale)?



We found that existing research looking at **policy documents** is often:

- Narrow in scope (e.g., focused on scientific references only),
- Policy domain-specific (e.g., climate, health),

We asked ourselves: Is there a *plausible vehicle* to study this (at scale)?

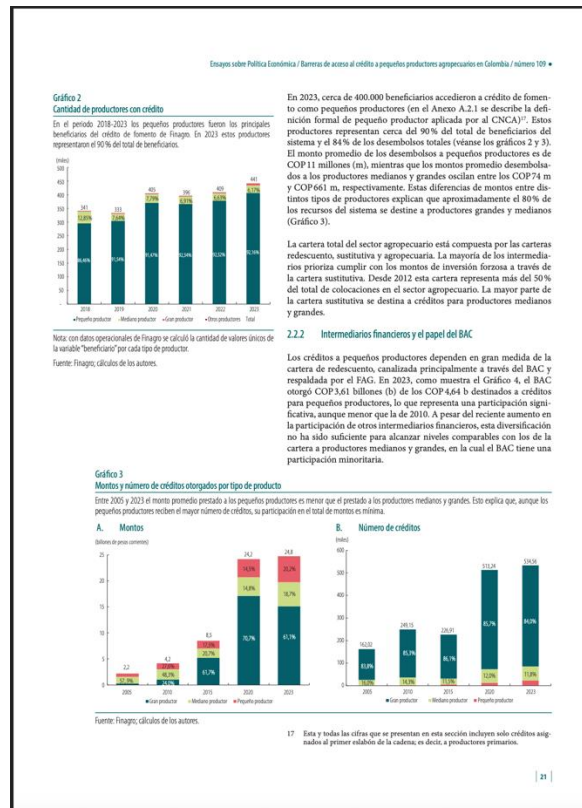


We found that existing research looking at **policy documents** is often:

- Narrow in scope (e.g., focused on scientific references only),
- Policy domain-specific (e.g., climate, health),
- Heavily U.S.-centric.

(e.g., Isett, Hicks, and Kingsley, [2025](#); Furnas, LaPira, and Wang, [2025](#); Ma and Cheng, [2025](#); Bornmann, Haunschild, and Marx, [2016](#)).

We asked ourselves: Is there a *plausible vehicle* to study this (at scale)?



We found that existing research looking at **policy documents** is often:

- Narrow in scope (e.g., focused on scientific references only),
- Policy domain-specific (e.g., climate, health),
- Heavily U.S.-centric.

(e.g., Isett, Hicks, and Kingsley, 2025; Furnas, LaPira, and Wang, 2025; Ma and Cheng, 2025; Bornmann, Haunschild, and Marx, 2016).

A potentially valuable evidence point missing is a **global, cross-sectoral perspective**—a way to capture patterns of evidence use **at scale** across countries and policy domains.

What did we do?

What did we do? We built a **relational database** of government documents globally.

What did we do?

We built a **relational database** of government documents globally.

Policy documents (Core)

n = 1,243,768

policy documents from 185 countries

Data from Overton (Szomszor and Adie, 2022) focused on documents produced by **governments** between 2000 and 2024 citing either *policy* or *academic* references.

From Overton, we extract:

- Document IDs
- Country of origin
- Policy sources referenced
- Academic sources referenced
- Summaries of document content

We employ the summaries to classify documents into Comparative Agendas Project (CAP) policy domains using an LLM (Llama3.3 70B).

- We collected data on **government-authored policy documents** (2000-2024) from Overton—the largest global database of policy documents.

What did we do?

We built a **relational database** of government documents globally.

Policy documents (Core)

n = 1,243,768



policy documents from 185 countries

Data from Overton (Szomszor and Adie, 2022) focused on documents produced by **governments** between 2000 and 2024 citing either *policy* or *academic* references.

From Overton, we extract:

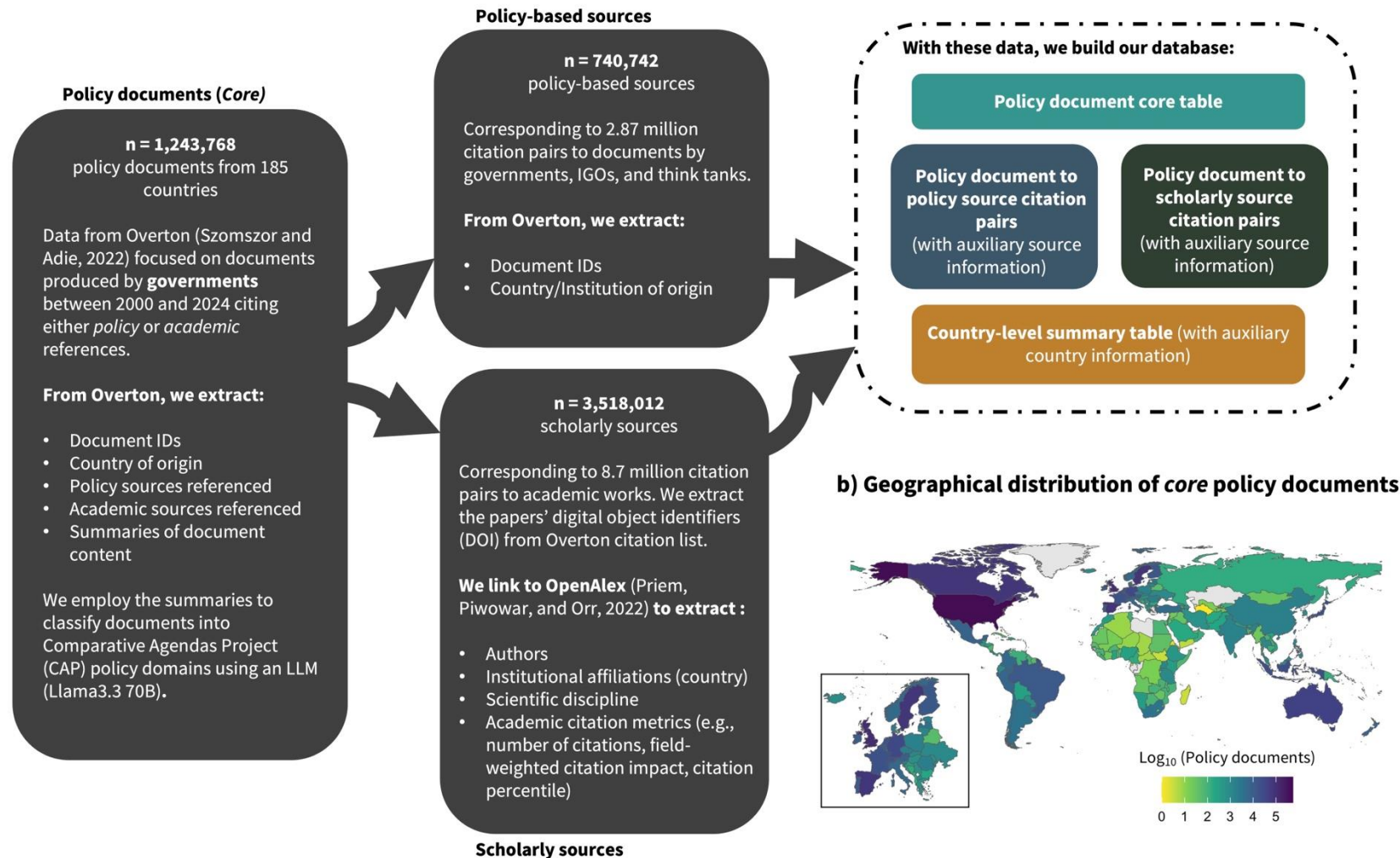
- Document IDs
- Country of origin
- Policy sources referenced
- Academic sources referenced
- Summaries of document content

We employ the summaries to classify documents into Comparative Agendas Project (CAP) policy domains using an LLM (Llama3.3 70B).

- We collected data on **government-authored policy documents** (2000-2024) from Overton—the largest global database of policy documents.
- Gathered the sources they referenced, dividing them into:
 -  **Policy-based** (govs, IGOs, think tanks)
 -  **Scholarly** (journals, working papers)

What did we do?

We built a **relational database** of government documents globally.



What did we do?

Equipped with these data, we conduct a systematic, large-scale analysis of citation dynamics within this global corpus of policy documents (guided by three sequential questions)

What did we do?

Equipped with these data, we conduct a systematic, large-scale analysis of citation dynamics within this global corpus of policy documents (guided by three sequential questions)

- 1) To what extent do governments rely on **domestic** versus **foreign** sources of evidence?

What did we do?

Equipped with these data, we conduct a systematic, large-scale analysis of citation dynamics within this global corpus of policy documents (guided by three sequential questions)

- 1) To what extent do governments rely on **domestic** versus **foreign** sources of evidence?
- 2) Which countries' **policy** and **scientific outputs** are **most** frequently **cited** and therefore **visible**?

What did we do?

Equipped with these data, we conduct a systematic, large-scale analysis of citation dynamics within this global corpus of policy documents (guided by three sequential questions)

- 1) To what extent do governments rely on **domestic** versus **foreign** sources of evidence?
- 2) Which countries' **policy** and **scientific outputs** are **most** frequently **cited** and therefore **visible**?
- 3) How do citation patterns vary across **policy domains** with **differing knowledge demands**?

What did we find?

1. To what extent do governments rely on **domestic** versus **foreign** sources of evidence?

What did we find?

1. To what extent do governments rely on **domestic** versus **foreign** sources of evidence?



For ***policy-based*** citations, we map:

- Country of the citing government vs. country of the cited institution.

What did we find?

1. To what extent do governments rely on **domestic** versus **foreign** sources of evidence?



For ***policy-based*** citations, we map:

- Country of the citing government vs. country of the cited institution.

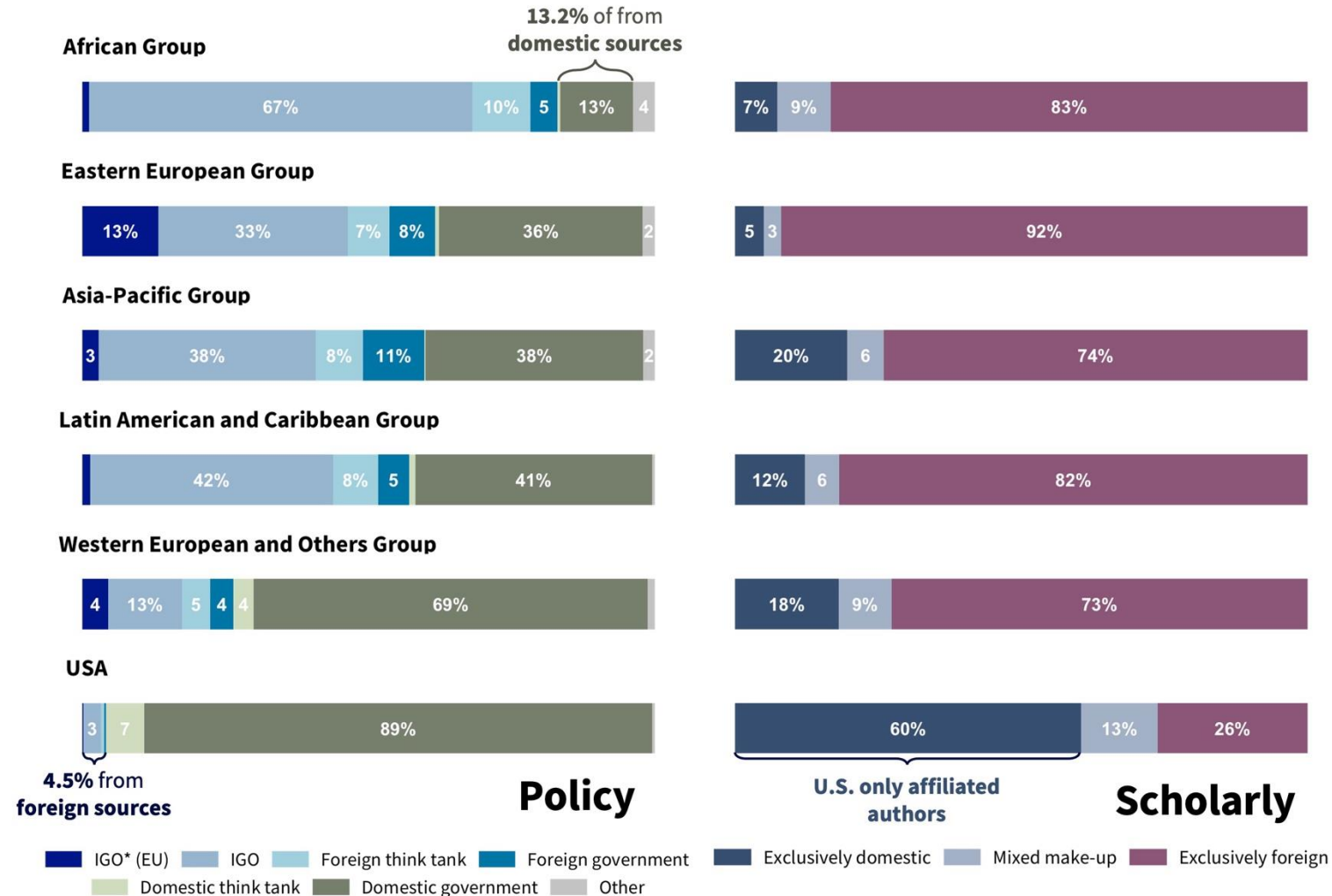


For ***scholarly*** citations, we identify institutional affiliation of authors and categorized them as:

- *Domestic*: All authors from the citing country
- *Mixed*: At least one domestic affiliation
- *Foreign*: No domestic affiliations

What did we find?

1. To what extent do governments rely on **domestic** versus **foreign** sources of evidence?



What did we find? 2. Which countries' **policy** and **scientific outputs** are **most** frequently **cited** and therefore **visible**?

What did we find?

2. Which countries' **policy** and **scientific outputs** are **most** frequently **cited** and therefore **visible**?



For ***policy-based*** citations, we construct a **government-to-government** citation matrix:

- We calculate an *H-index* to quantify international reach and also apply network centrality metrics (e.g., betweenness, eigenvector, and PageRank)

What did we find?

2. Which countries' **policy** and **scientific outputs** are **most** frequently **cited** and therefore **visible**?

 For ***policy-based*** citations, we construct a **government-to-government** citation matrix:

- We calculate an *H-index* to quantify international reach and also apply network centrality metrics (e.g., betweenness, eigenvector, and PageRank)

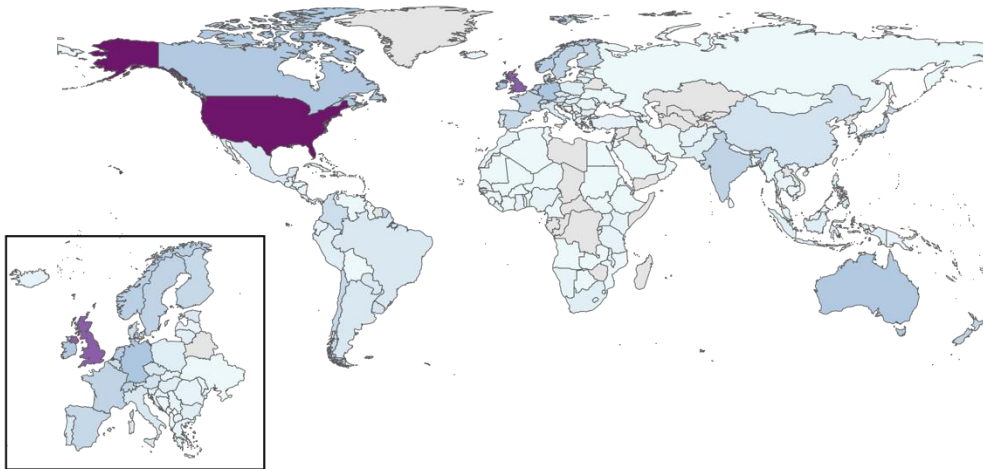
 For ***scholarly*** citations, we employ the **country affiliations of cited authors**:

- We compute the **share of total scholarly** references in the corpus that include at least one affiliated author from each country.

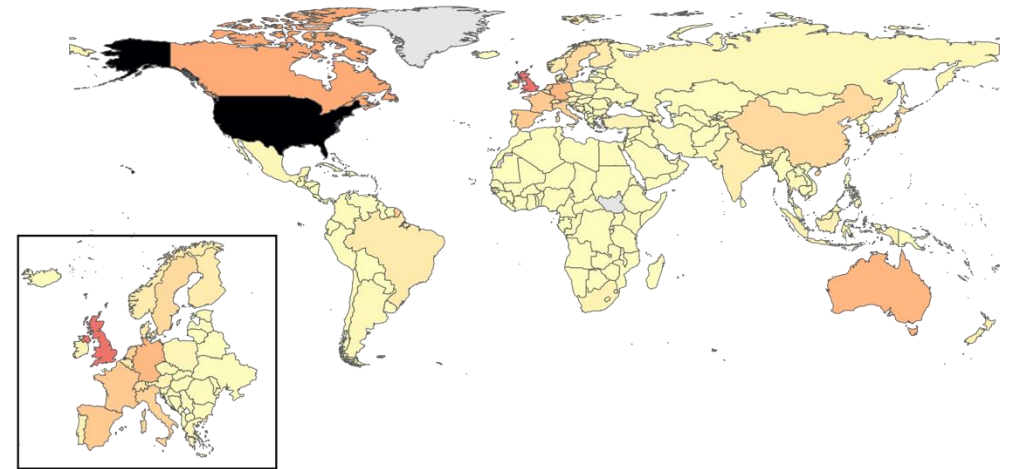
What did we find?

2. Which countries' **policy** and **scientific outputs** are **most** frequently **cited** and therefore **visible**?

a) Distribution of *H*-indexes government document references



b) Share of cited academic works with authors from country



What did we find?

2. Which countries' **policy** and **scientific outputs** are **most** frequently **cited** and therefore **visible**?

When we zoom-into policy 🏛️ and academic spaces 📖 :

What did we find?

2. Which countries' **policy** and **scientific outputs** are **most** frequently **cited** and therefore **visible**?

c) Countries with the highest *H*-indexes and share of academic affiliations

COUNTRY	H-I	N. REFS	GROUP
USA	44	10553	WEOG
UK	32	10508	WEOG
Germany	12	986	WEOG
Australia	11	1056	WEOG
Canada	11	1596	WEOG
Ireland	11	978	WEOG
Netherlands	10	663	WEOG
Norway	9	495	WEOG
France	8	706	WEOG
India	8	207	Asia-Pacific Group
New Zealand	8	533	WEOG
Sweden	8	967	WEOG
Switzerland	8	411	WEOG
Chile	7	171	GRULAC
Finland	7	502	WEOG
Japan	7	494	Asia-Pacific Group
Spain	7	628	WEOG
Belgium	6	412	WEOG
Colombia	6	283	GRULAC
Denmark	6	129	WEOG

Government-to-government references

COUNTRY	% DOCS	N. PAPERS	GROUP
USA	43.4%	1306064	WEOG
UK	13.3%	399890	WEOG
Canada	7.6%	227790	WEOG
Australia	6.6%	199213	WEOG
Germany	6.4%	191658	WEOG
Spain	4.9%	147406	WEOG
France	4.8%	143340	WEOG
Netherlands	4.2%	127181	WEOG
Italy	4%	120763	WEOG
China	3.3%	100288	Asia-Pacific Group
Sweden	3.3%	98589	WEOG
Japan	3.1%	94955	Asia-Pacific Group
Switzerland	2.4%	73578	WEOG
Belgium	1.9%	57785	WEOG
Denmark	1.8%	54042	WEOG
Brazil	1.8%	52669	GRULAC
Finland	1.7%	51935	WEOG
Norway	1.7%	50491	WEOG
India	1.6%	48793	Asia-Pacific Group
New Zealand	1.3%	39300	WEOG

Docs with author based at an insitution in country

When we zoom-into policy and academic spaces :

- 17 of the top 20 countries are shared across *policy-based* and *scholarly* outputs (largely *WEOG*)

What did we find? 3. How do citation patterns vary across **policy domains** with **differing knowledge demands**?

What did we find?

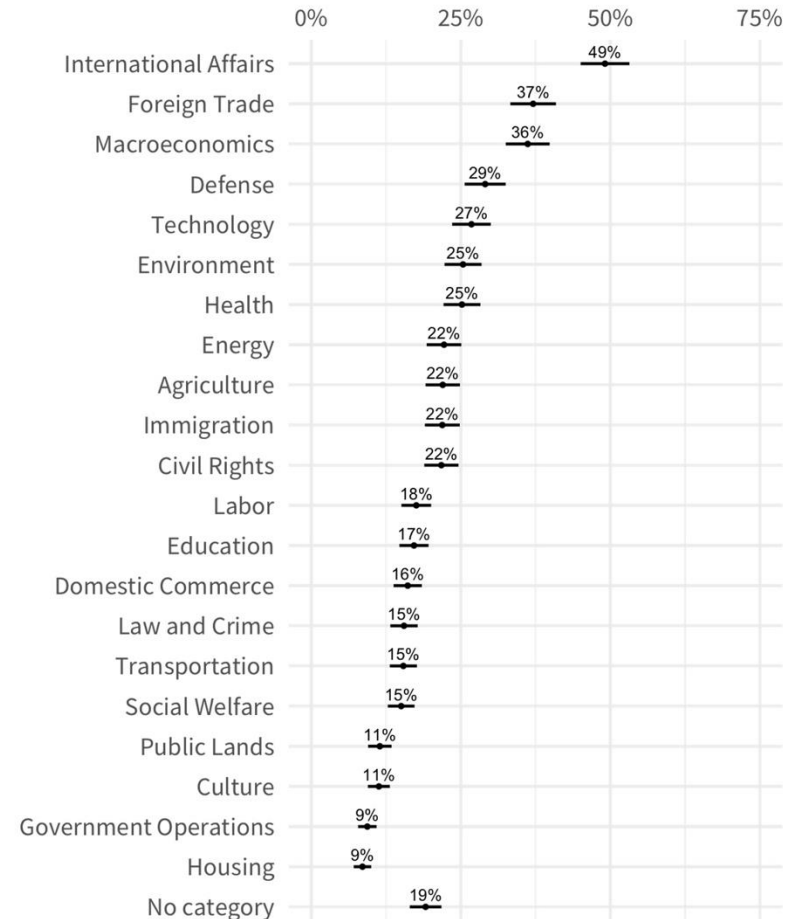
3. How do citation patterns vary across **policy domains** with **differing knowledge demands**?

 We employ policy document summaries to classify documents into their *most likely* **policy domain** from the *Comparative Agendas Project* (CAP) using an LLM (Llama3.3 70B).

What did we find?

3. How do citation patterns vary across **policy domains** with **differing knowledge demands**?

a) Policy-based sources

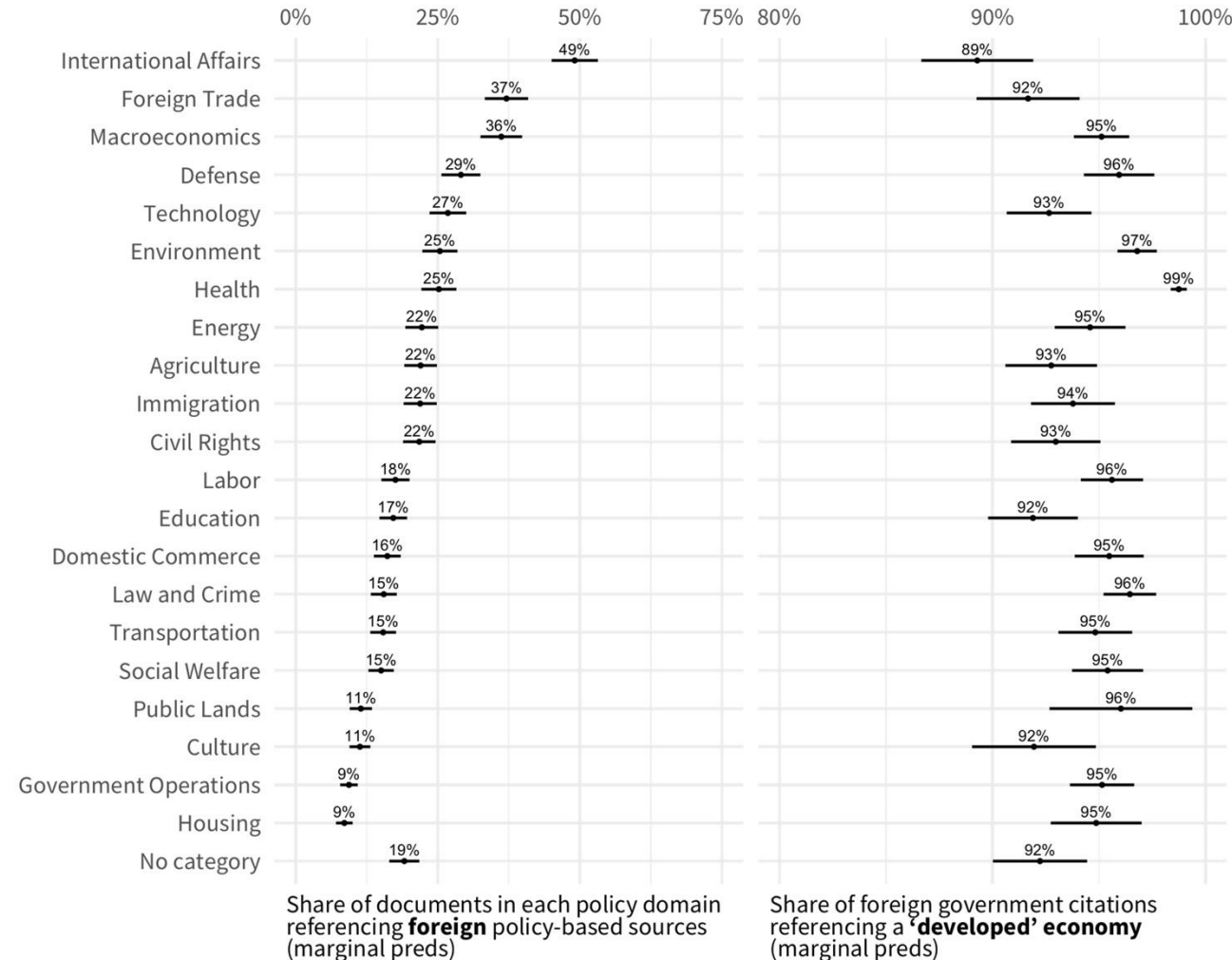


Share of documents in each policy domain
referencing **foreign** policy-based sources
(marginal preds)

What did we find?

3. How do citation patterns vary across **policy domains** with **differing knowledge demands**?

a) Policy-based sources



What did we find?

3. How do citation patterns vary across **policy domains** with **differing knowledge demands**?

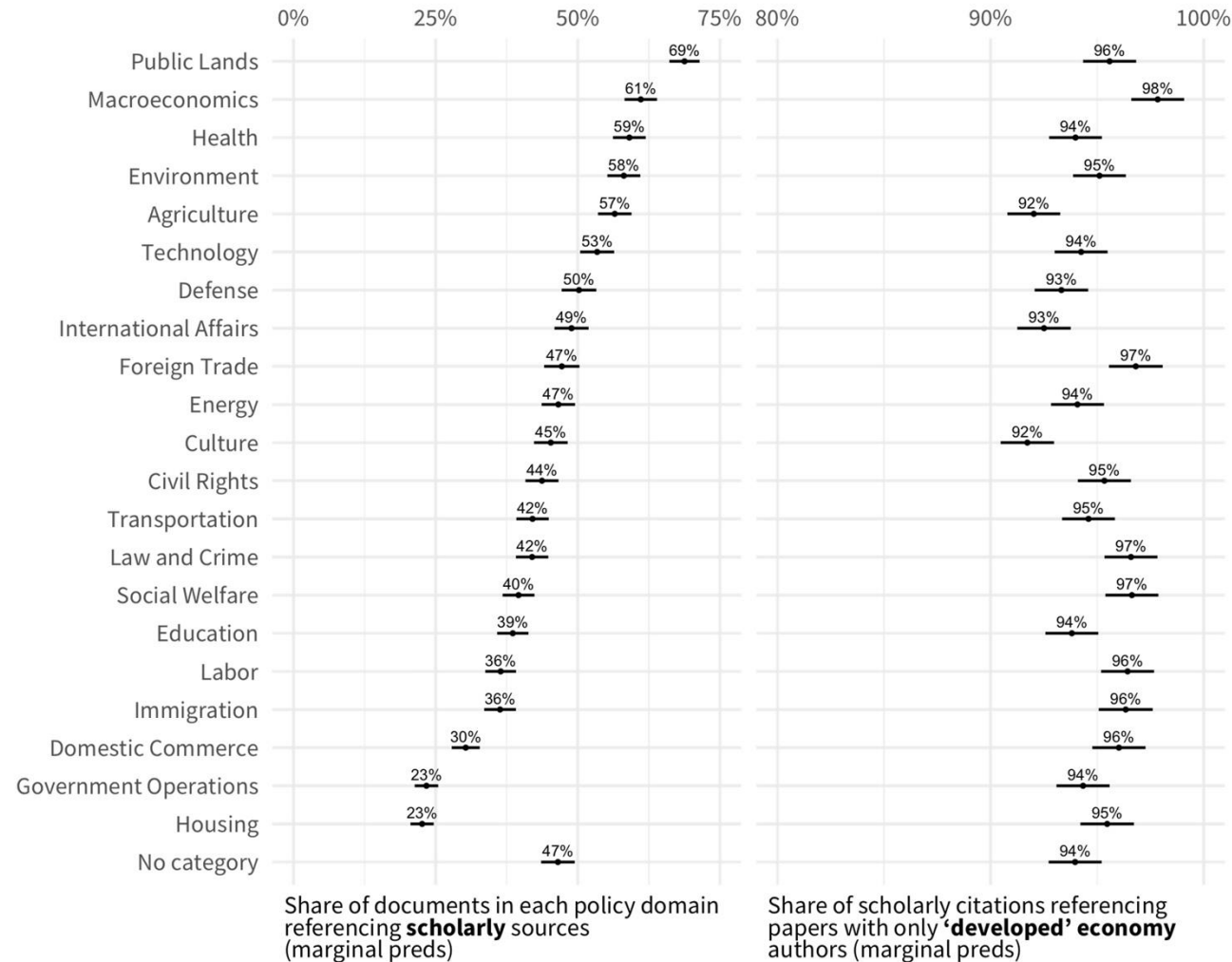
b) Scholarly sources



What did we find?

3. How do citation patterns vary across policy domains with differing knowledge demands?

b) Scholarly sources



In this project, we analyze data from **1.2 million government policy documents** from **185 countries** to find that:

In this project, we analyze data from **1.2 million government policy documents** from **185 countries** to find that:

- Governments in the global South more frequently cite **foreign materials**

In this project, we analyze data from **1.2 million government policy documents** from **185 countries** to find that:

- Governments in the global South more frequently cite **foreign** materials, while higher-income countries tend to draw from **domestic** expertise




In this project, we analyze data from **1.2 million government policy documents** from **185 countries** to find that:

- Governments in the global South more frequently cite **foreign** materials, while higher-income countries tend to draw from **domestic** expertise
- Countries in the global North (e.g., , , and ) lead in **production** of *policy-based & academic research* cited by governments worldwide.

In this project, we analyze data from **1.2 million government policy documents** from **185 countries** to find that:

- Governments in the global South more frequently cite **foreign** materials, while higher-income countries tend to draw from **domestic** expertise
- Countries in the global North (e.g., , , and ) lead in **production** of *policy-based & academic research* cited by governments worldwide.
- While some **policy domains** appear to be more *academic* (or *policy-based*) **evidence oriented**, the pattern holds

In this project, we analyze data from **1.2 million government policy documents** from **185 countries** to find that:

- Governments in the global South more frequently cite **foreign** materials, while higher-income countries tend to draw from **domestic** expertise
- Countries in the global North (e.g., , , and ) lead in **production** of *policy-based & academic research* cited by governments worldwide.
- While some **policy domains** appear to be more *academic* (or *policy-based*) **evidence oriented**, the pattern holds: evidence from economically powerful nations is consistently more prominent.

In the **draft**, we **engage** more with discussions relative to:

In the **draft**, we **engage** more with discussions relative to:

- *Documented, accessible, and digitally visible* policy evidence
(infrastructure of visibility)

In the **draft**, we **engage** more with discussions relative to:

- *Documented, accessible, and digitally visible* policy evidence (**infrastructure of visibility**)
- *Demand-side and supply-side* constraints (**preferences** for 'localized' evidence vs. **ability to act on them**)

In the **draft**, we **engage** more with discussions relative to:

- *Documented, accessible, and digitally visible* policy evidence (**infrastructure of visibility**)
- *Demand-side and supply-side* constraints (**preferences** for 'localized' evidence vs. **ability to act on them**)
- Fragility/robustness of academic and bureaucratic ecosystems of *global reference points* (**contestation** and **politicization**)

In the **draft**, we **engage** more with discussions relative to:

- *Documented, accessible, and digitally visible* policy evidence (**infrastructure of visibility**)
- *Demand-side and supply-side* constraints (**preferences** for 'localized' evidence vs. **ability to act on them**)
- Fragility/robustness of academic and bureaucratic ecosystems of *global reference points* (**contestation** and **politicization**)
- Concentration in the solution space -> **risk** of *local peaks*

In the **draft**, we **engage** more with discussions relative to:

- *Documented, accessible, and digitally visible* policy evidence (**infrastructure of visibility**)
- *Demand-side and supply-side* constraints (**preferences** for 'localized' evidence vs. **ability to act on them**)
- Fragility/robustness of academic and bureaucratic ecosystems of *global reference points* (**contestation** and **politicization**)
- Concentration in the solution space -> **risk** of *local peaks*

much more...

Thank you!



*Take a look at the
working draft!

We would really
appreciate your
feedback.



seramirezruiz.github.io



Sebastian.ramirezruiz@eui.eu



seramirezruiz.bsky.social

Auxiliary information

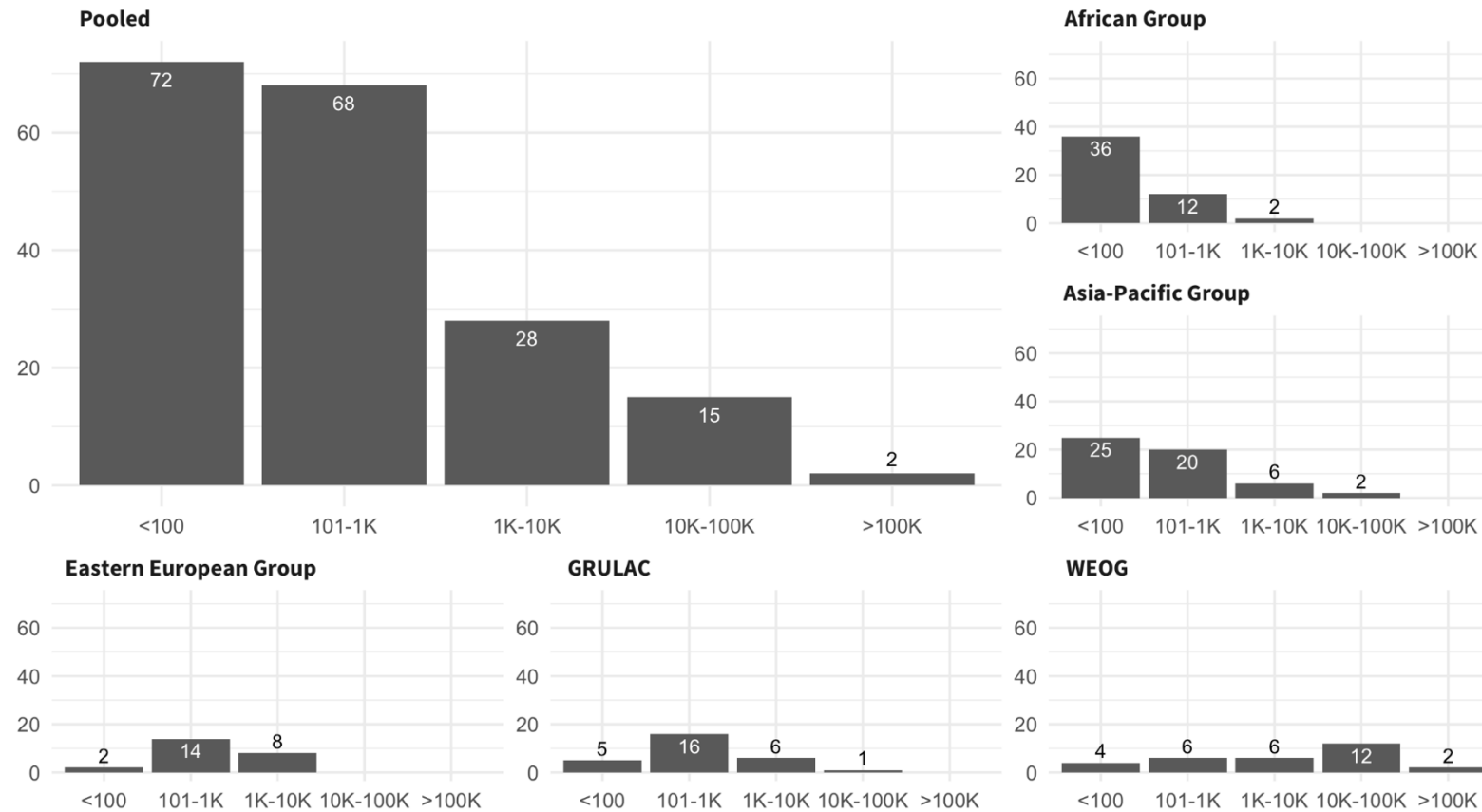


Figure A1: Overview of distribution of government authored policy documents. This figure shows the number of countries categorized by the total of policy documents collected, both in the full sample and grouped by UN Regional levels.

Auxiliary information

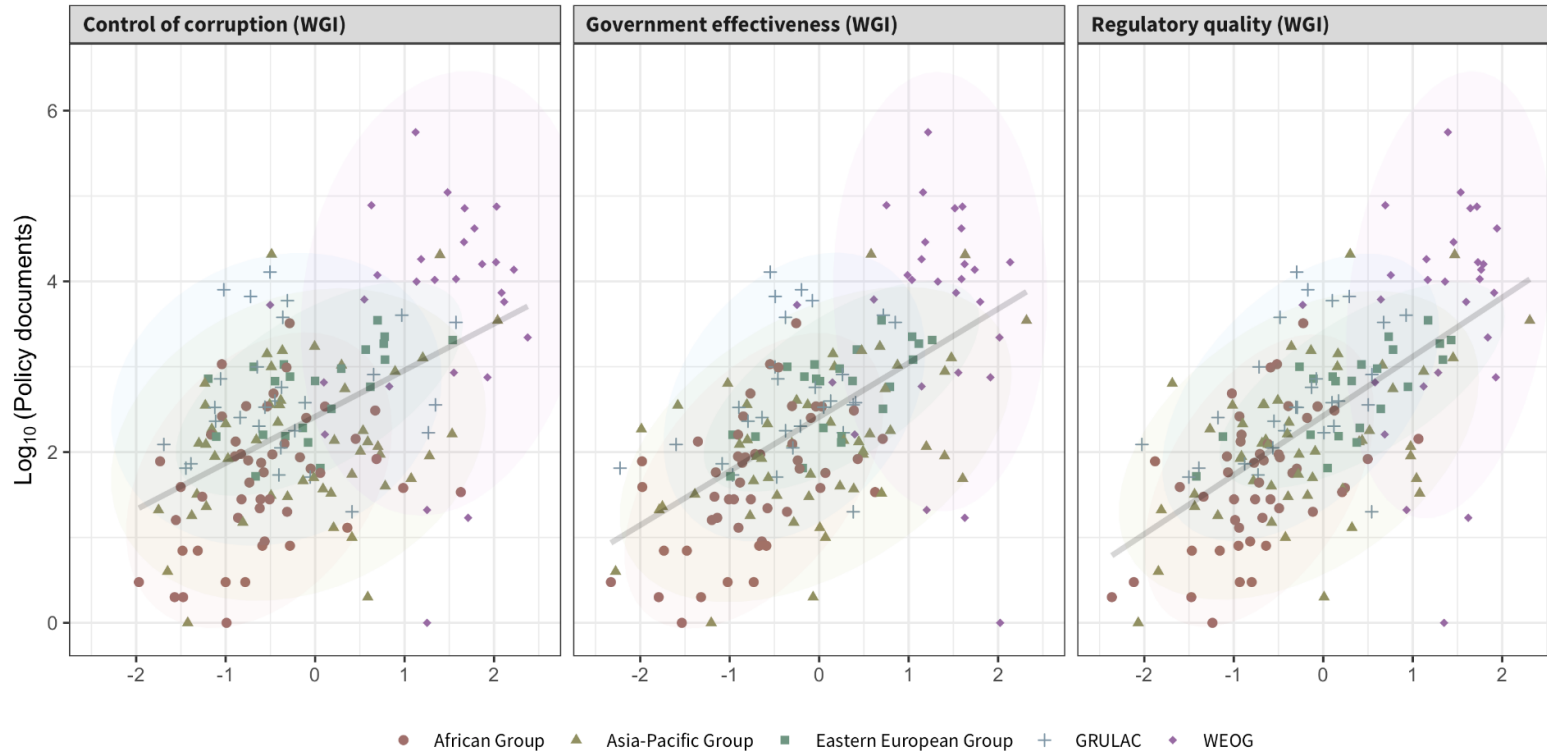


Figure A2: Relationship between Worldwide Governance Indicators (WGI) [1] and policy documents available across countries. Each dot represents individual countries, with ellipses at the UN Region-level.

Auxiliary information

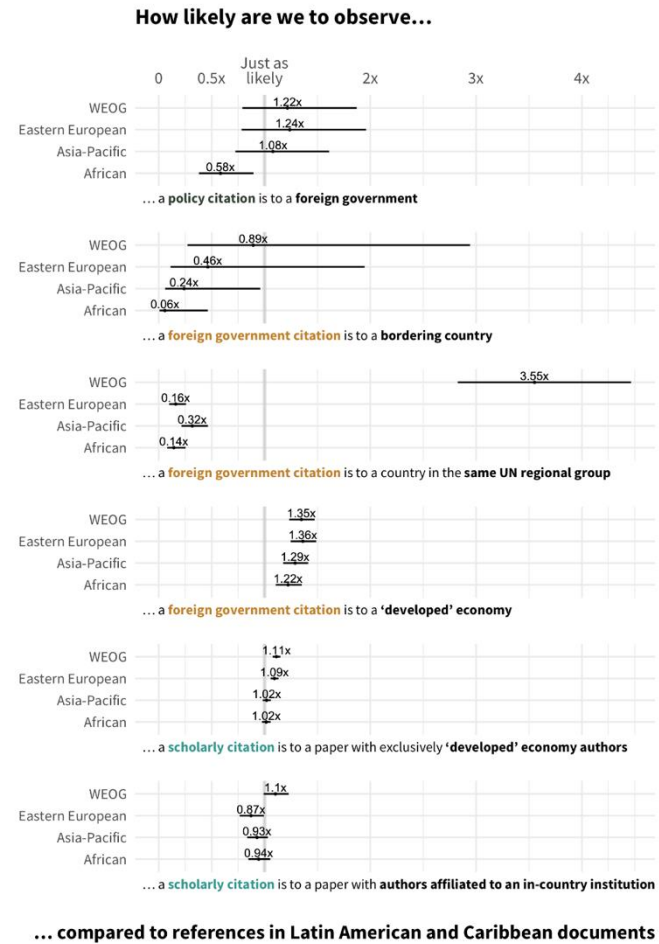


Figure A3: Marginal effects on observing a reference with a specified feature across groups. Results from six logistic mixed-effects models with document government author random effects. The estimates in the figure are relative risks representing the ratio of probability of observing a reference with the characteristic in a UN region group to the probability of observing it in documents authored by Latin American and Caribbean countries.

Auxiliary information

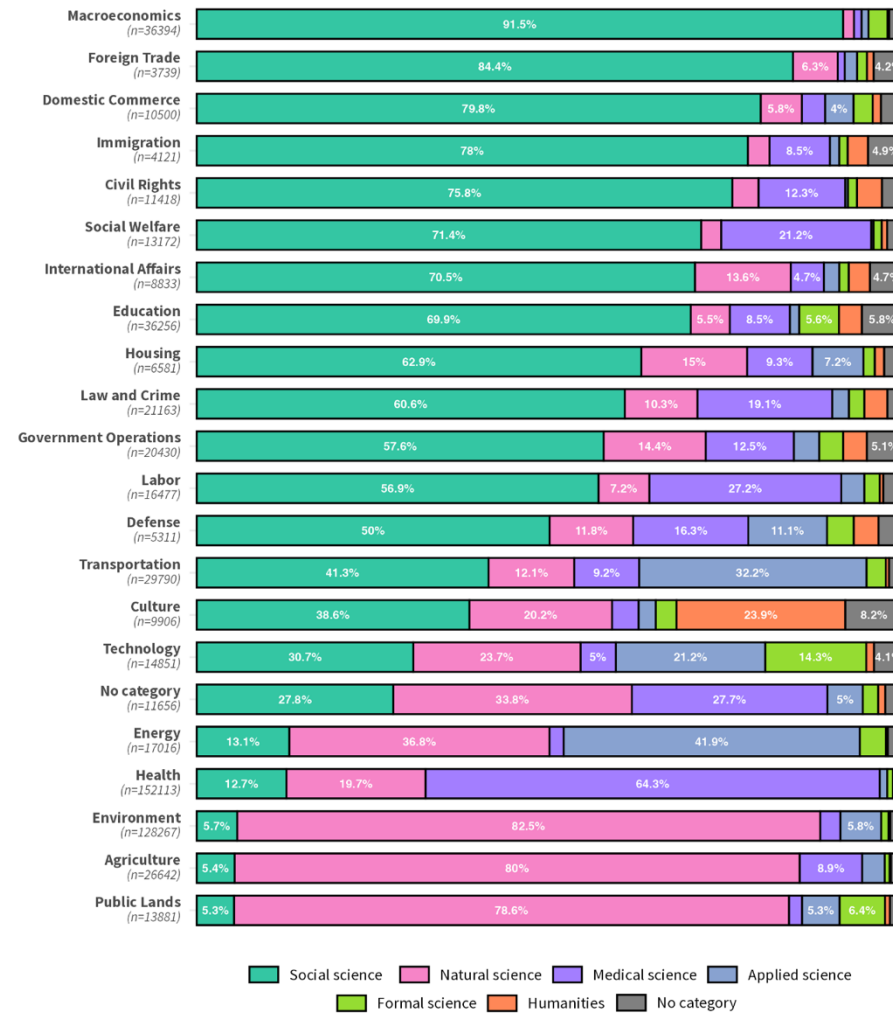


Figure B1: Share of references to scientific fields across policy domains.

Auxiliary information

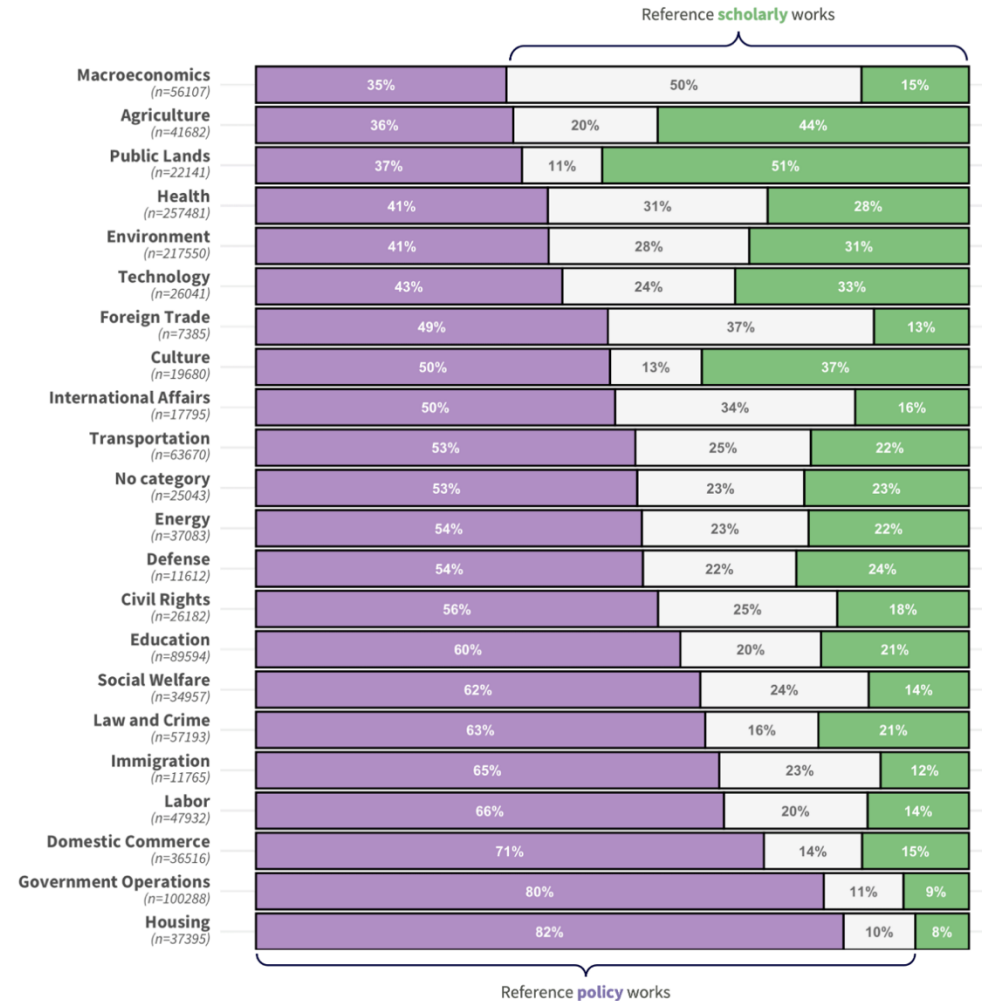


Figure B2: Share of policy documents citing policy, scholarly, or mixed sources across policy domains.