The Who, What, When and How of Industrial Policy
A Text-Based Approach

Réka Juhász¹,⁴,⁵, Nathan Lane², Emily Oehlsen², and Verónica C Pérez³
¹ UBC
² Oxford University
³ Columbia University
⁴ NBER
⁵ CEPR
Should states shape the composition of economic activity?

This idea—industrial policy—has been controversial since 18th century.
Should states shape the composition of economic activity?

A) Essential for promoting economic change, development  
   (Hamilton 1791, List 1856, Taussig 1914, Amsden 1989, Wade 1990)

B) Modern economics: likely harmful, impractical, prone to failure  
   (Krueger 1990, Pack 2000)

“The best industrial policy is none at all”

Gary Becker (1985)
"The best industrial policy is none at all."
Gary Becker, 1985
Motivation

Should states shape the composition of economic activity?

A) Essential for promoting economic change, development

B) Modern economics: harmful, impractical
   *(Krueger 1990, Pack 2000)*
   
   “The best industrial policy is none at all”
   *Gary Becker (1985)*

Despite importance: little systematic evidence on what industrial policy can achieve.
Thus, we focus on a bottleneck: measurement
Thus, we focus on a bottleneck: measurement

But why? Why is measurement fundamental?
Key empirical questions need measurement*

1. **One set of questions is about testing economic mechanisms (progress)**
   - E.g. Are market failures relevant?
   - Studied using careful natural experiments: Juhasz 2018, Hanlon 2020, Mitrunen 2021, Lane 2022, etc...
   - Use single episodes: Kalouptsidi 2018, Jia Barwick et al. 2020

2. **But another set of questions is about real-world implementation (little evidence)**
   - E.g. Are government failures relevant? E.g. political economy and informational issues.
   - Studied using comparing instances of IP.
   - Requires multiple episodes.

*Type-2 questions require measuring industrial policy in real world.*
Key empirical questions need measurement*

1. One set of questions is about testing economic mechanisms (progress)
   - E.g. are market failures relevant? Does it work as theory predicts?
   - Use single episodes: Kalouptsidi 2018, Jia Barwick et al. 2020

2. But another set of questions is about real-world implementation (little evidence)
   - E.g. Are government failures relevant? Why does success vary?
   - Studied by comparing across cases.
   - Requires multiple episodes.

*Type-2 questions require measuring industrial policy in real world.
Key empirical questions need measurement*

1. One set of questions is about testing economic mechanisms (progress)
   ○ E.g. are market failures relevant? Does it work as theory predicts?
   ○ Use single episodes: Kalouptsidi 2018, Jia Barwick et al. 2020

2. But another set of questions is about real-world implementation (little evidence)
   ○ E.g. Are government failures relevant? Why does success vary?
   ○ Studied by comparing across cases.
   ○ Requires multiple episodes.

*Type-2 questions require measuring industrial policy in real world.
Motivation

Problem:

No cross-country data on industrial policy.
Motivation

Problem:
No cross-country data on industrial policy.

Measurement challenge:
Industrial policy is not directly observable.
Consider tariffs:
classic instrument of infant industry promotion
Consider tariffs: classic instrument of infant industry promotion

But used for revenue (Cage and Gadenne 2018), ToT rationales (Broda, Limao, Weinstein 2008), and more.

Without additional information, hard to know if a tariff is an industrial policy.
This study delivers: measurement + basic facts.

1. Measurement proof of concept.
   - Develop algorithm to classify IP from text.
   - ... scale it to a global commercial policy database.
   - ... to construct new measures at the country-industry-year level.

2. Basic stylized facts for first time. Like,
   - Ex: IP is common in rich countries, less in poor.
   - ... IP is correlated with revealed comparative advantage.
Our approach: considers policy language, over policy measures per se.

Why text?

A. IP is state action meant to change the structure of an economy.

B. ... this goal-oriented nature of policy is conveyed in language (text).
Example:

Consider a summary of a Chinese subsidy programme:

"In the PRC Ministry of Industry and Information Technology's policy released on the 1st of March 2017, a plan is laid out to boost growth in the Chinese battery industry, specifically, batteries for automobiles [...]"

Policymaker: changing composition of economic activity by boosting a particular sector.
Examples:

“The Programme ‘National Champions’ was developed as part of the framework [...] of the President of the Republic of Kazakhstan, Mr. Nazarbayev.”

“The Thailand Board of Investment approved on June 10, 2009 further incentives aimed at making Thailand the automobile manufacturing hub in Asia.”

“Uzbekistan plans to invest within next 5 years about 1.4 billion US dollars to the domestic automobile industry. [...] The automobile industry was announced to be one of the priorities of the industrial development”

“In the New Growth Strategy of Japan (2010), the Japanese government announced that it would promote the use of Japanese wood instead of foreign wood.”

“On 6 June 2009, the Ministry of Information Industry (MII) of the People's Republic of China (PRC) issued a Planning Release [...] The release [...] seeks to provide guidance on maintaining and strengthening the PRC's position in the global ship-building industry.”

“On 29 May 2012, the Ministry of Textiles approved a INR 35,000 crores (USD 6 billion) restructuring package for the debt owed by the domestic textile industry [...] The stated goal of the scheme is to improve the participating companies' working capital positions.”
“On 29 September 2015, the Indonesian government announced a second stimulus package within weeks, which was supposed to boost the slowing economy and the country’s weakening currency.”

“The new Zambian authorities claimed that the sale of Zamtel was fraught with irregularities[...].”

“The new Zambian authorities claimed that the sale of Zamtel was fraught with irregularities[...].”

“On July 28th, 2010, the Brazilian government announced temporary tariff reduction on the ad-valorem import duty from 12% to 2% for flat-rolled products of iron[...]. It is important to note that this tariff reduction was implemented for reasons of shortage.”

“On August 21, 2013, the Government of the Russian Federation[...] approved the issuance in 2013 of 8 state guarantees fully covering the loans of 7 defence enterprises, as a part of the State Armament Programme 2011-2020.”

“[...] import duty exemption on several medical goods for prevention and control of the COVID-19 virus.”

“The Nigeria Ports Authority has introduced a Cargo Tracking Note System which it says will improve cargo security and safety.”
... As opposed to this Thai subsidy programme:

“The Thai Ministry of Commerce announced a reduction in rice-growing zones to stabilise rice prices [...] In addition, the Ministry has encouraged rice mills to purchase paddy directly from the farmers in an effort to supplement the farmers' income [...]”

Policymaker wants to stabilize rice prices in response to an oversupply of rice.
Our approach: uses NLP to discern policies at scale.

“In the PRC Ministry of Industry and Information Technology's policy released on the 1st of March 2017, a plan is laid out to boost growth in the Chinese battery industry, specifically, batteries for automobiles. [...]”

“The Thai Ministry of Commerce announced a reduction in rice-growing zones to stabilise rice prices [...]”

We develop + apply a classification algorithm to global database of policies to create indices of IP at the country-sector-year level.
1. **Empirical study of IP**
   - Individual case studies; IP is absent or unmeasured (Juhasz 2018; Criscuolo et al. 2019; Hanlon 2020; Mitrunen 2019; Lane 2021; Kalouptsidi 2018; Barwick, Kalouptsidi and Zahur 2019)
     ➢ We provide methods for measuring IP and shows basic stylized facts; Consider type 2 questions.

2. **Trade policy and development** (Harrison and Rodríguez-Clare, 2010; Goldberg and Pavcnik 2016; Verhoogen 2021)
   ➢ We address key deficits in literature: move beyond import tariffs and introduce a new way to capture NTMs

3. **Developmental state**
     ➢ We provide quantitative complement to seminal qualitative work

4. **Text-as-data to measure policy quantities through NLP + text** (Gentzkow, Kelly, Taddy 2019).
   - Dictionary-based approaches for policy uncertainty (Baker, Bloom, Davis 2016) and political risk (Hassan et al. 2019).
     ➢ We use supervised learning workflow to produce policy quantitued (ala Grimmer, Roberts, Stewart 2022).
Current Pipeline

Establishing basic facts

Comparative study of East Asian growth: Juhasz, Lane, Marczinek, Yang.

Comparative study of Latin America: (e.g. Ferraz, Juhasz, Lane).

This study *
A measurement proof of concept.

Validation, expanding: UNIDO, beyond

PE of current practice:
E.g. Juhasz, Lane, Rodrik (Annual. Rev.)

Expanding measurement and descriptive analysis.

Applying measures for comparative analysis.
Introduction

Defining IP

Data

Classifying IP

Model Results

Descriptive Statistics
Recall, our concept of IP:

*Concept:* IP is intentional state action directed at changing the structure of the domestic economy

A clear concept helps guide a supervised learning workflow.
1. **Stated goal**

A policy (i) seeks to change the relative prices *across sectors* or direct resources towards certain *selectively targeted activities* (e.g., exporting, R&D), with (ii) the purpose of *shifting the long-run composition of economic activity*.

2. **Level of government**

Policy action is taken by a *national or supranational government or organizations* that derive their operational scope and/or financing from those levels of government.

Not uniquely ours: (USITC 1980s; OECD 2000s)
Introduction

Defining IP

Data

Classifying IP

Model Results

Descriptive Statistics
Data: Lean on a public source of mass textual data.

We use Global Trade Alert database, an independent org. monitoring global commercial policy. Mass policy surveillance.

Key: tracks any economic policy that discriminates against foreign commercial interests.
Data: Advantages of GTA

Large global reach + comprehensive NTM source:
- Uses state sources, NGOs, media. Does not rely on reporting compliance.
- Most comprehensive data on non-tariff measures (~28,000 policies)

Suited for measuring IP:
- Covers scope of IP; IP typically impacts foreign commercial interests.
- Policy flows, not stock (some advantages).
- Gives short policy summaries in English.
3 Pieces of Data: GTA text + trade and income data

1. **Text summaries**
   - 100-200 word English summaries of policy in standardized format.

2. **Plus “meta-data” (country, year, sector)**
   - Country, year, measure type (e.g. tariff or subsidy)
   - Products, HS6 code, and eligible firms (e.g. firm v. sector)
   - Level of implementation (Part 2 of the definition).

3. **Merged with aggregates**
   - Trade data (HS2/6) from BACI (UN COMTRADE)
   - GDP per capita from World Bank, etc.
Introduction
Defining IP
Data

Classifying IP
Model Results
Descriptive Statistics
We want a classifier that maps text to a prediction,

“In the PRC Ministry of Industry and Information Technology's policy released on the 1st of March 2017, a plan is laid out to boost growth in the Chinese battery industry, specifically, batteries for automobiles. [...]”

“The Thai Ministry of Commerce announced a reduction in rice-growing zones to stabilise rice prices [...].”

In other words, we want $f : X_{text} \rightarrow Y$
Classifying IP: From text $\rightarrow$ prediction

Four step workflow (simplified)

1. Annotate $\rightarrow$ 2. Text-as-data $\rightarrow$ 3. Train model $\rightarrow$ 4. Predictions

We get our function $f : X_{text} \rightarrow Y$

using a 4-step supervised learning workflow.
Step 1: Annotate data

Hand label ~2,000 (~7%) examples of text from GTA database.
E.g. label observations as “IP” or “not IP.”

Yields training data: \( (X_{labeled}, Y_{labeled}) \subset (X, Y) \)
**Step 1: Annotators and Consistent Classification**

- Trained team of 4 annotators.
  - Oxford + Columbia students.
  - Codebook + initial 30 minute training with PIs
  - 200 trial entries. Feedback on coding for learning.

- Labeled >2,000 observations, stratified by measure type.

- Aggregated annotations with majority voting.
  - PI "expert-annotated" split decisions (n = 179)

- Result: Humans agree on what IP is using our formal definition.
  - Intercoder reliability (Krippendorff's alpha): 1st round: 0.57 - 0.59 → Next rounds: 0.76 - 0.77
Step 2: represent text as data.

Next, we pre-process text and represent it numerically.

We use Unigrams + Bi-Grams with Bag of Words (BoW)-style (actually TF-IDF, or, Term Frequency Inverse Document Frequency) to do this.
"In the PRC Ministry of Industry and Information Technology’s policy released on the 1st of March 2017, a plan is laid out to boost growth in the Chinese battery industry, specifically, batteries for automobiles [...]

Example: text summary ("document")

Bag-of-words vector (shown without weighting)

Can represent summaries as a matrix

(More processing)
Classifying IP: From text $\rightarrow$ prediction

**Step 3: train binary classifier.**

Use labeled data (Step 1) to train a classifier.

We use the training data $(X_{labeled}, Y_{labeled})$ to train or fit model $\hat{f}$.

Simplest $\hat{f}$: binary logistic regression classifier.
Step 3. Training: simplest setup

For simplicity, we use \textit{logistic regression} + regularization with binary outcomes,

\[
\ln \left( \frac{P_{IP}}{1 - P_{IP}} \right) = X\beta + \epsilon
\]

\textit{Odds that text summary is IP}

which we fit on our labeled data \((X_{labeled}, Y_{labeled}) \subset (X, Y)\) from step 1, to get \(\hat{f}\)

*Text representations, binary classification, and the classifier can be infinitely more complicated.*
3. Training: binary + logistic setup, interpretation

Train model using subset of labeled data.

Logistic model coefficients (\(\hat{\beta}_2\) hat, \(\hat{\beta}_8\) hat) tell us how terms contribute to prediction.

Matrix of labeled summaries.

\[
\begin{bmatrix}
1 \\
1 \\
0 \\
0 \\
\vdots \\
\vdots \\
y_n
\end{bmatrix}
\begin{bmatrix}
X_{labeled}
\end{bmatrix}
= Y_{labeled}
\]

"In the PRC Ministry of Industry and Information Technology's policy released on the 1st of March 2017, a plan is laid out to boost growth in the Chinese battery industry, specifically, batteries for automobiles [...]"

\(\hat{\beta}_2 = -.2\)
\(\hat{\beta}_8 = .2\)

[Cross-validated with test/train split, choose optimal parameters, deal with class imbalance, etc.]
Classifying IP: From text → prediction

**Step 4: predict IP in full data**


Use trained model (Step 3) to predictions each obs. in total dataset,

\[ \hat{f}(X_{text}) = \hat{Y} \]

These predictions \( \hat{Y} \) give counts at sector-year-country level.
Model Results

Descriptive Statistics
## Step 4. Prediction: Performance

### Performance - Simplest Classifier

<table>
<thead>
<tr>
<th></th>
<th>No IP Goal</th>
<th>IP Goal</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Accuracy</strong></td>
<td>0.93</td>
<td>0.93</td>
</tr>
<tr>
<td><strong>Precision</strong></td>
<td>0.96</td>
<td>0.87</td>
</tr>
<tr>
<td><strong>Recall</strong></td>
<td>0.95</td>
<td>0.88</td>
</tr>
<tr>
<td><strong>F1</strong></td>
<td>0.96</td>
<td>0.87</td>
</tr>
</tbody>
</table>

Baseline logistic model using unigrams and bigrams + cross validation and grid search.

**Accuracy** - Total correct predictions (true positives + true negatives) as share of all predictions.

**F1 Score** - Alternative performance measure; how good both precision and recall are together.

→ Both closer to 1 the better.
### Most Predictive Coefficients

<table>
<thead>
<tr>
<th>$\beta$ Coefficient</th>
<th>Feature name</th>
</tr>
</thead>
<tbody>
<tr>
<td>13.9</td>
<td>tech</td>
</tr>
<tr>
<td>13.7</td>
<td>green</td>
</tr>
<tr>
<td>12.7</td>
<td>project</td>
</tr>
<tr>
<td>10.9</td>
<td>export</td>
</tr>
<tr>
<td>10.2</td>
<td>million</td>
</tr>
<tr>
<td>10.1</td>
<td>plant</td>
</tr>
<tr>
<td>9.8</td>
<td>lobster</td>
</tr>
<tr>
<td>9.5</td>
<td>loan</td>
</tr>
<tr>
<td>9.4</td>
<td>technology</td>
</tr>
<tr>
<td>9.2</td>
<td>development</td>
</tr>
</tbody>
</table>

Rank of top-10 terms with largest coefficients.

Logistic classifier: allows us to “see” words most predictive of IP (and least)

Several word features - *e.g.*, *tech, technology, development, green* - seem reasonable for IP.
We believe IP has increased — We see this in data.
IP does not spike with pandemic response policies.

Step 4. Prediction: Smell tests
Introduction
Defining IP
Data
Classifying IP
Model Results

Descriptive Statistics
Descriptives: the contours of IP.

What can we say with the data? What does it reveal?
The contours of IP.

1. How much IP is there?

2. With what policy tools is IP conducted?

3. Which countries engage in IP?

4. Industrial policy is selective. Correlated with RCA and productivity.
The contours of IP.

1. How much IP is there?

2. With what policy tools is IP conducted?

3. Which countries engage in IP?

4. Industrial policy is selective. Correlated with RCA and productivity.
1. How common is IP?

~25% of policies in the database are classified as IP; IP ≠ NTM-s
The contours of IP

1. How much IP is there?

*IP is common (25%) and rising (trending up).*

2. With what policy tools is IP conducted?

3. Which countries engage in IP?

4. Industrial policy is selective. Correlated with RCA and productivity.
The contours of IP

1. How much IP is there?

2. With what policy tools is IP conducted?

3. Which countries engage in IP?

4. Industrial policy is selective. Correlated with RCA and productivity.
IP is overwhelmingly subsidies and export measures.

- Capital control measures
- Contingent trade-protective measures
- Export-related measures
- FDI Measures
- Government procurement restrictions
- Instrument unclear
- Intellectual property
- Migration measures
- Non-automatic licensing, quotas, prohibitions
- Price-control measures
- Subsidies (excluding export subsidies under P7)
- Tariff Measures
- Trade-related investment measures

... Tariffs are a poor measure of modern IP.
2. What

Sectoral IP is granular: skewed toward firm-specific policies v. sector-specific. Over half.
The contours of IP

1. How much IP is there?

2. **With what policy tools is IP conducted?**

   *IP tends to be export promotion/subsidies, and *firm-specific* (> 50% of IP)*

3. Which countries engage in IP?

4. Industrial policy is selective. Correlated with RCA and productivity.
Stylized facts

1. How much IP is there?

2. With what policy tools is IP conducted?

3. Which countries engage in IP?

4. Industrial policy is selective. Correlated with RCA and productivity.
3. Who?

Top countries engaged in IP

- Germany (5)
- Japan (5)
- Brazil (4)
- Canada (5)
- India (2)
- Russia (3)
- United States of America (5)
- Saudi Arabia (4)
- United Kingdom (5)
- Switzerland (5)
- Australia (5)
- China (3)
- Italy (5)
- Republic of Korea (5)
- Spain (5)
- France (5)
- South Africa (3)
- Argentina (5)
- Belarus (3)
- Turkey (4)

Not too dissimilar from OECD (2023) and CSIS (2022) analyses with smaller N.

(N) is income quintile; 5 being largest.
Richest countries dominate IP in our data

Total number of IP through time, by income quintile

3. Who

Quintiles based on 2010 GDP per capita. Higher quintiles are higher income.
Robust correlation between IP and income level

Controlling for # total policies

$$IP_c = \alpha + \sum_{i=2}^{5} \beta_i \text{IncomeQuintile}_i + \gamma X_c + \epsilon_c$$
3. Who

Richest countries dominate IP in our data

The correlation between IP and GDP per capita continues to hold when we control for the total number of policies captured by the GTA.

No controls

Controls

Notes: Controls include ln(total number of policies). ln(GDP per capita) measured in 2010.
The contours of IP.

1. How much IP is there?

2. With what policy tools is IP conducted?

3. Which countries engage in IP?

   Industrial policy is very common in wealthy countries

4. Industrial policy is selective. Correlated with RCA and productivity.
The contours of IP.

1. How much IP is there?

2. With what policy tools is IP conducted?

3. Which countries engage in IP?

4. Industrial policy is selective. Correlated with RCA and productivity.
IP is typically sectorally selective. Countries target a subset of sectors.

Top ranked for number of policies.

Share of sectors (HS codes) covered by IP

(N) is income quintile; 5 being highest.
4. What

IP is correlated with RCA, *but only for rich countries*

<table>
<thead>
<tr>
<th>Independent Variables</th>
<th>(1) IP = 1</th>
<th>(2) IP = 1</th>
<th>(3) IP = 1</th>
<th>(4) IP = 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>RCA</td>
<td>0.00011***</td>
<td></td>
<td>0.00002</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.00004)</td>
<td></td>
<td>(0.00002)</td>
<td></td>
</tr>
<tr>
<td>RCA &gt; 1</td>
<td></td>
<td>0.00933***</td>
<td></td>
<td>0.00080*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.00252)</td>
<td></td>
<td>(0.00043)</td>
</tr>
<tr>
<td>GDPpc &gt; Median × RCA</td>
<td></td>
<td></td>
<td>0.00041**</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.00018)</td>
<td></td>
</tr>
<tr>
<td>GDPpc &gt; Median × RCA &gt; 1</td>
<td></td>
<td></td>
<td></td>
<td>0.01662***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.00475)</td>
</tr>
</tbody>
</table>

CtryxYear FE | Yes | Yes | Yes | Yes
Observations | 195261 | 195261 | 168586 | 168586
R-squared | 0.347 | 0.348 | 0.346 | 0.347
Mean d.IP | 0.017 | 0.017 | 0.019 | 0.019
# of Countries | 183 | 183 | 158 | 158

Notes: Standard errors are clustered at the country level.
High-income countries target mass of products with *high comparative advantage*.

\[
IP_{kct} = \alpha + \sum_{i=2}^{5} \beta_i (RCA_{Quintile,kct}) + \delta_{ct} + \epsilon_{kct}
\]
4. What?

Targeting by Income Group

Lower Income Countries (Quintiles 1 and 2)

- Electrical machinery and equipment and parts ... etc.
- Nuclear reactors, boilers, machinery and mech... etc.
- Plastics and articles thereof
- Optical, photographic, cinematographic, measu... etc.
- Iron or steel articles
- Vehicles; other than railway or tramway roli... etc.
- Cotton
- Apparel and clothing accessories; not knitted... etc.
- Textiles, made up articles; sets; worn clothi... etc.
- Furniture; bedding, mattresses, mattress supp... etc.
- Apparel and clothing accessories; knitted or ... etc.
- Fabrics; special woven fabrics, tufted textile... etc.
- Articles of leather; saddlery and harness; tr... etc.
- Vegetable textile fibres; paper yarn and wove... etc.
- Fabrics; knitted or crocheted
- Copper and articles thereof
- Metal; miscellaneous products of base metal
- Glass and glassware
- Wadding, felt and nonwovens, special yarns; t... etc.
- Carpets and other textile floor coverings

Higher Income Countries (Quintiles 3 and 4)

- Nuclear reactors, boilers, machinery and mech... etc.
- Mineral fuels, mineral oils and products of t... etc.
- Electrical machinery and equipment and parts ... etc.
- Aircraft, spacecraft and parts thereof
- Optical, photographic, cinematographic, measu... etc.
- Vehicles; other than railway or tramway roli... etc.
- Ships, boats and floating structures
- Iron or steel articles
- Railway, tramway locomotives, rolling-stock a... etc.
- Plastics and articles thereof
- Iron and steel
- Organic chemicals
- Ores, slag and ash
- Furniture; bedding, mattresses, mattress supp... etc.
- Pharmaceutical products
- Tools, implements, cutlery, spoons and forks,... etc.
- Inorganic chemicals; organic and inorganic co...
- Chemical products n.e.s.
- Beverages, spirits and vinegar
- Dairy produce; birds' eggs; natural honey; ed... etc.
Some overlap: heavy and hi-tech industry.
4. What?

Targeting by Income Group

Lower Income Countries (Quintiles 1 and 2)
- 85 - Electrical machinery and equipment and parts ... etc.
- 84 - Nuclear reactors, boilers, machinery and mech... etc.
- 39 - Plastics and articles thereof
- 90 - Optical, photographic, cinematographic, measu... etc.
- 73 - Iron or steel articles
- 87 - Vehicles; other than railway or tramway rolli... etc.

- 52 - Cotton
- 62 - Apparel and clothing accessories; not knitted... etc.
- 63 - Textiles, made up articles; sets; worn clothi... etc.
- 94 - Furniture; bedding, mattresses, mattress supp... etc.
- 61 - Apparel and clothing accessories; knitted or ... etc.
- 58 - Fabrics; special woven fabrics, tufted textile... etc.
- 42 - Articles of leather; saddlery and harness; tr... etc.
- 53 - Vegetable textile fibres; paper yarn and wove... etc.
- 60 - Fabrics; knitted or crocheted
- 74 - Copper and articles thereof
- 83 - Metal; miscellaneous products of base metal
- 70 - Glass and glassware
- 56 - Wadding, felt and nonwovens, special yarns; t... etc.
- 57 - Carpets and other textile floor coverings

Higher Income Countries (Quintiles 3 and 4)

- 84 - Nuclear reactors, boilers, machinery and mech...
- 27 - Mineral fuels, mineral oils and products of t...
- 85 - Electrical machinery and equipment and parts 
- 88 - Aircraft, spacecraft and parts thereof
- 90 - Optical, photographic, cinematographic, measu...
- 87 - Vehicles; other than railway or tramway rolli...
- 89 - Ships, boats and floating structures
- 73 - Iron or steel articles
- 86 - Railway, tramway locomotives, rolling-stock a...
- 39 - Plastics and articles thereof
- 72 - Iron and steel
- 29 - Organic chemicals
- 26 - Ores, slag and ash
- 94 - Furniture; bedding, mattresses, mattress supp...
- 30 - Pharmaceutical products
- 82 - Tools, implements, cutlery, spoons and forks,...
- 28 - Inorganic chemicals; organic and inorganic co...
- 38 - Chemical products n.e.s.
- 22 - Beverages, spirits and vinegar
- 04 - Dairy produce; birds' eggs; natural honey; ed...

Differences: labour int. & textiles v. green IP and high-tech/heavy industry
The contours of IP.

1. How much IP is there?

2. With what policy tools is IP conducted?

3. Which countries engage in IP?

4. Industrial policy is selective. Correlated with RCA and productivity.
Conclusion

**Goal:** Proof of concept using text-as-data methods to consistently measure IP.

**Result:** Simple, off-the-shelf ML tools → usable count-based indices.

**Allows descriptive characterization of modern IP use:**

- IP is expensive: mostly subsidies and export measures, targeted towards firms and well-performing sectors.
- IP is almost non-existent in poor countries.
- Preliminary evidence that IP may be differently used in poorer countries

**Measurement matters:** To talk about comparative questions in IP.

**Next steps:** Doing this historically (Juhasz-Lane) in E.Asia and Lat. Am.
Establishing basic facts

Current Pipeline

This study *
A measurement proof of concept.

Validation, expanding: UNIDO, beyond

PE of current practice:
E.g. Juhasz, Lane, Rodrik (Annual. Rev.)

Comparative study of East Asian growth:
Juhasz, Lane, Marczinek, Yang.

Comparative study of Latin America:
(e.g. Ferraz, Juhasz, Lane).

Expanding measurement and descriptive analysis.

Applying measures for comparative analysis.
Thank you
Policies covered

Capital control measures
Contingent trade-protective measures
Export-related measures
FDI measures
Finance measures
Government procurement restrictions
Intellectual property
Migration measures
Non-automatic licensing, quotas, prohibitions and other
Price-control measures, including additional taxes and charges
Sanitary and phytosanitary measures
Subsidies (excluding export subsidies)
Tariff Measures
Technical barriers to trade
Trade-related investment measures

Back to talk
Developed a system for annotating policy descriptions with a i) codebook and ii) simple interface.

1. **Annotate**: Annotate entries from GTA database

Annotators Manual

**Codebook for identifying industrial policy intention from policy descriptions**

**Introduction**

You will be annotating, or coding, descriptions of economic policy. These policy descriptions you will code come from our Global Policy Alert (GTA) database. The following codebook introduces annotators (you) to the definitions and criteria used to code the intentionality of policy based on the measured description. Specifically, you will be coding whether or not policies show industrial policy intentions.

We take you through the coding process in two steps. First, we show you our working definition of “intentionality” and how to code the three types of policy intentionality. Second, we show you the rules for going about coding intentionality from policy descriptions.

The Ministry of Health has issued Decree No. 1010/08 regulating the registration and import of pharmaceutical products. The Decree establishes a separation between manufacturers and wholesalers to protect consumer health and the safety of pharmaceutical products. With regard to imports, initial registration must now be made through an Indonesian manufacturer. Once the registration process is completed a foreign company may directly sell to the wholesalers concerned.

- **IP intention**: 1
- **Other Intention**: 2
- **Not enough information**: 3

**PROJECT INFO**

**DATASET**

intentionality_all
textcat.manual
choice

**PROGRESS**

THIS SESSION

7

TOTAL

7

**ACCEPT**

0

**REJECT**

0

**IGNORE**

7

**HISTORY**

On 18 November 2008, India increased the...

The government increased the...

On August 30th, 2007, the Ind...

On November 19, 2007, the Ind...

On February 23, 2009, the Per...

© 2017-2021 Prodigy

**Back to talk**
1. How: Did we need a text-based measure?

Many policies are not IP. The type of policy does not predict IP.
2. **When:** Levels of IP for the U.S. and China?
2. When: Levels of IP without Green IP & R&D
3. Who: Few countries engage in IP

Many countries do no IP at all. A few countries account for the lion’s share of global IP.
3. Who: Who does Green IP and R & D

- United States of America (5)
- Australia (5)
- South Africa (3)
- China (3)
- Japan (3)
- Spain (3)
- France (3)
- Germany (3)
- Canada (3)
- Brazil (4)
- India (2)
- Sweden (5)
- United Kingdom (5)
- Norway (5)
- Russia (3)
- Republic of Korea (5)
- Denmark (5)
- Finland (5)
- Italy (5)
- Greece (5)

- Brazil (4)
- Japan (5)
- Russia (3)
- Canada (5)
- China (3)
- France (5)
- Spain (5)
- Italy (5)
- Germany (5)
- Kazakhstan (3)
- Sweden (5)
- Australia (5)
- United States of America (5)
- Finland (5)
- Denmark (5)
- Republic of Korea (5)
- India (2)
- Singapore (5)
- United Kingdom (5)
- Portugal (5)
3. **Who**: IP and GDP highly positively correlated

The correlation between IP and GDP per capita continues to hold when we control for the total number of policies captured by the GTA.

Notes: Controls include ln(total number of policies). ln(GDP per capita) measured in 2010.
3. Who: IP and GDP highly positively correlated

The correlation between IP and GDP per capita continues to hold when we control for the total number of policies captured by the GTA and dropping the Green & R&D IP.

Notes: Controls include ln(total number of policies). GDP per capita constant 2010.
3. Who: Correlation between IP and income level

The correlation between IP and GDP per capita continues to hold when we use coverage ratios to measure the magnitude of IP in the economy.
Total number of policies by income level
<table>
<thead>
<tr>
<th>Independent Variables</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>IP = 1</td>
<td>IP = 1</td>
<td>IP = 1</td>
<td>IP = 1</td>
</tr>
<tr>
<td>RCA</td>
<td>0.00011*** (0.00004)</td>
<td>0.00002 (0.00002)</td>
<td>0.00080* (0.00043)</td>
<td>0.01662*** (0.00475)</td>
</tr>
<tr>
<td>RCA &gt; 1</td>
<td>0.00933*** (0.00252)</td>
<td>0.00041** (0.00018)</td>
<td>0.00080* (0.00043)</td>
<td>0.01662*** (0.00475)</td>
</tr>
<tr>
<td>GDPpc &gt; Median × RCA</td>
<td></td>
<td>0.00041** (0.00018)</td>
<td>0.00080* (0.00043)</td>
<td>0.01662*** (0.00475)</td>
</tr>
<tr>
<td>GDPpc &gt; Median × RCA &gt; 1</td>
<td></td>
<td></td>
<td>0.00080* (0.00043)</td>
<td>0.01662*** (0.00475)</td>
</tr>
<tr>
<td>CtryxYear FE</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Observations</td>
<td>195261</td>
<td>195261</td>
<td>168586</td>
<td>168586</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.347</td>
<td>0.348</td>
<td>0.346</td>
<td>0.347</td>
</tr>
<tr>
<td>Mean d_IP</td>
<td>0.017</td>
<td>0.017</td>
<td>0.019</td>
<td>0.019</td>
</tr>
<tr>
<td># of Countries</td>
<td>183</td>
<td>183</td>
<td>158</td>
<td>158</td>
</tr>
</tbody>
</table>