Updated Descriptive Statistics for “The Who, What, When and How of Industrial Policy: A Text-Based Approach”

Réka Juhász (UBC, NBER & CEPR)
Nathan Lane (Oxford)
Emily Oehlsen (Oxford)
Verónica C. Pérez (Boston University)

First version: August, 2022
Version 2: July, 2023

This note provides an update to Juhász, Lane, Oehlsen, and Pérez (2022) “The Who, What, When and How of Industrial Policy: A Text-Based Approach”, henceforth JLOP (2022). Our model is evolving and our data—both source data and our output—are continuously updated. Any figures used from this note should cite the original working paper and include a note that the figures are produced using version 2 of the data and model.

This technical note has three sections. First, we describe changes to the input data and model relative to JLOP (2022). Second, we evaluate model performance. Third, we provide updated descriptive statistics through the end of year 2022.

1 Updates relative to the original model and source data used in JLOP (2022)

1.1 Changes to the Model

No changes were made to the model itself. We increased the size of the training dataset by adding new annotations. There are now in total 2,719 observations that have been hand-labeled. These are split between a training, testing and validation set.

1.2 Changes to the Source Dataset

The data were updated using a new version of the Global Trade Alert (GTA). The data were extracted April 4, 2023, and we use data for 2010-2022 (the complete years in the data).

The GTA is a living project and data is continuously updated. The source dataset now includes a substantially larger number of observations. The original dataset used in JLOP (2022) contained 28,333 observations; the new data contain 52,628 observations. The new data contain a larger number of observations both because of the addition of interventions from the years 2020 - 2022, and also because of updating and back filling of observations from earlier years.

Note that the text-based logistic classifier is trained and evaluated using annotations that were randomly drawn from the original GTA dataset we worked with. No new annotations were added for new years included in the updated version of the data.

2 Model performance

Table 1 reports updated performance metrics for the binomial logistic classifier. Table 2 contains the ten features with the highest coefficients. These are the features (N-grams) most predictive of industrial policy goals.

Table 1: Performance metrics for baseline model.

	Precision	Recall	F1-Score	Accuracy
No Industrial Policy Goal	0.98	0.97	0.98	0.96
Industrial Policy Goal	0.91	0.94	0.92	0.96

Table 2: Ten most predictive terms for baseline industrial policy classifier.

Feature Names	Coefficient Size
export	5.5
development	5.3
project	4.5
support	4.4
industry	3.4
million	3.6
energy	3.5
research	3.3
boost	3.3
technology	3.1

Notes: Coefficients come from the baseline classifier for a text-based, binomial logistic regression, and correspond to individual tokens. The text of these features are reported in the left column.

3 Descriptive Statistics

3.1 The time trend of industrial policy

These plots show the number of new industrial policy interventions by year. For the time-trend descriptives, we follow the guidance provided by the GTA; we use a version of the dataset that only includes policies entered into the dataset within the same calendar year. This is because the GTA is a growing, continuously updated dataset. This implies that at any given point in time, there is a more comprehensive inventory of policies for earlier years. To mitigate this type of reporting bias, the GTA advises the use of "like-for-like" comparisons whereby the researcher restricts the data to the same intra-year snapshots (Evenett & Fritz, 2018). We have chosen a cutoff date of December 31 for each calendar year. This means that for each year, only data entered by the GTA up to the end of year is included.[1]

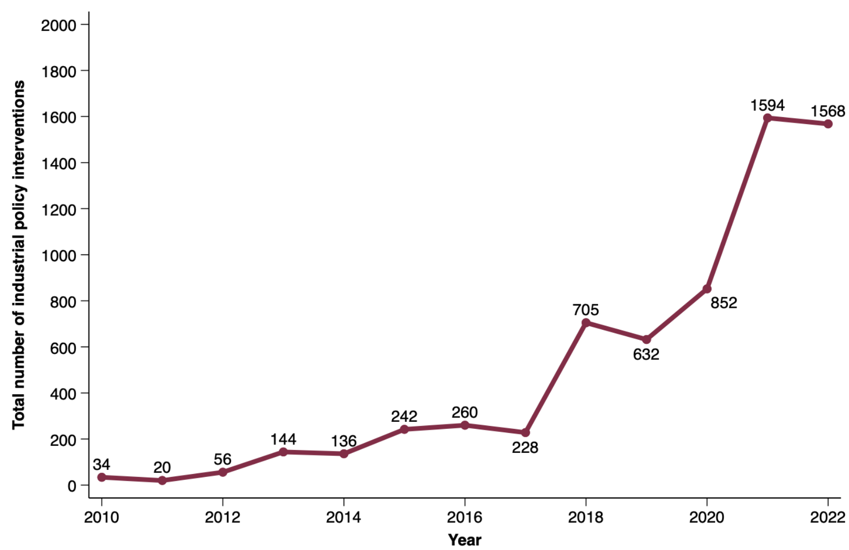


Figure 1: The Time Trend of Industrial Policy Interventions

Notes: The figure plots the total number of industrial policy interventions by year. Only entries entered in the same calendar year are included to ensure comparability over time. See text for details.

Figures 1-3 presents the time trend of total industrial policy interventions. Figure 1 shows that industrial policy has been growing through the 2010s with marked accelerations in 2018 and 2021. One concern is that the GTA may have become better at measuring policies over time. To this end, Figure 2 plots the share of industrial policy interventions among all interventions captured by the GTA. The increase in industrial policy is also evident using these share-based measures. Finally, Figure 3 shows that the recent surge in industrial policy is primarily driven by an increase in industrial policy interventions from the highest income quintile.

3.2 The income gradient of industrial policy

Figure 4 plots the total number of industrial policy interventions by income quintile and geographic region. We caution users of these data that high-income countries may be more readily observed because of more transparent and more granular reporting of interventions, as well as because of more attention from the GTA. For these reasons, we advise users of these data to focus on the findings from Figure 3, which examine the growth in the number of industrial policy interventions by income quintile. While measurement issues may not

[1]For example, observations for the year 2020 includes data recorded by the GTA between Jan. 1. 2020 until Dec. 31 2020.

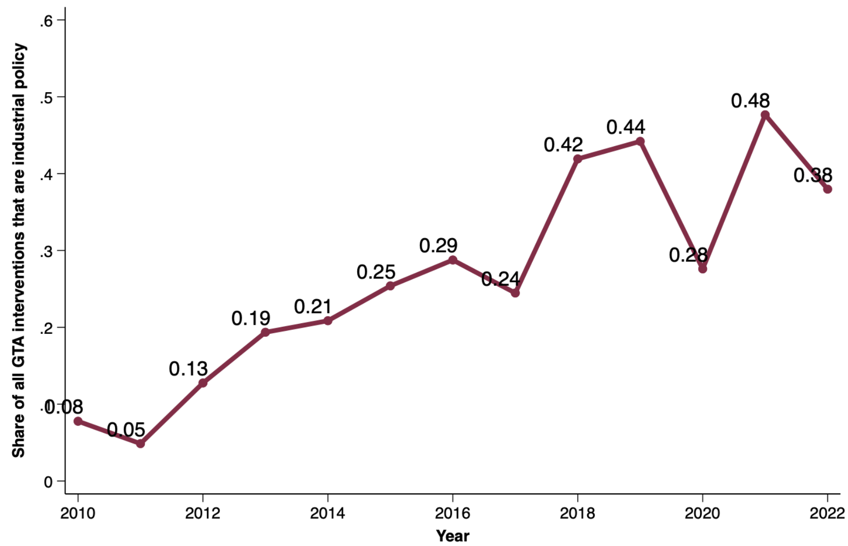


Figure 2: Time Trend of the Share of Industrial Policy Interventions

Notes: The figure plots the share of industrial policy interventions among all interventions in the GTA by year. Only entries entered in the same calendar year are included to ensure comparability over time. See text for details.

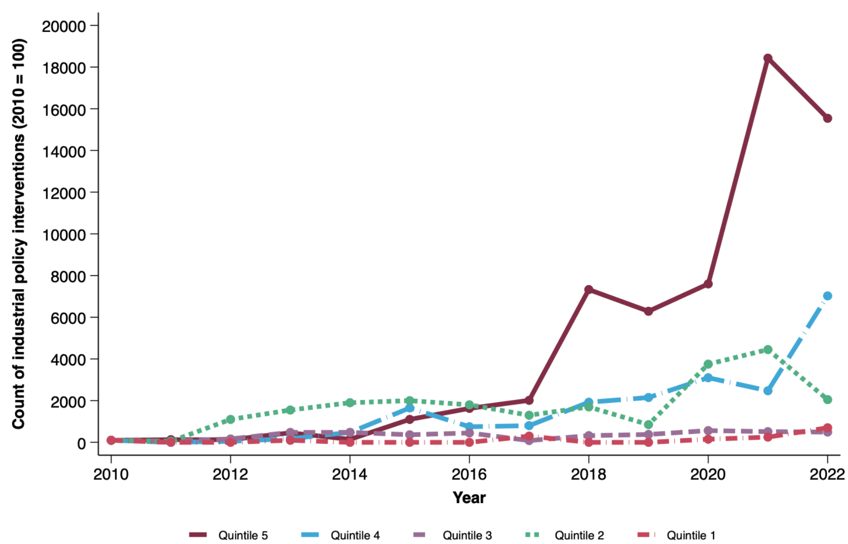
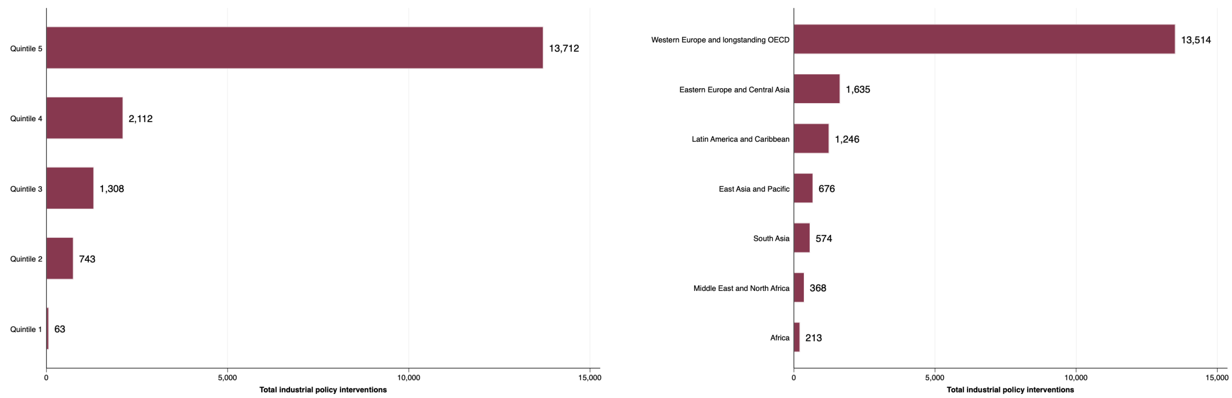


Figure 3: Time Trend of Industrial Policy Interventions by Income Quintile (2010 = 100)

Notes: The figure plots the count of industrial policy interventions by income quintile relative to 2010. Income quintiles based on GDP per capita in 2010 (PPP adjusted, year 2000 USD). Quintile 5 is the richest 20% of countries, quintile 1 is the poorest 20% of countries. Only entries entered in the same calendar year are included to ensure comparability over time. See text for details.

be entirely resolved when examining changes within an income group, they are likely much smaller.



(a) By Income

(b) By Region

Figure 4: Total Industrial Policy Interventions by Income level and Region.

Notes: The figures plot the total number of industrial policy interventions by income quintile and region. Income quintiles based on GDP per capita in 2010 (PPP adjusted). Quintile 5 is the richest 20% of countries, quintile 1 is the poorest 20% of countries.

3.3 The policy levers used to conduct industrial policy

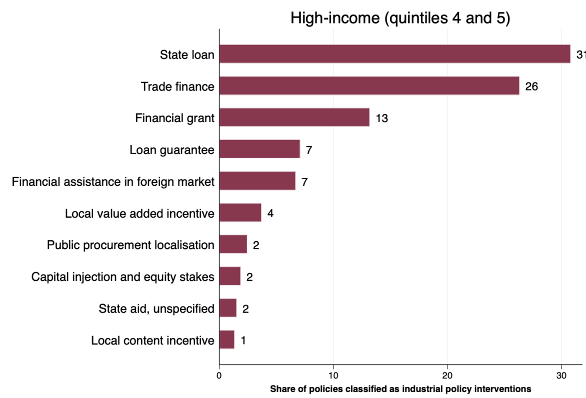
Figure 5 plots the policy measures most commonly used to deploy industrial policy, by income quintile. The policy categories are classified by the GTA using their in-house classification system. This classification system is more granular than the UNCTAD non-tariff measure codes. The GTA handbook contains the mapping between the two.

The figure shows substantial overlap among interventions used across the income distribution suggesting that countries use a similar set of tools to deploy industrial policy. A notable exception are import tariffs, which are more frequently used (in relative terms) in poorer countries, though even in these countries, they account for only 3% of industrial policies.

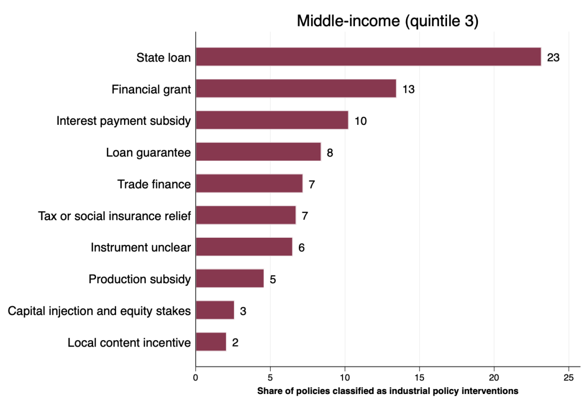
3.4 Sectors targeted by industrial policy

Figure 6 plots the Harmonized System 2 digit (HS2) categories most commonly used to deploy industrial policy by income quintile. It is important to note that HS 27 includes both fossil fuels and clean electricity generation. To inform the discussion about green industrial policies, we further distinguish these. In the highest income quintiles (1 and 2), 55% of policies in HS 27 target clean electricity generation. In middle income countries (quintile 3), 40% of policies target clean electricity generation. This same number is 33% in low-income countries (quintiles 1 and 2).

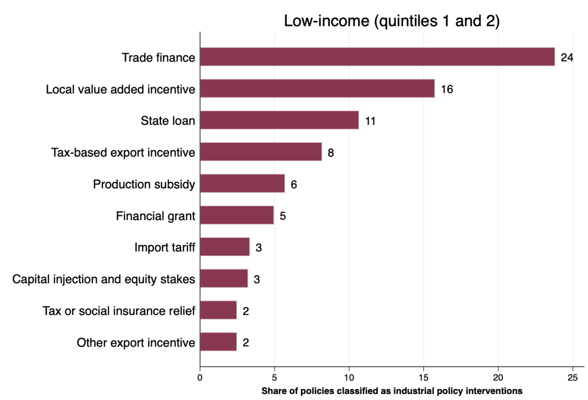
There is again substantial overlap amongst the sectors most intensively targeted by industrial policy. All country groups target capital and transportation equipment. In addition, high-income countries intensively promote green technologies, middle income countries target the food sector, and low-income countries target clothing and apparel.



(a) High-income countries



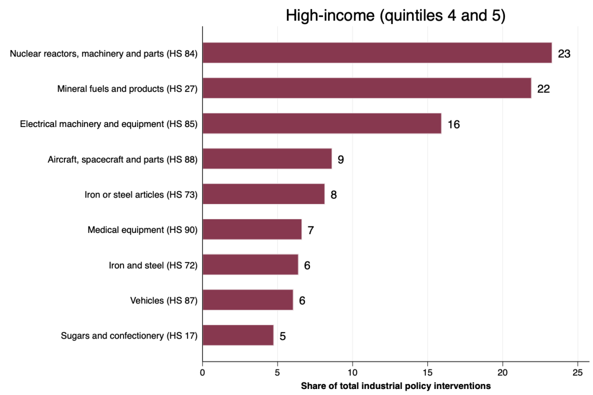
(b) Middle-income countries



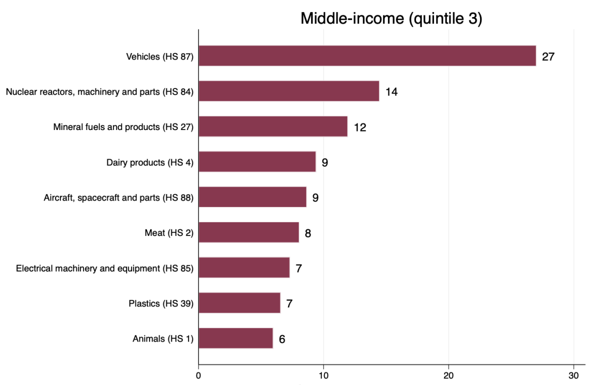
(c) Low-income countries

Figure 5: Top Policy Levers By Income Group

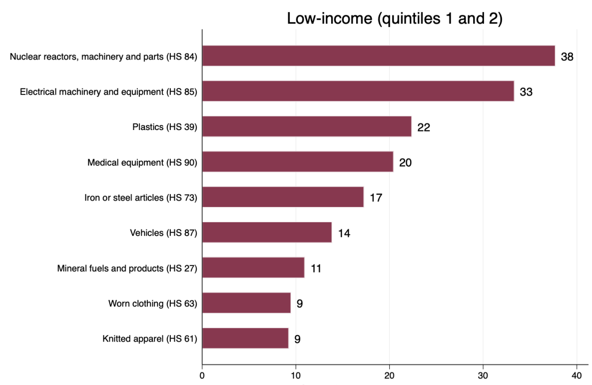
Notes: The figure plots the top 10 most common policy levers used to deploy industrial policy by income group. The bars represent the share of industrial policy interventions accounted for by each policy measure in a given income group. Income quintiles based on GDP per capita in 2010 (PPP adjusted). Quintile 5 is the richest 20% of countries, quintile 1 is the poorest 20% of countries.



(a) High-income countries



(b) Middle-income countries



(c) Low-income countries

Figure 6: Top 10 Sectors Targeted By Industrial Policy.

Notes: The figure plots the top 10 HS 2 digit sectors targeted by industrial policy by income group. The bars represent the share of industrial policy interventions accounted for by each sector in a given income group. Income quintiles based on GDP per capita in 2010 (PPP adjusted). Quintile 5 is the richest 20% of countries, quintile 1 is the poorest 20% of countries.

References

Evenett, S., & Fritz, J. (2018). *Working With a Growing Data Set: Technical Note on the Implications for Proper Reporting Lag Adjustment.*

Juhász, R., Lane, N., Oehlsen, E., & Pérez, V. C. (2022). *The Who, What, When, and How of Industrial Policy: A Text-Based Approach.* doi: doi:10.31235/osf.io/uyxh9