Democracy and Trade Policy at the Product Level: Evidence from a New Tariff-line Dataset

Soubhik Barari† In Song Kim‡

June 4, 2021

Abstract

Despite the fact that trade policies vary significantly across products and trading partners, granular trade policy data is often inaccessible due to difficulties in linking data from multiple sources. We overcome these challenges and construct a dataset of 5.7 billion observations of tariffs and non-tariff barriers between all pairs from 136 countries over 20 years. Using this data, we offer the first product-level analysis of the impact of political regime type on trade liberalization. Our study shows significant variation across products even within the same industry, revealing, for example, that democracies are more likely than non-democracies to protect consumer goods. Our findings suggest that international trade policy is driven by narrow policy preferences towards specific goods, therefore researchers may draw incomplete conclusions about trade politics if empirical studies are not done at the appropriate unit of analysis.

Keywords: tariff-line data, big data, democracy, trade liberalization, heterogeneous effects, agricultural protection

---

*We thank the IDB team at the World Trade Organization for their approval of the re-dissemination of tariff-line data. This research received financial support from the Amazon Web Services (AWS) Cloud Credits for Research Program. We thank Weihuang Wong for his valuable insights and contributions to improve the paper. Thanks also to Tim Büthe, Devin Caughey, Volha Charnysh, Raymond Hicks, David Lake, Lisa Cammer McKay, Asya Magazinnik, Edward Mansfield, Helen Milner, Rich Nielsen, Pablo Pinto, Peter Rosendorff, Sujeong Shim, Alastair Smith, Randy Stone, Jan Stuckatz, Rory Truex, Ariel White, and Bernardo Zacka as well as the seminar participants at Princeton University, Rutgers University, Technische Universität München, Washington University in St. Louis, University of Wisconsin-Madison, UCLA (Computational IR Workshop), Texas A&M, University of California, Riverside, the 2017 International Political Economy Society (IPES), and the 2018 European Political Science Association annual meetings. An earlier version of this article was circulated under the title “Trade Liberalization and Regime Type: Evidence from a New Tariff-line Dataset.” We are grateful for Samir Dutta and Hyungsuk Yoon for their excellent research assistance.

†Ph.D. Candidate, Department of Government, Harvard University, Cambridge, MA, 02138. Email: sbarari@g.harvard.edu URL: http://www.soubhikbarari.org

‡Associate Professor, Department of Political Science, Massachusetts Institute of Technology, Cambridge, MA, 02139. Email: insong@mit.edu URL: http://web.mit.edu/insong/www
1 Introduction

Despite the enormous heterogeneity in trade policies across products and trading partners, granular trade policy data have rarely been used in international political economy research. The primary contribution of this letter is to address this need for better data by constructing a dataset of over 5.7 billion observations of tariffs disaggregated to the unit at which trade policy is actually set: the \textit{tariff-line} level. We develop a replicable automated procedure to (1) retrieve tariff data from multiple web sources, (2) identify the partner-specific tariff rates for each product, and (3) resolve any discrepancies that arise.

We use this data to examine the question of whether and how political regime type interacts with between-industry and within-industry domestic trade preferences. Although the effects of domestic political institutions on trade policy have long been a source of theoretical debate, empirical studies often employ Most Favored Nation (MFN) applied tariff rates or non-tariff barrier “coverage ratios” that are averaged across a massive range of products and trading partners when evaluating how domestic trade preferences are translated into policy outcomes (Gawande and Hansen, 1999). Consequently, a single number for each importer-year has often been used to examine how trade policies differ across regime types (e.g. Milner and Kubota, 2005).

We depart from the existing studies by offering the first quantitative estimates of the effect of regime type on product-level trade policies. We find a high level of heterogeneity across products even within narrowly defined industries. Specifically, we find that democracies are more likely to protect consumer goods such as food, textile, and manufactured products than are non-democracies. In contrast, industries such as wood and metal as well as highly differentiated intermediate goods tend to get significantly lower tariff rates in democracies compared to non-
democracies. Our findings have important implications for understanding how political regimes interact differently with underlying preferences of political actors. We find that democratic institutions, with larger selectorates than autocracies, do not necessarily empower consumers (i.e., voters), thereby calling into question the underlying assumption of many studies in the literature that individual trade preferences (e.g. given by their factorial or sectoral interests) are translated into actual trade policies (see Rho and Tomz [2017] for this important debate). Furthermore, the heterogeneous effects of political institutions on product-level trade policies also suggests that democratic political systems may interact differently with producers depending on the types of products that they produce (Kim [2017]; Naoi and Kume [2011]; Betz [2017]; Baccini et al. [2018]; Betz and Pond [2019]).

2 A New Bilateral Product-level Tariffs Database

Our novel bilateral product-level tariffs database collects, parses, merges, and resolves more than 2,000 WTO importer-year pairs between 1989 and 2016 from the two major sources of applied tariffs, the WTO Integrated Database (IDB) and the UNCTAD Trade Analysis Information System (TRAiNS). We detail two major innovations of this database: (1) capturing the massive heterogeneities in tariff policies at the level they are set and (2) resolving the major inconsistencies that arise from the standard practice of using tariff data from only one of these two sources. Additional data collection, text parsing, and entity resolution steps are described in detail in Appendix S1.

Capturing the Massive Variation in Tariffs Across Products and Trading Partners. First, the need for a dataset to captures bilateral tariffs at the product level stems from the substantial heterogeneity in trade policies, as Figure 1 shows. This heterogeneity exists despite broad membership
Figure 1: Variations in Ad Valorem Applied Tariff Rates across Trading Partners and Industries. This figure demonstrates how our data captures the variation in tariff policies across HS 2-digit industries and trading partners. Exporters are plotted down each column and importers are plotted across each row. For a given country and partner, our data distinguishes precise average tariff rates across HS 2-digit industries (colored within plot) from 1989 to 2015.

in the WTO because WTO members are permitted to enter regional trade agreements under Article XXIV of GATT, Enabling Clause, and to lower tariffs for least developed countries with the GSP. That is, the rule of “non-discrimination” does not hold in practice.

Indeed, researchers may draw incomplete conclusions about trade politics if empirical studies are not done at the appropriate unit of analysis under the Harmonized System of hierarchical product classification. For instance, consider the HS 2-digit chapter code for Vegetables (07). The U.S. average applied MFN tariff rate for products in this chapter code is 4.35%, which looks comparable to the country-level average rate of 3.34% across all products from various industries. Scholars have used this fact to argue that the U.S. has mostly eliminated trade barriers in many industries, including the agricultural sector. Of course, if average tariff rates are falling, that does
<table>
<thead>
<tr>
<th>Issue</th>
<th>Example tariff-line  (Year–Importer–Exporter–HS)</th>
<th>WTO IDB report</th>
<th>UNCTAD TRAINS report (<em>≈ AVE</em>)</th>
<th>Solution</th>
<th>Instances (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Missing Report</td>
<td>2013–China–India–09041200 (Crushed or ground Piper pepper)</td>
<td>10% ✓</td>
<td>none</td>
<td>Use non-missing.</td>
<td>2.35 billion (41.8%)</td>
</tr>
<tr>
<td></td>
<td>1991–Japan–Korea–140490499 (Cod fish)</td>
<td>none</td>
<td>10% ✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conflicting</td>
<td>1997–Australia–Singapore–22082010 (Grape wine)</td>
<td>3%</td>
<td>3% + $31.12/L (≈ 127%) ✓</td>
<td>Use ad valorem equivalent (AVE) if computed by UNCTAD.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2005–Canada–Australia–22084010 (Rum)</td>
<td>24.56¢/litre of alcohol</td>
<td>24.56¢/litre of alcohol (≈ 1.43%) ✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2004–Argentina–Paraguay–87083110 (Motor vehicle brakes)</td>
<td>0% ✓</td>
<td>14%</td>
<td>Use lower (preferential) rate if no AVE.</td>
<td>0.24 billion (4.25%)</td>
</tr>
<tr>
<td></td>
<td>1996–USA–Mexico–87033100 (Cars of ≤ 1,500 cc cylinder capacity)</td>
<td>2.5%</td>
<td>0% ✓</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1: **Solutions to Tariff Data Inconsistencies.** This table illustrates examples of specific issues that arise when attempting to find the correct applied rate for a tariff-line using the IDB and TRAINS databases. See Appendix S1.2 for the full merging algorithm.

Tell us that some products are receiving less protection than before. But this statistic elides important information about tariff rates across products, which reveals that protection is indeed alive and well for numerous agricultural products. The 2016 U.S. applied MFN tariff on *Onions* identified by a HS 6-digit subheading code 071220 is 25.6%, for example.\(^1\) Compare this with *Asparagus* (070920) which receives roughly half the MFN duty at 13%. This more nuanced examination of tariff rates provides a clearer picture of the granular cleavages that emerge in trade politics.

**Resolving Major Inconsistencies in Existing Bilateral Tariff Data.** Secondly, and most crucially, there exist inconsistencies – missing or conflicting reports – between the two data sources. Missing reports are problematic since researchers tend to use one of these data sources for empirical research but usually not both. As Table \[\] shows, they are not a trivial problem either: at least 2.35 billion tariff-lines are missing from one of the databases. To solve this, we resolve each tariff-line to the available report whenever possible. Second, IDB returns duties as they are originally reported (e.g. 24.56¢/litre of alcohol), while TRAINS uses a method to estimate an ad

\(^1\) At the HS 4-digit heading level, onions are classified as *Dried vegetables* (0712), which has an average tariff rate of 6.51%.
valorem equivalent (AVE) for any reported non-ad valorem rate (e.g. 24.56¢/litre of alcohol \(\approx 1.43\%\)). In this case, our algorithm chooses TRAINS, since it is the more precise and informative source. Third, preferential rates may be reported in only one source. As shown in the last row of Table I, TRAINS shows the correct 1996 NAFTA duty-free rate for United States-Mexico trade in Cars of \(\leq 1,500\) cc cylinder capacity while IDB does not. Our algorithm picks the correct partner-specific preferential rate for both tariff-lines. After resolving issues of missing data and discrepancies between the two sources, we create a dataset of over 5.7 billion observations of bilateral trade policy at the product level. Appendix SI.2 details each step in our resolution algorithm.

In total, our dataset covers 2,476 WTO importer-years (3,080 importer-years overall) from 1989 to 2015. We then combine our product-level trade policy data with over 30,000 product-level non-tariff barriers (NTBs), as well as numerous country-, dyad-, and directed dyad-level datasets, such as GATT/WTO membership and bilateral trade volume at the Harmonized System (HS) 6-digit level. To the best of our knowledge, this is the first database that combines bilateral trade policies (encompassing both tariffs and non-tariff measures) and trade volume for 136 countries over 30 years at this level of granularity.

3 The Effects of Regime Type on Trade Policy

Do democratic political institutions facilitate unilateral trade liberalization? Milner and Kubota (2005) argue that democratization empowers the owners of factors with which their country is abundantly endowed, and therefore one should expect that trade liberalization will ensue, reflecting the median voter’s preferences. Using MFN tariff rates averaged across products, they find that democratization in labor-abundant developing countries is associated with lower trade barriers.
Other scholars have argued the reverse, however, suggesting that the need to win elections makes
democratic politicians sensitive to the demands of interest groups who offer support in exchange
for trade protection (Frieden and Rogowski, 1996). Autocracies, meanwhile, need appeal to only
a very small segment of society to secure their power, and therefore they might be less susceptible
to interest group pressures (Acemoglu and Robinson, 2005; Henisz and Mansfield, 2006).

We re-examine whether trade policy varies between democracies and non-democracies at the
product-level. To illustrate the severity of the ecological fallacy that may occur when researchers
employ aggregated measures of trade policies, we gradually disaggregate our units of analysis
from HS2 to HS4 and to HS6 levels. We find it valuable to disaggregate up to, but not beyond,
the HS6 level because it explains more variance in countries’ MFN tariff profiles than the HS2
or HS4 levels while remaining comparable across countries. For example, in the United States,
variance between HS2 groupings of products explained 38% of the variance in 2012 MFN applied
tariffs, while variance between HS4 groupings of products explained 52% and variance between
HS6 groupings of products explained 68%.

3.1 Methodology

To estimate the effects of regime type on trade policy, we introduce the following hierarchical Tobit
model of the observed MFN tariff rate $\tau_{ipt}$ for importer $i$ and product group $p$ in year $t$:

$$
\tau_{ipt}^* = \beta D_{it} + \gamma_p^T X_{it} + \delta^T Z_{it} + \lambda_1^T V_{ipt} + \lambda_2 W_{ih[p]t} + \eta_i + \theta_t + \epsilon_{ipt}
$$

$$
\tau_{ipt} = \begin{cases} 
\tau_{ipt}^* & \text{if } \tau_{ipt}^* \geq 0 \\
0 & \text{otherwise}
\end{cases}
$$ (1)
where $\tau_{ipt}^*$ is a latent tariff that we observe if it is greater than zero and is censored at zero otherwise. We perform separate analyses in which we index $p$ by its HS2 industry group, HS4 product group, and HS6 product group respectively. In each case, we compute the time-varying average tariff rate $\tau_{ipt}$ at the desired aggregation level from the new tariff-line dataset, described in Section 2, and use logged values to address the high skewness of tariffs. Following the literature, we use a binary measure of democracy where $D_{it}$ is unity if importer $i$’s Polity IV score is 6 or above in year $t$ and zero otherwise (e.g. Mansfield et al., 2000; Milner and Kubota, 2005). While we recognize that there exists substantial variation in political institutions even within democracies and autocracies (Rickard, 2015), this measurement strategy not only helps compare our findings to existing studies, but also facilitates conceptually cleaner comparisons across numerous estimates, which is the primary objective of this section.

Although the results presented in this section should not be interpreted as causal effects of democracy on trade policy, we include a large number of covariates to account for as much potential selection and confounding as possible. $X_{it}$ is a set of country-level covariates—democracy ($D_{it}$), log GDP per capita, and an intercept—for which we estimate product-specific coefficients. Furthermore, we control for the presence of 30,327 product-level NTBs for each country covering sanitary and phytosanitary measures, technical barriers to trade, anti-dumping, countervailing duties, safeguards, quantitative restrictions, and tariff-rate quotas. This allows us to account for (1) leaders’ incentives to use tariffs rather than NTBs may vary between regime types and across products, and the (2) substitutability/complementarity between tariffs and NTBs (Mansfield and Busch, 1995; Kono, 2006; Bown and Tovar, 2011; Anderson et al., 2013). $Z_{it}$ represents a vector of country-level time-varying confounders of regime type and trade policy: log GDP per capita (PPP basis), log population, an indicator for GATT/WTO membership, and an intercept. $V_{ipt}$ con-
contains log import volume and an indicator for whether the importer imposes any non-tariff barriers on the product in the given year (at the same level of product aggregation as the tariff rate). All covariates are lagged by 1 year. We also include the continuous Balassa index, $W_{ih[p]t}$, which captures technological differences across countries, industries $h$, and time $t$ for product $p$.\(^2\) To address missingness in covariate data, we create multiple imputed datasets and conduct estimation separately across them following Honaker et al. (2011).\(^3\)

Finally, $\eta_i$ and $\theta_t$ are importer- and year-varying intercepts respectively, and $\epsilon_{ipt}$ is idiosyncratic error assumed to be drawn from a Normal distribution:

$$
\eta_i \overset{i.i.d.}\sim \mathcal{N}(0, \Sigma_\eta), \quad \theta_t \overset{i.i.d.}\sim \mathcal{N}(0, \Sigma_\theta), \quad \epsilon_{ipt} \overset{i.i.d.}\sim \mathcal{N}(0, \sigma_\epsilon^2).
$$

(2)

To account for heterogeneous political processes across products, we model the product-varying effects hierarchically. Specifically, we allow the effects of political processes to vary across products $p$ (e.g. vegetables vs. fish) but incorporate the complex correlations within a broader sector $k$ (e.g. food sector) that operates differently from other sectors (e.g. textile sector):

$$
\gamma_p \sim \mathcal{N}(\phi_{k[p]}, \Sigma_\gamma), \quad \phi_k \sim \mathcal{N}(0, \Sigma_\phi).
$$

(3)

where the effect $\gamma_p$ for product $p$ belonging to sector $k$ is drawn from a multivariate-Normal distribution with a mean vector $\phi_{k[p]}$ and covariance matrix $\Sigma_\gamma$, and $\phi_k$ is drawn from a multivariate-Normal distribution with mean 0 and covariance matrix $\Sigma_\phi$. This means that the product-specific coefficients vary based on the sector $k$ to which the product belongs, which increases the plausibility of the exchangeability assumption for the product-specific effects.

\(^2\)The Balassa index of a given HS2 industry in a given country is the ratio of the industry’s share of the country’s total exports to the industry’s share of global exports.

\(^3\)At the HS4 and HS6 levels, we performed linear extrapolation to impute missing trade volume for a given importer’s products across years instead of using Amelia package, which is exceptionally computationally intensive at such scale.
Figure 2: Effect of Democracy on HS2 Log Tariffs. This plot presents posterior means and 95% credible intervals for the estimated effect of democracy on tariff rates for each HS2 chapter.

3.1.1 Empirical Results

Our quantity of interest is the effect of democracy on trade policy across products. The model given in equation (1) decomposes this quantity into two parts: (1) the main effect $\beta$ and (2) the product-specific partial effect of democracy $\gamma_p^{DEM}$ at the HS2, HS4, or HS6 level.

Figure 2 reports the posterior distribution of our quantity of interest, $\beta + \gamma_p^{DEM}$. The mean of the posterior distribution of the main effect of democracy, $\beta$ (marked by the dotted horizontal line), shows that across all industries, the MFN tariffs imposed by democracies are about 31\% ($\approx \exp(0.27) - 1$) lower than the MFN tariffs imposed by non-democracies, precisely replicating the findings from previous studies (Milner and Kubota, 2005; Chaudoin et al., 2015). Note that our study extend the previous studies by including a large number of developed countries. Thus, the conclusion that democracies impose lower tariffs on average than non-democracies does not

---

*Note that $\gamma_p$ is a vector of product-varying effects across covariates, and we denote the element corresponding to the democracy variable $D_{it}$ by $\gamma_p^{DEM}$. 

---
hinge on the abundant supply of labor in developing countries. Rather, our finding accords with the theoretical logic that the preferences of the median voter is generally in line with liberal trade policy in democracies. For comparability with Milner and Kubota (2005), we also conduct our analysis with only less-developed countries where we find similar results for the average effect of democracy across all products (see Appendix S3).

The HS2 results reveal significant heterogeneity in the effect of democracy across industries. Figure 2 shows several industries for which the effect of democracy differs notably from the posterior mean of the main effect, \( \beta \) (the dotted line). Animal, Vegetable, and Foodstuffs products have positive estimates; for the majority of industries in these sectors, the effects turn out to be statistically significant (marked by black vertical lines). That is, democracies are more protective of agricultural consumer goods than non-democracies are. We also find that democracies tend to protect Textiles industries. Conversely, Minerals, Chemicals, Wood, and Metals industries have estimates that are significantly lower than the main effect, suggesting that democracies engage in deeper liberalization of these industries than non-democracies do.

Next, our analysis at the HS4 level reveals further heterogeneities as we account for more nuanced product categories: for example, there are only two HS2 chapters for Machinery/Electrical goods, but they consist of 133 unique HS4 subheadings. Figure 3 shows that when we use HS4 product groups instead of HS2 groups, the estimated main effect of democracy (dotted line) becomes smaller (on average, MFN tariffs imposed by democracies are 8% lower than those imposed by non-democracies, i.e., much smaller effect size than the HS2 level analysis) while product-specific heterogeneity substantially increases. We find stronger effects (roughly double) of democratic political institutions on agricultural protection at the HS4 level than at the HS2 level. Notably, the highest tariffs are imposed on Meat of bovine animals (0201) and Cucumbers, fresh
Figure 3: **Effect of Democracy on HS4 Log Tariffs:** This plot presents posterior means and 95% credible intervals for the estimated effect of democracy on tariff rates for each HS4 product group. Boxes group together products belonging to a common HS2 industry with the chapter code given at the bottom of each box.

or chilled (0707). We also find stronger evidence that democracies are more protective of textile products than are non-democracies, nearly 40% larger effects than the HS2-level effects. On the other hand, intermediate goods tend to have lower tariffs in democracies than in non-democracies, as shown by the smaller estimates for the products that belong to **MINERALS**, **METALS**, and **MACHINERY/ELECTRICAL**. This is consistent with [Baccini et al. (2018)](), who show that the increasing importance of global value chains and intra-industry trade makes it easier for countries to liberalize intermediate goods than finished products. Our findings add nuance to this claim by showing that such effects are more pronounced among democracies than non-democracies, holding economic size, comparative advantages, institutional memberships, and trade volumes constant.

Finally, we present the estimated effects of democracy on trade policy at the HS6 level. Figure 4 present the estimates from two distinct types of products: (1) meat and (2) tobacco. Panel (a) shows that there exists substantial variations across similar products in the meat industry –
Figure 4: **Effect of Democracy on HS6 Log Tariffs, Meat and Tobacco Products.** This plot presents posterior means and 95% credible intervals for the estimated effect of democracy on tariff rates for HS6 products in the meat industry (panel a) and tobacco industry (panel b). Boxes indicate distinct product groupings.

Again, consumer products such as beef, pork, and lamb are more protected whereas animal organs are more liberalized in democracies than in non-democracies. Similarly in panel (b) we see that consumer goods in the tobacco industry, like cigarettes, are chosen for protection by democracies relative to autocracies. On average, MFN tariffs are about $43\% (\approx \exp(0.36) - 1)$ higher for democracies than non-democracies across meat products and while MFN tariffs exhibit no average difference between democracies and non-democracies across tobacco products. Appendix S4 presents a larger set of estimates from each Harmonized System Section such as the steel industry.

We conduct a systematic analysis of the differences in democratic protection between consumption goods and intermediate goods using two different product-level measures. First, we map each HS6 product to a discrete Broad Economic Category (BEC), which categorizes products based on their main end use (e.g. capital vs. consumption vs. intermediate). We find that the HS 6-level estimates of $\gamma_{p}^{DEM}$ across all 3,950 products with available BEC mappings are significantly

---

5 Estimates for HS 6-level product-varying coefficients are pooled at the industry rather than sector level. That is $k$ is replaced with $h$ in equation (3) for faster convergence and better Markov chain mixing.
smaller for intermediate goods than consumer products even within the same industry (a univariate regression model reveals a statistically significant coefficient of 0.37 at the $\alpha = 0.01$ level; the result remains statistically significant after controlling for HS 2-level fixed effects across products).

That is, we find that democracies are more likely than non-democracies to liberalize (protect) intermediate (consumer) goods. Second, we map each HS6 product to a continuous measure of its “downstreamness” in global value chains (GVCs) across all countries between 1995 and 2011 estimated by [Antras and Chor (2018)]. This GVC-based measure confirms the directional result of the discrete measure: after adjusting for sector and industry, more downstream products are predicted to be more highly protected by democracies relative to non-democracies.

4 Concluding Remarks

In this letter, we present a novel dataset with nearly 6 billion observations of product-level applied tariff rates that countries levy on their trading partners, incorporating the universe of preferential rates and the Generalized System of Preferences. To do so, we combine and augment datasets available from the WTO and UNCTAD, merging this data with non-tariff barriers and trade volume data along with numerous country-, dyad-, and directed dyad-level datasets available in the literature. Our data produces the first product-level estimates of how political regimes affect trade liberalization, revealing significant variations across products even within the same industry. Specifically, we show that democracies tend to impose higher tariffs on consumer products such as agricultural goods than non-democracies do, while democracies’ tariffs on intermediate goods are lower than non-democracies’. Our data and findings lay an important empirical foundation for investigating trade politics at a much more granular level than has historically been done.

---

*A nonparametric visualization of the differences is shown in Appendix Figure S2; the parametric estimates of the differences described here are shown in Appendix Figure S3.*
References


