Tariffs and State Capacity:
A Specific Story

Erik Madsen
NYU

Martin Rotemberg
NYU

Sharon Traiberman
NYU

Shizhuo Wang
NYU

July 2026

Abstract

Evasion poses a major challenge for raising tariff revenue, especially when administrative capacity is limited. We study how this challenge shapes the design of tariff codes. We develop an equilibrium model of tariff collection in which a government partitions goods into tariff lines, assigns each line an ad valorem or specific tariff, and assesses tariffs based on importers' strategically (mis)declared shipment values. Tariff code design involves several tradeoffs: Multiple tariff lines are costly to administer, ad valorem tariffs require verification of declared values, and specific tariffs distort relative prices. Using newly compiled records from the Early American Republic, we show that the model rationalizes several empirical patterns. Specific tariffs are concentrated on goods that are homogeneous, cheap, or lightly taxed. Over time, tariff codes become more complex and reliant on tailored specific tariffs, consistent with reduced administrative costs. We show that similar patterns are present in modern tariff schedules.

Keywords: Import tariffs, specific tariffs, tax evasion, state capacity, economic history

JEL Classification: D82, F10, H26, N41

Please contact emadsen@nyu.edu, mrotemberg@nyu.edu, sharon.traiberman@nyu.edu, and shizhuo.wang@nyu.edu with any questions or comments. We are particularly grateful to Jacob Carrasco, Eden Glickman, Xianru Han, Katie McDonald, Alejandra Perez, Stuti Saria, Marios Tsoukis, and Xuancheng Zhao for excellent research assistance. We are grateful to seminar participants at Cowles Trade Day, Brown, CUNY, Dartmouth, NBER ITI, NEUDC, NC State, SED, Stanford, and UC Davis, and to Miguel Acosta, Lydia Cox, Andrew Greenland, John Lopresti, and Christopher Meissner for helpful comments. This material is based upon work supported by the National Science Foundation under Grant No. 2314696. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF.

1 Introduction

Tariffs have historically been a key source of government revenue, accounting for around 90 percent of federal government revenue in the United States until the Civil War (Reamer, 2016; Irwin, 2020). They continue to play an important role in the public finances of developing countries, due to the relative ease of observing goods entering at ports (Besley and Persson, 2009; Bown and Irwin, 2019). Nevertheless, collecting import duties is challenging due to evasion: Importers may misrepresent their cargo in an effort to reduce assessments. Multilateral agencies emphasize customs fraud as a major constraint on revenue collection (World Bank, 2005; International Trade Centre, 2020; IMF, 2025).

In this paper, we argue that concerns over misvaluation and misclassification shape the design of tariff codes, focusing on the Early American Republic as a case study. We emphasize two features of tariff codes: *complexity* (the number of distinct tariff lines) and *regime* (which goods face specific tariffs and which face ad valorem tariffs).[1] To measure these features, we construct tariff classifications for the Early American Republic directly from the statutes. Because goods are often described differently from one act to the next, under headings that range from broad categories to narrow product-specific descriptions, we hand-match provisions across statutes using both the language of the acts and information about each article's characteristics. The result is a harmonized panel of goods, regimes, and rates from 1789–1842, which we use in our primary analysis.

In the spirit of Bates et al. (1998), we build on historical narratives to develop our theory of the design of the tariff code. Policymakers consistently argued that the structure of tariffs was central to raising revenue while limiting evasion and minimizing distortions. They understood that ad valorem tariffs were preferable to specific tariffs "theoretically," because specific tariffs distorted relative prices: The Polk administration, for example, viewed them as "oppressive" because they imposed higher effective rates on cheaper varieties. Nevertheless, many policymakers preferred specific tariffs "practically," because of the susceptibility of ad valorem tariffs to "deception" (Burrows, 1894).

In principle, policymakers can leverage complexity to mitigate the costs of specific tariffs, as tailored specific tariffs on disaggregated, narrowly defined goods limit relative price distortions. In particular, for goods with no variation in their unit values v, an ad valorem tariff τ is equivalent to a specific tariff $t = \tau \cdot v$ (Falvey, 1979). However, in practice, disaggregation required costly investment in Customs capacity. For example, Jones (1835) describes how British mills could produce rolled iron bars that mimicked hammered iron in order to evade

[1] Specific tariffs are charged as a function of quantities (such as weight, number of units, or volume), and ad valorem tariffs are charged as a function of value. In practice, some tariffs are a mix of both types, or have other features. We focus on specific and ad valorem tariffs, which comprise the vast majority of all tariffs, both historically and currently.

tariffs. Detecting such evasion required the Customs official overseeing iron imports to be a "skillful artist who has a knowledge of chemistry," capable of distinguishing iron bars based on their oxide content. Officials had to be trained to make many such subtle distinctions when determining the appropriate tariff rate.

In sum, policymakers face two enforcement frictions. First, a tariff schedule with distinct lines creates scope for misclassification across categories, requiring costly expertise to enforce those distinctions (Grant, 2025). Second, ad valorem duties give importers scope to misvalue their goods, potentially requiring customs officials to verify reported prices (Bhagwati, 1964).

We develop a model of tariff collection to analyze how misclassification and misvaluation frictions shape tariff design. Absent enforcement concerns, the standard result obtains: The optimal tariff schedule is ad valorem and uniform across goods (Costinot et al., 2015). With imperfect enforcement, this benchmark no longer applies. Misvaluation creates a cost of using ad valorem tariffs, while the possibility of misclassification makes it costly to tailor specific tariffs across imports.

We model the assessment of ad valorem tariffs as a game between customs officials and importers. Importers may underreport values to evade duties, while officials may pay for costly audits and fine importers caught misreporting. This game has a unique outcome within a natural class of equilibria with monotone reporting strategies. We compare the ad valorem regime to one in which the government instead uses specific tariffs, which are perfectly implemented but distort consumer prices. We model the cost of tariff code complexity as a classification cost the government must pay to distinguish among imports before assigning them customized tariffs.

An optimal tariff code balances several countervailing forces. First, the verification costs of ad valorem tariffs are traded off against the price distortions under specific tariffs. Second, the costs of complex classification schemes are traded off against the gains from creating narrow, disaggregated categories.

With the model in hand, we perform a series of empirical analyses using our historical dataset. To validate the practical relevance of misvaluation in the Early American Republic, we first evaluate the effects of a large policy change. In 1846, all specific tariffs were replaced with ad valorem ones for political reasons. Consistent with our modeling assumption that importers can misvalue shipments and that enforcement is imperfect, we document that Customs officials immediately reported a rise in fraudulently low valuations and a drop in reported unit values following the policy change (Meredith, 1849).

In principle, such a decline could merely reflect a compositional change in the spirit of Alchian and Allen (1964), since switching from specific to ad valorem tariffs relatively lowers the effective tariff on cheaper varieties. We use the model to decompose the observed change in unit values into two components: one due to the compositional distortion of the specific tariff, and another which affects unit values only if ad valorem tariffs induce misvaluation. We operationalize this decomposition in a difference-in-differences-style exercise. Even conditional on the model-implied Alchian and Allen (1964) controls, reported import prices fell sharply for goods that switched from specific to ad valorem tariffs, relative to those that had never been subject to specific tariffs.

We next consider the cross-sectional implications of tariff design in the shadow of verification and complexity frictions. Our model yields several testable predictions for which tariff regime is preferred across different categories within a tariff code, which we validate in our dataset.

First, the model correctly predicts that the government should choose specific tariffs when the expected customs revenue from a shipment is low, either because the good is inexpensive or because its effective tariff rate is modest. Intuitively, the government wants to avoid paying verification costs that would consume most or all of the tariff revenue.

Second, the model correctly predicts that specific tariffs are more likely to be imposed on goods with little variation in prices, such as commodities (Rauch, 1999). This may seem surprising, since a common argument is that prices of homogeneous products are harder to falsify (Javorcik and Narciso, 2008), seemingly favoring the use of ad valorem tariffs. In our framework, price verification entails administrative or legal costs, which must be paid even when price uncertainty is small. By contrast, specific tariffs distort relative prices, a welfare loss that diminishes when prices are relatively uniform.

Third, while we focus on the choice of specific versus ad valorem tariffs, the model also has implications for ad valorem tariff rates: The government charges higher rates on cheaper goods to tilt consumption toward relatively expensive shipments, reducing verification costs per dollar of revenue.

We then evaluate how the tariff code evolved during the Early American Republic, allowing us to distinguish between alternative hypotheses regarding the development of state capacity. Over the course of our sample period, Customs collection capacity improved dramatically. Expenditure on Customs collection increased by an order of magnitude (as did imports). This growth was accompanied by a tenfold increase in the Customs house workforce, which in turn became increasingly specialized. Conceptually, improved capacity could either lower the costs of verifying reported prices or lower the costs of complexity. We find that the data are consistent with state capacity lowering the cost of complexity. In particular, over our sample period the number of distinct tariff rates tripled and the share of goods with specific tariffs increased fivefold. The model predicts that both of these trends are consequences of reduced complexity costs, and not of reduced verification costs.

During our primary sample period, specific tariffs represented around half of total tariff revenue. While specific tariffs have become less common since then, especially following multilateral General Agreement on Tariffs and Trade (GATT) negotiations, the relationship between state capacity, tariff code complexity, and tariff specificity remains relevant in contemporary settings. High-capacity countries continue to invest in complex tariff schedules—for instance, Switzerland's tariff schedule contains over 1,600 distinct rates, all of which are specific. The United States currently has over a thousand different specific rates and has recently expanded efforts to enforce customs regulations (Koehler et al., 2025). Conversely, many less-developed countries such as Senegal and Peru have fewer than five distinct tariff rates, all of which are ad valorem. We analyze modern tariff systems both as a robustness check and because modern datasets contain information that is unavailable historically: For instance, we have direct measures of cross-country Customs house capacity (Farhad et al., 2024),[2] as well as direct measures of price dispersion using disaggregated data on unit values available in U.S. data from the early 1980s.

A growing empirical literature discusses the effects of specific tariffs in the United States (Crucini, 1994; Irwin, 1998; Acosta and Cox, 2024; Greenland and Lopresti, 2025; Klein and Meissner, 2025; Greenland et al., 2026b).[3] The trade literature on specific tariffs has typically focused on the role of competition, uncertainty, and price and quality dispersion (Das and Donnenfeld, 1987; Helpman and Krugman, 1989; Krishna, 1990).[4] To the best of our knowledge, ours is the first work connecting specific tariffs to state capacity or the tariff code's classification system.

We also build on a large literature considering how heterogeneity across goods in supply, demand, or political importance can generate heterogeneous tariffs (Grossman and Helpman, 1994; Lashkaripour and Lugovskyy, 2023; Bartelme et al., 2025; Grant, 2025). We demonstrate the potential value of a complex tariff code even with classical demand specifications and no industrial policy considerations.

Our research also relates to work on state administrative capacity (Kapon et al., 2024; Casey, 2026; Mastrorocco and Teso, 2026). We highlight a mechanism—the state's ability to classify imports—through which increased state capacity leads to improved revenue collection.[5] In doing so, we bridge broader studies of classification as a form of state capacity

[2]Using different data, Betz (2019) also shows that greater bureaucratic capacity is associated with more distinct tariff rates in a modern context.

[3]For a discussion of 19th-century European tariffs, see Bairoch (1989).

[4]A companion literature in public finance also broadly concludes that specific tariffs are preferable to ad valorem tariffs only under narrow circumstances, if ever (Cournot, 1838; Keen, 1998; Wang and Wright, 2017).

[5]The tradeoff between price (ad valorem) and quantity (specific) instruments at the heart of our mechanism also arises in a related literature on firm choices between such instruments (Klemperer and Meyer,

(Scott, 1998) and a literature studying revenue collection technologies (Rodrik, 2008; Jensen, 2022; D'Arcy et al., 2024; Goyal et al., 2025).

Finally, our findings complement a large literature on tax evasion (Allingham and Sandmo, 1972; Border and Sobel, 1987; Yitzhaki, 1987; Slemrod and Yitzhaki, 2002; Basri et al., 2021). Tariff evasion has been widely documented (Fisman and Wei, 2004; Fisman et al., 2008; Javorcik and Narciso, 2008; Mishra et al., 2008; Rotunno et al., 2013; Sequeira, 2016; Iyoha et al., 2025), though policies to reduce evasion have had mixed success (Yang, 2008; Javorcik and Narciso, 2017; Chalendard et al., 2023). Rather than focusing on improvements in valuation technology (Dar et al., 2025), we show that the design of the tariff code itself can help deter evasion.

2 Historical Context

Tariffs were extremely important for the Early American Republic. Indeed, a central motivation for the Constitutional Convention was the desire to impose tariffs federally (Elkins and McKitrick, 1993; Irwin, 2017). In this section, we provide historical context for our analysis.[6] First, we provide a short history of U.S. Customs House operations from 1789 to 1846, with a focus on increasing capacity. Second, we review major tariff legislation in the United States. Finally, we discuss the administrative history of tariffs, focusing on the debate between ad valorem and specific tariffs.

2.1 Roles and Capacity of the Customs House

The first law enacted by the Washington administration outlined the oath of office to be taken by the President and other federal officials. The second addressed the new Republic's most pressing need: It established a tariff to generate revenue. However, revenue generation requires more than legislation; it requires bureaucrats.

The fifth law, "An Act to regulate the Collection of the Duties imposed by law on the tonnage of ships or vessels, and on goods, wares and merchandises imported into the United States," established the administrative structure of customs collection. Customs officials oversaw inspection and quantification of cargo and calculated duties. The surveyors and those under them typically engaged in one of four tasks: *weighing* of bulk goods, *gauging* of liquid goods, *measuring* of goods where duties were calculated on dimension, and *proving* and marking of spirits. These tasks were regulated by Congress, which determined official definitions of weights, measures, and proofs. For calculating ad valorem duties, inspectors and collectors would compare manifests to the cargo on board. The larger Customs houses eventually hired appraisers, but for those without in-house expertise, Customs officials would

1986; Flynn et al., 2026).

[6]See Young (1877), Goss (1897), Taussig (1910), and Irwin (2017) for more detailed descriptions of tariff policy during the period.

hire qualified merchants to adjudicate values in case of dispute.

The operations of the Customs houses, and the revenue they were able to collect, grew rapidly. Figure 1, Panel A plots the large increase in expenditures on Customs houses, which tracks the growth in gross imports. To further document the increasing capacity at Customs houses, we digitized the *Official Register of the United States*. The *Register* contains data on all workers at each Customs house, and it additionally reports the tasks they performed, including whether they performed multiple tasks.

In the wake of a scandal revealing the low rate of customs collection, in 1818 Congress invested in tariff collection by hiring "well qualified" staff (Rao, 2016). Figure 1, Panel B shows the rapid increase in Customs house employment, from around 200 in 1816 to 2,000 by the 1840s.[7] In tandem with this growth, employees became more specialized. Figure 1, Panel C shows that the fraction of workers engaged in more than one task declined from nearly 20 percent in 1821 to 7.5 percent by the end of our sample. Moreover, even as workers became more specialized in single tasks, more total tasks were added. In Figure 1, Panel D we plot the number of occupation titles over time. In 1821, workers focused on general tasks such as weighing, measuring, and gauging. By 1850, the number of distinct occupations nearly tripled, to around 50. The majority of the documented specialization is in measuring distinct goods, as workers had ever-narrower remits.

2.2 Overview of Tariff Legislation

Initially, tariffs served almost entirely to generate revenue (Taussig, 1910). After the War of 1812, policymakers also used tariffs as a tool to protect key industries from foreign competition (Irwin, 2017). One approach was the introduction of so-called "minimum" tariffs, beginning with cotton manufactures in 1816 (Taussig, 1910). Minimum tariffs imposed a statutory minimum valuation when assessing ad valorem duties. For instance, in 1816 the minimum valuation for cotton manufactures was 25 cents per square yard. Given the ad valorem rate of 25 percent, cheap cotton manufactures imported under the 1816 law effectively paid a specific tariff of 6.25 cents per square yard. Over the 1820s, minimum tariffs were applied to a variety of goods, from broad categories such as woolen manufactures to narrow ones such as straw hats.

The Tariff of 1824 broadened protection for iron and wool, reflecting mounting pressure from Northern producers (Taussig, 1910). Controversy deepened with the Tariff of 1828 (the "Tariff of Abominations"), which sparked fierce objections from the South due to its high rates. The Tariff of 1832 was passed with some Southern support, but not enough to mollify all objectors, and the ensuing rancor culminated in the Nullification Crisis (Irwin,

[7]Total employment in 1816 is likely underestimated, as only the collector, the surveyor, and the naval officer were recorded as federal employees.

2008). In response, under President Andrew Jackson, Congress passed the Compromise Tariff of 1833, simplifying the tariff code and lowering rates to mitigate tensions (Taussig, 1910). The Tariff of 1842 (the "Black Tariff") dramatically raised duties after the economic downturn of the late 1830s (Irwin, 2017). In 1846, the Walker Tariff broadly lowered tariff rates and simplified the tariff code, cutting the number of distinct rates by over 90 percent and eliminating all specific tariffs. Additionally, Congress passed many smaller tariff acts throughout the period.

2.3 Customs Enforcement and Tariff Design

As tariffs evolved to serve both protective and fiscal purposes, administration and enforcement posed ongoing challenges. From the early days of the Republic, there were concerns over "frauds of the revenue" and calls for rigorous enforcement against smugglers and unscrupulous traders (Hamilton, 1790).

Policymakers argued that to prevent evasion, tariffs should be "simple in their provisions and easily understood" (Hayes, 1877). They understood that ad valorem rates, while theoretically ideal, invited misrepresentation of goods' values (Rao, 2016). To that end, Congress instructed Customs officials to inspect a random set of packages (often a tenth) to partially deter evasion. It was understood that the costs of collection would be non-trivial, and so Treasury reports consistently presented tariff revenue *net* of collection costs. During the 19th century, five Treasury Secretaries reported to Congress that specific duties were preferable to ad valorem to prevent evasion (Burrows, 1894).[8] President Taylor made the same argument during his only State of the Union Address (Taylor, 1849). Policymakers also understood the downsides of specific duties: They were not "practicable" for goods with substantial price dispersion (Walker, 1847; Burrows, 1894).

3 Model

In this section, we set up and analyze a stylized model of tariff code design in the shadow of evasion. In Section 3.1, we describe the economic environment and the government's problem. In Section 3.2, we study optimal tariff design when the government is constrained to assess a uniform tariff on all imported goods. This setting parsimoniously illustrates the model's main economic forces. In Section 3.3, we extend our analysis to permit disaggregation of goods in order to customize tariffs. We will use this version of the model to obtain testable predictions in Section 4.

3.1 Economic Environment

The government of a small open economy operates a customs house with the goal of raising revenue through tariffs. In this economy, a representative consumer purchases a continuum

[8] They were Secretaries Hamilton, Gallatin, Dallas, Crawford, and Meredith.

of domestic and imported varieties, with nested CES preferences and a unit elasticity of substitution between domestic and imported goods, i.e. Cobb-Douglas preferences. As a result, consumption of domestic goods is independent of the tariff code, and total expenditure on imports is a fixed share of national income. We therefore normalize expenditure on imports to 1 and drop domestic goods from our analysis. Meanwhile, the consumer has elasticity of substitution $\sigma > 1$ between varieties of imported goods.

Each import variety is produced and transported at a constant marginal cost in a perfectly competitive industry. Its equilibrium price is therefore its (production and transportation) cost, which we refer to as its *value*, plus expected tariff payments.[9] Letting $r(v)$ be the expected per-unit tariff payment for a variety of value v, the equilibrium price of each variety is $p(v) = v + r(v)$. We assume that values of import varieties have distribution $F \in \Delta(\mathbb{R}_+)$, where F has full support on $\mathbb{R}_+$ and is absolutely continuous with density f. We further assume that $\mathbb{E}[v] < \infty$ and $\mathbb{E}[v^{-\sigma-1}] < \infty$.

Our assumption of a representative consumer with CES preferences over imports implies that the price index

$$P = \left(\mathbb{E}\left[p(v)^{1-\sigma}\right]\right)^{1/(1-\sigma)}$$

is a sufficient statistic for consumer welfare, where $\mathbb{E}$ denotes averages over the values of import varieties under distribution F.

3.1.1 Government's Problem

The government's objective is to raise an exogenously specified tariff revenue $R > 0$, net of any tariff administration and enforcement costs, while minimizing the impact of tariffs on consumer welfare. It has no additional motives for imposing tariffs.[10] Since we have normalized consumer expenditures on imports to 1, R can be interpreted as the revenue raised *per import dollar*. It can also be interpreted as the equivalent ad valorem tariff burden (i.e., the percentage of consumer expenditure collected by the government) in a frictionless benchmark.

To raise revenue, the government can levy either an ad valorem tariff, charged as a percent of declared import value; or a specific tariff, assessed as a fixed payment per unit imported. It faces two frictions when administering tariffs. First, importers can misdeclare values used to compute ad valorem tariffs. The customs house can attempt to verify these declarations,

[9] We assume that importers set prices and quantities before shipping their goods. As a result, these variables cannot adjust to random variation in realized tariff payments. In equilibrium, importers correctly predict the expected tariff they will pay at the dock when choosing supply.

[10] Our assumption of a small open economy ensures that tariffs do not impact economic outcomes beyond their direct incidence on the prices consumers face for imported goods. Of course, real-world policymakers may impose tariffs for reasons not captured by our model. In our view, these additional motives are plausibly orthogonal to the revenue-raising and consumer-welfare concerns we model.

but verification is costly and the penalties that can be levied for fraudulent declarations are limited. We model this verification process formally in Section 3.2. Second, the government cannot assess tailored tariff regimes and rates without administrative investment. It must train customs officials to distinguish among categories of varieties when assessing tariffs, and this training is costly. We model these classification costs formally in Section 3.3.

Our model makes a fundamental distinction between import values, which require effort to verify; and import categories and quantities, which we take to be costlessly observable (once training costs have been sunk). This modeling distinction reflects a real-world dichotomy. Categories and quantities are objective features of a shipment that can be codified and rapidly verified by a trained inspector. By contrast, shipment value is inherently subjective and cannot be fully codified ex ante. As a result, assessing value requires costly ex post verification effort exerted on a per-shipment basis.[11]

Even when importers are *obviously* lying, an expert and a judge still need to be involved. Indeed, during the time period covered by our sample, Customs officials reported difficulties punishing importers who submitted invoices containing prices that "no one believes," particularly since "the accused party (especially with French goods) can bring witnesses to prove almost anything" (Walker, 1845; Meredith, 1849). Even today, it remains costly and difficult to bring charges of misvaluation through the courts (United States Court of International Trade, 2018).

3.2 Optimal Single-Category Tariff Codes

We begin our analysis in a setting in which the customs house cannot directly distinguish any subsets of varieties. As a result, the government must assign a uniform tariff across all imported goods. We refer to the combination of tariff regime and rate as the *tariff code*.

The government first posts a tariff code, following which all importers simultaneously post prices. The representative consumer then purchases a bundle of imports at the posted prices. Finally, all importers simultaneously import the purchased quantities, and the customs house assesses tariffs.

3.2.1 Tariff Assessment and Verification

Specific tariff: Imported quantities are directly observed by the customs house. As a result, under a specific tariff, the expected tariff revenue on each imported unit is exactly the posted rate.

Ad valorem tariff: Ad valorem tariffs are assessed on each imported unit in two stages. First, the importer declares a non-negative value for the imported unit and pays the posted

[11] Policymakers have contemplated standardized rules for valuing all shipments. However, the proposed methods were complicated and politically contentious due to their imprecision, and they have never been implemented (Goss, 1897; Smith, 1948).

tariff on the declared value. Next, the customs house decides whether to verify each unit's value, at a cost per unit of $c_{AV} > 0$. We assume that each unit is assessed separately and units are unbundled from shipments before assessment. As a result, customs officials cannot make inferences regarding a unit's value based on the other goods with which it arrives. Verification allows the customs house to perfectly observe a unit's value and collect, at a penalty rate of up to $\kappa > 1$ times the scheduled rate, any tariff revenue that an importer attempted to evade by fraudulent declaration (as in Yitzhaki 1987). We interpret the maximum penalty rate as an exogenous restriction that cannot be tailored by the government, for instance because of jurisprudential limits on civil penalties or practical difficulties with collecting large fines from importers.

3.2.2 Equilibrium Tariff Collection

We assume that each importer acts to maximize expected profits net of tariff payments, while the customs house acts to maximize its expected tariff revenue from each importer, net of verification costs. Our solution concept for this importation game is perfect Bayesian equilibrium, with no precommitment by the customs house to verify declarations.

Specific tariff: Given our assumptions, there is no opportunity for evasion under a specific tariff. As a result:

Proposition 1. *Suppose the government imposes a specific tariff t. Then there exists a perfect Bayesian equilibrium of the importation game. In every equilibrium, no verification takes place and the expected tariff payment on each unit is t, while the customs house's net tariff revenue is*

$$\Pi_S(t) = t \cdot \frac{\mathbb{E}[(v+t)^{-\sigma}]}{\mathbb{E}[(v+t)^{1-\sigma}]}$$

and the aggregate price index is

$$\bar{P}_S(t) = \left(\mathbb{E}[(v+t)^{1-\sigma}]\right)^{1/(1-\sigma)}.$$

(All proofs of formal results are collected in the Online Appendix.)

For expositional simplicity and technical convenience, we restrict attention to value distributions under which revenue is globally increasing in the specific tariff rate.[12] Going forward, we maintain the following assumption:

Assumption 1. Π_S *is increasing in* t.

This assumption amounts to a mild regularity condition on the distribution of import values. As the following lemma illustrates, it holds under a wide class of commonly used

[12] A rising specific tariff tilts consumption toward high-value goods, reducing the total quantity of imports and counteracting the higher revenue from each unit. In pathological environments, this substitution effect can locally dominate the direct effect of a rising tariff.

parametric distributions.

Lemma 1. *Assumption 1 holds if f is differentiable and $v \cdot f'(v)/f(v)$ is decreasing in v. This condition is satisfied by the log-normal, generalized gamma (including gamma, Weibull, and half-normal), inverse gamma, Lomax, and log-logistic distributions.*

Ad valorem tariff: Under an ad valorem tariff, some verification must take place in equilibrium to deter fraudulent declarations, and some misdeclaration must take place to motivate verification. For tractability and plausibility, we restrict attention to *monotone* equilibria, in which the importer's equilibrium declaration policy is a pure strategy $\rho : \mathbb{R}_+ \to \mathbb{R}_+$ that is non-decreasing and (strictly) increasing whenever positive. Here $\rho(v)$ is the value declared by an importer with true value v. We let $\phi : \mathbb{R}_+ \to [0,1]$ denote the customs house's (potentially mixed) verification policy, with $\phi(v')$ the probability that a declared value of v' is verified. (Trivially, whenever verification takes place, the maximum permissible penalty rate is levied on any evasion.)

The following proposition establishes that there is a unique equilibrium outcome within the class of monotone equilibria, and it characterizes this outcome. Let $\bar{v} \equiv \mathbb{E}[v^{1-\sigma}]/\mathbb{E}[v^{-\sigma}]$ be the *unit value of consumption*, i.e., the average value of a unit of imported goods. (Under an ad valorem tariff, this value is independent of the tariff rate.) Also let $\bar{c}_{AV} \equiv c_{AV}/(\kappa \cdot \bar{v})$ be the verification cost per import dollar.

Proposition 2. *Suppose the government imposes an ad valorem tariff τ.*

If $\tau < \bar{c}_{AV}$, then there exists a perfect Bayesian equilibrium of the importation game under a tariff satisfying $\phi(v') = 0$ for every v' and $\rho(v) = 0$ for every v. In this equilibrium, the expected tariff payment on each unit is 0, while the customs house's net tariff revenue is $\Pi_{AV}(\tau) = 0$ and the aggregate price index is $\bar{P}_{AV}(\tau) = \left(\mathbb{E}[v^{1-\sigma}]\right)^{1/(1-\sigma)}$. Further, every monotone equilibrium is outcome equivalent to this one.[13]

If $\tau > \bar{c}_{AV}$, then there exists an equilibrium satisfying $\phi(v') = 1/\kappa$ for every v' and

$$\rho(v) = \begin{cases} v - c_{AV}/(\kappa \cdot \tau), & v \geq v^* \\ 0, & v < v^*, \end{cases}$$

where $v^* \in (c_{AV}/(\kappa \cdot \tau), \infty)$ uniquely satisfies

$$\frac{\mathbb{E}[v^{1-\sigma} \mid v < v^*]}{\mathbb{E}[v^{-\sigma} \mid v < v^*]} = c_{AV}/(\kappa \cdot \tau).$$

In this equilibrium, the expected tariff payment on each unit is $\tau \cdot v$, while the customs house's

[13] In the knife-edge case $\tau = \bar{c}_{AV}$, it can be shown that this outcome remains an equilibrium, but there are additional monotone equilibria which differ in the verification rate when $v' = 0$. In particular, any choice of $\phi(0) \in [0, 1/\kappa]$ is consistent with equilibrium, and all choices generate the same net tariff revenue. The choice $\phi(0) = 0$ maximizes consumer welfare.

net tariff revenue is

$$\Pi_{AV}(\tau) = \frac{\tau - \bar{c}_{AV}}{1+\tau}$$

and the aggregate price index is

$$\bar{P}_{AV}(\tau) = (1+\tau) \cdot \left(\mathbb{E}[v^{1-\sigma}]\right)^{1/(1-\sigma)}.$$

Further, every monotone equilibrium is outcome equivalent to this one.

When the tariff rate is sufficiently high, importers shade their declared value down by a fixed *absolute* amount $c_{AV}/(\kappa \cdot \tau)$, except for importers with values close to zero, who shade all the way down to zero.[14] Meanwhile, the customs house verifies all declarations at a fixed rate. Perhaps surprisingly, importers do not gain from underdeclaring on average—their declarations are verified just often enough that their expected tariff payment (including penalties) is the same as if they declared truthfully. While verification effectively dissipates any gains from underdeclaring, it also entails costs relative to a benchmark without misdeclaration. These costs are substantial enough that the customs house does not generate revenue from an ad valorem tariff unless the tariff rate is sufficiently high.

Our result that importers shade their declarations by a constant value stands in contrast to standard assumptions in the literature on tariff evasion, where importers are usually assumed to evade a constant *share* of tariff payments (Mishra et al., 2008; Jean and Mitaritonna, 2010). In line with the lessons of a venerable literature on policy adaptations to strategic behaviors (Mirrlees, 1971; Björkegren et al., 2026), if importers choose to evade a constant share of tariffs, then the government can simply increase the ad valorem rate to compensate (Haan et al., 2012). However, if importers evade tariffs non-proportionally, the possibility of evasion has non-trivial implications for the efficiency of raising revenue through ad valorem tariffs.

3.2.3 Optimal Tariff Regime

We next explore the tradeoff between ad valorem and specific tariffs as tools for raising revenue. The following result establishes basic properties of the specific and ad valorem tariffs required to achieve any given revenue. (Recall that the revenue goal is net of tariff enforcement costs, and so gross collections must cover any verification costs under an ad valorem tariff.)

Proposition 3. *For every $R \in (0,1)$, there exists a unique specific tariff $t^*(R)$ such that $\Pi_S(t^*(R)) = R$. This tariff is increasing in R and satisfies $t^*(0+) = 0$ and $t^*(1-) = \infty$.*

[14] The construction of Proposition 2 can be straightforwardly adapted to accommodate a lower bound $\underline{v} > 0$ on unit values. If $\underline{v} > c_{AV}/(\kappa \cdot \tau)$, then $v^* = \underline{v}$ and all declared values are non-zero and shaded equally.

For every $R \in (0,1)$, there exists a unique ad valorem tariff $\tau^*(R)$ such that $\Pi_{AV}(\tau^*(R)) = R$. This tariff is increasing in R and satisfies $\tau^*(0+) > 0$ and $\tau^*(1-) = \infty$.

A key distinction between specific and ad valorem tariffs is that the specific tariff converges to zero as the revenue goal vanishes, while the ad valorem tariff does not. Intuitively, the ad valorem tariff must cover verification costs on every unit imported even when very little (net) revenue is required, and these costs do not vanish as revenue approaches zero.

Let $P_S(R) \equiv \bar{P}_S(t^*(R))$ and $P_{AV}(R) \equiv \bar{P}_{AV}(\tau^*(R))$ be the price indices under, respectively, the specific and ad valorem tariffs which collect revenue R. The following result characterizes their relative growth rates and establishes a single-crossing condition.

Proposition 4. *The following relationships hold:*

- $P_S(0+) < P_{AV}(0+)$,
- $\frac{d}{dR}\log P_S(R) > \frac{d}{dR}\log P_{AV}(R)$ *for all* $R > 0$,
- *There exists a* $c^*_{AV} > 0$ *such that* $\lim_{R\to 1} P_S(R)/P_{AV}(R) > 1$ *if and only if* $c_{AV} < c^*_{AV}$.

The behavior of the price index under specific and ad valorem tariffs is visualized in Figure 2. Panel A plots the price index distortion by revenue goal for different tariff regimes. As a benchmark, we also plot the price index if the government had access to perfect (costlessly enforced) ad valorem tariffs. Panel B plots the lower envelope of the price indices, tracing out the distortion under the optimal tariff.

As shown in Figure 2, the first property of Proposition 4 establishes that specific tariffs are less distortionary than ad valorem tariffs for small revenue goals. This result follows directly from the non-vanishing ad valorem verification costs discussed above. The second property establishes that the price distortion from specific tariffs grows more quickly than from ad valorem tariffs. Intuitively, specific tariffs distort relative prices and push consumers to substitute toward higher-value goods, reducing the total quantity of goods consumed. This substitution effect reduces the rate at which revenue can be raised by increasing the specific tariff, entailing larger marginal price distortions than an ad valorem tariff that preserves relative prices.

These two properties taken together imply that P_S and P_{AV} cross at most once. Whether they cross depends on the cost of verification, as established by the third property. When verification costs are low, crossing occurs, and there exists a revenue goal $R^* \in (0,1)$ such that the specific tariff is more efficient when $R < R^*$, while the ad valorem tariff is more efficient otherwise. If verification costs are high, then a specific tariff is *always* more efficient.

3.3 Optimal Multi-Category Tariff Codes

We now extend our model to allow the customs house to partly disaggregate import varieties. We provide a brief overview of the extended model here, while deferring a detailed technical description to Appendix D.

We assume that the government can pay a cost to classify goods into a finite set of distinguishable *categories*. Classification is costly because it requires training customs inspectors to distinguish goods. Further, finer-grained classification schemes entail higher training costs. We model these costs as fixed, i.e., independent of the volume of imports or quantity of tariffs collected.

When the government posts a multi-category tariff code, it commits to a categorization of goods alongside tariff regimes and rates. Formally, goods are grouped into a finite number of *products* represented by a finite set $\mathcal{G}$, with each product $g \in \mathcal{G}$ containing a proportion $\mu(g) \in (0,1)$ of varieties in the economy. A product is an irreducible category within which varieties cannot be further distinguished by a customs agent (or a court) regardless of training.[15] We will write F_g for the distribution of unit values associated with each product $g \in \mathcal{G}$, with $\mathbb{E}_g$ denoting expectations with respect to F_g. We impose the same regularity assumptions on F_g as we did on the distribution of values in our single-category model.

A *categorization* $\mathcal{K}$ is a partition of the set $\mathcal{G}$. Each category $K \in \mathcal{K}$ represents a set of products that customs inspectors have been trained to distinguish from other products. The government may tailor both the tariff regime and rate across (but not within) categories. Within each category, tariff assessment occurs as in the single-category analysis of Section 3.2.

The government may choose any categorization $\mathcal{K}$ of $\mathcal{G}$ subject to payment of a classification cost $c_S \cdot H(\mathcal{K})$. The cost function H satisfies $H(\mathcal{K}^0) = 0$ for the trivial categorization $\mathcal{K}^0 = \{\mathcal{G}\}$ which does not distinguish any products; and $H(\mathcal{K}') > H(\mathcal{K})$ whenever $\mathcal{K}'$ is a (proper) refinement of $\mathcal{K}$. The cost parameter $c_S > 0$ is exogenously given and can be interpreted as an inverse measure of state capacity. Any classification costs must be recouped by tariff receipts. That is, the revenue goal R is net of the classification cost $c_S \cdot H(\mathcal{K})$.

In Appendix D, we formally state the multi-category tariff code design problem, establish that an optimal tariff code exists, and characterize structural features of an optimal code. These technical developments underpin the theoretical predictions we obtain in Section 4.

4 Theoretical Predictions

In this section, we apply the model of tariff-code design developed in Section 3 to derive a set of testable predictions regarding optimal tariff codes. In Section 4.1, we develop an empirical test for the presence of misvaluation under ad valorem tariffs. In Section 4.2, we predict how tariff regimes and rates vary across categories within an optimal code. In

[15] Finiteness of the set of products is imposed for technical convenience, to ensure that an optimal tariff code exists and that the regularity conditions maintained on the universe of varieties continue to apply within each category even when products can be completely disaggregated.

Section 4.3, we predict how features of an optimal code change with two dimensions of state capacity, namely the cost of training customs inspectors to classify goods and the cost of verifying declared values. In Section 4.4, we collect these predictions in one place and assign them numerical labels for ease of reference.

4.1 Testing for Misvaluation

Before considering how state capacity shapes the structure of the tariff code, we derive a model-based empirical test for the presence of misvaluation. When the government switches a tariff from specific to ad valorem, reported unit values change for two reasons. First, under ad valorem tariffs, importers misdeclare values in equilibrium (Proposition 2). Second, specific tariffs distort relative prices across varieties within a category. Since the implied ad valorem rate t/v is higher for low-v varieties, a switch to an ad valorem tariff lowers the tariff on cheaper varieties proportionally more, shifting consumption toward them and lowering average (honestly) reported unit values (Alchian and Allen, 1964).

The net change in reported unit values following a switch from specific to ad valorem tariffs may be quantitatively decomposed as a sum of these two effects. Given a category of goods and a specific tariff rate t, we define the ad valorem equivalent rate $\bar{\tau}(t)$ as the ad valorem rate that would preserve the share of consumer category spending raised as tariff revenue, in a benchmark setting where importers cannot misdeclare values. The share of expenditure raised as tariff revenue under a (frictionless) ad valorem tariff is $\tau/(1+\tau)$, while under a specific tariff it is $t \cdot \mathbb{E}[(v+t)^{-\sigma}]/\mathbb{E}[(v+t)^{1-\sigma}]$. Equating these expressions and solving for τ yields the ad valorem equivalent rate

$$\bar{\tau}(t) = \frac{t \cdot \mathbb{E}[(v+t)^{-\sigma}]}{\mathbb{E}[v \cdot (v+t)^{-\sigma}]}.$$

Proposition 5. *Fix a tariff code in which a specific tariff $t > \bar{\tau}^{-1}(\bar{c}_{AV})$ is levied on some category. If the code is modified so that the category receives the ad valorem tariff $\bar{\tau}(t)$, then reported log (average) unit value within the category falls by an amount*

$$\log\left(\frac{\bar{\tau}(t)}{\bar{\tau}(t) - \bar{c}_{AV}}\right) + \Delta(\bar{\tau}(t)),$$

where $\Delta(\tau)$ is independent of c_{AV} and satisfies

$$\Delta(\tau) = \beta \cdot \tau + o(\tau),$$

with $\beta > 0$.

The first term in this decomposition measures the change in reported values due to fraudulent declarations. The second term measures the Alchian–Allen compositional effect. Importantly, it vanishes at $\tau = 0$ and has a first-order dependence on the ad valorem equivalent rate. This decomposition therefore suggests a test for misvaluation: Once we control for the linear dependence of unit values on the ad valorem equivalent, any additional decline in reported unit values following a switch from specific to ad valorem tariffs reflects misvaluation.

4.2 Patterns of Tariff Incidence

We now study the distribution of tariff burdens and regimes within an optimal tariff code.

Tariff burdens: To consistently measure and compare levels of taxation across categories, we focus on the *tariff burden* within each category. Two notions are useful. The *gross* tariff burden measures the percentage of category spending paid as tariffs or fines, while the *net* tariff burden measures the percentage of spending collected by the government as revenue, net of verification costs.

In a category subject to a specific tariff t, the two notions of burden coincide and are equal to $\bar{\tau}(t)/(1+\bar{\tau}(t))$, where $\bar{\tau}(t)$ is the ad valorem equivalent rate defined in Section 4.1. In a category subject to an ad valorem tariff τ, the gross burden is higher than the net burden due to verification costs. The gross tariff burden is $\tau/(1+\tau)$, so gross tariff burdens map directly onto tariff rates within ad valorem categories.

Under an ad valorem tariff, the optimal gross tariff burden (and therefore tariff rate) is dictated by the unit value of consumption in the category. Recall that the unit value of consumption is the average value of a unit of imported goods (for any level of ad valorem tariff). It is given by $\bar{v}_j = \mathbb{E}[v_j^{1-\sigma}]/\mathbb{E}[v_j^{-\sigma}]$, where expectations are taken with respect to the distribution of values within category j.

Proposition 6. *Under an optimal tariff code, across categories subject to ad valorem tariffs, the gross tariff burden (and therefore tariff rate) is lower in categories with higher unit values of consumption.*

This result contrasts with the classic finding in frictionless environments that a *uniform* ad valorem tariff is optimal. In an environment with verification costs, total costs rise with the number of imported units that need verification. As a result, the government prefers a non-uniform ad valorem tariff to tilt consumption toward higher-value categories requiring less verification.

Meanwhile, the *net* tariff burden varies systematically between categories subject to different regimes.

Proposition 7. *Under an optimal tariff code, the net tariff burden is higher in categories subject to ad valorem tariffs than in those subject to specific tariffs.*

This result relates to the finding of Proposition 4 that marginal price distortions are lower under ad valorem than specific tariffs. As a result, the government is more willing to trade price distortions for incremental revenue in ad valorem categories than in categories subject to specific tariffs. Recall that gross tariff burdens are equal to net tariff burdens in categories subject to specific tariffs, but are higher than net burdens under ad valorem tariffs. Therefore:

Corollary 1. *Under an optimal tariff code, the gross tariff burden is higher in categories subject to ad valorem tariffs than in those subject to specific tariffs.*

Tariff regimes: A category's optimal tariff regime depends on both the average unit value of consumption in the category and the dispersion of values around that average. Higher average unit values and greater dispersion both make ad valorem tariffs more likely. We formally establish these claims for *balanced* category pairs, which attract equal spending from the representative consumer in a zero-tariff benchmark. Comparing balanced categories avoids complications that arise when categories attract different levels of consumer spending.

To vary unit values while controlling for dispersion, we compare categories with a fixed proportional variation in values. Formally, we say that two categories j and k with value distributions F_j and F_k are *value-ordered* if $F_k(v) = F_j(v/\alpha)$ for some $\alpha > 1$ and all v. In other words, the distribution of values in category k is a scaled-up version of the distribution in category j. Value-ordered categories differ in their (average) unit values of consumption but exhibit the same relative dispersion of unit values around their means.

Proposition 8. *Under an optimal tariff code, whenever two balanced value-ordered categories use different tariff regimes, the cheaper category uses a specific tariff.*

Intuitively, the costs of administering an ad valorem tariff rise with the number of units requiring verification. Since fewer units are imported in high-value categories, administrative costs are lowest there, making an ad valorem tariff more attractive.

To vary dispersion while controlling for (average) unit values, we compare categories with a fixed unit value of consumption. Formally, given a value distribution F, we define the quantity-weighted value distribution F^Q by $F^Q(v) \equiv \int_0^v w^{-\sigma}\, dF(w) / \int_0^\infty w^{-\sigma}\, dF(w)$. We say that two categories j and k with value distributions F_j and F_k are *dispersion-ordered* if F_k^Q is a mean-preserving spread of F_j^Q. By construction, any two dispersion-ordered categories have identical (average) unit values, but they differ in the heterogeneity of unit values around the average.

Proposition 9. *Under an optimal tariff code, whenever two balanced dispersion-ordered categories use different tariff regimes, the more-heterogeneous category uses an ad valorem tariff provided that its net tariff burden is not too high.*

Intuitively, specific tariffs distort relative prices because the ad valorem equivalent rate is lower for expensive varieties. As a result, when values are dispersed, specific tariffs impose a large consumption distortion. In a single-category setting, this force tends to favor ad valorem tariffs under more-dispersed value distributions. (Recall that consumption distortions drive the higher price index growth rates under specific tariffs in Proposition 4.) Proposition 9 is a natural multi-category analog of that logic.

The assumption of a low tariff burden in Proposition 9 plays a technical role. We measure dispersion using quantity weights under a zero-tariff benchmark. However, a specific tariff itself shifts consumption away from cheap varieties. As a result, under high tariffs our measure of dispersion need not track the true quantity-weighted dispersion metric, and therefore the true marginal price distortion. Focusing on low tariff burdens avoids such complications. More generally, the conclusion of Proposition 9 holds whenever the *aggregate* price distortion under the empirically observed tariff burden is larger in the more-dispersed category.

4.3 Improvements in State Capacity

The United States invested heavily in state capacity throughout the 19th century. Ex ante, it is plausible that these investments reduced the cost of either classification (c_S) or verification (c_{AV}). We now establish that these two channels have opposing implications for the structure of an optimal tariff code.

Declines in c_S: As classification costs fall, it becomes worthwhile to draw increasingly fine distinctions among goods. This force spurs a transition from coarse categories toward more disaggregated ones. For expositional simplicity, we formally establish this claim in environments where every product would be worth distinguishing in the absence of classification costs:

Definition 1. *A model is* non-degenerate *if no two products share the same tariff regime and rate under any optimal tariff code when $c_S = 0$.*

Let $\mathcal{K}^{FB}$ be the finest possible categorization, in which every product can be distinguished and assigned a customized tariff regime and rate. At the other extreme, let $\mathcal{K}^0$ be the coarsest possible categorization, in which all products must be assigned a uniform tariff regime and rate. The following result establishes that every optimal tariff code uses the coarsest possible categorization when classification costs are high, and the finest possible one when costs are low.

Proposition 10. *Hold all model parameters fixed except for c_S. For c_S sufficiently large, every optimal tariff code uses the categorization $\mathcal{K}^0$. If the model is non-degenerate, then for c_S sufficiently small, every optimal tariff code uses the categorization $\mathcal{K}^{FB}$.*

The transition toward finer classification schemes established in Proposition 10 is associated with a greater reliance on specific tariffs. Since individual products do not have perfectly homogeneous values, we establish this claim formally for products that are sufficiently homogeneous:

Definition 2. *A set of products $\mathcal{G}$ is* δ-homogeneous *if*

$$\sum_{g\in\mathcal{G}} \mu(g)\mathbf{1}\left\{\mathbb{E}_g[v] - \left(\mathbb{E}_g[v^{1-\sigma}]\right)^{1/(1-\sigma)} \leq c_{AV}/\kappa\right\} \geq 1-\delta.$$

If a product g were comprised of homogeneous varieties with no variation in values, then $\mathbb{E}_g[v] = (\mathbb{E}_g[v^{1-\sigma}])^{1/(1-\sigma)}$ and the indicator on the left-hand side of this definition would be 1 for that product. In actuality, every product exhibits some variation in values. Then since $\sigma > 1$, Jensen's inequality implies that $\mathbb{E}_g[v] > (\mathbb{E}_g[v^{1-\sigma}])^{1/(1-\sigma)}$. When the value distribution F_g is close to a point mass on a single value, the gap between these two quantities becomes small. A δ-homogeneous set of products satisfies this homogeneity condition for most products.[16]

Proposition 11. *Hold all model parameters fixed except for c_S. Suppose that the set of products is δ-homogeneous for some $\delta \in [0, 1)$. Then for c_S sufficiently small, the fraction of varieties assigned a specific tariff is at least $1-\delta$ under every optimal tariff code.*

When the set of products is δ-homogeneous, at least a fraction $1-\delta$ of varieties would optimally receive a specific tariff under any net tariff burden, supposing that the tariff code distinguishes all products. (Recall Proposition 4, which obtains conditions in a single-category setting under which specific tariffs are optimal for all revenue goals.) If the model is non-degenerate, then Proposition 10 ensures that an optimal tariff code distinguishes all products when classification costs are low. Hence, when δ is small, most varieties must receive specific tariffs in this limit. In fact, this conclusion holds even for degenerate models, as the proof of the proposition demonstrates. While an optimal code under a degenerate model may not distinguish all products in the limit of low classification costs, it must still eventually use specific tariffs for all categories.

Declines in c_{AV}: Declines in verification costs naturally make ad valorem tariffs more attractive and complex tariff codes less profitable. Declines in verification and classification costs therefore have opposing implications. As the following Proposition formally demonstrates, if state capacity improvements primarily lower verification costs, the tariff code should become simpler and more ad valorem-reliant.

Proposition 12. *Hold all model parameters fixed except for c_{AV}. For c_{AV} sufficiently small, every optimal tariff code consists of a single category which is assessed an ad valorem tariff. For c_{AV} sufficiently large, every optimal tariff code uses only specific tariffs.*

[16] As a special case, one could insist that *all* products satisfy this condition, i.e., $\delta = 0$. We view such an assumption as typically unrealistic, since most real-world classification schemes contain at least one "not otherwise classified" category with potentially substantial heterogeneity in values. Our more general result accommodates such residual categories.

4.4 Summary of Model Predictions

Our model of tariff code design is built on the assumptions that unscrupulous importers may shade down their declared values, and that it is prohibitively costly for the government to partition imports into categories that perfectly identify unit values. It delivers several testable predictions regarding how verification and classification costs shape the design and effects of an optimal tariff code:[17]

1. Switching from a specific to an ad valorem tariff will lower reported unit values, even after controlling for substitution toward cheaper varieties (Proposition 5).
2. Specific tariffs will be assessed on goods with relatively homogeneous values (Proposition 9).
3. Specific tariffs will be assessed on goods with relatively low gross tariff burdens (Corollary 1).
4. Among goods assessed ad valorem tariffs, cheaper goods will face higher rates (Proposition 6).
5. Specific tariffs will be assessed on cheaper goods (Proposition 8).

Our model also provides tests for whether improvements in state capacity reduce classification or verification costs, as the two channels yield contrasting predictions for tariff design:

6. *Tariff complexity:* As c_S falls, the tariff code will become more complex, with more distinct categories receiving tailored tariff regimes and/or rates (Proposition 10). Conversely, as c_{AV} falls, the tariff code will become simpler (Proposition 12).
7. *Tariff regimes:* As c_S falls, the tariff code will utilize specific tariffs for more goods (Proposition 11). Conversely, as c_{AV} falls, the tariff code will utilize ad valorem tariffs for more goods (Proposition 12).

5 Empirical Results

In this section, we describe our data and test the predictions of the model. We start with the model's cross-sectional implications, focusing on the characteristics that determine whether goods receive specific or ad valorem tariffs. We then evaluate the historical evolution of the U.S. tariff code, comparing the alternative hypotheses that increased state capacity reduced the cost of classification versus verification.

[17] In our empirical analysis, we use the term "good" to refer to a standardized import category which is comparable across tariff codes. We therefore phrase our predictions in terms of goods in anticipation of testing them empirically.

5.1 Historical Data

There was no formal classification system for imports during our sample period. Tariff acts instead listed goods in statutory prose, with each tariff line applying to a verbal category such as "all woolen manufactures." We refer to each such line of text as a *label*. A central difficulty for analysis is that, across tariff acts, labels are related many-to-many: Some define a narrow set of imports, while others are much broader.

For our empirical analysis, we define *goods*: harmonized product categories that can be consistently followed across tariff acts. Specifically, a good is defined as a set of varieties such that (i) over our sample all varieties in the set always face the same tariff; and (ii) the set is materially distinct from any other set that follows the same sequence of tariffs over that period. Within each good, there may nevertheless be many latent *varieties* with different unit values. For instance, a good may be "high-quality woolen flannels": While high-quality woolen flannels of different colors may have different values, they all share the same label in every year.

Intuitively, a good is the smallest product category that can be assigned a distinct tariff history in our data. Appendix A describes the full matching procedure, examples, and validation exercises. Appendix Table A.1 illustrates the linking procedure for flannels, and Figure A.1 shows the statutory language for woolen manufactures in 1828. Some labels cover only one good, while others cover many. There is variation within tariff acts in how many goods each label refers to, and variation over time within goods in how many other goods share the same label.

We also concord the tariff acts to the 1843-1849 editions of the *Commerce and Navigation of the United States*, which report import values and, for some imports, quantities. Because the import statistics were reported inconsistently over time, they do not support a harmonized panel: Reported categories change across years, so few import lines can be tracked. We calculate repeated cross-sectional measures of ad valorem equivalents for goods with specific tariffs.[18]

In the aftermath of the 1846 Walker Tariff, when *all* specific tariffs were converted to ad valorem, Treasury reports and letters from Customs officials highlighted increases in fraudulent declarations. Meredith (1849) documented 1843-1849 prices and quantities for 13 goods that initially had specific tariffs and then transitioned to ad valorem.[19] Meredith

[18] In the model, cross-category comparisons are stated in terms of tariff burdens. In the data, we instead calculate ad valorem equivalents (AVEs): the statutory rate τ for ad valorem goods, and $\bar{\tau}(t)$ for goods receiving specific tariffs, as defined in Section 4.1. The AVE in the latter case is readily computed from the data as $t/v_S(t)$, where $v_S(t)$ is the reported average unit value in the import data, i.e., total value divided by total quantity. AVEs are easier to interpret than tariff burdens, and they preserve cross-category rankings: The gross tariff burden equals $\text{AVE}/(1+\text{AVE})$ under either tariff regime.

[19] The goods were English beer, Scottish beer, brandy, cigars, claret, grain spirits, Madeira, other red

compiled these data expressly to corroborate officials' accounts by quantifying the decline in reported unit values, in effect using what we would now call an interrupted time-series design. We augment Meredith's data in order to run difference-in-differences regressions, using as a control group the five goods that initially did not have a purely specific tariff but nevertheless had values and quantities consistently recorded in the *Commerce and Navigation* tables.

A key component of the model is price heterogeneity across varieties within a particular good. We do not directly observe the quantities and values by shipments in the historical data, so we cannot directly measure price dispersion. To proxy for within-good dispersion, we follow a large literature in the spirit of Rauch (1999) and classify goods as homogeneous if they are commodities. To create our measure of commodities, we digitized a list of commodities from the 1863–64 *Executive Documents* printed by the House of Representatives.[20] We also run analyses where we consider goods as homogeneous if they are associated with singleton labels: ones that only cover one good.

5.2 Empirical Strategy

Our main estimating equation is of the form

$$Y_{it} = \beta Z_{it} + \Lambda_{it} + \gamma_t + \gamma_i + \epsilon_{it}, \tag{1}$$

where Y_{it} is the outcome for good i at time t. In addition to the covariate of interest, Z_{it}, we may include other potentially time varying controls Λ_{it}, as well as, when feasible, time or good fixed effects. Standard errors are clustered by good. For the analysis on historical tariffs, the good is defined in Appendix A, though for our analysis on imports the unit of observation is what is listed in the import tables (which typically covers multiple goods from the perspective of the tariff acts).[21]

For our country-level analysis, we estimate the following equation:

$$y_c = \beta X_c + \epsilon_c, \tag{2}$$

where y_c is the outcome for country c. X_c represents characteristics at the country level, which is the level of Customs house capacity. We calculate robust standard errors, and show a variety of robustness checks in Appendix B.

5.3 Direct Tests of the Model

In this subsection, we test the model's direct predictions. We begin with evidence that importers misvalue their goods under ad valorem tariffs, following which we examine which

wine, port, sherry, Sicilian wine, West India spirits, and white wine.

[20] We also show our results are robust to using Rauch's modern classification.

[21] Across all good-level regressions, we include only goods with specific or ad valorem tariffs, for instance, dropping those on the free list.

types of goods receive ad valorem versus specific and high versus low tariffs. Because our predictions follow from the structure of an optimal tariff code under limited state capacity, confirming them validates the model's central mechanism.

5.3.1 Prediction 1: Evidence for Misvaluation

To test whether ad valorem tariffs induce importers to shade declared values, as in Prediction 1, we study changes around the 1846 Walker Tariff. Our estimating equation is a relatively standard difference-in-differences regression, augmented with controls for ad valorem equivalents to control for Alchian and Allen (1964) effects:

$$\begin{aligned} Y_{it} = &\ \beta\,\text{Specific}_i \times \text{Post-1846}_t \\ &+ \gamma_1\,\text{Post-1846}_t \times \text{AVE}_{1842,i} + \gamma_2\,\text{Specific}_i \times \text{Post-1846}_t \times \text{AVE}_{1842,i} \\ &+ \gamma_3\,\text{Post-1846}_t \times \text{AVE}_{1846,i} + \gamma_4\,\text{Specific}_i \times \text{Post-1846}_t \times \text{AVE}_{1846,i} \\ &+ \gamma_i + \gamma_t + \epsilon_{it}. \end{aligned} \tag{3}$$

where Y_{it} is the reported unit value, and Specific_i measures if good i had a specific tariff pre-1846. To control for Alchian and Allen (1964)-type effects, we additionally control for Post times the goods' ad valorem equivalents (separately for the 1842 and 1846 acts), interacted with initially having a specific tariff. Each t indexes an import table, roughly corresponding to a year; the 1846 act took effect in December 1846, so the 1846/47 fiscal year is split into two tables. To estimate elasticities, we use Poisson Pseudo Maximum Likelihood (PPML) regressions (Silva and Tenreyro, 2006). We also estimate event study regressions, where we estimate a separate coefficient for Specific_i for each year.

The switch from specific to ad valorem tariffs was followed by a sharp decline in reported unit values. Table 1 estimates a decline of roughly 40 percent. Figure A.2 shows that relative prices were stable before the reform and collapsed afterward, consistent with contemporaneous reports from Customs officials.[22]

5.3.2 Prediction 2: Homogeneity and Specific Tariffs

A key prediction of the model is that specific tariffs are more likely to be assessed on relatively homogeneous goods. We test this prediction using the structure of the tariff acts, exploiting the fact that some goods in some acts are covered by singleton labels. Table 2, Panel A shows that these narrow labels are more likely to have specific tariffs. Columns 1-3 pool across all years in our data. Relative to the baseline regression in Column 1, Column 2 adds year-by-SITC (3-digit) fixed effects, and Column 3 uses within-good variation in the

[22] A feature of the Alchian and Allen (1964) effect is that we would expect the total value of a good imported to weakly rise after its tariff switches from specific to ad valorem, as the effective price of the good falls. However, in Table A.3 we show that total value instead weakly falls for the goods that switch from specific to ad valorem tariffs.

narrowness of their labels by including good fixed effects. Column 4 instead includes only 1842, the final tariff act in the sample. Across all specifications, singleton labels are around 10 percentage points more likely to be specific (the baseline mean in the whole sample is about 10 percent, and was 50 percent in 1842).

As an alternative, we also test Prediction 2 using properties of the goods themselves. Table 2, Panel B shows that goods listed as commodities are around 15 percentage points more likely to have specific tariffs. The result is robust to including year-by-SITC (2-digit) fixed effects, or only including 1842.

5.3.3 Predictions 3, 4, and 5: Revenue, Rates, and Specific Tariffs

An important feature of our environment is that governments wish to avoid paying verification costs when the revenue from a particular shipment is low. This concern yields several distinct testable predictions.

Prediction 3 implies that imports with lower ad valorem equivalents should be more likely to receive specific tariffs. Table 3 tests this prediction by relating tariff regime to bins of the 1842 ad valorem equivalent (AVE). Panel A shows a clear gradient: Relative to imports with AVEs above 25 percent, imports with AVEs below 5 percent are around 20 percentage points more likely to receive specific tariffs; imports with AVEs between 5 and 15 percent are around 15 percentage points more likely; and imports with AVEs between 15 and 25 percent are around 5 percentage points more likely. Panel B adds fixed effects for the import's label in the 1832 tariff act, the last major act before 1842. The comparison is therefore among imports that were grouped together in the same statutory category but were later assigned different tariffs. Consistent with the theory, as these labels were subdivided, the subcategories with the lowest (gross) tariff burdens were more likely to receive specific tariffs.

Prediction 4 states that, conditional on having an ad valorem tariff, imports with lower unit values should have higher tariff rates. Table 4 shows that a one-standard-deviation unit value increase is associated with an ad valorem rate around 1 percentage point lower. We observe this pattern whether we use only unit values reported under the 1842 act, only those under the 1846 act, or those reported in all years.

Prediction 5 states that goods with low unit values are more likely to have specific tariffs. Unfortunately, testing this prediction with regressions on 19th century data is difficult, as under the 1842 act unit values were rarely reported for goods with ad valorem tariffs. However, it is well understood that cheaper goods within particular categories were more likely to face specific tariffs (Irwin, 2017). In particular, the minimum tariffs discussed in Section 2.2 exactly fit the logic of the model: The cheaper varieties of a particular type of import (such as cotton cloths worth less than 25 cents a square yard in 1816) paid what was effectively a specific tariff, as the tariff act was a function only of the quantity (Taussig, 1910). We test Prediction 5 more systematically in modern data below.

5.4 The Evolution of Tariffs

We next use time-series data to evaluate whether increased state capacity reduced the cost of classification or verification. The model predicts that declines in the two types of costs push the tariff code in opposite directions. We can therefore use the actual evolution of the tariff code to distinguish which margin of state capacity improved. We also show that complexity and the share of specific tariffs comove, as the model predicts under either type of cost reduction.

5.4.1 Prediction 6: Complexity over Time

Figure 3 shows that tariff legislation in the United States became increasingly complex over time. Panel A displays the number of goods that had been defined by each year. That is to say, it records how many categories of imports would have had a unique path of tariffs, had we stopped our analysis in a particular year. The number of distinct goods initially grew modestly, starting with 105 goods in 1789. In the aftermath of the professionalization and expansion of the Customs workforce in 1818, the number of listed goods increased by more than 25 percent. There were 1,696 distinct goods defined by 1842.

Alongside the increase in goods, the number of distinct tariff rates grew dramatically. Panel B shows that the number of rates in each year tripled to around 150 over our sample period. However, it is worth noting that in the aftermath of the 1832–33 Nullification Crisis, the tariff code was simplified, lowering the number of distinct tariff rates by a sixth. The number of rates increased again only in the 1840s.

The model predicts that a decrease in c_S should increase tariff code complexity, as occurred in our sample. By contrast, the model predicts that a decrease in c_{AV} should lead to a simpler tariff code, a prediction at odds with the data. The data therefore suggest that improved state capacity mostly reduced classification costs rather than verification costs.

This conclusion is also consistent with the composition of the U.S. Customs workforce throughout this time period. Despite the large expansion in Customs headcount, there were only 24 appraisers by the end of our sample, suggesting that increased state capacity was mostly not directed toward deterring misvaluation.

5.4.2 Prediction 7: Specificity over Time

The increasing complexity of the tariff code just documented was broadly accompanied by a shift toward specific tariffs. Figure 3, Panel C shows that, in 1789, approximately one-tenth of goods had specific tariffs and, by the 1840s, half of all goods had specific tariffs. There was a sharp rise in the specific tariff share after the Customs workforce was professionalized in 1818, an event which we have seen was accompanied by an uptick in tariff code complexity.

In the model, it is not capacity *per se* that leads to an increase in specific tariffs, but rather an increase in complexity. Consistent with this mechanism, when the tariff code was simplified in the 1830s, the share of specific tariffs fell by a third, rising again only in the 1840s alongside complexity.

This comovement of complexity and usage of specific tariffs validates our model, which predicts such comovement as a consequence of increased state capacity. As with the complexity of the tariff code, the pattern of increasing reliance on specific tariffs suggests that state capacity mostly reduced classification costs rather than verification costs.

5.5 Empirical Results: The Modern Era

We supplement our historical analysis with data on modern tariffs. First, we use data on American imports for 1981–1985, from the U.S. Census in the Annual Import Data Bank.[23] From 1981 to 1985, quantities, values, and (estimated) tariff revenue for imports were reported by exporting country and 7-digit Tariff Schedules of the United States Annotated (TSUS) product. The tariff schedule was disaggregated to the 5-digit level at the time (Acosta and Cox, 2023; Greenland et al., 2026a). This source lets us measure unit values and ad valorem equivalents for all imports, which we use to construct a direct measure of price dispersion within tariff labels: Within each 5-digit code, we calculate the standard deviation of log unit values across exporter-by-7-digit-code imports.

We also use data on current tariffs from the World Trade Organization (WTO)'s *Integrated Database* (IDB). We focus on the applied Most Favored Nation (MFN) rates in the most recent year available for each country.[24] Tariffs are recorded at the level of a tariff line. Each tariff line is a 6 to 10 digit code, with the first 6 digits harmonized across countries according to the World Customs Organization's Harmonized System (HS). The trailing digits are defined by each importing country—for example, the United States uses a 10-digit system maintained by the U.S. International Trade Commission, while the European Union maintains the 8-digit Combined Nomenclature. Importantly, the database provides a classification that allows us to identify whether duties are ad valorem.[25] We obtain ad valorem equivalents for specific tariffs from the UN TRAINS dataset. We merge ad valorem equivalents into our primary WTO tariff data using tariff line codes.

To test the model's predictions regarding how lower enforcement costs affect tariff design, we use a direct measure of cross-country differences in Customs house capacity. In particular,

[23] See Feenstra (1996), Schott (2008), and Greenland et al. (2026a) for more detail. To avoid measurement issues, we use Column 1 tariffs.

[24] We use MFN rates in the modern data to avoid measurement issues; see Teti (2024) for more discussion of the data.

[25] The database classifies duties using 5 tariff regimes: Ad valorem, Specific, Compound, Mixed, and Other. Our main analysis focuses solely on tariffs that are either explicitly ad valorem or specific.

we use trade misreporting indices derived by Farhad et al. (2024). These indices quantify the extent of trade misreporting by comparing each country's reports to their trading partners'. Our perspective is that more capable Customs houses are better at measuring trade flows. Our main analysis uses both an export misreporting index and a standardized overall trade misreporting index, where lower values indicate less misreporting. We prefer the former measure, because export misreporting is a measure of capacity that is not directly endogenous to the applied tariffs. We use the most recent year available for each country in the database as our measure of bureaucratic customs capacity.

5.5.1 Revenue, Homogeneity and Specific Tariffs: Evidence from the 1980s United States

A key advantage of 1980s U.S. import statistics is that we have information on values and quantities for almost all imports, allowing us to directly test Prediction 5: Cheaper goods are more likely to have specific tariffs. Table 5, Panel A shows that doubling unit values leads to around a two percentage point decrease in the likelihood of specific tariffs. The unit of analysis is TSUS 5-digit codes, and the results are robust to controlling for 4-digit fixed effects.

The 1980s tariff code is also consistent with the model's other predictions regarding which goods have specific tariffs. Table 5, Panel B tests Prediction 2, that more variance makes specific tariffs less attractive. Instead of using commodity status as a proxy for homogeneity, we directly measure the dispersion of log prices across 7-digit varieties within each 5-digit TSUS code. A one standard deviation increase is associated with around a 10 percentage point decrease in the probability of a specific tariff across all goods, and around 2-4 percentage points within 4-digit codes. Table 5, Panel C tests Prediction 3, showing that doubling the ad valorem equivalent leads to around a 15 percentage point decrease in the likelihood of specific tariffs.[26]

5.5.2 Revenue, Homogeneity and Specific Tariffs: Evidence Across Countries

We now turn to replicating and extending our tests of the model's predictions for the modern era, comparing tariffs across countries. To study how tariff regimes vary across goods, we use the same regression specification as in the historical data, Equation (1), with goods as the unit of observation.

Table 6, Panel A tests Prediction 2, that price homogeneity makes specific tariffs more attractive. Commodities are about 0.8 percentage points more likely to be specific. Although the coefficient is lower than in our historical analysis, specific tariffs are also currently less

[26] Acosta and Cox (2024) show Prediction 4 holds in the 1980s: Cheaper goods had higher ad valorem rates. It is worth noting that Acosta and Cox (2024) offer an explanation distinct from the theory we present: They find that in the 20th century, the negative correlation between unit values and ad valorem rates emerged in the 1950s, and they argue that protectionism was a driving force.

common, covering around 2 percent of global tariffs. While we do not observe how the tariff acts are written and therefore lack a direct modern analog to singleton labels, we can take advantage of the modern tariff hierarchy. Goods that do not share the same tariff with any of the other tariff lines within the same HS 4-digit code (referred to as "non-grouped" goods) are around one percentage point more likely to have specific tariffs. We also find that the interaction of the two, homogeneity and not being grouped, predicts specificity.

Table 6, Panel B tests Prediction 3, that goods with lower (gross) tariff burdens are more likely to be specific. Indeed, we find that goods with low (below 5%) ad valorem equivalents are around five percentage points more likely to be specific. Figure A.3 visualizes the relationship, showing that there are almost no goods with ad valorem tariffs well below 5%, but many with specific tariffs below that equivalent rate.[27]

5.5.3 The Role of Improved Capacity: Evidence Across Countries

We now turn to exploring how state capacity predicts the overall structure of the tariff code. We use simple bivariate cross-country regressions, where the outcomes are features of each country's tariff code and the independent variable is a measure of how effectively each country keeps track of its trade flows (Farhad et al., 2024).

Table 7, Panel A shows, consistent with Prediction 6, that countries with less capability use coarser partitions in the tariff code. They use fewer distinct tariff rates overall, and they are less likely to split HS 6-digit categories into subcategories with different tariff rates. Consistent with Prediction 7, Panel B shows that these countries have a lower share of specific tariffs, either at the level of the tariff line or grouping to HS 6-digit level. Overall, a one standard deviation increase in Customs house misreporting lowers the share of tariff lines with specific tariffs by around two percentage points. As with the historical evidence, the evidence in the modern data is consistent with countries with higher state capacity having lower classification costs.

5.6 Alternative Explanations

We conclude our empirical analysis by discussing several alternative ways of thinking about the economic environment.

A justification for specific tariffs outside our model is that they stabilize prices, because their ad valorem equivalent is negatively correlated with a good's unit value (Lampe, 2020). While 19th-century policymakers were aware of this phenomenon (Woodbury, 1835), it is rarely mentioned in policy documents as a motivation for specific tariffs. Rather, policymakers argued that Congress could stabilize effective duties by updating specific tariffs as

[27] Another pattern likely reflecting an endogenous response to state capacity is that non-OECD countries exhibit substantial heaping in their ad valorem tariff rates, especially at values ending in 5 or 0. Hamilton (1795) encouraged heaping, writing "inconvenient fractions ... serve to perplex the calculation of the duties."

prices adjusted (Burrows, 1894). While we do not observe measures of price volatility over time in the 19th century, we find little evidence that this mechanism matters for policies in the 1980s: Table A.2 shows that goods with more price variation over time were not disproportionately likely to have specific tariffs.

An additional evasion tactic outside our model is misclassification of goods by importers, in the face of an established classification system run by trained customs inspectors. Although misclassification does occur in practice,[28] the historical record suggests that it is of secondary importance compared to misvaluation. Historically, Customs officials acknowledged that "false weights and measures" occurred but argued they were easy to detect and "not practiced to any great extent" (Walker, 1845; Burrows, 1894).[29] In line with this evidence, our model singles out verification of shipment values as entailing substantially higher variable costs than verification of classifications, once inspectors have been trained to classify goods. We believe that misvaluation concerns play a central role in shaping tariff codes even when goods might occasionally be misclassified.

An alternative explanation for the increasing complexity of the tariff code in the 19th century is the Industrial Revolution: Over time, there were more goods to classify. Industrial policy motivations for tariffs also became more important, and some specific tariffs were designed to protect relatively cheap, low-quality American products (Irwin, 2017). Tables A.4 and A.5 address these concerns using restricted samples.

In both tables, Panel A restricts the sample to non-manufactured goods, which are less directly affected by the expansion of manufacturing during the Industrial Revolution.[30] Panel B drops goods that ever received different tariffs for varieties with different prices, such as cheap and expensive wool (Irwin and Temin, 2001). Panel C drops goods whose industry was mentioned in Hamilton's (1791) *Report on Manufactures* (Irwin, 2004; Sylla, 2024). Across these restricted samples, the results are similar to those in our main exhibits.

A complementary perspective on misvaluation incentives is that importers might work harder to falsify values when tariff rates are high (Fisman and Wei, 2004). While we have

[28]For instance, in 2024, Ford paid a $365 million fine for allegedly misclassifying cargo vans by installing "sham rear seats and other temporary features to make the vans appear to be passenger vehicles" (Shepardson, 2024).

[29]As additional support for this perspective in a modern context, Anne et al. (2023) find that misvaluation drives most tariff revenue losses from evasion in Madagascar.

[30]Manufactured goods are around half of all goods, so estimates that both drop manufactured goods and restrict attention to 1842 rely on a substantially smaller sample. Our findings on the development of complex classification schemes are also not driven by differences in product composition, either historically or today. Figure A.5 shows that the tariff code became progressively more complex and reliant on specific tariffs over the 19th century, even when the sample is restricted to non-manufactured imports. Table A.6 shows that the positive relationship between capacity and tariff complexity/specificity in the modern cross-country data remains when the sample is limited to goods that are broadly imported (by all of the twenty non-OPEC countries with tariff data under HS 2022 and with the highest levels of misreported trade).

limited data, we do not find evidence for such a relationship between tariff rates and misvaluation in our historical setting. Figure A.4 plots reported unit values of unmanufactured wool—the sole good for which we observe unit prices, duties which are not purely specific, and a change in the tariff treatment of those duties—in the 1840s. When the duty changed in 1846, there was no systematic change in wool's reported unit price. In line with this finding, 19th-century Treasury officials opined that misvaluation was "inherent" in ad valorem tariffs, regardless of the rate (Manning, 1885).

6 Discussion

States design policy around their administrative capacity. We show how this consideration shapes tariff design. High-capacity customs houses are able to implement complex, disaggregated tariff codes. Such codes are suited to specific tariffs, especially for homogeneous goods and those with low tariff burdens. Our theory rationalizes both the historical evolution of the U.S. tariff code as well as modern patterns. Tariffs are the subject of active policy debates, and our results contribute to these discussions.

Until recently, the United States imposed no tariffs on low-value imports intended for consumption (Fajgelbaum and Khandelwal, 2025). Many countries maintain some form of *de minimis* exemption, as the administrative costs of collecting revenue on such imports typically outweigh the fiscal benefits. In February 2025, the United States attempted to eliminate the exemption for goods originating in China, but was forced to reverse the decision shortly thereafter when Customs authorities became overwhelmed (Lawder et al., 2025). Like their 19th-century counterparts, policymakers quickly recognized how to avoid wasting resources on verifying low-revenue shipments. In May 2025, the exemption was replaced by a new specific component: Importers can pay a fixed rate per package (Tobin and Chang, 2025).

Recent technological advances are increasing the capacity of customs houses around the world (Budimirovic et al., 2022) and improving tax collection more broadly (Dzansi et al., 2025). These technologies are especially effective at preventing misclassification (Davies et al., 2024). Our results suggest that these tools are most effective when paired with more complex tariff schedules. Our theory additionally sheds light on how complexity should be utilized to minimize price distortions while preserving government revenue.

References

Acosta, Miguel and Lydia Cox, "HTS to TSUS Concordance," *Working Paper*, 2023.

— **and** — , "The Regressive Nature of the U.S. Tariff Code: Origins and Implications," *Working Paper*, 2024.

— , — , **Andrew Greenland, John Lopresti, Christopher M. Meissner, Martin Rotemberg, and Sharon Traiberman**, "US Tariff Policy Since 1789," *Working Paper*, 2026.

Alchian, Armen A. and William R. Allen, *University Economics*, Belmont, CA: Wadsworth Publishing, 1964.

Allingham, Michael G. and Agnar Sandmo, "Income Tax Evasion: A Theoretical Analysis," *Journal of Public Economics*, 1972, *1* (3–4), 323–338.

Anne, Clément Xavier Frederik, Cyril Romain Chalendard, Ana Margarida Fernandes, Bob Rijkers, and Vincent Vicard, "Containing Tariff Evasion," Policy Research Working Paper 10606, World Bank 2023.

Azcárraga, Augusto Azael Pérez, José García Sanjinés, Rossana San Juan, Selvin Lemus, Philip Wood, and Robert Kokoli, "Developing a Risk-Based Compliance Improvement Plan for Customs Administrations," IMF Technical Notes and Manuals 2025.

Bairoch, Paul, "European Trade Policy, 1815–1914," in Peter Mathias and Sidney Pollard, eds., *The Cambridge Economic History of Europe, Volume 8: The Industrial Economies: The Development of Economic and Social Policies*, Cambridge, U.K.: Cambridge University Press, 1989, pp. 1–60.

Bartelme, Dominick, Arnaud Costinot, Dave Donaldson, and Andrés Rodríguez-Clare, "The Textbook Case for Industrial Policy: Theory Meets Data," *Journal of Political Economy*, 2025, *133* (5), 1527–1573.

Basri, M. Chatib, Mayara Felix, Rema Hanna, and Benjamin A. Olken, "Tax Administration versus Tax Rates: Evidence from Corporate Taxation in Indonesia," *American Economic Review*, 2021, *111* (12), 3827–3871.

Bates, Robert H., Avner Greif, Margaret Levi, Jean-Laurent Rosenthal, and Barry R. Weingast, *Analytic Narratives*, Princeton, NJ: Princeton University Press, 1998.

Besley, Timothy and Torsten Persson, "The Origins of State Capacity: Property Rights, Taxation, and Politics," *American Economic Review*, 2009, *99* (4), 1218–1244.

Betz, Timm, "Tariff evasion and trade policies," *International Studies Quarterly*, 2019, *63* (2), 380–393.

Bhagwati, Jagdish N., "On the Underinvoicing of Imports," *Oxford Bulletin of Economics and Statistics*, 1964, *26* (4), 389–397.

Björkegren, Daniel, Joshua E Blumenstock, and Samsun Knight, "Manipulation-Robust Prediction," *American Economic Review*, 2026. Forthcoming.

Border, K. C. and Joel Sobel, "Samurai Accountant: A Theory of Auditing and Plunder," *The Review of Economic Studies*, 1987, *54* (4), 525–540.

Bown, Chad P. and Douglas A. Irwin, "Mainly Poor Countries Use Tariffs as a Major Source of Government Revenue," July 16 2019.

Budimirovic, Milena, Vyara Filipova, Tejo Kusuma, Özlem Soysanli, Kian Cassehgari Posada, Emmanuelle Ganne, and Sheri Rosenow, *The Role of Advanced Technologies in Cross-Border Trade: A Customs Perspective*, World Trade Organization, 2022.

Burrows, Julius C., *The Tariff: Speech of Hon. Julius C. Burrows, of Michigan, in the House of Representatives, Tuesday, January 9, 1894*, Washington, D.C.: Government Printing Office, 1894.

Casey, Katherine, "Politics and Policy Where State Capacity Is Low," in Pascaline Dupas, Pinelopi Koujianou Goldberg, and Rohini Pande, eds., *Handbook of Development Economics*, Vol. 6, Amsterdam: Elsevier, 2026, chapter 4.

Chalendard, Cyril, Ana M. Fernandes, Gael Raballand, and Bob Rijkers, "Corruption in Customs," *The Quarterly Journal of Economics*, February 2023, *138* (1), 575–636.

Costinot, Arnaud, Dave Donaldson, Jonathan Vogel, and Ivan Werning, "Comparative Advantage and Optimal Trade Policy," *Quarterly Journal of Economics*, 2015, *130* (2), 659–702.

Cournot, Antoine, *Recherches sur les principes mathématiques de la théorie des richesses*, Paris: L. Hachette, 1838.

Crucini, Mario J., "Sources of Variation in Real Tariff Rates: The United States, 1900-1940," *American Economic Review*, 1994, *84* (3), 732–743.

Dar, Manzoor, Kyle Emerick, Davis Gimode, Erin Kelley, and Gregory Lane, "Market Unraveling and Quality Control in Agriculture.," *Pre-Analysis Plan*, 2025.

D'Arcy, Michelle, Marina Nistotskaya, and Ola Olsson, "Cadasters and Economic Growth: A Long-Run Cross-Country Panel," *Journal of Political Economy*, 2024, *132* (11), 3786–3826.

Das, Satya P. and Shabtai Donnenfeld, "Trade Policy and Its Impact on the Quality of Imports: A Welfare Analysis," *Journal of International Economics*, 1987, *23* (1–2), 77–95.

Davies, Ronald B., Kyle McNabb, and Miroslav Palanský, "You Can't Tax What You Can't See: Using Fixed Cargo Scanners to Combat Tax Evasion," *Working Paper*, 2024.

Dzansi, James, Anders Jensen, David Lagakos, and Henry Telli, "Technology and Tax Capacity: Evidence from Local Governments in Ghana," *Working Paper*, 2025.

Elkins, Stanley M. and Eric L. McKitrick, *The Age of Federalism*, New York: Oxford University Press, 1993.

Fajgelbaum, Pablo D. and Amit Khandelwal, "The Value of De Minimis Imports," *Working Paper*, 2025.

Falvey, Rod, "The Composition of Trade within Import-Restricted Product Categories," *Journal of Political Economy*, 1979, *87*, 1105–1114.

Farhad, Mohammad, Jakob B. Madsen, Michael Jetter, and Andrew Williams, "The Mystery of Misreported Trade: A Novel Method to Identify Distorted Trade Figures," *The World Economy*, 2024, *47* (2), 748–778.

Feenstra, Robert C, "US imports, 1972-1994: Data and concordances," 1996.

Fisman, Raymond and Shang-Jin Wei, "Tax Rates and Tax Evasion: Evidence from "Missing Imports" in China," *Journal of Political Economy*, April 2004, *112* (2), 471–496.

_ , Shang-Jin Moustakerski, and Shang-Jin Wei, "Outsourcing Tariff Evasion: A New Explanation for Entrepôt Trade," *The Review of Economics and Statistics*, 2008, *90* (3), 587–592.

Flynn, Joel P., George Nikolakoudis, and Karthik A. Sastry, "Organizational Targets in General Equilibrium," *Working Paper*, 2026.

Goss, John Dean, *The History of Tariff Administration in the United States: From Colonial Times to the McKinley Administrative Bill*, 2nd ed., New York: Columbia University, 1897.

Goyal, Radhika, Anders Jensen, David Lagakos, and Abdoulaye Ndiaye, "Tax Productivity and Development: A Quantitative Macroeconomic Analysis," *Working Paper*, 2025.

Grant, Matthew, "Where Do the Data Come From? Endogenous Classification in Administrative Data," *Working Paper*, 2025.

Greenland, Andrew and John Lopresti, "Trade and Structural Change: Focusing on the Specifics," *Working Paper*, 2025.

_ , James Lake, and John Lopresti, "AVE Tariffs are a Misleading Measure of Tariff Policy," *Working Paper*, 2026.

_ , _ , and _ , "Trade and US Inequality in the Tokyo Round," *Working Paper*, 2026.

Grossman, Gene M. and Elhanan Helpman, "Protection for Sale," *The American Economic Review*, 1994, *84* (4), 833–850.

Haan, Marco A., Pim Heijnen, Lambert Schoonbeek, and Lars A. Toolsema, "Sound Taxation? On the Use of Self-Declared Value," *European Economic Review*, 2012, *56* (2), 205–215.

Hamilton, Alexander, "Report Relative to a Provision for the Support of Public Credit," 1790. Treasury Department, January 9. Communicated to Congress on January 14, 1790.

— , "Alexander Hamilton's Final Version of the Report on the Subject of Manufactures," *The Papers of Alexander Hamilton*, 1791, *10*, 230–340. Edited by Harold C. Syrett. New York: Columbia University Press, 1966.

— , "Improvement of the Revenue: Communicated to the House of Representatives, February 2, 1795," in Henry Cabot Lodge, ed., *The Works of Alexander Hamilton, (Federal Edition), vol. 3*, New York and London: G.P. Putnam's Sons, 1904, pp. 515–541.

Hayes, Rutherford B., "First Annual Message," Presidential Address December 1877.

Helpman, Elhanan and Paul R. Krugman, *Trade Policy and Market Structure*, Cambridge, MA: MIT Press, 1989.

International Trade Centre, *Getting down to business: Making the Most of the WTO Trade Facilitation Agreement*, Geneva: ITC, 2020.

Irwin, Douglas A., "Changes in U.S. Tariffs: The Role of Import Prices and Commercial Policies," *The American Economic Review*, 1998, *88* (4), 1015–1026.

— , "The Aftermath of Hamilton's Report on Manufactures," *Journal of Economic History*, 2004, *64* (3), 800–821.

— , "Antebellum Tariff Politics: Regional Coalitions and Shifting Economic Interests," *Journal of Law and Economics*, 2008, *51* (4), 715–742.

— , *Clashing over Commerce: A History of US Trade Policy*, Chicago: University of Chicago Press, 2017.

— , "Trade Policy in American Economic History," *Annual Review of Economics*, 2020, *12*, 23–44.

— **and Peter Temin**, "The Antebellum Tariff on Cotton Textiles Revisited," *The Journal of Economic History*, 2001, *61* (3), 777–798.

Iyoha, Ebehi, Edmund Malesky, Jaya Wen, and Sung-Ju Wu, "Exports in Disguise? Trade Rerouting During the U.S.-China Trade War," *Working Paper*, 2025.

Javorcik, Beata S. and Gaia Narciso, "Differentiated Products and Evasion of Import Tariffs," *Journal of International Economics*, December 2008, *76* (2), 208–222.

— **and** — , "WTO Accession and Tariff Evasion," *Journal of Development Economics*, 2017, *125*, 59–69.

Jean, Sébastien and Cristina Mitaritonna, "Determinants and Pervasiveness of the Evasion of Customs Duties," *Working Paper*, 2010.

Jensen, Anders, "Employment Structure and the Rise of the Modern Tax System," *American Economic Review*, 2022, *112* (1), 213–234.

Jones, Andrew A., *Jones's Digest: Being a Particular and Detailed Account of the Duties Performed by the Various Officers Belonging to the Custom-House Departments of the United States*, New York: G. F. Hopkins & Son, Printers, 1835.

Kapon, Samuel, Lucia Del Carpio, and Sylvain Chassang, "Using Divide-and-Conquer to Improve Tax Collection," *The Quarterly Journal of Economics*, 2024, *139* (4), 2475–2523.

Keen, Michael, "The balance between specific and ad valorem taxation," *Fiscal studies*, 1998, *19* (1), 1–37.

Klein, Alexander and Christopher M. Meissner, "Did Tariffs Make American Manufacturing

Great? New Evidence from the Gilded Age," *Working Paper*, 2025.

Klemperer, Paul and Margaret Meyer, "Price Competition vs. Quantity Competition: The Role of Uncertainty," *RAND Journal of Economics*, 1986, *17* (4), 618–638.

Koehler, Kristin Graham, Ted Murphy, Craig Francis Dukin, and Aaron M. Applebaum, "Tariff Evasion Is Within DOJ's Crosshairs: Potential Criminal and Civil Consequences for Noncompliance With Increased U.S. Tariffs," *Insights, Sidley Austin LLP*, 2025.

Krishna, Kala, "Protection and the Product Line: Monopoly and Product Quality," *International Economic Review*, 1990, *31* (1), 87–102.

Lampe, Markus, "European Trade Policy in the 19th Century," in Jonathan H. Hamilton, ed., *Oxford Research Encyclopedia of Economics and Finance*, Oxford, U.K.: Oxford University Press, 2020.

Lashkaripour, Ahmad and Volodymyr Lugovskyy, "Profits, Scale Economies, and the Gains from Trade and Industrial Policy," *American Economic Review*, 2023, *113* (10), 2759–2808.

Lawder, David, Helen Reid, Lisa Baertlein, and Lisa Barrington, "Trump pauses de minimis repeal as packages pile up at US customs," *Reuters*, February 2025.

Manning, Daniel, *Report of the Secretary of the Treasury on the State of the Finances for the Year Ending June 30, 1885*, Washington, D.C.: Government Printing Office, 1885. Submitted to Congress, December 7, 1885.

Mastrorocco, Nicola and Edoardo Teso, "State Capacity as an Organizational Problem," *Review of Economic Studies*, 2026. Forthcoming.

Meredith, William M., "Report on Finances," Annual Report of the Secretary of the Treasury 1849.

Mirrlees, James A., "An Exploration in the Theory of Optimum Income Taxation," *Review of Economic Studies*, 1971, *38* (2), 175–208.

Mishra, Prachi, Arvind Subramanian, and Petia Topalova, "Tariffs, Enforcement, and Customs Evasion: Evidence from India," *Journal of Public Economics*, 2008, *92* (10-11), 1907–1925.

Political Risk Services (PRS) Group, "International Country Risk Guide," Technical Report 2022.

Rao, Gautham, *National Duties: Custom Houses and the Making of the American State*, University of Chicago Press, 2016.

Rauch, James E., "Networks versus markets in international trade," *Journal of International Economics*, 1999, *48* (1), 7–35.

Reamer, Andrew, "Before the U.S. Tariff Commission: Congressional Efforts to Obtain Statistics and Analysis for Tariff Setting, 1789–1916," in Paul R. Bardos, ed., *A Centennial History of the United States*, International Trade Commission, 2016, chapter 2, pp. 33–70.

Rodrik, Dani, "Second-best institutions," *American economic review*, 2008, *98* (2), 100–104.

Rotunno, Lorenzo, Pierre-Louis Vézina, and Zheng Wang, "The Rise and Fall of (Chinese) African Apparel Exports," *Journal of Development Economics*, 2013, *105*, 152–163.

Schott, Peter K, "The Relative Sophistication of Chinese Exports," *Economic policy*, 2008, *23* (53), 6–49.

Scott, James C., *Seeing Like a State: How Certain Schemes to Improve the Human Condition Have Failed*, New Haven, CT: Yale University Press, 1998.

Sequeira, Sandra, "Corruption, Trade Costs, and Gains from Tariff Liberalization: Evidence from Southern Africa," *American Economic Review*, 2016, *106* (10), 3029–3063.

Shepardson, David, "Ford to Pay $365 Million in US Import Tariff Evasion Case," *Reuters*, March 2024.

Silva, João M. C. Santos and Silvana Tenreyro, "The Log of Gravity," *The Review of Economics and Statistics*, 2006, *88* (4), 641–658.

Slemrod, Joel and Shlomo Yitzhaki, "Tax Avoidance, Evasion, and Administration," in Alan J. Auerbach and Martin Feldstein, eds., *Handbook of Public Economics*, Vol. 3, Elsevier, 2002, chapter 22, pp. 1423–1470.

Smith, R. Elberton, *Customs Valuation in the United States: A Study in Tariff Administration*, Chicago: University of Chicago Press, 1948.

Sylla, Richard, "Alexander Hamilton's Report on Manufactures and Industrial Policy," *Journal of Economic Perspectives*, 2024, *38* (4), 111–130.

Taussig, Frank W., *The Tariff History of the United States*, 6th ed., New York: G.P. Putnam's Sons, 1910.

Taylor, Zachary, "Annual Message," Presidential Address December 1849.

Teti, Feodora A., "Missing Tariffs," 2024. Working Paper.

The Merchants' and Bankers' Almanac, "The Merchants' and Bankers' Almanac," New York: Office of the Bankers' Magazine and Statistical Register 1869.

Tobin, Meaghan and Agnes Chang, "What to Know About Trump's Latest Changes to Taxes on Small Packages From China," *The New York Times*, May 2025.

United Nations Conference on Trade and Development, "Trade Analysis and Information System (TRAINS)," 2025.

United States Congress, "Executive Documents Printed by Order of the House of Representatives, During the First Session of the Thirty-Eighth Congress, 1863–'64," Technical Report 1194, Government Printing Office, Washington, D.C. 1864.

United States Court of International Trade, "The Origins and Creation of the Board of General Appraisers," Technical Report 2018.

U.S. Census Bureau, "Annual Import Data Bank Files (IA-245/IM-145), 1964–2007," Accessed 2025. National Archives Catalog, `https://catalog.archives.gov/id/566340`.

U.S. Department of State, "Official Register of the United States," various. Washington, DC: Government Printing Office.

Walker, Robert J., "Report on Finances," Annual Report of the Secretary of the Treasury 1845.

_ , "Report on Finances," Annual Report of the Secretary of the Treasury 1847.

Wang, Zhu and Julian Wright, "Ad valorem platform fees, indirect taxes, and efficient price discrimination," *RAND Journal of Economics*, 2017, *48* (2), 467–484.

Woodbury, Levi, *Report on the Finances, December 1835*, Washington, D.C.: U.S. Department of the Treasury, 1835. Annual Report of the Secretary of the Treasury on the State of the Finances.

World Trade Organization, "Integrated Database (IDB)," 2025. `https://tao.wto.org`.

Wulf, Luc De and José B. Sokol, *Customs Modernization Handbook* Trade and Development, Washington, DC: World Bank, 2005.

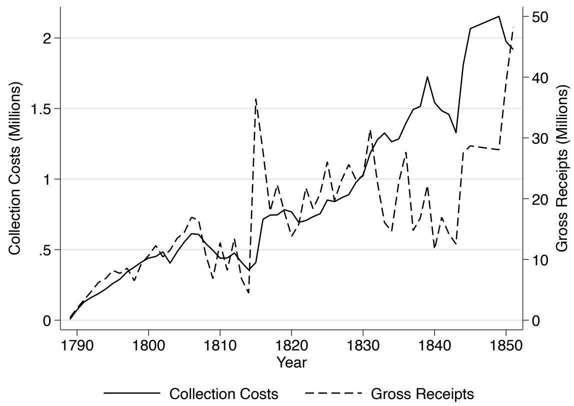
Yang, Dean, "Can Enforcement Backfire? Crime Displacement in the Context of Customs Reform in the Philippines," *The Review of Economics and Statistics*, 2008, *90* (1), 1–14.

Yitzhaki, Shlomo, "On the Excess Burden of Tax Evasion," *Public Finance Quarterly*, 1987, *15* (2), 123–137.

Young, Edward, *Special Report on the Customs-Tariff Legislation of the United States, with Appendixes*, Washington: Government Printing Office, 1877.

Figure 1: The Expansion of the Customs House

Panel A: Customs Over Time



Panel B: Customs Workers Over Time

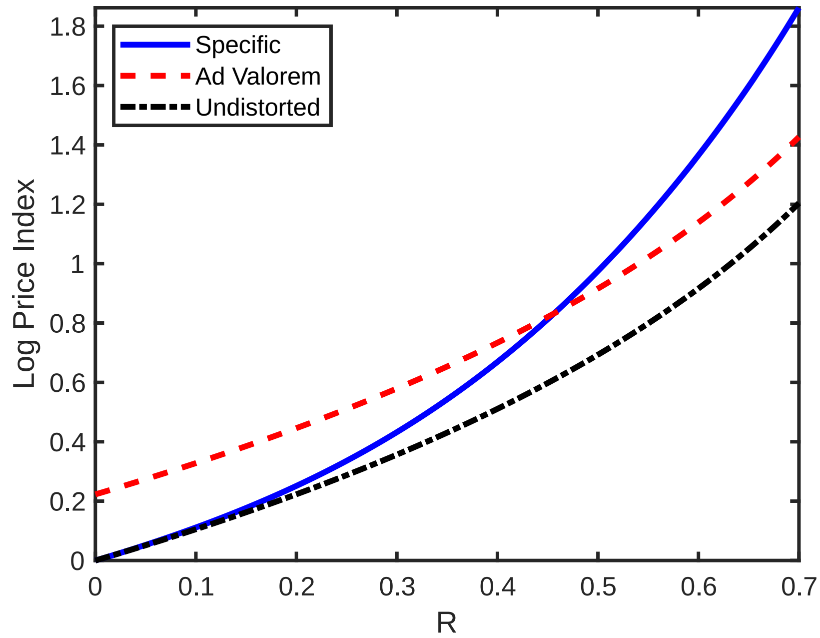
Panel C: Share of Employees with Many Jobs

Panel D: Customs Occupations Over Time

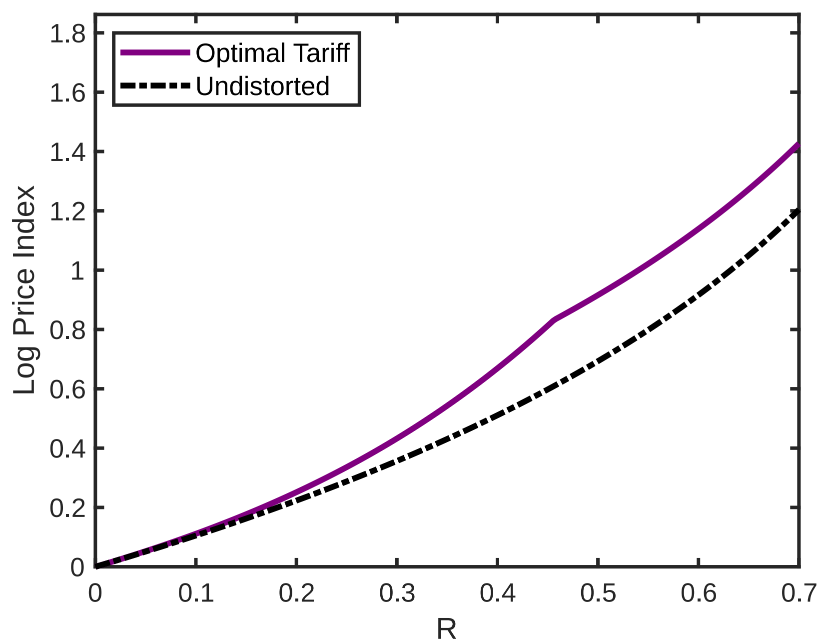
Notes: This figure shows the expansion of Customs over time. Panel A shows gross Customs revenue and collection costs (in nominal dollars). Panel B shows the number of workers at the Customs houses. Panel C shows the share of Customs workers who are listed as having multiple distinct jobs. Panel D shows the number of distinct occupations at Customs houses over time, where we only include occupations that appear in at least two different houses. Source: *Commerce and Navigation of the United States* and *Annual Reports of the Secretary of the Treasury on the State of the Finances.*

Figure 2: Revenue and Price Index Distortions for Different Tariffs

Panel A: Specific, Ad Valorem, and Undistorted Ad Valorem



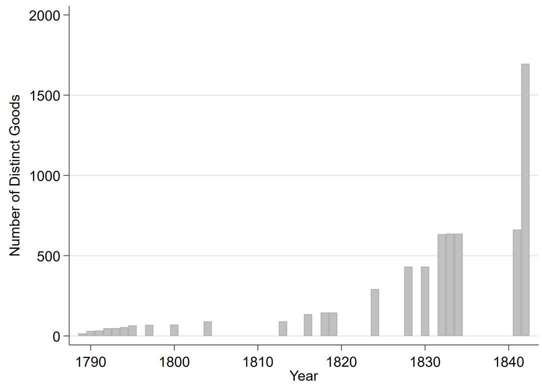
Panel B: Optimal Tariff and Ad Valorem



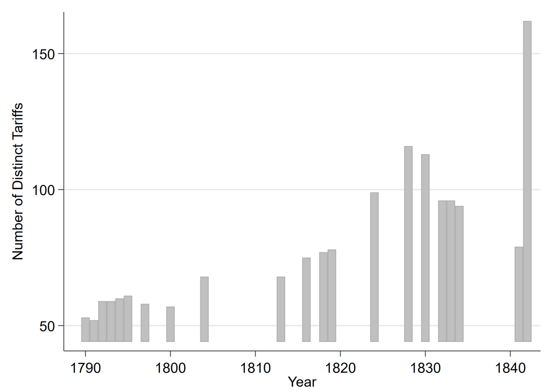
Notes: This figure shows how the price index changes with the share of expenditure collected as tariff income (net of any verification costs). The first panel plots how the price index responds with specific tariffs (in solid blue), ad valorem tariffs with enforcement costs (dashed red), and perfectly and freely enforced ad valorem tariffs (black dot dashed). The second panel compares the undistorted ad valorem to the lower envelope of specific and ad valorem tariffs with enforcement costs (in solid purple).

Figure 3: Increasing Tariff Disaggregation Over Time

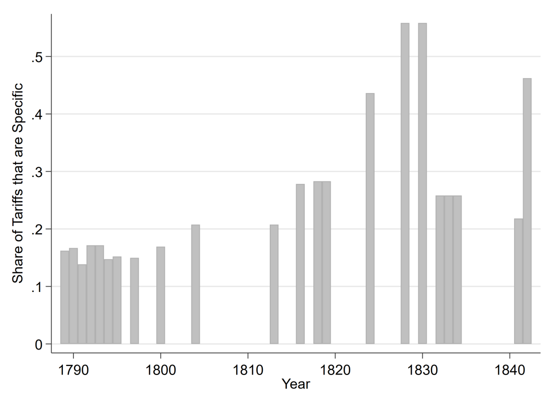
Panel A: Number of Goods Defined by Each Year



Panel B: Number of Distinct Tariffs Over Time



Panel C: Share of Goods with Specific Tariffs Over Time



Notes: This figure plots the evolution of the complexity of the American tariff code over time. Panel A plots the number of goods defined for each year t. The set of goods expands over time. A good is defined as a set of varieties such that: (i) in every year from 1789 through t, all varieties in the set face the same tariff; and (ii) the set is physically distinct from any other set that follows the same sequence of tariffs over the time period. Panel B plots the number of distinct tariff rates that are in effect for each year. Panel C plots the share of all goods that have a specific tariff. Source: *Tariff Acts of the United States*.

Table 1: Tariff Regime and Reported Unit Values

	(1)	(2)	(3)	(4)
Post × Specific in 1842	-0.56	-0.56	-0.53	-0.53
	(0.18)	(0.18)	(0.15)	(0.15)
Include Wool		✓		✓
Full Years Only			✓	✓
Number of Goods	17	18	17	18
Number of Observations	133	141	84	89

Notes: This table shows the effect of switching from specific to ad valorem tariffs on reported unit values. The outcome for all regressions is reported unit values. We use the initially-specific goods listed in the 1849 Annual Report of the Secretary of the Treasury for 1843–1849, and compare them to the goods without specific tariffs for which unit values were reported (two types of coffee, two types of tea, and wool). Because under the 1842 act different types of raw wools had different tariffs, they are a bit unusual. To show this shift is not driving our results, the odd columns exclude them from the control group. Columns 3 and 4 show results only for the import statistics with 12 months of reporting (dropping 1843 and two reports for 1847, one for each tariff act). All regressions include good and year fixed effects, as well as Post-1846 interacted with both the ad valorem equivalents implied by the 1842 and 1846 tariff acts, allowing these effects to differ for initially-specific goods. Standard errors are clustered by good. Source: *Commerce and Navigation of the United States* and *Annual Reports of the Secretary of the Treasury on the State of the Finances.*

Table 2: Homogeneity Predicts Specific Tariffs

Panel A. Singletons

	(1)	(2)	(3)	(4)
Singleton Label	0.053	0.080	0.088	0.14
	(0.0096)	(0.0097)	(0.011)	(0.028)
Mean Outcome	0.11	0.11	0.11	0.49
Number of Goods	1692	1692	1692	1588
Number of Observations	34639	34639	34639	1588
Specification	Year FE	Year × SITC FE	Year & Good FE	Only 1842

Panel B. Commodities

	(1)	(2)	(3)
Homogeneous	0.11	0.16	0.27
	(0.0092)	(0.041)	(0.032)
Mean Outcome	0.11	0.11	0.49
Number of Goods	1692	1692	1588
Number of Observations	34639	34639	1588
Specification	Year FE	Year × SITC FE	Only 1842

Notes: This table shows how price dispersion affects if a good has a specific tariff. Each good is included only in the years in which it is covered by a tariff and it has either a specific or ad valorem tariff. Each observation is a good-by-year. The outcome for all regressions is if the good has a specific tariff. In Panel A, the independent variable is if the label for the good is a singleton (only covers one good). In Panel B, the independent variable is if the good was listed as a commodity by *The Executive Documents* of the House of Representatives. All regressions control for year. Column 2 additionally controls for year interacted with SITC classifications (three-digit in Panel A, and two-digit in Panel B, as three-digit absorbs the variation in commodity classifications). Column 3 in Panel A controls for good fixed effects (which are collinear with the measure of being a commodity). Column 4 in Panel A, and Column 3 in Panel B, only include 1842. All regressions cluster by good. Source: *Tariff Acts of the United States*.

Table 3: Low AVEs Predict Specific Tariffs

Panel A. Baseline Specification

	(1)	(2)	(3)
AVE < 5	0.26	0.21	0.23
	(0.074)	(0.11)	(0.076)
5 < AVE < 15	0.18	0.10	0.15
	(0.064)	(0.086)	(0.059)
15 < AVE < 25	0.083	0.010	0.052
	(0.046)	(0.059)	(0.039)
Mean Outcome	0.679	0.431	0.573
Number of Imports	759	521	1209
Number of Observations	969	727	1696
Tariff Regime	1842	1846	Both

Panel B. Within 1832 Labels

	(1)	(2)	(3)
AVE < 5	0.15	0.17	0.16
	(0.088)	(0.14)	(0.056)
5 < AVE < 15	0.15	0.086	0.099
	(0.064)	(0.081)	(0.050)
15 < AVE < 25	0.064	-0.13	-0.000058
	(0.044)	(0.053)	(0.033)
Mean Outcome	0.745	0.620	0.694
Number of Imports	716	470	1170
Number of Observations	967	708	1675
Tariff Regime	1842	1846	Both

Notes: This table shows the share of imports with specific tariffs, for each bin of ad valorem equivalent. Across all columns the outcome is whether the import had a specific tariff in 1842, and the key independent variables are dummies for the ad valorem equivalent level. Panel A presents the baseline specification, while Panel B adds fixed effects for the 1832 classification label of the import (and only includes imports that share an 1832 classification label with another import). Regressions are clustered by import and include year fixed effects. The sample is restricted to imports whose ad valorem equivalent was below the maximum ad valorem rate in 1842 (50 percent). Each observation is a row in the import tables of *Commerce and Navigation of the United States*. Additional source: *Tariff Acts of the United States*.

Table 4: Lower Unit Values Predict Higher Ad Valorem Tariffs

	(1)	(2)	(3)
Unit Value (Standardized)	-1.34	-0.62	-0.89
	(0.32)	(0.28)	(0.28)
Mean Outcome	28.312	27.972	28.175
Number of Imports	854	405	1163
Number of Observations	1069	605	1674
Tariff Regime	1842	1846	Both

Notes: This table shows the relationship between unit values and ad valorem duties. The outcome in all regressions is the ad valorem rate for the import after 1846. The sample is the imports with specific tariffs in the 1842 tariff act, as we can calculate unit values for those goods. All regressions additionally control for each import's 1842 ad valorem equivalent. Column 1 includes only observations from the 1842 tariff act, Column 2 only from the 1846 tariff act, and Column 3 includes all observations. Results are clustered by good. Each observation is a row in the import tables of *Commerce and Navigation of the United States.* Additional source: *Tariff Acts of the United States.*

Table 5: 1980s American Imports: Drivers of Specificity

Panel A. Unit Values

	(1)	(2)	(3)	(4)
ln(Unit Values)	-0.028	-0.011	-0.031	-0.014
	(0.0020)	(0.0025)	(0.0019)	(0.0039)
Mean Outcome	0.20	0.20	0.20	0.20
Number of Goods	4248	4248	4248	4248
Number of Observations	18459	18459	4288	4288
4-Digit FEs		✓		✓
Sample	Annual	Annual	Pooled	Pooled

Panel B. Price Dispersion Across Exporters

	(1)	(2)	(3)	(4)
Price Dispersion	-0.12	-0.016	-0.14	-0.038
	(0.0081)	(0.0047)	(0.010)	(0.011)
Mean Outcome	0.19	0.19	0.19	0.19
Number of Goods	4109	4109	4109	4109
Number of Observations	17442	17442	4219	4219
4-Digit FEs		✓		✓
Sample	Annual	Annual	Pooled	Pooled

Panel C. Ad Valorem Equivalents

	(1)	(2)	(3)	(4)
ln(Ad Valorem Equivalent)	-0.15	-0.10	-0.15	-0.11
	(0.0060)	(0.0076)	(0.0067)	(0.012)
Mean Outcome	0.17	0.17	0.17	0.17
Number of Goods	4872	4872	4872	4872
Number of Observations	21419	21419	4910	4910
4-Digit FEs		✓		✓
Sample	Annual	Annual	Pooled	Pooled

Notes: The outcome in this table is if the import has a specific tariff. Each observation is a TSUS 5-digit import. Columns 1 and 2 include each good by year separately, and Columns 3 and 4 average across years within goods. Columns 2 and 4 additionally include fixed effects for each 4-digit TSUS code. All regressions are clustered by 5-digit TSUS code. Panel A considers the unit values of the goods (the value divided by the quantity). Panel B considers the dispersion of unit values across shipments, as described in the text. Panel C considers the ad valorem equivalents (the tariff revenue divided by the value). Source: US Census Annual Import Data Bank (IDB).

Table 6: Modern Low AVEs and Homogeneous Goods Predict Specificity

Panel A. Homogeneity

	Specific Tariff			
	(1)	(2)	(3)	(4)
Homogeneous	0.0079		0.0083	0.0022
	(0.001)		(0.001)	(0.001)
Non-Grouped		0.011	0.011	0.0044
		(0.001)	(0.001)	(0.001)
Homogeneous ×				0.017
Non-Grouped				(0.003)
Mean Outcome	0.017	0.017	0.017	0.017
Number of Observations	801016	801016	801016	801016

Panel B. Ad Valorem Equivalents

	Specific Tariff		
	(1)	(2)	(3)
AVE < 5	0.045	0.044	0.033
	(0.003)	(0.003)	(0.004)
Homogeneous		0.0071	0.0052
		(0.001)	(0.001)
AVE < 5 ×			0.023
Homogeneous			(0.005)
Mean Outcome	0.017	0.017	0.017
Number of Observations	801016	801016	801016

Notes: This table shows modern predictors of specific tariffs. Each panel includes coefficients from three predictors, one of which is always if the good is homogeneous. Panel A considers if the good is non-grouped, while Panel B considers the tariff burden, calculated as an indicator for the ad valorem equivalent being below 5%. Each panel also shows the interaction of the two variables. Each observation is a tariff line, and the outcome is if the tariff is specific. Homogeneous goods are those that are sold on an organized exchange or have a reference price, following the conservative classification of Rauch (1999). Non-grouped goods are those that do not share the same tariff with all of the other tariff lines within the same HS 4-digit code. All regressions include importer fixed effects, and standard errors are clustered by HS 6-digit code. We use the most recently available data for each country. Source: WTO Integrated Database and UN Trade Analysis Information System.

Table 7: Modern Customs House Capacity Predicts More Complex Tariffs and More Specific Tariffs

Panel A. Tariff Complexity

	ln(Number of Distinct Tariff Rates)		Share of HS 6-digit Codes with Variation in Tariffs	
	(1)	(2)	(3)	(4)
Overall Export Misreporting Index	-0.41 (0.094)		-0.012 (0.0055)	
Overall Trade Misreporting Index		-0.46 (0.10)		-0.014 (0.0060)
Mean Outcome	3.04	3.04	0.073	0.073
Number of Observations	125	125	125	125

Panel B. Specific Tariffs

	Share of Tariff Lines with Specific Tariffs		Average of HS 6-digit Share with Specific Tariffs	
	(1)	(2)	(3)	(4)
Overall Export Misreporting Index	-0.018 (0.0079)		-0.016 (0.0079)	
Overall Trade Misreporting Index		-0.019 (0.0091)		-0.017 (0.0091)
Mean Outcome	0.035	0.035	0.029	0.029
Number of Observations	125	125	125	125

Notes: This table shows the relationship between modern Customs house capacity and the tariff code. Customs house capacity is measured with a (normalized) index of each country's agreement with its trading partners, following Farhad et al. (2024), where the first row considers only exports and the second row considers all trade flows. Panel A considers measures of the complexity of the tariff code: the (ln) number of distinct tariff rates, and the share of HS 6-digit codes with variation in tariffs. Panel B considers the specificity of the tariff code. The outcome in Columns 1 and 2 is the share of tariff lines that are specific. Columns 3 and 4 take the unweighted average tariff lines within each HS-6 that are specific, and then take the average across all HS-6 codes. Each observation is a country in its most recently available data, and we consider MFN tariff rates. Robust standard errors are reported in parenthesis. Source: WTO Integrated Database.

Online Appendix for
Tariffs and State Capacity: A Specific Story

Erik Madsen
NYU

Martin Rotemberg
NYU

Sharon Traiberman
NYU

Shizhuo Wang
NYU

July 2026

A Data

In this section, we describe in detail how we constructed the datasets used in the paper.

A.1 Historical Tariff Acts

Our primary sources of data are the texts of the 1789–1846 tariff acts, which we harmonize over time. Creating and applying concordances is an inescapable aspect of modern empirical research for trade economists (Pierce and Schott, 2012; Fort and Klimek, 2018).[31]

Because there was no formal classification system for goods during our sample period, these tariff acts present additional challenges for data construction: In addition to concording tariffs over time, we need to define our unit of observation. In particular, we define a *good* as the set of possible shipments that (a) in any particular year have exactly one tariff; and (b) at some point face a different tariff from any potentially observationally similar shipments.

To define the set of goods, we first parse the tariff acts into *labels*, where a label is the string that is attached to a tariff line. We match labels across years. Matching labels is complicated by the many-to-many relationship between labels and goods: Each label may cover many goods, and while each good has only one label in a given year, its label may change over time (though not every good is mentioned in every tariff act). To obtain a consistent set of goods with unique tariff lines, we define goods by the finest partition obtained from *intersecting* labels across acts.

We think that this notion of a good is the correct unit of analysis. First, it reflects lines in modern tariff code: It is the unit to which distinct tariffs apply. Second, the tariff code kept track of imports once they had been disaggregated, even if their tariffs were later combined. For instance, in 1842 there were three different tariffs for different kinds of olive oil. The

[31] Matching physical goods to their tariff classification is also difficult in practice. For instance, the Customs Broker License Examination has a pass rate of around 10-25 percent (Customs and Protection, 2024). Bilateral tariff rates are also often measured with error (Teti, 2024).

1846 tariff act explicitly includes "olive oil in casks, other than salad oil," "olive salad oil," and "all other olive oil, not otherwise provided for," even though they all receive the same tariff rate.

Importantly, while a good is the smallest object that appears in the tariff code, each good nevertheless comprises many latent *varieties*: distinct product types that may differ in value (and other characteristics). Some differences across varieties within a good may be innocuous, such as color (for example, blue versus red flannels). The key force for tariff policy is within-good price dispersion: Varieties of the same good may have different prices. Because policymakers often classify imports using physical characteristics that are only imperfectly correlated with price, tariff categories generally group together varieties with different values (Grant, 2025). In 1846, the three olive oil goods presumably had different prices, though they shared a label.[32]

Appendix Table A.1 illustrates how we match labels to create goods, using the example of cotton and woolen flannels. Ultimately, there are twelve different flannel goods, distinguished by their material, price, and coloring. Panel A shows the dynamics for woolen flannels. In 1789 and 1790, they were not disaggregated at all and were covered by the label "all other goods." In 1791 and 1794, there were no labels that would cover woolen flannels, so their tariffs remain unchanged. In 1816, there was a tariff for all "wool manufactures." In 1824, there was a tariff for woolen manufactures that is somewhat complicated, but flannels were exempt from the complications so they all paid the same rate. The 1828 act splits woolen manufactures into five groups by their price per square yard. Had we stopped our analysis in 1828, we would not consider there to be any woolen flannel goods in particular, as any woolen flannel import would have the same history of tariffs as many other woolen manufactures. But that choice would obscure further developments. The 1832 (and 1842) acts give a tariff to all woolen flannels. This change means that, starting from 1832, the woolen flannels all had a different path of tariffs from the other woolen manufactures. On the whole, in our dataset we have five different woolen flannel goods (one for each 1828 price group).

Panel B shows the analogous process for cotton flannels. The linking for the cotton flannels was complicated by the minimum tariffs, which imposed a minimum valuation for computing ad valorem duties. Each change in the minimum cutoff, as well as the 1842 act's distinct tariffs for colored and uncolored cotton manufactures, effectively created new

[32] Another example is "refined sugars," which had a common tariff throughout our sample period. While we therefore consider them to be one good, refined sugars with different "Dutch Standard" color classifications had different prices. Indeed, in 1862 (after our sample ends), the tariff code introduced four distinct tariffs for different types of refined sugars. Relatedly, the United States historically assigned tariffs to spirits by proof, understanding that higher-proof spirits commanded higher prices but that the correlation was not perfect.

partitions. For instance, the cutoff was 25 cents a square yard in 1816 and 30 cents a square yard in 1824. This evolution creates an implicit group of flannels that are between 25 and 30 cents a square yard, which is never explicitly enumerated, but which has a different tariff history than both the cheaper and more expensive varieties.

Note that linking the cotton and flannel labels requires external knowledge. There is never a tariff for "cotton flannels": We link the "flannels" and "cotton manufactures" labels because we know what flannels are. There are seven cotton flannel goods.

We repeat a similar process for all labels in all of the tariff acts. For each label, we find its matches in the other tariff acts. While most labels are disaggregated over time, the flannels example shows that the process can be non-monotonic. In our analysis, when we consider if a good is associated with a singleton label or not, we use the number of goods that currently are sharing the same label.

Process of Linking : We made all matches by hand. We describe the process below.

Broadly, we split the labels into three types. First, there are *everything* labels, which apply to all imports that are not explicitly enumerated in a tariff act. Second, there are *general* labels, which apply to a broad set of imports (such as all "cotton manufactures"). Finally, there are *narrow* labels, which apply to a small number of goods (such as "flannels"). The distinction between general and narrow labels is not important for any of our final analyses. Many narrow labels do cover more than one good; for instance we classify the 1832 "flannels" tariff as narrow even though there are many flannel goods. The distinction is relevant only because it affects how we process and harmonize the tariff lines.

We started with a list of goods from the appendix to the 1832 tariff act. We then merged all of the narrow labels in each year to the goods in the list. Some narrow labels had no match, since the 1832 list was incomplete; in those cases we added new goods to the list. For our regression analysis on the tariff acts we completed the matching process through 1842, though we also linked the 1846 tariffs. This process required many passes: Every time we added a new entry because of a novel label, we needed to check it against the labels in all previous years.

The language in the acts is inconsistent, as different labels can mean the same thing, and they are sometimes anachronistic. We were not surprised by tariffs for "quicksilver" or "brimstone," but we were all a little surprised to find a tariff in 1846 for "dragon's blood." We made many passes over the list to ensure synonyms were linked consistently.

Once we had the full preliminary set of goods through 1842, we turned to the general tariffs, grouping them into the following categories: brass; bronze; carpets; copper; cotton; dye; flax; fruits; all glass; framed glass; silvered glass; gold; grass, straw, etc.; hemp; iron; iron and brass wires; lead; leathers; linens; marble; millinery; nuts; oils; paper; pastes, tinctures, etc.; pearl; pewter; plants; raisins; raw hides, furs, and skins; salt; silk; silver; steel; tin; wood; wool; and worsted. The general tariffs sometimes had exemptions (such as the woolen-manufactures tariff which exempted flannels), which we also tracked.

One complication with the general tariffs is that they are sometimes associated with particular prices (such as the minimum tariffs, for instance the cotton manufactures less than 25 cents a square yard). For instance, Figure A.1 reports five different tariffs for wool manufactures in the 1828 act, depending on their price. We kept track of these prices separately. Ultimately, our list of general tariffs was defined by category, price, and year. Unfortunately, we could track only nominal prices, not real prices.

For each good in our preliminary list, we noted whether it fell into any of these categories. Many of the goods matched to multiple categories, which required further processing.

When goods matched to multiple categories, sometimes there was a clear hierarchy, with one category naturally taking precedence. In particular, when general labels intersected, the more targeted (use-based) label was chosen over any broader material-based ones. As an example, consider cotton bonnets. In a year with general tariffs for both "millinery" and "cotton manufactures," only the former would apply: There is no cotton bonnet that is not millinery. Some years had a cotton manufactures tariff but not a millinery tariff, in which case the cotton general tariff did apply to the relevant goods. Thankfully, no good had more than one potential high-priority label in a single act. For each good, we created a list of high-priority labels and low-priority labels.

In other cases, when a good matched to multiple labels, we treated them as equally important. For instance, flannels could be either cotton or wool. As a result, we split the "flannel" entry in our preliminary set of goods into two: cotton flannels and woolen flannels. We then further split the flannel categories by price, as described in Appendix Table A.1.[33]

We then merged the lists. First, we performed a many-to-many merge between the preliminary list of goods and the list of general tariffs (many-to-many because some preliminary goods matched to multiple general labels, and all of the general labels matched to multiple preliminary goods). This created a long list of goods.

For each tariff act, we then undertook the following steps for all of the goods and labels:

1. If a good was associated with a narrow label, then that was its assigned label and tariff for that act.
2. If a good was not yet associated with a tariff, we checked whether it had a high-priority general match (making sure that it was not exempt). If so, we assigned the good that

[33] Carpets provide another example of unranked categories. Some carpets, such as "Wilton carpets," are always made of wool. Others, such as "Aubusson carpets," though typically wool, could also be made of silk, for example `https://archive.md/WhuXk`. As a result, we have one good for Wilton carpets in our final dataset, and two for Aubusson carpets.

label and tariff.

3. If a good was still not yet associated with a tariff, we then checked if it had a low-priority general match (again making sure that it was not exempt). If so, we assigned the good that label and tariff.
4. If the good was still not yet associated with a tariff, and the act had an everything tariff, then we assigned the good that label and tariff.[34]

When harmonizing goods to the labels, we faced two main concerns: making a false match and missing a valid one. False matches were relatively straightforward to identify. We constructed a comprehensive two-column list of every good–label pair and manually verified that each combination was sensible. Missed matches were more difficult to detect, since the space of potential matches is large. To address this issue, as a final step we merged the list of goods to SITC Revision 2 codes. For each tariff act, we identified all labels that matched any good within each 3-digit SITC heading. We then verified that we had not missed any additional potential matches between these labels and any other good within the same heading.

Deduplication: Multiple labels in the same tariff act may refer to what we ultimately consider to be one good, because the labels are not very different. For instance, ales, beers, ciders, and porters (in bottles) are often all mentioned explicitly and separately in the tariff acts, so for many tariff acts we parsed four distinct labels, one for each drink. However, they all always have the same tariff rate. Because we think they are also physically similar, we consider them to jointly be one good. This deduplication approach required some judgment calls. For instance, while we group ales and beers, we do not group "mats of straw and grass" and "milk of roses" (a cosmetic product made from rosewater and almond oil), even though the latter two always share a tariff. This is because we consider them to be physically different.[35]

For our main dataset, we merge labels that a reasonable person would consider to capture similar concepts. In robustness exercises, we consider stricter approaches, where we merge labels that are less obviously similar.

Other sorts of deduplication were more straightforward. When we merged goods to

[34] We had to perform these steps in chronological order, as some everything labels applied to all goods that currently were paying a particular rate. For instance, the 1832 tariff act has an everything tariff of 15 percent for "articles, all not herein specified either as free – or as liable to a different duty, and which, by the existing laws, pay an ad valorem duty higher than 15." So first we had to determine what each good's tariff was until 1832, and if it was not explicitly mentioned in the 1832 act and otherwise would have had a high ad valorem duty, we knew that its rate was lowered to 15 percent.

[35] What exactly distinguishes goods is a philosophical question (for instance, beers are brewed but ciders are not, so it might be defensible to consider them different goods, and indeed they do not share an HS code in modern classifications). The set of goods is likely a function of the economic environment (Grant, 2025). We think that a principled approach is to consider goods as distinguished by their nature and by their tariff.

their categories, sometimes those categories were not binding, for instance if the good had a narrow label in every tariff act. As a result, while conceptually the good could be made of different things, its construction was not relevant for tariffs, and so we dropped those essentially duplicated goods. Ultimately, for our baseline analysis we consider 1,696 goods. We undertake a strict deduplication, where we merged all plausibly similar options (which turned out not to lower the sample much, as there are still 1,616 goods under the strict approach). We also undertook an extremely strict deduplication approach, where we merged all options with the same path of tariffs, regardless of physical characteristics, which led to 1,169 goods.

A.1.1 Classification of Commodities

In the main analysis, our notion of if a good is a commodity comes from a table in the 1864 *The Executive Documents* of the House of Representatives, which in the Report on the Finances includes prices for a range of goods in New York.[36] From our perspective, the list is incomplete, even by 19th century standards. For instance, it includes several but not all kinds of sugar, some but not all kinds of tea, and entries for wheat, rye, oats, and corn but not barley (which was also widely traded on the Chicago Board of Trade by the 1850s). As a result, we classify a good as a commodity if it is similar to an entry explicitly enumerated in the congressional list. For example, we consider all sugars to be commodities.[37]

We also use the list of commodities in Rauch (1999). To merge the historical goods to the modern list, we used our hand-link of the historical goods to the SITC Revision 2 classification, which was typically possible at the 3 or 4 digit level. Rauch (1999) provides a measure for that classification, so no further concordances were needed. A few goods link to different SITC codes. When goods can plausibly match to multiple SITC codes, we treat them as homogeneous if any of the codes are labeled as homogeneous.

In the late 18th and early 19th centuries, actual organized exchanges and reference prices were rare for the goods classified as homogeneous by Rauch (1999), since these markets were not fully developed (Taylor, 1917; Banner, 1998; Dunlavy, 2024). While we therefore use contemporaneously documented commodities as our main measure, we think the Rauch (1999) measure is reasonable as well. What matters from the perspective of the model is not that there is a *known* price, but that there is little price variation. We think a reasonable logic is that these goods were plausibly homogeneous during our sample period, even if their

[36] Congress did not use the word "commodity" to describe the goods contained in the list, but the *Merchants & Bankers' Almanac* included a similar list and did call the entries commodities. We use the goods reported for 1842, though the list does not appear to change over time.

[37] The categories are breadstuffs, candles, coal, coffee, copper, cotton, fish, fruit, furs, glass, gunpowder, hides, hops, indigo, iron, lead, leather, liquors, molasses, nails, naval stores, oils, paints, provisions, rice, salt, seeds, sheetings, soap, spices, spirits, sugars, tallow, teas, tobacco, whalebone, wine, wool, and zinc.

prices were not as formally documented and they were not traded anonymously on organized markets.

In addition to classifying if the goods were commodities, we also manually kept track of if a good was manufactured, which we use for robustness checks.

A.1.2 Alternative Panels of Tariffs

Over the 19th century, the U.S. government created several tables covering the tariffs during our sample period (U.S. Congress, 1832; Walker, 1845; United States Treasury Department, 1871, 1888). These government documents highlight the difficulties of keeping track of goods, and we cannot use them for our primary analysis. First, the documents contain many errors: The text of an act imposes one import duty but the later-compiled list records something else. Second, these retrospective documents are often incomplete: They do not report the duty for every enumerated good. The incompleteness is probably due to a mix of error and the fact that it is difficult to list all the imports, since there are many and some are relatively complicated. We found that they were particularly likely to describe the minimum tariffs incompletely, and they rarely did a good job with the general tariffs such as "manufactures of wool."

Consistent with a central friction in our model, even the government found it difficult to perfectly track its complex tariff code. Indeed, the authors of the 1871 report discuss the "great labor in its preparation," noting that it was "almost" accurate.

Occasionally, Customs officials would compile the set of current tariffs, as this was not typically a document officially provided by the government (Jones, 1835, 1854). These lists are much more user-friendly than the tariff acts (for instance, they list goods in alphabetical order), but unfortunately they do not provide the complete linking of tariffs across years.

A.2 Statistics on Imports

Beginning in the early 1820s, somewhat disaggregated import data became available in the *Commerce and Navigation of the United States*, published by the U.S. Treasury Department, initially titled the *Annual Report on Commerce and Navigation* (Acosta et al., 2026). We digitized the editions from 1843 through 1849.[38] Most of the tables cover full years (ending on June 30), though 1843 covers nine months, and 1847 is split in two, one for imports under the 1842 act (until December 1846) and the other for imports under the 1846 act (until June 30, 1847).

We merged the import statistics to the tariff acts by hand. The import statistics were typically more aggregated than the acts, so most observations linked to multiple goods.

[38] The coverage of the *Commerce and Navigation of the United States* improved over time (though even in the 1840s, it only covered imports for around half of the goods).

In addition to using this resource for checking our work,[39] we also used it to calculate ad valorem equivalents for the specific tariffs. First, we calculated the implied tariff revenue by multiplying the total imported quantity by the import's specific tariff (calculated as the unweighted average of the goods that match to each import). We then divided the implied tariff revenue by the total value of imports.[40] Total import values are always reported. Quantities, and therefore unit values, are consistently available for goods with specific tariffs and for some goods with ad valorem tariffs that historically had specific tariffs. Important for our purposes, around half of the goods with specific tariffs in 1842 continue to have import quantities reported after they switched to ad valorem tariffs in 1846. We collected additional data on tariff collections, Customs costs, and other government revenues from the *Annual Reports of the Treasury Secretary of the United States*.

We tried to create a panel from the import data, but unfortunately the imports were broadly inconsistently defined and incompletely reported. While in each year we can match imports to goods, matching imports to each other did not work well: Creating time-consistent observations required a lot of aggregation, due to inconsistent reporting. The resulting observations did not seem very reliable. For instance, there was a lot of price variation within most import groupings within a given tariff act. Due to these data-quality issues, we mostly focus on cross-sectional analysis. Indeed, for the analysis on unit values in the 1849 *Annual Report*, Treasury officials went directly to Customs houses to compile their data, instead of relying only on the *Commerce and Navigation of the United States*.

There were five goods that before 1846 did not have purely specific tariffs but nevertheless had quantities consistently reported in the import statistics and for which we could create a panel. Those imports are teas and coffees imported from their place of production (free before and after the reform), teas and coffees not imported from their place of production (20 percent before and after the reform), and unmanufactured wool. The tariff for unmanufactured wool was complicated. Before 1846, the tariff depended on the price: If the wool was seven cents or under per pound, the duty was 5 percent; otherwise the duty was three cents per pound plus 30 percent. The average ad valorem equivalent tariff before 1846 was 38 percent. After 1846, all unmanufactured wool had an ad valorem tariff of 30 percent.

[39] While the reports do not include the tariff rate, they do say if a good had a specific, ad valorem, or free tariff.

[40] Calculated ad valorem equivalents for specific tariffs can be difficult to interpret, as the ex ante rate (as a function of all potential imports) and the ex post rate (as a function of the goods that are actually imported) are typically different (Bouët et al., 2008; Greenland et al., 2026a). In our setting, the theoretically correct object to measure is the ex post measure, which is what can be calculated in the data. As in modern data, we find some ad valorem equivalents that are extremely high (Teti, 2020). For all of our analysis, we drop goods with specific tariffs whose ad valorem equivalent is higher than the highest ad valorem tariff. Why there are such high specific tariffs is an interesting question. We leave to future research the job of distinguishing between mismeasurement, political-economy forces, and other more exotic possibilities.

A.3 Statistics on Customs House Operations

Beginning in 1816, Congress mandated the Secretary of State to compile a list "containing correct lists of all the officers and agents, civil, military, and naval, in the service of the United States, made up to the last day of September, of each year in which a new Congress is to assemble, [to] be compiled and printed." These lists became *The Official Register of the United States*. We digitized records for Customs, the Light Houses, the Revenue Cutters, and the Marine Hospital for 1816, 1821, 1829, 1835, 1843, and 1849 editions.[41] See also Mastrorocco and Teso (2026), who digitized and linked the *Registers* through 1905.

The organization of each book changes slightly over time. For example, in 1816, officials are separated by their title (e.g., all Collectors are listed, followed by surveyors, and so forth). Later employees are organized by Customs districts, and then by districts and department (e.g., collections, appraisal, etc.). Despite these small changes, the information is consistent: name, occupation(s), port and district, place of birth, and compensation. Figure A.6 provides an example of how the data are organized.

The only change we make from the raw data is cleaning occupational titles. To economize on space, the publishers of *The Register* made use of shorthand that was neither consistent within years nor across years. We decoded this shorthand manually. For example we replaced "W." "We'r", "Wr." with Weigher.[42]

In the raw data, the number of occupational titles rises dramatically over time. Much of this is driven by the New York Customs House giving particularly narrow titles. For our main analysis, we include only occupation titles mentioned by at least two houses. We present two sets of occupation titles for robustness. In the first, we take a more lenient classification and keep all titles. In the second, we focus only on core tasks and remove all clerks, bookkeepers, etc. Here the expansion is driven by the introduction of appraisers, examiners, a separation of proving and marking, as well as the growth of hierarchy in these roles (e.g., "assistant appraisers").

A.4 Modern Cross-Country Tariff Data

We use the WTO's IDB tariff data to construct four country-level measures, two each for complexity and specificity. First, we measure the (ln) number of distinct tariff rates for each

[41] In early editions, all of these operations were listed under Customs. The original Customs houses were also involved in operating the Revenue Cutters (the predecessor to the Coast Guard, which acted to prevent smuggling), the Light Houses, and the Marine Hospital. These tasks were separated from the Customs houses over time, and are not closely related to duty collection, so we do not focus on these departments, though we still digitized their worker registers to ensure comparability over time.

[42] The other change is that the string 'do' was used for ditto. With multiple job titles, it is difficult to interpret what is shared across workers. Typically "do" was written multiple times, presumably when multiple job titles were shared. As a result, we followed the rule that we copied all job titles that were in alignment with the "do" (or, if only one "do" was used and it was unaligned, we assumed that also meant to copy the whole title).

country. A tariff rate is considered distinct if it differs from others, regardless of whether it is ad valorem or specific. For example, 4 percent and 10 percent (both ad valorem) and 1.4 cents/kg and 4.4 cents/kg (both specific) represent four distinct tariff rates. Second, for each country we compute the share of HS six-digit codes with variation in tariff rates. If all tariff lines within an HS six-digit category (i.e., tariff lines sharing the same first six digits) have the same tariff rate, then the category has no variation. If at least two distinct tariff rates exist within the category, we classify it as having variation. We use these two measures to capture tariff complexity.

Our third measure is a direct average across tariff lines: the share of tariff lines with a specific tariff in each country. Our fourth measure is the average across HS6 codes: For each HS six-digit category, we compute the share of lines with a specific tariff, and then take the average of these shares across categories within each country. The two measures are related: The latter downweights categories with many tariff lines and gives equal weight to each HS6 group.

For our analysis at the tariff-line level, we construct four indicators. These include whether: (a) a tariff line has a specific tariff; (b) its ad valorem equivalent (or the duty rate, if ad valorem) falls below a certain threshold—5 percent in the main analysis; (c) it is classified as homogeneous according to Rauch (1999); and (d) it is "non-grouped." The non-grouped indicator equals one if not all products within the HS four-digit category to which the tariff line belongs in that country share the same duty rate.

A.5 American Imports, 1981–1985

The data in the American IDB include (estimates of) total duties collected, the rate provision, as well as quantities and values. We are grateful to Andrew Greenland for pointing us to this source, and we followed the approach in Greenland et al. (2026a) as much as possible.[43] We include only imports where the rate provision has a second digit of 1, which implies that the tariff is MFN normal trade relations. We restrict attention to specific tariffs (first digit of the rate provision code is 2) and ad valorem tariffs (first digit of 3).

Each observation in the resulting data is a year × 7-digit TSUS code × exporter triplet. One complication for calculating the unit values is that many products are reported with different types of units across different exporters. For each row, we calculate unit values by dividing the value by the quantity, assuming that the unit reported in the data corresponds to the size of the package that would be inspected for each shipment in the model (therefore allowing the unit size to vary across exporters within a good). We then calculate the total value weighted average unit value, either within each year or for the entire dataset. We

[43] We start in 1981 because Greenland et al. (2026a) note that the data quality in 1980 was particularly low, and the relevant information stopped being recorded in 1986.

calculate ad valorem equivalents by dividing total (estimated) tariff revenue by the total value.

B Robustness Exercises

In this section, we show that our results are broadly robust to alternative empirical approaches, including different sample restrictions and alternative measures of our main variables.

B.1 Historical Context

Many occupations appear in only one Customs house. In our main analysis, we include all occupation titles that appear in at least two houses. Figure A.7 shows a similar pattern of increasing occupations for alternative measures of Customs house sophistication, either including all occupations or only those involving core tasks. We discuss different classifications of occupational titles in Appendix A.3.

A key mechanism in the model is that complexity allows for the state to create relatively narrowly defined labels, which in turn have specific tariffs. We find this pattern empirically, as the tariff labels were disproportionally narrow. Figure A.8, Panel A shows that initially, even when there were very few labels, many targeted narrow sets of imports that would never be further disaggregated. Over subsequent decades, labels became increasingly specialized. By 1842, half of all labels applied exclusively to individual goods. Panel B shows a similar pattern for tariff rates.

B.2 Homogeneity and Specific Tariffs

We undertake a variety of robustness checks for our historical tests of Prediction 2.

When considering the mapping between labels and goods, we consider in our primary analysis if the label is a singleton: if a good shares its label with nothing else. Table A.7 shows that our results are similar if we instead define narrow labels as those covering no more than five (or ten) goods.

For our main analysis, we digitized a 19th century Congressional list of commodities. However, that list was not intended to be a comprehensive list of all commodities for academic use. Table A.8 shows our results are robust to using the commodity classification in Rauch (1999), which was designed for that purpose.

One concern with using a commodity classification for this type of analysis comes from Broda and Weinstein (2006), who point out that we will characterize a good as homogeneous even when it includes multiple distinct (and known) varieties with varying prices. Their example is that "tea" is categorized as a commodity by Rauch (1999), but in practice includes many varieties. A similar issue arises using our 19th century commodity classification: The tariff acts distinguish 18 types of tea, all of which are listed as commodities. However, in some years there are only two labels for tea, and in those years each tea label likely covers more price heterogeneity than a single homogeneous commodity would. We address this potential concern by considering narrowly-defined commodities. Table A.9 shows that while singleton labels and homogeneous goods are each more likely to receive specific tariffs, goods that meet both criteria are especially likely to do so.

Our baseline regressions include no additional controls or weights. However, a good's tariff burden may be an important driver of its tariff regime (as described in Corollary 1). In Panel A of Tables A.10 and A.11, we control for the 1842 ad valorem equivalent for all goods for which we have import data.

Many of our goods are narrowly defined, and so likely reflect a small fraction of imports. In Panel B of Tables A.10 and A.11, we weight the regressions by the inverse of the number of goods within each three-digit SITC classification, so as to raise the importance of relatively more distinct goods for our analysis. Our results are robust to these changes.

The data for Table 2 are unbalanced for two reasons. First, we exclude goods in those years in which they appear on the free list, as they have neither a specific nor an ad valorem tariff in those years. Panel A of Tables A.12 and A.13 show that our results are robust to including only goods that are never on the free list. Second, each observation is a good by tariff act. However, some tariff acts are relatively minor, affecting a small share of goods. Panel B of Tables A.12 and A.13 show that our results are robust to limiting the sample to major tariff acts (those that describe at least half of the goods).[44]

As described in Appendix A, our definition of a good requires some judgment to distinguish between physically distinct imports that nevertheless share a path of tariffs. Tables A.14 and A.15 show our results are robust to alternative approaches, both a strict deduplication where we group together goods that are even plausibly similar, and an extremely strict deduplication where we group together all goods with the same path of tariffs, regardless of any physical similarity.

B.3 Revenue and Specific Tariffs

One potential concern with a good-level analysis is that some goods do not matter much for aggregate revenue and may be overweighted in our analysis. Table A.16 shows that the relationship between unit values and rates is in fact slightly stronger when goods are weighted by the total value of imports. Note that the model predicts unit values are not misdeclared under specific tariffs but are shaded down under ad valorem tariffs, which if anything would push against our finding that cheaper goods have specific tariffs.

[44]Those acts are in 1789, 1790, 1792, 1794, 1816, 1824, 1832, and 1842.

B.4 Modern Cross-Country Analysis

Table 7 shows that, using our misreporting-based measure of capacity, more capable customs houses have more complex and more specific-heavy tariff schedules. We can use alternative measures as well. Figure A.9 shows that countries with higher GDP per capita also have a greater number of distinct tariff rates. Table A.17 shows that our results are robust to instead using measures of Bureaucracy Quality from the *International Country Risk Guide*, which cover broad state capacity, not necessarily that of the Customs houses specifically. See also Betz (2019), who reaches a similar conclusion.

Table 6, Panel A shows that goods with lower AVEs are more likely to have specific tariffs, using a 5% ad valorem equivalent cutoff. Table A.18 shows similar results using cutoffs of 2.5% or 10%. In current cross-country tariffs, 10% is around the median ad valorem equivalent, 5% is around the 10th percentile, and 2.5% is around the 5th percentile.

Table 6, Panel B shows that price homogeneity makes specific tariffs more attractive. Rauch (1999) provides both a conservative and liberal classification to account for ambiguities in product categorization. The conservative version minimizes the number of products classified as low dispersion, while the liberal one maximizes it. Our main analysis in Table 6 uses the conservative classification, and Table A.19 shows that our results are similar using the liberal definition of homogeneous.

A natural concern is that our modern results might simply reflect hysteresis in tariff codes rather than conscientious design (Acosta and Cox, 2024). To address this concern, Table A.20 shows that our results are similar if we use only goods in sectors that do not appear in the historical tariff code.[45]

B.5 Analysis of the 1980s United States

An alternative driver of tariffs is lobbying (Grossman and Helpman, 1994). One potential cause of differential lobbying is the concentration of imports, as potentially importers with a larger share of the market are able to lobby more. To evaluate this concern, we measure the HHI of imports by origin within each 5-digit code. Table A.21 shows that our results on the drivers of specificity are similar when we include only goods with an HHI below 0.5, roughly the median.

We can also visualize the drivers of specific tariffs in the 1980s. Figure A.10 shows a lowess curve of the specific-tariff share against unit values. The share of goods with specific tariffs monotonically falls as unit values increase. Figure A.11 visualizes the patterns in Table 5, Panels A and C, by plotting a heatmap of the specific-tariff share across unit values

[45]Specifically, we use goods in the SITC three-digit sectors that do not have a label only for that sector, at any point in our period. Sectors that we include in this analysis include for example synthetic fibers, petroleum products, and electric machines. A few sectors, including rice, never had a dedicated label but unambiguously existed in the nineteenth century; we exclude these as well.

and ad valorem equivalents. Consistent with the theory and the regression analysis, cheap goods and those with low protection levels are both more likely to have specific tariffs. Their interaction matters as well, since cheap goods with low tariff burdens are particularly likely to have specific tariffs.

C Proofs of Single-Category Results

C.1 Proof of Proposition 1

Existence and form of the equilibrium are immediate, because values do not impact payments. The expression for the price index follows directly from its definition plus the equilibrium import prices $p(v) = v + t$. As for the customs house's profit, equilibrium demand for an import variety of value v is proportional to $(v+t)^{-\sigma}$ under CES preferences. Since we have normalized import spending to 1, equilibrium import demands must be

$$q_S(v;t) = \frac{(v+t)^{-\sigma}}{\mathbb{E}[(v+t)^{1-\sigma}]}.$$

Net tariff revenue is therefore $\Pi_S(t) = t \cdot \mathbb{E}[q_S(v;t)]$, which is the expression in the proposition statement.

C.2 Proof of Lemma 1

The expression for net tariff revenue obtained in Proposition 1 can be rearranged to read

$$\Pi_S(t)^{-1} = 1 + \int \frac{v}{t} \cdot \frac{(v/t+1)^{-\sigma} f(v)}{\int (v'/t+1)^{-\sigma} f(v')\, dv'}\, dv.$$

Making the substitution $v = t \cdot \hat{v}$ in the integrals appearing in this expression yields

$$\Pi_S(t)^{-1} = 1 + \int \hat{v}\, dG(\hat{v};t),$$

where $G(\hat{v};t)$ is the distribution function defined by the density

$$g(\hat{v};t) \equiv \frac{(\hat{v}+1)^{-\sigma} f(t \cdot \hat{v})}{\int (\hat{v}'+1)^{-\sigma} f(t \cdot \hat{v}')\, d\hat{v}'}.$$

Suppose that g exhibits the strict monotone likelihood ratio property (SMLRP) in $\hat{v}$ with respect to $-t$. Then $G(\cdot;t')$ strictly first-order stochastically dominates $G(\cdot;t)$ for $t' < t$, implying that $\Pi_S(t')^{-1} > \Pi_S(t)^{-1}$ for $t' < t$. In other words, Π_S is increasing and Assumption 1 holds.

The SMLRP requires that

$$\frac{g(\hat{v}';t')}{g(\hat{v};t')} > \frac{g(\hat{v}';t)}{g(\hat{v};t)}$$

for every $\hat{v}' > \hat{v}$ and $t' < t$. Since

$$\frac{g(\hat{v}';t)}{g(\hat{v};t)} = \left(\frac{\hat{v}'+1}{\hat{v}+1}\right)^{\sigma} \cdot \frac{f(t\cdot\hat{v}')}{f(t\cdot\hat{v})},$$

it is sufficient that

$$\frac{\partial}{\partial t}\log\left(\frac{f(t\cdot\hat{v}')}{f(t\cdot\hat{v})}\right) = t^{-1}\cdot(\zeta(t\cdot\hat{v}')-\zeta(t\cdot\hat{v})) < 0$$

for all t and $\hat{v}' > \hat{v}$, where $\zeta(v) \equiv v\cdot f'(v)/f(v)$. The desired result therefore follows if ζ is decreasing everywhere.

For the log-normal distribution location parameter m and scale parameter $s > 0$,

$$\zeta(v) = \frac{m-\log v}{s^2} - 1.$$

For the generalized gamma distribution with scale parameter $\alpha > 0$ and shape parameters $\beta > 0$ and $\gamma > 0$,

$$\zeta(v) = \beta - 1 - \gamma\cdot\left(\frac{v}{\alpha}\right)^{\gamma}.$$

For the inverse gamma distribution with shape parameter $\alpha > 0$ and scale parameter $\beta > 0$,

$$\zeta(v) = -1-\alpha+\frac{\beta}{v}.$$

For the Lomax distribution with shape parameter $\alpha > 0$ and scale parameter $\lambda > 0$,

$$\zeta(v) = -(\alpha+1)\cdot\frac{v}{v+\lambda}.$$

For the log-logistic distribution with shape parameter $\gamma > 0$ and scale parameter $s > 0$,

$$\zeta(v) = -1-\gamma+\frac{2\gamma}{1+(v/s)^{\gamma}}.$$

In all of these cases, ζ is decreasing everywhere.

C.3 Proof of Proposition 2

In any equilibrium, it must be that $\rho(v) \leq v$ for all v, since any importer who overstates his value strictly reduces his expected tariff payment by deviating to a truthful declaration. We therefore ignore the possibility of overdeclaring in what follows.

In any monotone equilibrium, there must exist a threshold $v^* \in \mathbb{R}_+ \cup \{\infty\}$ such that $\rho(v) = 0$ if $v < v^*$ and $\rho(v) > 0$ if $v > v^*$. If $v^* = 0$, then all declarations in $\rho((0,\infty))$ must uniquely reveal the importer's value. Further, declarations in $\rho((0,\infty))$ close to $\rho(0+)$ must be associated with values close to 0. There must therefore exist a declaration $v' \in \rho((0,\infty))$ for which verification is not worthwhile and the probability of verification is 0. But then every importer whose value satisfies $\rho(v) > v'$ would reduce their expected tariff payment by deviating downward to declare v', a contradiction of equilibrium. So it must be that $v^* > 0$.

Whenever $v^* > 0$, the customs house infers from a declaration of 0 that the importer's true value lies on either the interval $[0, v^*]$ or $[0, v^*)$, depending on whether $\rho(v^*) > 0$. Since the distribution of values is continuous, which interval is conditioned upon does not impact the inferred value distribution. By convention, we condition on $[0, v^*)$ going forward.

Suppose first that $\tau < \bar{c}_{AV}$. Since all importers declaring 0 face the same net ad valorem tariff rate $\kappa \cdot \tau \cdot \phi(0)$, relative prices for goods with values in $[0, v^*]$ are undistorted. The customs house therefore believes that the conditional density of values is

$$\pi(v \mid 0) = \frac{v^{-\sigma} \cdot f(v)}{\int_0^{v^*} w^{-\sigma} \, dF(w)}$$

on $[0, v^*)$ and 0 elsewhere. Hence, the customs house's expected payoff from verifying a declaration of 0 is

$$\kappa \cdot \tau \cdot \int_0^{v^*} v \cdot \pi(v \mid 0) \, dv - c_{AV} = \kappa \cdot \tau \cdot \frac{\mathbb{E}[v^{1-\sigma} \mid v < v^*]}{\mathbb{E}[v^{-\sigma} \mid v < v^*]} - c_{AV}.$$

This expression must be negative, since the right-hand side is increasing in v^* and equals $\kappa \cdot \tau \cdot \bar{v} - c_{AV} < 0$ when $v^* = \infty$. It is therefore uniquely optimal for the customs house not to verify a declaration of 0 no matter the value of v^*. It must then be uniquely optimal for all importers to declare zero value, since any other declaration leads to a positive expected tariff payment. Hence, $\phi(0) = 0$ and $\rho(v) = 0$ for all v in any monotone equilibrium. Further, there exists an equilibrium with the property that $\phi(v') = 0$ for all v', which can be supported by the off-path belief that $v = v'$ whenever $v' > 0$.

For the remainder of the proof, suppose that $\tau > \bar{c}_{AV}$. We first verify that the stated strategies can be supported as an equilibrium. For off-path declarations $v' \in (0, v^*)$, let the customs house believe that $v = v' + c_{AV}/(\kappa \cdot \tau)$ with probability 1. In that case, for each declaration $v' > 0$ the customs house believes that the importer's value is $v' + c_{AV}/(\kappa \cdot \tau)$, and its payoff is $\tau \cdot v'$ whether it verifies or not. Meanwhile, for the declaration $v' = 0$ the customs house believes that the conditional density of values is $\pi(v \mid 0)$ on $[0, v^*)$ and 0 elsewhere, where $\pi(v \mid 0)$ is as defined above. Hence, the customs house's expected payoff from verifying the importer's declaration is

$$\kappa \cdot \tau \cdot \int_0^{v^*} v \cdot \pi(v \mid 0) \, dv - c_{AV} = \kappa \cdot \tau \cdot \frac{\mathbb{E}[v^{1-\sigma} \mid v < v^*]}{\mathbb{E}[v^{-\sigma} \mid v < v^*]} - c_{AV} = 0.$$

For every declaration v', the customs house is therefore indifferent over verification and willing to choose $\phi(v') = 1/\kappa$. Meanwhile, for any value v and declaration $v' \le v$, the importer's expected tariff payment is $r(v, v') = \tau \cdot v$, and so the importer is indifferent over all declarations. In particular, the stated declaration policy is optimal.

Since each importer is indifferent over all declarations, his expected tariff payment can be calculated assuming no underdeclaring, yielding a payment $\tau \cdot v$. Meanwhile, since the customs house is indifferent over verification following every declaration, its net revenue can be calculated assuming it never verifies declarations, yielding

$$\Pi_{AV}(\tau) = \int_{v^*}^{\infty} (\tau \cdot v - c_{AV}/\kappa) \cdot q(v; \tau) \, dF(v),$$

where

$$q(v; \tau) \equiv (1 + \tau)^{-1} \cdot \frac{v^{-\sigma}}{\mathbb{E}[v^{1-\sigma}]}$$

is the quantity of imports of each value. Now, the equation characterizing v^* can be rearranged to read

$$\int_{0}^{v^*} (\tau \cdot v - c_{AV}/\kappa) \cdot q(v; \tau) \, dF(v) = 0.$$

Hence

$$\Pi_{AV}(\tau) = \int_{0}^{\infty} (\tau \cdot v - c_{AV}/\kappa) \cdot Q(v; \tau) \, dF(v),$$

which simplifies to the expression in the proposition statement.

We now establish on-path uniqueness within the class of monotone equilibria. Fix a monotone equilibrium (ρ, ϕ). We first show that $\phi(\rho(v)) < 1$ for every v. Suppose by way of contradiction that $\phi(\rho(v)) = 1$ for some v. Let $v' = \rho(v)$. Then expected tariff payments satisfy $r(v, v') < r(v, v)$ unless $v' = v$. Equilibrium therefore requires that $\rho(v) = v$, and further that $\rho(v'') \neq v$ for every $v'' \neq v$. But in that case, the customs house places probability 1 on the importer's declaration being truthful following a declaration of v. Verification is then not optimal, contradicting $\phi(\rho(v)) = \phi(v) = 1$.

We next establish that $\phi(v') > 0$ for every v'. Suppose by way of contradiction that $\phi(v') = 0$ for some v'. Then expected tariff payments satisfy $r(v, v') < r(v, v'')$ for every $v \geq v'' > v'$, implying that $\rho(v) \leq v'$ for every v. Now, if $\rho(\infty) = 0$, then every importer declares a value of 0 and the effective ad valorem tariff on each unit is $\kappa \cdot \tau \cdot \phi(0)$. Relative prices are therefore undistorted, and so the expected value of an imported unit is $\mathbb{E}[v^{1-\sigma}]/\mathbb{E}[v^{-\sigma}] > c_{AV}/(\kappa \cdot \tau)$. In other words, verification is strictly optimal when $v' = 0$, contradicting $\phi(\rho(v)) < 1$ for every v. So $\rho(\infty) > 0$, meaning that $\rho(v) > 0$ for sufficiently large v. Monotonicity of the equilibrium implies that, for all such v, in equilibrium the customs house places probability 1 on the importer's value being v following a declaration of $\rho(v)$.

Since $\rho(v) \leq v'$ for all v, we have

$$\kappa \cdot \tau \cdot (v - \rho(v)) > c_{AV}$$

for sufficiently large v. In other words, verification is strictly optimal for the declarations of importers with large values, contradicting $\phi(\rho(v)) < 1$ for all v.

Since $\phi(\rho(v)) \in (0, 1)$ for every v, it must be that the customs house is indifferent over verification whenever $\rho(v) > 0$. Additionally, monotonicity of the equilibrium implies that the customs house believes that the importer's value is v with probability 1 whenever $v' = \rho(v) > 0$. In other words,

$$\kappa \cdot \tau \cdot (v - \rho(v)) = c_{AV}$$

must hold whenever $\rho(v) > 0$. Equivalently,

$$\rho(v) = v - c_{AV}/(\kappa \cdot \tau)$$

whenever $\rho(v) > 0$. Monotonicity of the equilibrium therefore implies existence of a cutoff v^* such that

$$\rho(v) = \begin{cases} v - c_{AV}/(\kappa \cdot \tau), & v \geq v^* \\ 0, & v < v^* \end{cases}$$

Since the customs house must additionally be indifferent over verification when $\rho(v) = 0$, and since the conditional density of values given a declaration of 0 is $\pi(v \mid 0)$ no matter the choice of v^* and $\phi(0)$, the cutoff v^* is uniquely pinned down by the condition in the proposition statement.

Finally, we establish that $\phi(\rho(v)) = 1/\kappa$ for each v. (For $v' \in (0, v^*)$ the choice of $\phi(v')$ is off-path and does not impact outcomes.) If $\phi(\rho(v)) > 1/\kappa$ for some $v > 0$, then expected tariff payments satisfy $r(v, \rho(v)) > \tau \cdot v$, contradicting the optimality of $v' = \rho(v)$. Hence $\phi(\rho(v)) \leq 1/\kappa$ for every $v > 0$. And since $\rho(0) = \rho(v^*/2)$, this inequality also holds at $v = 0$. Now, suppose that $\phi(\rho(v)) < 1/\kappa$ for some v. Then for every $v' > \max\{v^*, \rho(v)\}$, expected tariff payments satisfy

$$r(v', \rho(v)) = \tau \cdot \Big(\rho(v) + \phi(\rho(v)) \cdot \kappa \cdot (v' - \rho(v))\Big) = \tau \cdot \Big(v' - (1 - \phi(\rho(v)) \cdot \kappa) \cdot (v' - \rho(v))\Big)$$

and

$$r(v', \rho(v')) = \tau \cdot \Big(v' - (1 - \phi(\rho(v')) \cdot \kappa) \cdot \frac{c_{AV}}{\kappa \cdot \tau}\Big).$$

Hence

$$r(v', \rho(v')) - r(v', \rho(v)) = \tau \cdot (1 - \phi(\rho(v)) \cdot \kappa) \cdot (v' - \rho(v)) - (1 - \phi(\rho(v')) \cdot \kappa) \cdot \frac{c_{AV}}{\kappa}.$$

Since $1 - \phi(\rho(v)) \cdot \kappa > 0$, this expression is positive for sufficiently large v', contradicting the optimality of $v' = \rho(v')$. So it must be that $\phi(\rho(v)) = 1/\kappa$ for every v.

C.4 Proof of Proposition 3

Consider first a specific tariff. Since $\mathbb{E}[v] < \infty$ and $\mathbb{E}[v^{-\sigma-1}] < \infty$, it must be that $\mathbb{E}[v^{-\sigma}] < \infty$ and $\mathbb{E}[v^{1-\sigma}] < \infty$. The dominated convergence theorem therefore ensures that $\mathbb{E}[(v+t)^{-\sigma}]$ and $\mathbb{E}[(v+t)^{1-\sigma}]$ are continuous in t, and hence so is $\Pi_S(t)$. Assumption 1 further ensures that Π_S is increasing. And $\Pi_S(0) = 0$ while

$$\lim_{t\to\infty} \Pi_S(t) = \lim_{t\to\infty} \frac{\mathbb{E}[(v/t+1)^{-\sigma}]}{\mathbb{E}[(v/t+1)^{1-\sigma}]} = 1$$

by the monotone convergence theorem. The stated properties of t^* follow immediately.

Now consider an ad valorem tariff. For each $R \in (0,1)$, the tariff $\tau^*(R)$ satisfying $\Pi_{AV}(\tau^*(R)) = R$ can be characterized by solving

$$R = \frac{\tau}{1+\tau} - \frac{\bar{c}_{AV}}{1+\tau}$$

to obtain the unique solution

$$\tau^*(R) = \frac{R + \bar{c}_{AV}}{1-R}.$$

It is immediate that this solution satisfies the stated properties.

C.5 Proof of Proposition 4

The price indices under specific and ad valorem tariffs are

$$\bar{P}_S(t) = \left(\mathbb{E}\left[(v+t)^{1-\sigma}\right]\right)^{1/(1-\sigma)}, \quad \bar{P}_{AV}(\tau) = (1+\tau) \cdot \left(\mathbb{E}[v^{1-\sigma}]\right)^{1/(1-\sigma)}.$$

Also, $P_S(R) = \bar{P}_S(t^*(R))$ while $P_{AV}(R) = \bar{P}_{AV}(\tau^*(R))$ for every $R \in (0,1)$. The dominated convergence theorem ensures that $\bar{P}_S$ is continuous in t, and it is immediate that $\bar{P}_{AV}$ is continuous in τ. The first result of the proposition then follows from the fact that $t^*(0+) = 0$ while $\tau^*(0+) > 0$.

We next establish that τ^* and t^* are differentiable and obtain implicit expressions for their derivatives. The explicit expression for τ^* obtained in the proof of Proposition 3 verifies that it is differentiable with derivative

$$\frac{d\tau^*}{dR} = \frac{1 + \bar{c}_{AV}}{(1-R)^2}.$$

Meanwhile, t^* solves the equation $\Pi_S(t^*(R)) = R$. In other words, t^* is the inverse of Π_S. We wish to invoke the inverse function theorem, which requires establishing that Π_S is continuously differentiable. The assumptions $\mathbb{E}[v^{-\sigma-1}] < \infty$ and $\mathbb{E}[v] < \infty$ ensure that the Leibniz integral rule applies and

$$\frac{d}{dt}\mathbb{E}[(v+t)^{-\sigma}] = -\sigma \cdot \mathbb{E}[(v+t)^{-\sigma-1}], \quad \frac{d}{dt}\mathbb{E}[(v+t)^{1-\sigma}] = (1-\sigma) \cdot \mathbb{E}[(v+t)^{-\sigma}].$$

Hence Π_S is differentiable and

$$\Pi_S'(t) = \frac{\Pi_S(t)}{t} \cdot (1 - \sigma \cdot \omega(t) + (\sigma - 1) \cdot \Pi_S(t)),$$

where

$$\omega(t) \equiv t \cdot \frac{\mathbb{E}[(v+t)^{-\sigma-1}]}{\mathbb{E}[(v+t)^{-\sigma}]}.$$

Assumption 1 ensures that $\omega(t) \leq (1 - \sigma^{-1}) \cdot \Pi_S(t) + \sigma^{-1}$ for all t.

The dominated convergence theorem ensures that Π_S' is continuous. Assumption 1 ensures further that $\Pi_S'(t) \geq 0$. Whenever the derivative is positive, the inverse function theorem implies that t^* is differentiable and

$$\frac{dt^*}{dR} = \frac{1}{\Pi_S'(t^*(R))}.$$

Meanwhile, whenever $\Pi_S'(t^*(R)) = 0$, the identity $\Pi_S(t^*(R)) = R$ combined with the monotonicity of t^* implies that $dt^*/dR = \infty$.

We next obtain expressions for the derivatives of P_S and P_{AV}. For the same reasons that Π_S is differentiable, the derivative of $\bar{P}_S$ exists everywhere and

$$\bar{P}_S'(t) = \bar{P}_S(t) \cdot \frac{\mathbb{E}[(v+t)^{-\sigma}]}{\mathbb{E}[(v+t)^{1-\sigma}]} = \bar{P}_S(t) \cdot \frac{\Pi_S(t)}{t}.$$

Let $\Omega(R) \equiv \omega(t^*(R))$. Then by the chain rule,

$$P_S'(R) = \bar{P}_S'(t^*(R)) \cdot \frac{dt^*}{dR} = \frac{P_S(R)}{1 - \sigma \cdot \Omega(R) + (\sigma - 1) \cdot R},$$

with this expression taken to be $+\infty$ whenever $\Omega(R) = (1 - \sigma^{-1}) \cdot R + \sigma^{-1}$. Meanwhile, direct computation yields

$$P_{AV}'(R) = \bar{P}_{AV}'(\tau^*(R)) \cdot \frac{d\tau^*}{dR} = \frac{P_{AV}(R)}{1 + \tau^*(R)} \cdot \frac{1 + \bar{c}_{AV}}{(1-R)^2} = \frac{P_{AV}(R)}{1 - R}.$$

We now compare these two derivatives. The Cauchy–Schwarz inequality implies that

$$\mathbb{E}[(v+t)^{-\sigma-1}] \cdot \mathbb{E}[(v+t)^{1-\sigma}] > \left(\mathbb{E}[(v+t)^{-\sigma}]\right)^2$$

and therefore

$$\omega(t) > t \cdot \frac{\mathbb{E}[(v+t)^{-\sigma}]}{\mathbb{E}[(v+t)^{1-\sigma}]} = \Pi_S(t)$$

for all $t > 0$. Hence $\Omega(R) > R$ and

$$\frac{1}{P_S(R)}\frac{dP_S}{dR} > \frac{1}{P_{AV}(R)}\frac{dP_{AV}}{dR}$$

for all $R > 0$, establishing the second result from the proposition statement.

The results so far establish that the ratio P_S/P_{AV} is increasing in R and initially below 1. It remains only to determine whether it crosses 1. Let

$$\Phi(t) \equiv \left(\mathbb{E}[(v+t)^{1-\sigma}]\right)^{1/(1-\sigma)} \cdot (1 - \Pi_S(t)),$$

and write P_S/P_{AV} as

$$\frac{P_S(R)}{P_{AV}(R)} = \frac{\Phi(t^*(R))}{(1+\bar{c}_{AV}) \cdot (\mathbb{E}[v^{1-\sigma}])^{1/(1-\sigma)}}.$$

Since P_S/P_{AV} and t^* are each increasing, it must be that Φ is as well, and so P_S/P_{AV} crosses 1 iff

$$\Phi(\infty) > (1+\bar{c}_{AV}) \cdot (\mathbb{E}[v^{1-\sigma}])^{1/(1-\sigma)}.$$

Writing Φ as

$$\Phi(t) = \left(\mathbb{E}[(v/t+1)^{1-\sigma}]\right)^{\sigma/(1-\sigma)} \cdot \mathbb{E}[v \cdot (v/t+1)^{-\sigma}]$$

allows the limit $t \to \infty$ to be evaluated using the monotone convergence theorem, yielding $\Phi(\infty) = \mathbb{E}[v]$. Hence, a crossing point exists iff

$$\mathbb{E}[v] > (1+\bar{c}_{AV}) \cdot (\mathbb{E}[v^{1-\sigma}])^{1/(1-\sigma)}.$$

This condition holds iff c_{AV} is sufficiently small.

C.6 Proof of Proposition 5

The (average) reported unit value under a specific tariff t is

$$\frac{\mathbb{E}[v \cdot (v+t)^{-\sigma}]}{\mathbb{E}[(v+t)^{-\sigma}]}.$$

Meanwhile, the same quantity under an ad valorem tariff τ is

$$\bar{v} - \frac{c_{AV}}{\kappa \cdot \tau}.$$

Hence, the drop in the reported log (average) unit value following a switch from the specific tariff t to the ad valorem tariff $\bar{\tau}(t) > \bar{c}_{AV}$ is

$$\log\left(\frac{\mathbb{E}[v \cdot (v+t)^{-\sigma}]}{\mathbb{E}[(v+t)^{-\sigma}]}\right) - \log \bar{v} + \log\left(\frac{\bar{\tau}(t)}{\bar{\tau}(t) - \bar{c}_{AV}}\right).$$

Let

$$\tilde{\Delta}(t) \equiv \log\left(\frac{\mathbb{E}[v \cdot (v+t)^{-\sigma}]}{\mathbb{E}[(v+t)^{-\sigma}]}\right) - \log \bar{v}.$$

Then $\Delta(\tau) = \tilde{\Delta}(\bar{\tau}^{-1}(\tau))$. It is immediate that $\Delta(0) = \tilde{\Delta}(0) = 0$. Meanwhile, straightforward computation yields

$$\tilde{\Delta}'(0) = \frac{\sigma}{\bar{v}} \cdot \left(\frac{\mathbb{E}[v^{1-\sigma}] \cdot \mathbb{E}[v^{-\sigma-1}]}{(\mathbb{E}[v^{-\sigma}])^2} - 1\right), \quad \bar{\tau}'(0) = \bar{v}^{-1}.$$

Hence,

$$\Delta'(0) = \frac{\tilde{\Delta}'(0)}{\bar{\tau}'(0)} = \sigma \cdot \left(\frac{\mathbb{E}[v^{1-\sigma}] \cdot \mathbb{E}[v^{-\sigma-1}]}{(\mathbb{E}[v^{-\sigma}])^2} - 1\right).$$

Taylor's theorem ensures that

$$\Delta(\tau) = \Delta'(0) \cdot \tau + o(\tau).$$

The Cauchy–Schwarz inequality implies that $\Delta'(0) > 0$, yielding the claims in the proposition statement with $\beta = \Delta'(0)$.

D Optimal Multi-Category Tariff Codes

In this appendix, we formally state and analyze the multi-category tariff code design problem.

As described in Section 3.3, we partition the universe of varieties into a finite, exogenous set of *products* $\mathcal{G}$. The varieties within each product $g \in \mathcal{G}$ have values distributed as F_g, each of which satisfies the same smoothness and support conditions as F. The proportion of varieties contained within each product is $\mu(g) > 0$, with $\sum_{g \in \mathcal{G}} \mu(g) = 1$.

A *tariff code* consists of a tuple $\mathcal{C} = (\mathcal{K}, \boldsymbol{x}, \boldsymbol{T})$, where:

- $\mathcal{K} = (K_j)_{j \in \mathcal{J}}$ is a partition of $\mathcal{G}$, which we refer to as a *categorization* of products. The set $\mathcal{J} = \{1, ..., |\mathcal{K}|\}$ indexes categories within the partition, and we will generally refer to a category $K_j \in \mathcal{K}$ by its index j.
- $\boldsymbol{x} = (x_j)_{j \in \mathcal{J}}$ records a tariff regime for each category, with $x_j \in \{S, AV\}$.
- $\boldsymbol{T} = (T_j)_{j \in \mathcal{J}}$ records a tariff rate for each category.

Given a tariff code, we write:

- $\mu_j \equiv \sum_{g \in K_j} \mu(g)$ for the proportion of varieties in each category,

- $F_j \equiv \mu_j^{-1} \sum_{g \in K_j} \mu(g) F_g$ for the distribution of values in each category,
- $\mathbb{E}_j$ for expectations under F_j,
- $\bar{v}_j \equiv \mathbb{E}_j[v^{1-\sigma}]/\mathbb{E}_j[v^{-\sigma}]$ and $\bar{c}_{AV,j} \equiv c_{AV}/(\kappa \cdot \bar{v}_j)$ for average unit values and unit inspection costs in each category,
- p_j for category price indices, where[46]

$$p_j \equiv \begin{cases} \left(\mathbb{E}_j[(v+T_j)^{1-\sigma}]\right)^{1/(1-\sigma)}, & x_j = S \\ (1 + T_j \cdot \mathbf{1}\{T_j \geq \bar{c}_{AV,j}\}) \cdot \left(\mathbb{E}_j[v^{1-\sigma}]\right)^{1/(1-\sigma)}, & x_j = AV \end{cases}$$

- ι_j for unnormalized category revenues, where

$$\iota_j \equiv \begin{cases} T_j \cdot \mathbb{E}_j[(v+T_j)^{-\sigma}], & x_j = S \\ (1+T_j)^{-\sigma} \cdot \mathbb{E}_j[v^{1-\sigma}] \cdot \max\{T_j - \bar{c}_{AV,j}, 0\}, & x_j = AV \end{cases}$$

- $R_j \equiv \iota_j / p_j^{1-\sigma}$ for net tariff burdens.

(All of these objects are code-specific. To streamline notation, we suppress explicit dependence on the code.)

To ensure regularity, we maintain a multi-category version of Assumption 1: For every categorization $\mathcal{K}$ and every category value distribution F_j, Assumption 1 holds under value distribution F_j.

Let P denote the aggregate price index under a given tariff code. Since $\sigma > 1$, an optimal tariff code maximizes $Q \equiv P^{1-\sigma}$ subject to achieving net tariff revenue R. Using the notation set out above, Q may be written

$$Q = \sum_{j \in \mathcal{J}} \mu_j p_j^{1-\sigma},$$

while aggregate tariff revenue (net of verification and classification costs) may be written

$$\Pi = \frac{\sum_{j \in \mathcal{J}} \mu_j \iota_j}{\sum_{j \in \mathcal{J}} \mu_j p_j^{1-\sigma}} - c_S \cdot H(\mathcal{K}).$$

The government's problem may therefore be written

$$\max_{\mathcal{C}} \sum_{j \in \mathcal{J}} \mu_j p_j^{1-\sigma} \quad \text{s.t.} \quad \sum_{j \in \mathcal{J}} \mu_j p_j^{1-\sigma} \cdot (R_j - c_S \cdot H(\mathcal{K}) - R) \geq 0.$$

We will say that a tariff code $(\mathcal{K}, \boldsymbol{x}, \boldsymbol{T})$ is *optimal* if it solves this problem. We will say

[46] In the boundary case $T_j = \bar{c}_{AV,j}$, there are multiple equilibria of the verification game, and the price index is not well-defined, as discussed in footnote 13. To ensure a continuous objective over a closed choice set, we select the equilibrium with the highest verification rate. The choice of equilibrium does not impact the set of optimal tariff codes, since every optimal code imposes a positive net tariff burden on every category.

that a code is $\mathcal{K}$-*optimal* if $(\boldsymbol{x}, \boldsymbol{T})$ solves the problem just stated when the categorization $\mathcal{K}$ is held fixed. And we will say that a code is $(\mathcal{K}, \boldsymbol{x})$-*optimal* if $\boldsymbol{T}$ solves the problem just stated when both the categorization $\mathcal{K}$ and tariff regimes $\boldsymbol{x}$ are held fixed. A given $(\mathcal{K}, \boldsymbol{x})$ may not admit any choices of tariff rates satisfying the constraints, in particular if $R + c_S \cdot H(\mathcal{K}) \geq 1$. We say that a pair $(\mathcal{K}, \boldsymbol{x})$ is *feasible* if $R + c_S \cdot H(\mathcal{K}) < 1$.

As a preliminary simplification, we restrict the domain of tariff rates to $T_j \geq \bar{c}_{AV,j}$ when $x_j = AV$. This restriction is without loss of generality. When $x_j = AV$, any rate $T_j \in (0, \bar{c}_{AV,j})$ is equivalent to the rate $T_j = 0$, so it is sufficient to rule out $T_j = 0$. If $T_j = 0$, then the tariff regime can be swapped to $x_j = S$ without affecting the objective or constraint. Hence, if there existed an optimal tariff code with this feature, there would exist an optimal code in the restricted problem involving a zero specific tariff in some category. But we establish below (in Lemma D.2) that optimal tariff burdens (and therefore rates) are always positive in the restricted problem. So no such optimal code can involve such a choice. Our restriction therefore does not change the set of optimal tariff codes. Going forward, we maintain the restriction without comment.

We further perform a change of variables and write tariff codes in terms of the net tariff burdens R_j rather than the tariff rates T_j. Observe that $R_j = \iota_j / p_j^{1-\sigma}$ is exactly the revenue raised in a single-category problem with value distribution F_j. Proposition 3 therefore implies that, for each category and tariff regime, there exists a unique tariff rate in $(0, \infty)$ if $x_j = S$ or $(\bar{c}_{AV,j}, \infty)$ if $x_j = AV$ inducing any net tariff burden in $(0, 1)$. By additionally identifying $R_j = 0$ with $T_j = 0$ when $x_j = S$ and with $T_j = \bar{c}_{AV,j}$ when $x_j = AV$, each vector of tariff rates can be uniquely identified with a corresponding vector of net tariff burdens, and vice versa.

Let $P_j(R_j; x_j)$ denote the category price index induced by raising revenue $R_j \in (0, 1)$ in a single-category problem with value distribution F_j and tariff regime x_j. Define $Q_j(R_j; x_j) \equiv P_j(R_j; x_j)^{1-\sigma}$. We continuously extend Q_j to $R_j = 0$ by setting $Q_j(0; x_j) = Q_j(0+; x_j)$. To compactify the choice set, we additionally allow $R_j = 1$ to be chosen by setting $Q_j(1; x_j) = Q_j(1-; x_j) = 0$. Note that Q_j is continuous and (strictly) decreasing in R_j for each category j and tariff regime x_j.

In terms of net tariff burdens, the government's problem is

$$\max_{\mathcal{C}} \sum_{j \in \mathcal{J}} \mu_j Q_j(R_j; x_j) \quad \text{s.t.} \quad \sum_{j \in \mathcal{J}} \mu_j Q_j(R_j; x_j) \cdot (R_j - c_S \cdot H(\mathcal{K}) - R) \geq 0,$$

where a tariff code is now a tuple $\mathcal{C} = (\mathcal{K}, \boldsymbol{x}, \boldsymbol{R})$ of categories, tariff regimes, and net tariff burdens. We will refer to

$$V(\mathcal{C}) \equiv \sum_{j\in\mathcal{J}} \mu_j Q_j(R_j; x_j)$$

as the *objective* of the government's optimization problem.

Lemma D.1. *There exists an optimal tariff code, and for each feasible $(\mathcal{K}, \boldsymbol{x})$ there exists a $(\mathcal{K}, \boldsymbol{x})$-optimal tariff code. Further, for each feasible $(\mathcal{K}, \boldsymbol{x})$, every $(\mathcal{K}, \boldsymbol{x})$-optimal code $(\mathcal{K}, \boldsymbol{x}, \boldsymbol{R}^*)$ satisfies $\boldsymbol{R}^* \ll \mathbf{1}$ and saturates the revenue constraint.*

Proof. To prove existence, it is sufficient to establish existence of a $(\mathcal{K}, \boldsymbol{x})$-optimal tariff code $(\mathcal{K}, \boldsymbol{x}, \boldsymbol{R}^{**}(\mathcal{K}, \boldsymbol{x}))$ given an arbitrary feasible categorization and pattern of tariff regimes. An optimal tariff code can then be constructed by choosing $(\mathcal{K}^*, \boldsymbol{x}^*)$ such that $(\mathcal{K}^*, \boldsymbol{x}^*, \boldsymbol{R}^{**}(\mathcal{K}^*, \boldsymbol{x}^*))$ maximizes the objective among the (finite, non-empty) feasible set of categorizations and tariff regimes.

Fix a feasible $(\mathcal{K}, \boldsymbol{x})$. The objective and constraint are both continuous in $\boldsymbol{R}$, since $Q_j(\cdot; x_j)$ is continuous for each j and x_j. The feasible set is additionally non-empty, since the choice of $\boldsymbol{R} = (R + c_S \cdot H(\mathcal{K})) \cdot \mathbf{1}$ satisfies the constraint. And since each net tariff burden must lie in $[0, 1]$, the feasible set is compact. Hence, there must exist an optimal vector of net tariff burdens $\boldsymbol{R}^*$.

Suppose that some $R_j^* = 1$. Consider modifying the tariff code by setting $R_j = R + c_S \cdot H(\mathcal{K}^*)$. This modification leaves the value of the constraint unchanged, and so the modified code is feasible. Additionally, it raises the objective, contradicting the hypothesized optimality of the original code. Next, suppose that the revenue constraint is slack under $(\mathcal{K}, \boldsymbol{x}, \boldsymbol{R}^*)$. Then it must be that $\boldsymbol{R}^* = \mathbf{0}$, or else some R_j could be lowered slightly to raise the objective without violating the constraint. But $\mathbf{0}$ violates the revenue constraint, another contradiction. ☐

Given a categorization $\mathcal{K}$ and category $j \in \mathcal{J}$, let $\Omega_j(R_j; AV) \equiv R_j$ and

$$\Omega_j(R_j; S) \equiv T_j^*(R_j; S) \cdot \frac{\mathbb{E}_j[(v + T_j^*(R_j; S))^{-\sigma-1}]}{\mathbb{E}_j[(v + T_j^*(R_j; S))^{-\sigma}]},$$

where $T_j^*(R_j; S)$ is the unique T_j satisfying

$$R_j = T_j \cdot \frac{\mathbb{E}_j[(v + T_j)^{-\sigma}]}{\mathbb{E}_j[(v + T_j)^{1-\sigma}]}.$$

The proof of Proposition 4 obtains expressiones for the derivative of the price index under specific and ad valorem tariffs. These expressions can be used to establish that

$$Q'_j(R_j; x_j) = \frac{(1-\sigma) \cdot Q_j(R_j; x_j)}{1 - \sigma \cdot \Omega_j(R_j; x_j) + (\sigma - 1) \cdot R_j}$$

for every j and x_j.

Lemma D.2. *Suppose that $(\mathcal{K}, \boldsymbol{x}, \boldsymbol{R}^*)$ is a $(\mathcal{K}, \boldsymbol{x})$-optimal tariff code. Then $\boldsymbol{R}^* \gg 0$ and $\Omega_j(R_j^*; x_j)$ is the same across all categories.*

Proof. Fix a $(\mathcal{K}, \boldsymbol{x})$-optimal tariff code $(\mathcal{K}, \boldsymbol{x}, \boldsymbol{R}^*)$. The result is trivial if $\mathcal{K}$ has only a single category, so assume that there are at least two categories in the code. Recall also from Lemma D.1 that $\boldsymbol{R}^* \ll 1$.

Suppose first that $Q'_j(R_j^*; x_j) < \infty$ for all categories. In that case, the objective and constraint are locally smooth in the net tariff burdens around the optimum. Since the gradient of the constraint function is non-vanishing everywhere, there must exist a Lagrange multiplier $\lambda > 0$ such that $\boldsymbol{R}^*$ is a stationary point of the Lagrangian

$$\mathscr{L}(\boldsymbol{R}) \equiv \sum_{j \in \mathcal{J}} \mu_j \mathscr{L}_j(R_j),$$

where

$$\mathscr{L}_j(R_j) \equiv Q_j(R_j; x_j) \cdot \Big(1 + \lambda \cdot (R_j - R - c_S \cdot H(\mathcal{K}))\Big).$$

The slopes of these component Lagrangians are

$$\mathscr{L}'_j(R_j) = \lambda \cdot Q_j(R_j; x_j) \cdot \frac{(\sigma - 1) \cdot (R + c_S \cdot H(\mathcal{K}) - \lambda^{-1}) + 1 - \sigma \cdot \Omega_j(R_j; x_j)}{1 - \sigma \cdot \Omega_j(R_j; x_j) + (\sigma - 1) \cdot R_j}$$

for every $R_j < 1$.

If $\lambda \leq (R + c_S \cdot H(\mathcal{K}) + 1/(\sigma - 1))^{-1}$, then $\mathscr{L}'_j(R_j) < 0$ for every j and $R_j \in (0, 1)$, implying that $\boldsymbol{R}^* = \boldsymbol{0}$. These net tariff burdens violate the constraint and so cannot be optimal. It must therefore be that $\lambda > (R + c_S \cdot H(\mathcal{K}) + 1/(\sigma - 1))^{-1}$. In that case, $\mathscr{L}'_j(0) > 0$ and satisfaction of the first-order conditions requires that $\boldsymbol{R}^* \gg \boldsymbol{0}$ and (recalling that $\boldsymbol{R}^* \ll \boldsymbol{1}$) $\mathscr{L}'_j(R_j^*) = 0$ for every category j. This stationarity condition may be equivalently written

$$\Omega_j(R_j^*; x_j) = \sigma^{-1} + (1 - \sigma^{-1}) \cdot (R + c_S \cdot H(\mathcal{K}) - \lambda^{-1}).$$

Since this expression does not depend on the category, $\Omega_j(R_j^*; x_j)$ is the same across categories.

Suppose instead that some category j satisfies $Q'_j(R_j^*; x_j) = \infty$. This choice is consistent with optimality only if $Q'_k(R_k^*; x_k) = \infty$ for all other categories k; for otherwise an increase in the objective could be achieved at no cost to revenue by slightly reducing the net tariff burden in category j and increasing the net tariff burden in some category satisfying $Q'_k(R^*_k; x_k) < \infty$. Note that in this case $\boldsymbol{R}^* \gg 0$, since $1 - \sigma \cdot \Omega_j(0; x_j) + (\sigma - 1) \cdot 0 = 1 > 0$ and hence $Q'_j(0; x_j) < \infty$ for every category.

To apply first-order conditions in this environment, we perform a change of variables and let $\boldsymbol{q} = (q_j)_{j \in \mathcal{J}}$ be the choice variables, with $R_j = \eta_j(q_j)$, where η_j is the (well-defined, since Q_j is continuous and strictly monotone) inverse function of $Q_j(\cdot; x_j)$. Since $\eta'_j(q_j) = 1/Q'_j(\eta_j(q_j); x_j)$, it follows that $\eta'_j(q^*_j) = 0$ for all categories, where $q^*_j \equiv Q_j(R^*_j; x_j)$. Hence, there must exist a multiplier $\lambda > 0$ for which $\boldsymbol{q}^*$ is a stationary point of the transformed Lagrangian

$$\hat{\mathscr{L}}(\boldsymbol{q}) \equiv \sum_{j \in \mathcal{J}} \mu_j \hat{\mathscr{L}}_j(q_j),$$

where

$$\hat{\mathscr{L}}_j(q_j) \equiv q_j \cdot \Big(1 + \lambda \cdot (\eta_j(q_j) - R - c_S \cdot H(\mathcal{K}))\Big).$$

Stationarity requires that

$$\hat{\mathscr{L}}'_j(q^*_j) = 1 + \lambda \cdot (\eta_j(q^*_j) - R - c_S \cdot H(\mathcal{K})) + q^*_j \cdot \lambda \cdot \eta'_j(q^*_j) = 0$$

for every j. Since $\eta'_j(q^*_j) = 0$ for all j by hypothesis, these identities reduce to

$$\eta_j(q^*_j) = R + c_S \cdot H(\mathcal{K}) - \lambda^{-1}.$$

But the revenue constraint is

$$\sum_{j \in \mathcal{J}} \mu_j q_j \cdot \Big(\eta_j(q_j) - R - c_S \cdot H(\mathcal{K})\Big) \geq 0,$$

which is violated at this stationary point. So it must be that $Q'_j(R^*_j; x_j) < \infty$ for every category, meaning that our analysis for that case applies universally. $\square$

Given a tariff code $\mathcal{C} = (\mathcal{K}, \boldsymbol{x}, \boldsymbol{R})$, for each subset of categories $\mathcal{J}' \subset \mathcal{J}$ we define

$$E(\mathcal{J}'; \mathcal{C}) \equiv \frac{\sum_{j \in \mathcal{J}'} \mu_j Q_j(R_j; x_j)}{\sum_{k \in \mathcal{J}} \mu_k Q_k(R_k; x_k)}$$

to be the representative consumer's *expenditure share* on products in $\mathcal{J}'$. We further define the $\mathcal{J}'$-aggregate net tariff burden

$$\bar{R}(\mathcal{J}'; \mathcal{C}) \equiv \frac{\sum_{j \in \mathcal{J}'} E(j; \mathcal{C}) R_j}{E(\mathcal{J}'; \mathcal{C})}.$$

To streamline notation, we will write $E(j; \mathcal{C}) = E(\{j\}; \mathcal{C})$ and $\bar{R}(j; \mathcal{C}) = \bar{R}(\{j\}; \mathcal{C})$ for single-category expenditure shares and net tariff burdens.

Note that $\bar{R}(j;\mathcal{C}) = R_j$ for each category $j \in \mathcal{J}$, and further $\bar{R}(\mathcal{J};\mathcal{C})$ is the revenue raised by $\mathcal{C}$, gross of classification costs. If $\mathcal{C}$ is $(\mathcal{K}, \boldsymbol{x})$-optimal, then Lemma D.3 implies that $\bar{R}(\mathcal{J};\mathcal{C}) = R + c_S \cdot H(\mathcal{K})$.

Lemma D.3. *Suppose that $\mathcal{C}^* = (\mathcal{K}, \boldsymbol{x}^*, \boldsymbol{R}^*)$ is a $\mathcal{K}$-optimal tariff code. Fix any subset of categories $\mathcal{J}' \subset \mathcal{J}$, and let $\mathcal{K}' = (K_j)_{j\in\mathcal{J}'}$ and $\boldsymbol{x}' = (x_j^*)_{j\in\mathcal{J}'}$ and $\boldsymbol{R}' = (R_j^*)_{j\in\mathcal{J}'}$. Then the tariff code $\mathcal{C}' = (\mathcal{K}', \boldsymbol{x}', \boldsymbol{R}')$ must be a $\mathcal{K}'$-optimal tariff code in the problem with products $\mathcal{G}' = \mathcal{K}'$ in proportions $\mu_j' = \mu_j / \sum_{k\in\mathcal{J}'} \mu_k$, no classification costs, and revenue goal $\bar{R}(\mathcal{J}';\mathcal{C}^*)$.*

Proof. By way of contradiction, suppose that there exists a subset of categories $\mathcal{J}'$ for which the claim does not hold. Let $\bar{\mathcal{J}}' \equiv \mathcal{J} \setminus \mathcal{J}'$ and $M' \equiv \sum_{j\in\mathcal{J}'} \mu_j$ and $\bar{\mu}_j' \equiv \mu_j/(1-M')$ for each $j \in \bar{\mathcal{J}}'$. By hypothesis, there must exist a tariff code $\mathcal{C}^\dagger$ in the auxiliary problem achieving a higher objective than $\mathcal{C}'$ while raising at least as much revenue. Raise all net tariff burdens in $\mathcal{C}^\dagger$ simultaneously to preserve expenditure shares until the resulting code $\mathcal{C}'' = (\mathcal{K}', \boldsymbol{x}'', \boldsymbol{R}'')$ achieves the same objective as $\mathcal{C}'$. Then $\mathcal{C}''$ must raise more revenue than $\mathcal{C}'$.

Define a new tariff code $\mathcal{C}^{**} = (\mathcal{K}, \boldsymbol{x}^{**}, \boldsymbol{R}^{**})$ in the original problem by letting

$$x_j^{**} = \begin{cases} x_j'', & j \in \mathcal{J}' \\ x_j^*, & j \in \bar{\mathcal{J}}' \end{cases}$$

with $\boldsymbol{R}^{**}$ defined analogously. Since $\mathcal{C}'$ and $\mathcal{C}''$ achieve the same objective in the auxiliary problem, it follows that $\mathcal{C}^{**}$ achieves objective

$$\begin{aligned} & M' \cdot \sum_{j\in\mathcal{J}'} \mu_j' Q_j(R_j''; x_j'') + (1-M') \cdot \sum_{j\in\bar{\mathcal{J}}'} \bar{\mu}_j' Q_j(R_j^*; x_j^*) \\ &= M' \cdot \sum_{j\in\mathcal{J}'} \mu_j' Q_j(R_j^*; x_j^*) + (1-M') \cdot \sum_{j\in\bar{\mathcal{J}}'} \bar{\mu}_j' Q_j(R_j^*; x_j^*) \\ &= \sum_{j\in\mathcal{J}} \mu_j Q_j(R_j^*; x_j^*). \end{aligned}$$

in the original problem. In other words, $\mathcal{C}^{**}$ achieves the same price index as $\mathcal{C}^*$. Further, equalization of objectives in the auxiliary problem implies that $E(\mathcal{J}';\mathcal{C}^{**}) = E(\mathcal{J}';\mathcal{C}^*)$ in the original problem. Then since $\mathcal{C}''$ raises more revenue than $\mathcal{C}'$ in the auxiliary problem, it follows that revenue raised under $\mathcal{C}^{**}$ in the original problem is

$$\begin{aligned} & E(\mathcal{J}';\mathcal{C}^{**}) \cdot \frac{\sum_{j\in\mathcal{J}'} \mu_j Q_j(R_j''; x_j'') R_j''}{\sum_{j\in\mathcal{J}'} \mu_j Q_j(R_j''; x_j'')} + (1 - E(\mathcal{J}';\mathcal{C}^{**})) \cdot \frac{\sum_{j\in\bar{\mathcal{J}}'} \mu_j Q_j(R_j^*; x_j^*) R_j^*}{\sum_{j\in\bar{\mathcal{J}}'} \mu_j Q_j(R_j^*; x_j^*)} \\ &= E(\mathcal{J}';\mathcal{C}^*) \cdot \frac{\sum_{j\in\mathcal{J}'} \mu_j' Q_j(R_j''; x_j'') R_j''}{\sum_{j\in\mathcal{J}'} \mu_j' Q_j(R_j''; x_j'')} + (1 - E(\mathcal{J}';\mathcal{C}^*)) \cdot \frac{\sum_{j\in\bar{\mathcal{J}}'} \bar{\mu}_j' Q_j(R_j^*; x_j^*) R_j^*}{\sum_{j\in\bar{\mathcal{J}}'} \bar{\mu}_j' Q_j(R_j^*; x_j^*)} \end{aligned}$$

$$
\begin{aligned}
&> E(\mathcal{J}';\mathcal{C}^*)\cdot\frac{\sum_{j\in\mathcal{J}'}\mu'_j Q_j(R^*_j;x^*_j)R^*_j}{\sum_{j\in\mathcal{J}'}\mu'_j Q_j(R^*_j;x^*_j)} + (1-E(\mathcal{J}';\mathcal{C}^*))\cdot\frac{\sum_{j\in\bar{\mathcal{J}}'}\bar{\mu}'_j Q_j(R^*_j;x^*_j)R^*_j}{\sum_{j\in\bar{\mathcal{J}}'}\bar{\mu}'_j Q_j(R^*_j;x^*_j)}\\
&= \frac{\sum_{j\in\mathcal{J}}\mu_j Q_j(R^*_j;x^*_j)R^*_j}{\sum_{j\in\mathcal{J}}\mu_j Q_j(R^*_j;x^*_j)}.
\end{aligned}
$$

In other words, $\mathcal{C}^{**}$ raises more revenue than $\mathcal{C}^*$ in the original problem. Slightly lower all net tariff burdens in $\mathcal{C}^{**}$ simultaneously to preserve expenditure shares. The resulting tariff code satisfies the revenue goal while achieving a higher objective than $\mathcal{C}^*$, contradicting the hypothesized optimality of $\mathcal{C}^*$. □

E Proofs of Multi-Category Results

E.1 Proof of Proposition 6

Fix an optimal tariff code $\mathcal{C}^* = (\mathcal{K}^*, \boldsymbol{x}^*, \boldsymbol{R}^*)$. Lemma D.2 establishes that $\Omega_j(R^*_j; x^*_j)$ is the same across all categories. Since $\Omega_j(R^*_j; AV) = R^*_j$, net tariff burdens are optimally equalized across ad valorem categories. And as established in the proof of Proposition 3, whenever an ad valorem tariff is used, the tariff rate can be related to the net tariff burden via the expression

$$
T^*_j(R_j; AV) = \frac{R_j + \bar{c}_{AV,j}}{1 - R_j}.
$$

As a result, tariff rates (i.e., gross tariff burdens) are lower the lower is $\bar{c}_{AV,j} = c_{AV}/(\kappa\cdot\bar{v}_j)$, i.e., the higher is the unit value of consumption.

E.2 Proof of Proposition 7

Fix an optimal tariff code $\mathcal{C}^* = (\mathcal{K}^*, \boldsymbol{x}^*, \boldsymbol{R}^*)$. Lemma D.2 establishes that $\Omega_j(R^*_j; x^*_j)$ is the same across all categories. By definition, $\Omega_j(R_j; AV) = R_j$; meanwhile, the proof of Proposition 4 establishes that $\Omega_j(R_j; S) > R_j$. As a result, net tariff burdens must be higher in ad valorem categories than in specific categories.

E.3 Proof of Proposition 8

It suffices to prove that the result holds for an arbitrary categorization $\mathcal{K}$ and $\mathcal{K}$-optimal tariff code $\mathcal{C}^* = (\mathcal{K}, \boldsymbol{x}^*, \boldsymbol{R}^*)$. Suppose that categories $j, k \in \mathcal{J}$ are value-ordered. Lemma D.3 establishes that the tariff regimes and burdens applied to categories j and k in code $\mathcal{C}^*$ must also be optimal in the auxiliary problem with categories j and k acting as products, no classification costs, and revenue goal $\bar{R}(\{j,k\};\mathcal{C}^*)$. It is therefore sufficient to establish the claim for $\{1,2\}$-optimal codes in 2-product problems involving no classification costs and value-ordered categories.

We restrict attention to such problems going forward, letting category 2 be the higher-valued category for concreteness. When referring to category-specific expressions, we will drop subscripts for expressions involving expectations with respect to values in category 1, and use a subscript α for category 2.

We first observe that the price index and tariff revenue under scale factor α and specific tariff $T = \alpha \cdot \bar{T}$ are

$$\bar{P}_\alpha(T;S) = \alpha \cdot \left(\mathbb{E}[(v+\bar{T})^{1-\sigma}]\right)^{1/(1-\sigma)}, \quad \Pi_\alpha(T;S) = \bar{T} \cdot \frac{\mathbb{E}[(v+\bar{T})^{-\sigma}]}{\mathbb{E}[(v+\bar{T})^{1-\sigma}]}.$$

It follows immediately that the specific tariff inducing net tariff burden R in the scaled-ordered category is $T^*_\alpha(R;S) = T^*(R;S)$. The corresponding price index is

$$P_\alpha(R;S) = \alpha \cdot \left(\mathbb{E}[(v+T^*(R;S))^{1-\sigma}]\right)^{1/(1-\sigma)}.$$

Meanwhile, the same expressions under an ad valorem tariff $T > 0$ are

$$\bar{P}_\alpha(T;AV) = \alpha \cdot (1+T) \cdot \left(\mathbb{E}[v^{1-\sigma}]\right)^{1/(1-\sigma)}, \quad \Pi_\alpha(T;AV) = \frac{T}{1+T} - \frac{\bar{c}_{AV}/\alpha}{1+T}.$$

Hence

$$P_\alpha(R;AV) = \alpha \cdot \frac{1+\bar{c}_{AV}/\alpha}{1-R} \cdot \left(\mathbb{E}[v^{1-\sigma}]\right)^{1/(1-\sigma)}.$$

We make free use of these expressions in what follows.

By way of contradiction, suppose that $\boldsymbol{x}^* = (AV, S)$. Let $\boldsymbol{R}^{**} \equiv (R^*_2, R^*_1)$ and $\boldsymbol{x}^{**} \equiv (S, AV)$ and $\mathcal{C}^{**} \equiv (\mathcal{K}, \boldsymbol{x}^{**}, \boldsymbol{R}^{**})$. Then

$$\frac{E(1;\mathcal{C}^*)}{E(2;\mathcal{C}^*)} = \frac{\mu_1 Q(R^*_1;AV)}{\mu_2 Q_\alpha(R^*_2;S)} = \frac{\mu_1}{\mu_2} \cdot \left(\frac{1+\bar{c}_{AV}}{\alpha \cdot (1-R^*_1)}\right)^{1-\sigma} \cdot \frac{\mathbb{E}[v^{1-\sigma}]}{\mathbb{E}[(v+T^*(R^*_2;S))^{1-\sigma}]},$$

while

$$\frac{E(2;\mathcal{C}^{**})}{E(1;\mathcal{C}^{**})} = \frac{\mu_2 Q_\alpha(R^*_1;AV)}{\mu_1 Q(R^*_2;S)} = \frac{\mu_2}{\mu_1} \cdot \left(\frac{\alpha+\bar{c}_{AV}}{1-R^*_1}\right)^{1-\sigma} \cdot \frac{\mathbb{E}[v^{1-\sigma}]}{\mathbb{E}[(v+T^*(R^*_2;S))^{1-\sigma}]}.$$

The ratio of these two quantities is

$$\frac{E(1;\mathcal{C}^*)/E(2;\mathcal{C}^*)}{E(2;\mathcal{C}^{**})/E(1;\mathcal{C}^{**})} = \left(\frac{\alpha+\alpha\cdot\bar{c}_{AV}}{\alpha+\bar{c}_{AV}}\right)^{1-\sigma} \cdot \left(\frac{\mu_1}{\alpha^{1-\sigma}\mu_2}\right)^2.$$

Since $\sigma > 1$, we have

$$\left(\frac{\alpha+\alpha\cdot\bar{c}_{AV}}{\alpha+\bar{c}_{AV}}\right)^{1-\sigma} < 1.$$

Meanwhile, the assumption that the two categories are balanced is equivalent to $\mu_1 = \alpha^{1-\sigma}\mu_2$.

Hence,

$$\frac{E(1;\mathcal{C}^*)/E(2;\mathcal{C}^*)}{E(2;\mathcal{C}^{**})/E(1;\mathcal{C}^{**})} < 1.$$

Since expenditure shares sum to one under both codes, the previous inequality is equivalent to $E(1;\mathcal{C}^*) < E(2;\mathcal{C}^{**})$.

Since $\mathcal{C}^*$ is $\mathcal{K}$-optimal by hypothesis, it must exactly satisfy the revenue goal:

$$R = E(1;\mathcal{C}^*) \cdot R_1^* + (1 - E(1;\mathcal{C}^*)) \cdot R_2^*.$$

Recall from Lemma D.2 that a $\mathcal{K}$-optimal tariff code must satisfy $\Omega_1(R_1^*; x_1^*) = \Omega_2(R_2^*; x_2^*)$. Since $\Omega_1(R; AV) = R$ while $\Omega_2(R; S) > R$, it follows that $R_1^* > R_2^*$. Since also $E(1;\mathcal{C}^*) < E(2;\mathcal{C}^{**})$, revenue under the tariff code $\mathcal{C}^{**}$ is

$$E(2;\mathcal{C}^{**}) \cdot R_1^* + (1 - E(2;\mathcal{C}^{**})) \cdot R_2^* > R.$$

In other words, $\mathcal{C}^{**}$ also achieves the revenue goal. Meanwhile, the objective under $\mathcal{C}^*$ is

$$\begin{aligned} V^* &= \mu_1 Q(R_1^*; AV) + \mu_2 Q_\alpha(R_2^*; S) \\ &= \mu_1 \cdot \left(\frac{1+\bar{c}_{AV}}{1-R_1^*}\right)^{1-\sigma} \cdot \mathbb{E}[v^{1-\sigma}] + \mu_2 \cdot \alpha^{1-\sigma} \cdot \mathbb{E}[(v + T^*(R_2^*; S))^{1-\sigma}], \end{aligned}$$

while under $\mathcal{C}^{**}$ it is

$$\begin{aligned} V^{**} &= \mu_2 Q_\alpha(R_1^*; AV) + \mu_1 Q(R_2^*; S) \\ &= \mu_2 \cdot \alpha^{1-\sigma} \cdot \left(\frac{1+\bar{c}_{AV}/\alpha}{1-R_1^*}\right)^{1-\sigma} \cdot \mathbb{E}[v^{1-\sigma}] + \mu_1 \cdot \mathbb{E}[(v + T^*(R_2^*; S))^{1-\sigma}]. \end{aligned}$$

So

$$\begin{aligned} V^{**} - V^* = &\left(\mu_1 - \mu_2 \cdot \alpha^{1-\sigma}\right) \cdot \mathbb{E}[(v + T^*(R_2^*; S))^{1-\sigma}] \\ &+ \left(\mu_2 \cdot \alpha^{1-\sigma} \cdot \left(\frac{1+\bar{c}_{AV}/\alpha}{1+\bar{c}_{AV}}\right)^{1-\sigma} - \mu_1\right) \cdot \left(\frac{1+\bar{c}_{AV}}{1-R_1^*}\right)^{1-\sigma} \cdot \mathbb{E}[v^{1-\sigma}]. \end{aligned}$$

Since $\sigma > 1$, we have

$$\left(\frac{1+\bar{c}_{AV}/\alpha}{1+\bar{c}_{AV}}\right)^{1-\sigma} > 1$$

and therefore

$$V^{**} - V^* > \left(\mu_2 \cdot \alpha^{1-\sigma} - \mu_1\right) \cdot \left\{\left(\frac{1+\bar{c}_{AV}}{1-R_1^*}\right)^{1-\sigma} \cdot \mathbb{E}[v^{1-\sigma}] - \mathbb{E}[(v + T^*(R_2^*; S))^{1-\sigma}]\right\}.$$

The balance condition $\mu_1 = \alpha^{1-\sigma}\mu_2$ implies a vanishing right-hand side, i.e., $V^{**} > V^*$. In other words, the code $\mathcal{C}^{**}$ achieves a higher value of the objective than $\mathcal{C}^*$. Since it also satisfies the revenue goal, it cannot be that $\mathcal{C}^*$ is $\mathcal{K}$-optimal, which is the desired contradiction.

E.4 Proof of Proposition 9

As in the proof of Proposition 8, it is sufficient to establish the claim for $\{1,2\}$-optimal codes in 2-product problems involving no classification costs and dispersion-ordered categories. We restrict attention to such problems going forward, letting category 2 be the more-dispersed category for concreteness. When referring to category-specific expressions, we will drop subscripts for expressions involving expectations with respect to values in category 1, and use a subscript Δ when referring to values in category 2. We will further write $\mathbb{E}^Q$ for expectations wrt the quantity-weighted value function F^Q.

Since the density f^Q of F^Q satisfies $f^Q \sim v^{-\sigma} f$, it follows that

$$F(v) = \int_0^v w^\sigma \, dF^Q(w) \Big/ \int_0^\infty w^\sigma \, dF^Q(w).$$

Hence, for any measurable function h we have $\mathbb{E}[h(v)] = \mathbb{E}^Q[v^\sigma h(v)]/\mathbb{E}^Q[v^\sigma]$. In particular, since the two categories are dispersion ordered, we have

$$\bar{v} = \mathbb{E}[v^{1-\sigma}]/\mathbb{E}[v^{-\sigma}] = \mathbb{E}^Q[v] = \mathbb{E}^Q_\Delta[v] = \mathbb{E}_\Delta[v^{1-\sigma}]/\mathbb{E}_\Delta[v^{-\sigma}] = \bar{v}_\Delta.$$

Additionally, define $\xi(r) \equiv \Omega(r;S)/r$, with ξ_Δ defined similarly with respect to Ω_Δ. The definition of $\Omega(r;S)$ implies that

$$\xi(r) = \frac{\mathbb{E}[(v+T^*(r;S))^{-\sigma-1}] \cdot \mathbb{E}[(v+T^*(r;S))^{1-\sigma}]}{\left(\mathbb{E}[(v+T^*(r;S))^{-\sigma}]\right)^2}.$$

Since $T^*(0;S) = 0$, we have

$$\xi(0) = \frac{\mathbb{E}[v^{-\sigma-1}] \cdot \mathbb{E}[v^{1-\sigma}]}{\left(\mathbb{E}[v^{-\sigma}]\right)^2} = \mathbb{E}^Q[v^{-1}] \cdot \mathbb{E}^Q[v].$$

Similarly, $\xi_\Delta(0) = \mathbb{E}^Q_\Delta[v^{-1}] \cdot \mathbb{E}^Q_\Delta[v]$. Since F^Q_Δ is a mean-preserving spread of F^Q and v^{-1} is a strictly convex function of v, it follows that $\xi_\Delta(0) > \xi(0)$. The dominated convergence theorem implies that ξ and ξ_Δ are both continuous in r, so that also $\xi_\Delta(r) > \xi(r)$ for sufficiently small $r > 0$. Equivalently, $\Omega_\Delta(r;S) > \Omega(r;S)$ for sufficiently small $r > 0$.

Let $\mathcal{K} = \{1,2\}$ and $\mathcal{C}^* = (\mathcal{K}, \boldsymbol{x}^*, \boldsymbol{R}^*)$ be a $\mathcal{K}$-optimal tariff code. By way of contradiction, suppose that $x^* = (AV, S)$. Let $\boldsymbol{R}^{**} \equiv (R_2^*, R_1^*)$ and $x^{**} \equiv (S, AV)$ and $\mathcal{C}^{**} \equiv (\mathcal{K}, \boldsymbol{x}^{**}, \boldsymbol{R}^{**})$. Additionally, define $\zeta(r) \equiv 1 - \sigma \cdot \Omega(r;S) + (\sigma - 1) \cdot r$, with ζ_Δ defined analogously with respect to Ω_Δ. Then, integrating the derivative of Q stated in Appendix D, we obtain

$$\frac{E(1;\mathcal{C}^*)}{E(2;\mathcal{C}^*)} = \frac{\mu_1 Q(R_1^*;AV)}{\mu_2 Q_\Delta(R_2^*;S)}$$
$$= \frac{\mu_1 \mathbb{E}[v^{1-\sigma}]}{\mu_2 \mathbb{E}_\Delta[v^{1-\sigma}]} \cdot \left(\frac{1+\bar{c}_{AV}}{1-R_1^*}\right)^{1-\sigma} \cdot \exp\left((\sigma-1)\cdot \int_0^{R_2^*} \zeta_\Delta(r)^{-1}\, dr\right)$$

while

$$\frac{E(2;\mathcal{C}^{**})}{E(1;\mathcal{C}^{**})} = \frac{\mu_2 Q_\Delta(R_1^*;AV)}{\mu_1 Q(R_2^*;S)}$$
$$= \frac{\mu_2 \mathbb{E}_\Delta[v^{1-\sigma}]}{\mu_1 \mathbb{E}[v^{1-\sigma}]} \cdot \left(\frac{1+\bar{c}_{AV}}{1-R_1^*}\right)^{1-\sigma} \cdot \exp\left((\sigma-1)\cdot \int_0^{R_2^*} \zeta(r)^{-1}\, dr\right)$$

The ratio of these two quantities is therefore

$$\frac{E(1;\mathcal{C}^*)/E(2;\mathcal{C}^*)}{E(2;\mathcal{C}^{**})/E(1;\mathcal{C}^{**})} = \left(\frac{\mu_1 \mathbb{E}[v^{1-\sigma}]}{\mu_2 \mathbb{E}_\Delta[v^{1-\sigma}]}\right)^2 \cdot \exp\left(\sigma\cdot(\sigma-1)\cdot \int_0^{R_2^*} \frac{\Omega_\Delta(r;S)-\Omega(r;S)}{\zeta(r)\zeta_\Delta(r)}\, dr\right).$$

Whenever R_2^* is sufficiently small, we have $\Omega_\Delta(r;S) > \Omega(r;S)$ for all $r \in (0, R_2^*)$. In that case,

$$\frac{E(1;\mathcal{C}^*)/E(2;\mathcal{C}^*)}{E(2;\mathcal{C}^{**})/E(1;\mathcal{C}^{**})} > \left(\frac{\mu_1 \mathbb{E}[v^{1-\sigma}]}{\mu_2 \mathbb{E}_\Delta[v^{1-\sigma}]}\right)^2.$$

Additionally, the assumption that the two categories are balanced is equivalent to $\mu_1 \mathbb{E}[v^{1-\sigma}] = \mu_2 \mathbb{E}_\Delta[v^{1-\sigma}]$. Then the previous inequality implies

$$\frac{E(1;\mathcal{C}^*)/E(2;\mathcal{C}^*)}{E(2;\mathcal{C}^{**})/E(1;\mathcal{C}^{**})} > 1,$$

which is equivalently $E(1;\mathcal{C}^*) > E(2;\mathcal{C}^{**})$.

Define another code $\mathcal{C}^{***} = (\mathcal{K}, \boldsymbol{x}^{**}, \boldsymbol{R}^{***})$, where $\boldsymbol{R}^{***} = (R_2^{***}, R_1^*)$ and $R_2^{***} > R_2^*$ is chosen so that $E(2;\mathcal{C}^{***}) = E(1;\mathcal{C}^*)$. The revenue generated by code $\mathcal{C}^{***}$ is

$$\begin{aligned} E(2;\mathcal{C}^{***})\cdot R_1^* + E(1;\mathcal{C}^{***})\cdot R_2^{***} &> E(2;\mathcal{C}^{***})\cdot R_1^* + E(1;\mathcal{C}^{***})\cdot R_2^* \\ &= E(1;\mathcal{C}^*)\cdot R_1^* + E(2;\mathcal{C}^*)\cdot R_2^*, \end{aligned}$$

so code $\mathcal{C}^{***}$ generates more revenue than the (feasible, by hypothesis) code $\mathcal{C}^*$. It is therefore feasible. And the code $\mathcal{C}^*$ achieves the objective

$$\begin{aligned} V^* &= \mu_1 Q(R_1^*;AV) + \mu_2 Q_\Delta(R_2^*;S) \\ &= \mu_1 Q(R_1^*;AV)\cdot\left(1+\frac{E(2;\mathcal{C}^*)}{E(1;\mathcal{C}^*)}\right), \end{aligned}$$

while the code $\mathcal{C}^{***}$ achieves the objective

$$\begin{aligned}V^{***} &= \mu_2 Q_\Delta(R_1^*; AV) + \mu_1 Q(R_2^{***}; S)\\ &= \mu_2 Q_\Delta(R_1^*; AV) \cdot \left(1 + \frac{E(1; \mathcal{C}^{***})}{E(2; \mathcal{C}^{***})}\right)\\ &= \mu_2 Q_\Delta(R_1^*; AV) \cdot \left(1 + \frac{E(2; \mathcal{C}^{*})}{E(1; \mathcal{C}^{*})}\right).\end{aligned}$$

The hypothesis that the two categories are balanced implies

$$\mu_1 Q(R_1^*; AV) = \mu_1 \mathbb{E}[v^{1-\sigma}] \left(\frac{1+\bar{c}_{AV}}{1-R_1^*}\right)^{1-\sigma} = \mu_2 \mathbb{E}_\Delta[v^{1-\sigma}] \left(\frac{1+\bar{c}_{AV}}{1-R_1^*}\right)^{1-\sigma} = \mu_2 Q_\Delta(R_1^*; AV).$$

Hence, $V^{***} = V^*$. Slightly lowering both net tariff burdens in $\mathcal{C}^{***}$ to preserve expenditure shares yields a code that is feasible (because $\mathcal{C}^{***}$ generated a revenue strictly larger than R) and generates a higher objective than $\mathcal{C}^*$, contradicting the hypothesized optimality of $\mathcal{C}^*$.

E.5 Proof of Proposition 10

Let V^0 be the value of the objective achieved under a $\mathcal{K}^0$-optimal tariff code. Let V^{FB} be the value achieved by an ideal uniform ad valorem tariff raising tariff revenue R, which entails no administrative or classification costs. Since any feasible tariff code entails such costs and/or distorts relative prices, the value of any feasible code is bounded strictly below V^{FB}. In particular, $V^{FB} > V^0$.

Consider the dual problem of maximizing tariff revenue subject to achieving an objective of at least V^0. Let R^{FB} be the tariff revenue raised by an (infeasible) uniform ad valorem tariff that requires no verification costs. (Since a uniform tariff requires no classification costs, R^{FB} does not vary with c_S.) Since such an instrument achieves a value larger than V^0 in the primal problem, it follows that $R^{FB} > R$ in the dual problem.

Now, fix a categorization $\mathcal{K} \neq \mathcal{K}^0$, and suppose that $c_S > (R^{FB} - R)/H(\mathcal{K})$. Then in the dual problem, any feasible tariff code utilizing categorization $\mathcal{K}$ must raise net revenue bounded below R, since refining the partition beyond $\mathcal{K}^0$ cannot do better than raising (gross) tariff revenue from R to R^{FB}. As a result, in the primal problem, any feasible tariff code utilizing categorization $\mathcal{K}$ must achieve a value below V^0. In particular, this is true of any $\mathcal{K}$-optimal code. Let $\hat{c}_S \equiv \max_{\mathcal{K}\neq\mathcal{K}^0}(R^{FB} - R)/H(\mathcal{K})$. Since the set of categorizations is finite, $\min_{\mathcal{K}\neq\mathcal{K}^0} H(\mathcal{K}) > 0$ and therefore $\hat{c}_S$ is finite. Whenever $c_S > \hat{c}_S$, it must be that every optimal tariff code uses the categorization $\mathcal{K}^0$.

In the opposite direction, fix a non-degenerate model. Let $V^*(\mathcal{K}; c_S)$ be the objective achieved by a $\mathcal{K}$-optimal tariff code when classification costs are c_S. By assumption, every optimal tariff code involves a distinct tariff regime and/or rate for each product when $c_S = 0$. An immediate implication is that $V^*(\mathcal{K}^{FB}; 0) > V^*(\mathcal{K}; 0)$ for every $\mathcal{K} \neq \mathcal{K}^{FB}$. By the maximum theorem, $V^*(\mathcal{K}; c_S)$ is continuous in c_S for every $\mathcal{K}$. As a result, for each $\mathcal{K}$ there exists a $\underline{c}_S(\mathcal{K}) > 0$ such that $V^*(\mathcal{K}^{FB}; c_S) > V^*(\mathcal{K}; c_S)$ when $c_S < \underline{c}_S(\mathcal{K})$. Let $\underline{c}_S \equiv \min_{\mathcal{K} \neq \mathcal{K}^{FB}} \underline{c}_S(\mathcal{K})$. Finiteness of the set of categorizations ensures that $\underline{c}_S > 0$. Then $V^*(\mathcal{K}^{FB}; c_S) > \max_{\mathcal{K} \neq \mathcal{K}^{FB}} V^*(\mathcal{K}; c_S)$ whenever $c_S < \underline{c}_S$, implying that every optimal tariff code uses the finest possible categorization for small c_S.

E.6 Proof of Proposition 11

Lemma D.3 implies that, for each product $g \in \mathcal{G}$, a specific tariff is optimal for a category comprising only that good if and only if it is optimal in the one-category problem involving the value distribution for that product and the same net tariff burden. The proof of Proposition 4 therefore establishes that, for each product $g \in \mathcal{G}$, it is (uniquely) optimal to impose a specific tariff for all net tariff burdens whenever

$$\frac{\mathbb{E}_g[v]}{\left(\mathbb{E}_g[v^{1-\sigma}]\right)^{1/(1-\sigma)}} \leq 1 + \frac{c_{AV}}{\kappa} \cdot \frac{\mathbb{E}_g[v^{-\sigma}]}{\mathbb{E}_g[v^{1-\sigma}]}.$$

This bound is equivalently

$$\frac{c_{AV}}{\kappa} \geq \left(\mathbb{E}_g[v] - \left(\mathbb{E}_g[v^{1-\sigma}]\right)^{1/(1-\sigma)}\right) \cdot \frac{\left(\mathbb{E}_g[v^{1-\sigma}]\right)^{-\sigma/(1-\sigma)}}{\mathbb{E}_g[v^{-\sigma}]}.$$

Since $\sigma > 1$, the function $v^{1-\sigma}$ is a concave transformation of $v^{-\sigma}$. Hence, by Jensen's inequality,

$$\left(\mathbb{E}_g[v^{1-\sigma}]\right)^{-\sigma/(1-\sigma)} \leq \mathbb{E}_g[v^{-\sigma}].$$

A sufficient condition for the bound above is therefore

$$\frac{c_{AV}}{\kappa} \geq \mathbb{E}_g[v] - \left(\mathbb{E}_g[v^{1-\sigma}]\right)^{1/(1-\sigma)}.$$

It follows that, under any δ-homogeneous set of products, at least a fraction $1-\delta$ of varieties must be assigned a specific tariff in any $\mathcal{K}^{FB}$-optimal tariff code, no matter the optimal net tariff burdens—and, therefore, no matter the value of c_S.

Now suppose, by way of contradiction, that there exists a sequence of classification costs declining to zero, and corresponding optimal tariff codes $\mathcal{C}_1^*, \mathcal{C}_2^*, \ldots$, such that each code uses specific tariffs for a fraction less than $1-\delta$ of varieties. Since the set of tariff codes is compact, there exists a convergent subsequence of these codes with limit code $\mathcal{C}_\infty^*$. And since the set of categorizations and regimes is discrete, this subsequence must eventually exhibit constant categorizations and tariff regimes. By assumption, all codes along the sequence use specific tariffs for a fraction less than $1-\delta$ of varieties. Since this property applies to the eventually constant portion of the subsequence, it must also apply to the limit code $\mathcal{C}_\infty^*$. By the maximum theorem, the set of maximizers is upper hemicontinuous in c_S, and so $\mathcal{C}^*_\infty$ is an optimal tariff code in the model with $c_S = 0$. But since categories can be disaggregated for free when $c_S = 0$, this code can be used to construct another optimal tariff code with categorization $\mathcal{K}^{FB}$ and a fraction less than $1-\delta$ of varieties assigned specific tariffs. This conclusion contradicts δ-homogeneity of the set of products. So it must be that every optimal tariff code uses specific tariffs for at least $1-\delta$ fraction of varieties when c_S is small.

E.7 Proof of Proposition 12

We first establish that, for sufficiently small c_{AV}, every optimal tariff code must utilize a single category. By way of contradiction, suppose that there exists a sequence of verification costs $c^1_{AV}, c^2_{AV}, \ldots$ declining to zero, and an associated sequence of optimal tariff codes $\mathcal{C}^*_1, \mathcal{C}^*_2, \ldots$, such that $\mathcal{K}^*_i \neq \mathcal{K}^0$ for every i. Let V^*_i be the value of the objective achieved by $\mathcal{C}^*_i$. Since the set of tariff codes is compact, we may assume the sequence converges, passing if necessary to a subsequence. Since the set of categorizations is discrete, along this subsequence the categorization is eventually constant, say equal to $\mathcal{K}^\infty$. Let V^{**} be the value of the objective achieved by an (infeasible) uniform ad valorem tariff requiring no verification costs and raising tariff revenue (gross of classification costs) $R + c_S \cdot H(\mathcal{K}^\infty)$. Since any feasible tariff code entails relative price distortions and/or verification costs, it must be that eventually $V^{**} \geq V^*_i$.

Now, let $\tilde{V}_i$ be the value of the objective achieved by a (feasible) uniform ad valorem tariff which raises net tariff revenue R under verification costs c^i_{AV}. And let $\hat{V}$ be the objective achieved by an (infeasible) uniform ad valorem tariff requiring no verification costs and raising net tariff revenue R. Note that $\hat{V} > V^{**}$, because these two values are achieved under the same verification costs using the same tariff structure, while the latter value is depressed by a larger net tariff goal. The proof of Proposition 4 derives an explicit expression for the objective under a uniform ad valorem tariff raising a fixed tariff revenue, which by inspection is continuous in the verification cost. Hence, $\tilde{V}_i \to \hat{V}$. As an implication, $\tilde{V}_i > V^*_i$ for sufficiently large i. This conclusion contradicts the hypothesized optimality of the tariff codes $\mathcal{C}^*_i$.

We next establish that the $\mathcal{K}^0$-optimal tariff code utilizes an ad valorem tariff for sufficiently small c_{AV}. Let V^S be the value achieved by a uniform specific tariff raising tariff revenue R, with $V^{AV}(c_{AV})$ defined similarly for an ad valorem tariff. The proof of Proposition 4 established that

$$\frac{V^S}{V^{AV}(c_{AV})} = \frac{\Phi(T^*(R;S))^{1-\sigma}}{(1+\bar{c}_{AV})^{1-\sigma} \cdot \mathbb{E}[v^{1-\sigma}]}$$

for a function Φ that is increasing and independent of c_{AV} and that satisfies $\Phi(0)^{1-\sigma} =$

$\mathbb{E}[v^{1-\sigma}]$. Since $T^*(R;S) > 0$ is also independent of c_{AV}, we have

$$\lim_{c_{AV}\to 0} \frac{V^S}{V^{AV}(c_{AV})} = \frac{\Phi(T^*(R;S))^{1-\sigma}}{\mathbb{E}[v^{1-\sigma}]} < 1.$$

In other words, for sufficiently small c_{AV} the unique $\mathcal{K}^0$-optimal tariff code involves a single category which is assessed an ad valorem tariff.

Finally, we consider the limit of large c_{AV}. Proposition 4 implies that, for each categorization $\mathcal{K}$ and category $j \in \mathcal{J}$, there exists a $\hat{c}_{AV}(\mathcal{K}, j) < \infty$ such that a specific tariff is (uniquely) optimal no matter the net tariff burden in category j if $c_{AV} \geq \hat{c}_{AV}(\mathcal{K}, j)$. Letting $\hat{c}_{AV}(\mathcal{K}) \equiv \max_{j\in\mathcal{J}} \hat{c}_{AV}(\mathcal{K}, j)$, it follows from Lemma D.3 that any $\mathcal{K}$-optimal tariff code uses only specific tariffs whenever $c_{AV} \geq \hat{c}_{AV}(\mathcal{K})$. Since categorizations are finite, $\hat{c}_{AV}(\mathcal{K}) < \infty$. Let $\hat{c}_{AV} \equiv \max_{\mathcal{K}} \hat{c}_{AV}(\mathcal{K})$. Finiteness of the set of possible categorizations implies that $\hat{c}_{AV} < \infty$, and an optimal tariff code uses only specific tariffs whenever $c_{AV} \geq \bar{c}_{AV}$.

Appendix References

Acosta, Miguel, and Lydia Cox. 2024. "The Regressive Nature of the U.S. Tariff Code: Origins and Implications." *Working Paper.*

Banner, Stuart. 1998. "The Origin of the New York Stock Exchange, 1791–1860." *The Journal of Legal Studies*, 27(1): 113–140.

Betz, Timm. 2019. "Tariff evasion and trade policies." *International Studies Quarterly*, 63(2): 380–393.

Bouët, Antoine, Yvan Decreux, Lionel Fontagné, Sébastien Jean, and David Laborde. 2008. "Assessing Applied Protection across the World." *Review of International Economics*, 16(5): 850–863.

Broda, Christian, and David E Weinstein. 2006. "Globalization and the Gains from Variety." *The Quarterly Journal of Economics*, 121(2): 541–585.

Customs, U.S., and Border Protection. 2024. "Customs Brokers." *https://www.cbp.gov/trade/programs-administration/customs-brokers.*

Dunlavy, Colleen A. 2024. *Small, Medium, Large: How Government Made the U.S. into a Manufacturing Powerhouse.* Polity.

Farhad, Mohammad, Jakob B. Madsen, Michael Jetter, and Andrew Williams. 2024. "The Mystery of Misreported Trade: A Novel Method to Identify Distorted Trade Figures." *The World Economy*, 47(2): 748–778.

Fort, Teresa C., and Shawn D. Klimek. 2018. "The Effects of Industry Classification Changes on US Employment Composition." *Center for Economic Studies Working Paper 18-28.*

Grant, Matthew. 2025. "Where Do the Data Come From? Endogenous Classification in Administrative Data." *Working Paper.*

Greenland, Andrew, James Lake, and John Lopresti. 2026. "AVE Tariffs are a Misleading Measure of Tariff Policy." *Working Paper.*

Grossman, Gene M., and Elhanan Helpman. 1994. "Protection for Sale." *The American Economic Review*, 84(4): 833–850.

Hamilton, Alexander. 1791. "Alexander Hamilton's Final Version of the Report on the Subject of Manufactures." *The Papers of Alexander Hamilton*, 10: 230–340. Edited by Harold C. Syrett. New York: Columbia University Press, 1966.

Jones, Andrew A. 1835. *Jones's Digest: Being a Particular and Detailed Account of the Duties Performed by the Various Officers Belonging to the Custom-House Departments of the United States.* New York:G. F. Hopkins & Son, Printers.

Jones, Andrew A. 1854. *The Revenue Book: Containing the New Tariff of 1846, Together with the Tariff of 1842.* New York:Geo. H. Bell.

Mastrorocco, Nicola, and Edoardo Teso. 2026. "State Capacity as an Organizational Problem." *Review of Economic Studies.* Forthcoming.

Pierce, Justin R., and Peter K. Schott. 2012. "A Concordance Between Ten-Digit U.S. Harmonized System Codes and SIC/NAICS Product Classes and Industries." *Journal of Economic and Social Measurement*, 37(1-2): 61–96.

Political Risk Services (PRS) Group. 2022. "International Country Risk Guide."

Rauch, James E. 1999. "Networks versus markets in international trade." *Journal of International Economics*, 48(1): 7–35.

Taylor, Charles Henry. 1917. *History of the Board of Trade of the City of Chicago: Volume 1.* Illinois:R. O. Law.

Teti, Feodora A. 2020. "30 years of trade policy: Evidence from 5.7 billion tariffs." *ifo Working Paper.*

Teti, Feodora A. 2024. "Missing Tariffs." Working Paper.

The Merchants' and Bankers' Almanac. 1869. "The Merchants' and Bankers' Almanac." *New York: Office of the Bankers' Magazine and Statistical Register.*

United States Treasury Department. 1871. *Annual Report of the Bureau of Statistics on the Commerce and Navigation of the United States ("Foreign Commerce and Navigation of the United States").* Washington, DC:Government Printing Office.

United States Treasury Department. 1888. *The Existing Tariff on Imports into the United States, etc.; Together with Comparative Tables of Present and Past Tariffs, Under the Several Acts from 1789 to 1883, Inclusive.* Washington, DC:Government Printing Office.

U.S. Congress. 1832. "An Act to Alter and Amend the Several Acts Imposing Duties on Imports." Act of July 14, 1832, ch. 227, 4 Stat. 583.

Walker, R.J. 1845. *Report from the Secretary of the Treasury, on the state of the finances. E. Detailed statement of the several tariffs enacted by Congress, and of the duties prescribed by each.* Washington, DC:John C. Rives.

Figure A.1: Manufactured Wool Tariffs, 1828

Second. On manufactures of wool, or of which wool shall be a component part, (except carpetings, blankets, worsted stuff goods, bombazines, hosiery, mits, gloves, caps, and bindings,) the actual value of which, at the place whence imported, shall not exceed fifty cents the square yard, shall be deemed to have cost fifty cents the square yard and be charged thereon with a duty of forty per centum ad valorem, until the thirtieth day of June, eighteen hundred and twenty-nine, and from that time a duty of forty-five per centum ad valorem: *Provided*, That on all manufactures of wool, except flannels and baizes, the actual value of which, at the place whence imported, shall not exceed thirty-three and one third cents per square yard, shall pay fourteen cents per square yard.

Third. On all manufactures of wool, or of which wool shall be a component part, except as aforesaid, the actual value of which, at the place whence imported, shall exceed fifty cents the square yard, and shall not exceed one dollar the square yard, shall be deemed to have cost one dollar the square yard, and be charged thereon with a duty of forty per centum ad valorem, until the thirtieth day of June, eighteen hundred and twenty-nine, and from that time a duty of forty-five per centum ad valorem.

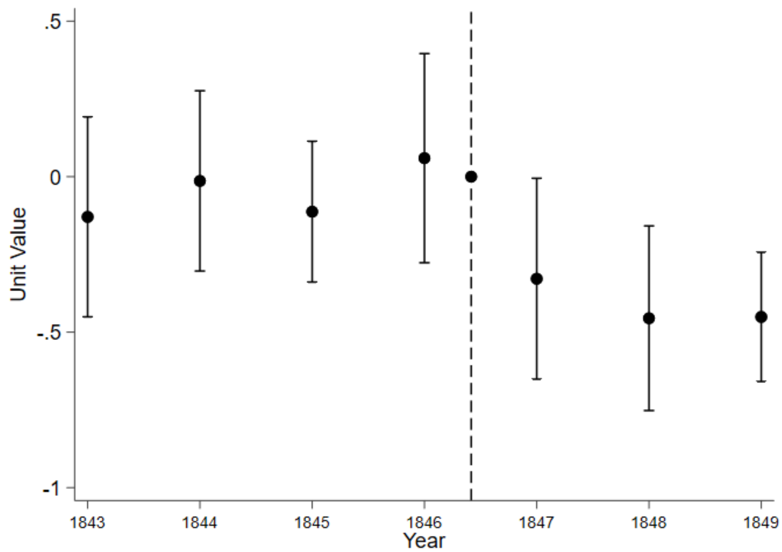
Fourth. On all manufactures of wool, or of which wool shall be a component part, except as aforesaid, the actual value of which, at the place whence imported, shall exceed one dollar the square yard, and shall not exceed two dollars and fifty cents the square yard, shall be deemed to have cost two dollars and fifty cents the square yard, and be charged with a duty thereon of forty per centum ad valorem, until the thirtieth day of June, eighteen hundred and twenty-nine, and from that time a duty of forty-five per centum ad valorem.

Fifth. All manufactures of wool, or of which wool shall be a component part, except as aforesaid, the actual value of which, at the place whence imported, shall exceed two dollars and fifty cents the square yard, and shall not exceed four dollars the square yard, shall be deemed to have cost, at the place whence imported, four dollars the square yard, and a duty of forty per cent. ad valorem, shall be levied, collected, and paid, on such valuation, until the thirtieth day of June, one thousand eight hundred and twenty-nine, and from that time a duty of forty-five per centum ad valorem.

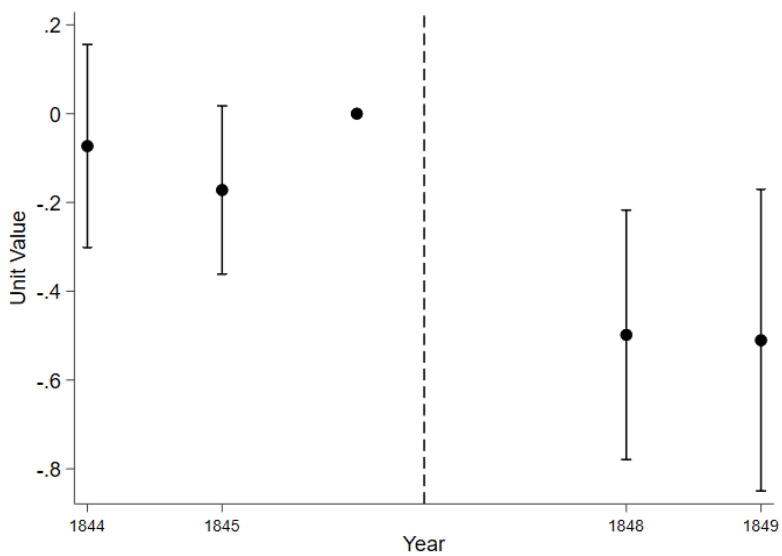
Notes: This figure shows the language of the May 19, 1828 Tariff act, relating to manufactured wool.

Figure A.2: Unit Values: Event Study Around Removing All Specific Tariffs

Panel A: Full Sample



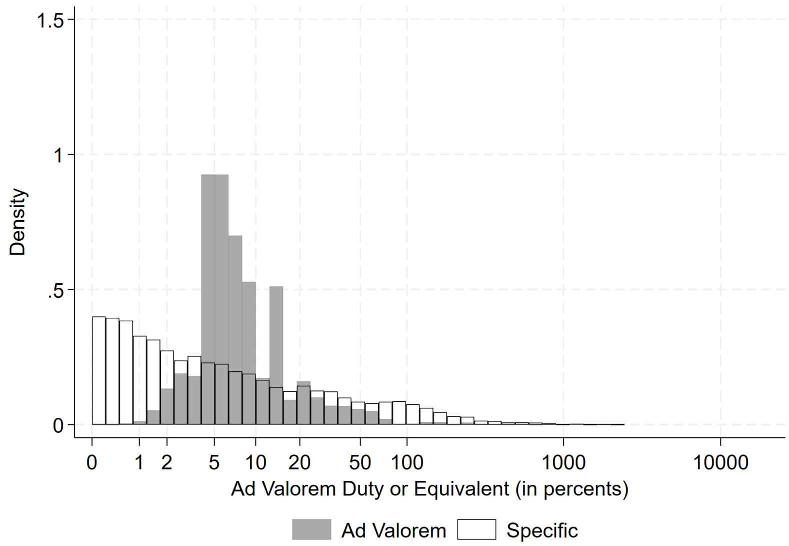
Panel B: Full Years Only



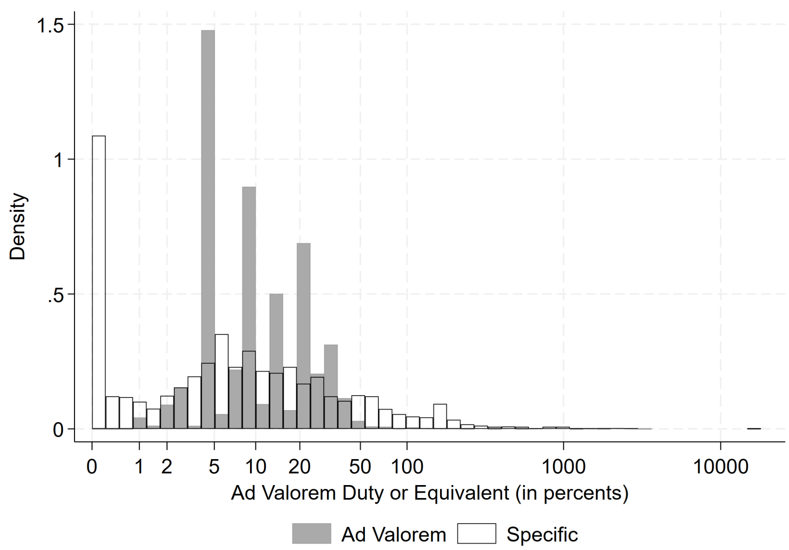
Notes: This figure plots the coefficients of a regression comparing the unit values for the initially-specific goods listed in the 1849 Annual Report of the Secretary of the Treasury for 1843–1849 relative to the goods without specific tariffs for which unit values were consistently reported (two types of coffee, two types of tea, and wool). The vertical line represents the implementation of the Walker Tariff in 1846, which removed all specific tariffs. Panel A shows data from all of the import statistics 1843–1849, and Panel B shows data only for the import statistics with 12 months of reporting (dropping 1843 and two reports for 1847, one for each tariff act). Results clustered by good. Data: *Commerce and Navigation of the United States.*

Figure A.3: Modern Distribution of Tariff Rates by Tariff Type

Panel A: OECD



Panel B: Non-OECD



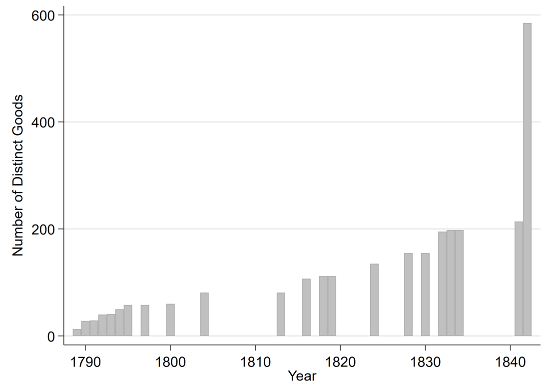
Notes: This figure plots the distribution of ad valorem equivalent tariff rates by specific and ad valorem tariffs. Panel A plots the distributions for countries in the OECD, and Panel B plots the distributions for countries not in the OECD. We only consider specific and ad valorem tariffs (dropping goods on the free list, and dropping extremely large ad valorem equivalent rates, where ln(ave)> 10). All countries are included in their most recent year. Source: WTO Integrated Database and UN Trade Analysis Information System.

Figure A.4: Unit Values for Wool

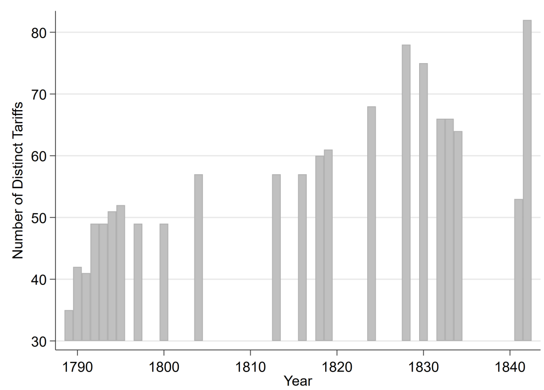
Notes: This figure plots the average reported unit values for unmanufactured wool. Before 1846, there were two tariffs, depending on the price: if the wool was seven cents or under per pound, the duty was 5%; otherwise the duty was three cents per pound plus 30%. The average ad valorem equivalent tariff was 38%. After 1846, all unmanufactured wool had an ad valorem tariff of 30%. Data: *Commerce and Navigation of the United States* and *Tariff Acts of the United States*

Figure A.5: Increasing Tariff Complexity Over Time, Non-Manufactured Goods

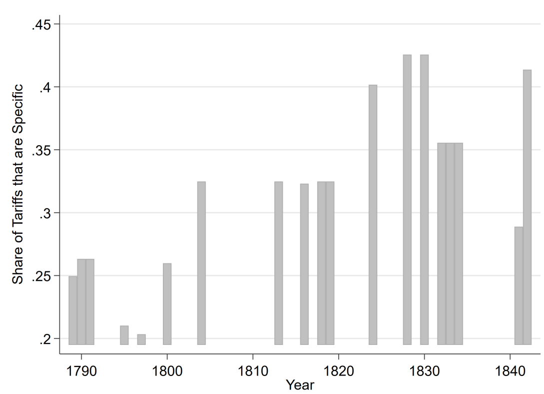
Panel A: Number of Goods Defined by Each Year



Panel B: Number of Distinct Tariffs Over Time



Panel C: Share of Goods with Specific Tariffs Over Time



Notes: This figure plots the evolution of the complexity of the American tariff code over time. In this figure, we limit the sample to goods that are not manufactured. Panel A plots the number of goods defined for each year t. The set of goods expands over time. A good is defined as a set of varieties such that: (i) in every year from 1789 through t, all varieties in the set face the same tariff; and (ii) the set is physically distinct from any other set that follows the same sequence of tariffs over the time period. Panel B plots the number of distinct tariffs that are in effect for each year. Panel C plots the share of goods that have a specific tariff. Source: *Tariff Acts of the United States.*

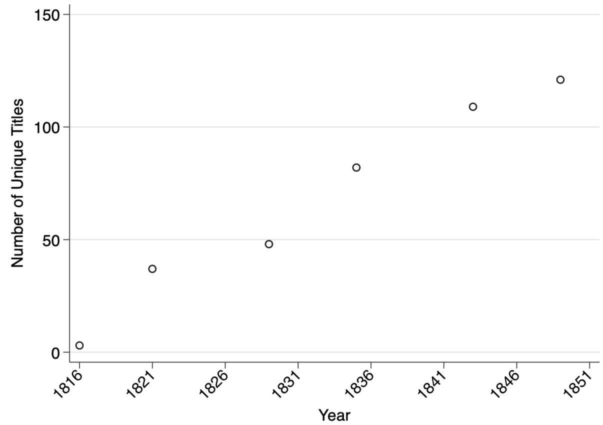
Figure A.6: Directory of Customs Officials in Boston, 1839

Names and Offices.	Where Employed.	Where born.	Compen-sation.
Boston and Charlestown.			
George Bancroft...Collector & Sup't of Light Houses.	Boston..........	Mass.	4400 00
Adams Bailey, Dept. Collector	.......do.......	..do..	1500 00
E. L. Frothingham, Cash Clerk and Book keeper..	.......do.......	..do..	1300 00
John Bingham..Bond Acc'ant	.......do.......	Conn.	1300 00
W. A. Wellman Clearance Cl'k	.......do.......	Mass.	1000 00
R. M. Gay....Debenture Cl'k	.......do.......	Maine	1000 00
S. Andrews......Marine Cl'k	.......do.......	Mass.	900 00
J. T. Prince.......Bond Cl'k	.......do.......	Maine	800 00
W. Alline........Impost Cl'k	.......do.......	Mass.	750 00
H. D. Clary, Ass't Impost Cl'k	.......do.......	..do..	700 00
N. M. Cutler..Ass't Deb. Cl'k	.......do.......	Maine	700 00
G. B. Wellman Coastwise Cl'k	.......do.......	Mass.	700 00
W. Palfrey..Ass't Clear. Cl'k	.......do.......	..do..	704 00
S. Draper, Statistical B'k k'per	.......do.......	..do..	704 00
R. Abell.....Temporary Cl'k	.......do.......	N. H.	184 00
A. H. Ward, Weig'r & Gauger	.......do.......	Mass.	1500 00
John Champney.........do..	.......do.......	..do..	1500 00
S. Ward.. Weigher & Gauger	Boston..........	Mass.	1500 00
Luther Hamilton........do..	.......do.......	..do..	1500 00
Samuel Walker.........do..	.......do.......	..do..	1500 00
Chauncey Clark........do..	.......do.......	..do..	1500 00
Fred. Robinson.........do..	.......do.......	N. H.	1500 00
J. Grafton, Measurer of Salt, Coal, &c................	.......do.......	Mass.	1500 00
Andrew Green..........do..	.......do.......	..do..	1500 00
Joseph Hall............do..	.......do.......	..do..	1500 00
Wm. B. Pike...........do..	.......do.......	..do..	1500 00
N. Hawthorne..........do..	.......do.......	..do..	1500 00

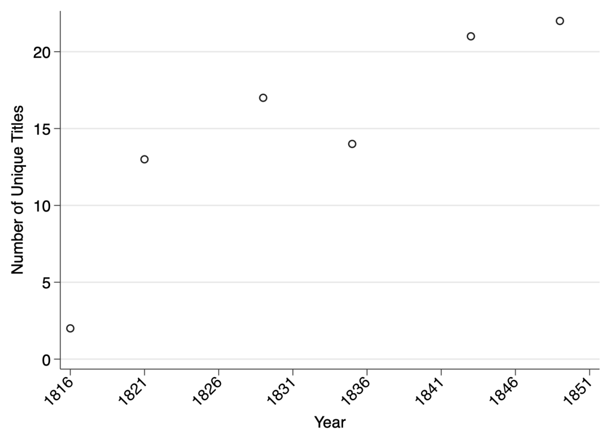
Notes: This figure shows excerpted scans from the 1839 *Official Register of the United States*, showing some of the officials in the Boston Customs House. Many had prominent careers beyond Customs. George Bancroft, the Collector and Superintendent of Light Houses, was also a historian and politician. Nathaniel Hawthorne, a Measurer of Salt & Coal, was also a writer.

Figure A.7: Alternative Measures of Customs House Expansion

Panel A: All Customs Occupations Over Time



Panel B: Core Customs Occupations Over Time

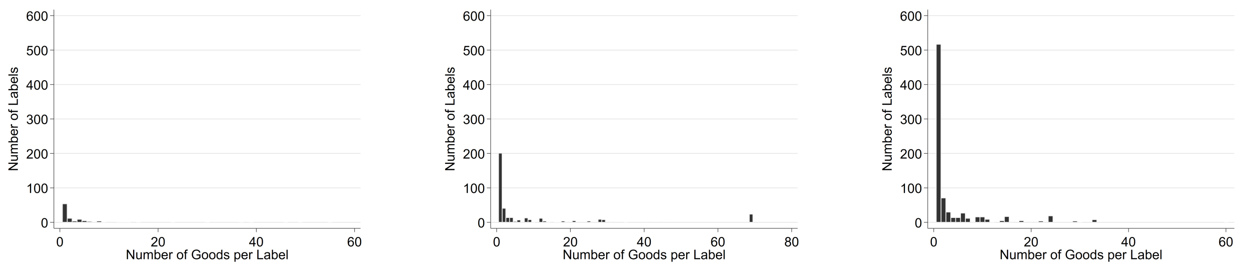


Notes: This figure plots the number of distinct occupational titles at the Customs house. Panel A includes any distinct occupational title. Panel B only focuses on core tasks: those engaged in collections, inspections, quantification, appraisal/examination, marking, and proving.

Figure A.8: Concentration of Groupings Over Time

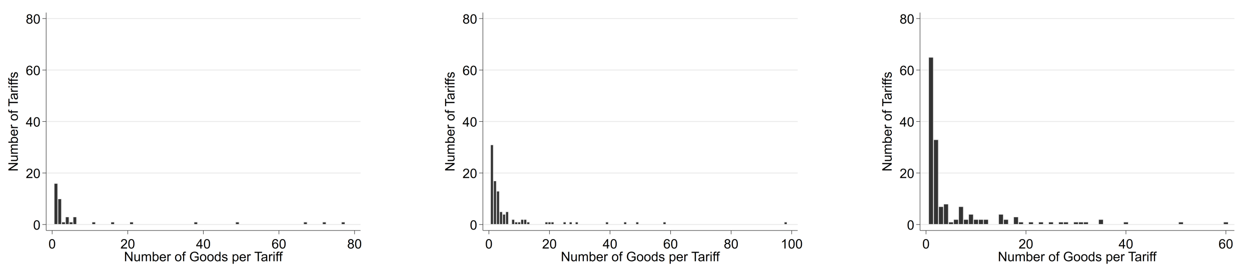
Panel A: Distribution of Goods per Label

1789 1824 1842



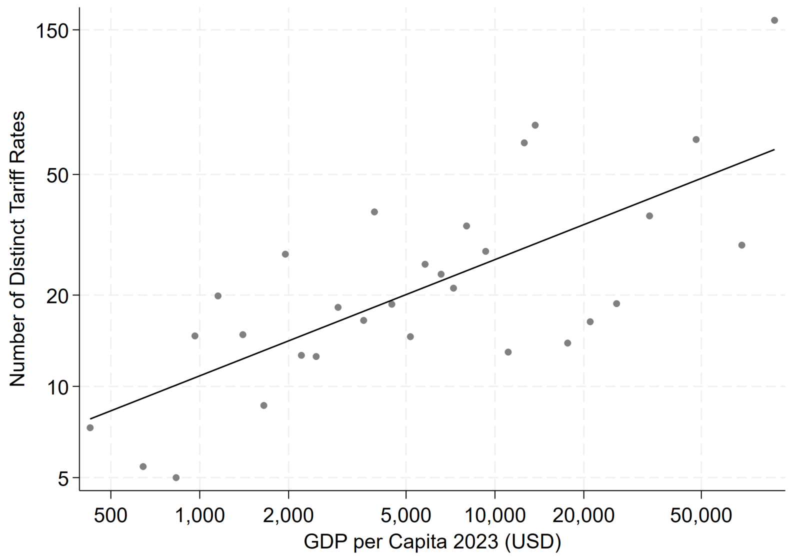
Panel B: Distribution of Goods per Tariff Rate

1789 1824 1842



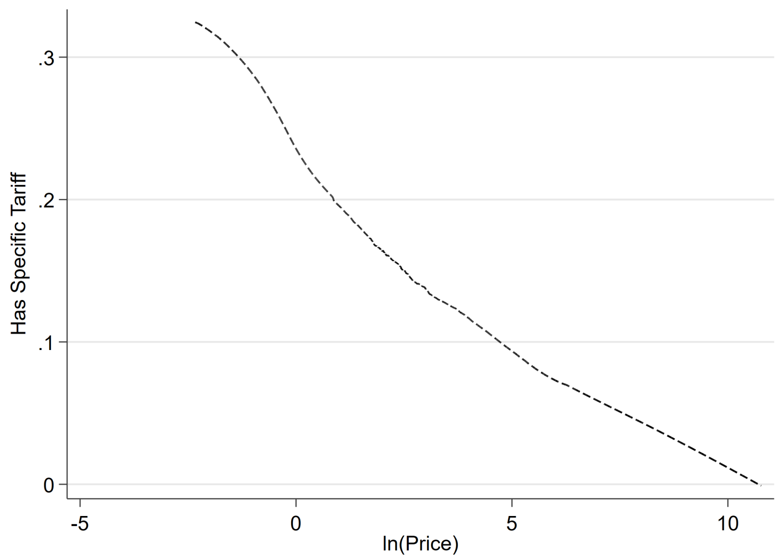
Notes: This figure plots the evolution of the disaggregation of the American tariff code over time, by plotting the number of goods associated with each grouping. Panel A shows the distribution of how many goods are associated with each label, and Panel B shows the distribution of how many goods are associated with each tariff rate, in 1789, 1824, and 1842. In 1789, there are 920 goods with the same label and tariff ("All other goods, wares, and merchandise," with a 5% tariff), which are not shown in the graphs. Source: *Tariff Acts of the United States.*

Figure A.9: Development and Distinct Tariff Rates



Notes: This figure shows a binscatter of the relationship between modern development and the complexity of the tariff code. Source: WTO Integrated Database and WDI.

Figure A.10: 1980s American Imports: Lowess of Unit Values and Specificity



Notes: This figure shows the relationship between unit values, ad valorem equivalents, and specificity in American imports 1980-1985. The independent variable is ln(Price) and the outcome is if an import has a specific tariff. We only include imports whose ad valorem equivalent tariff is below the maximum ad valorem rate, and we trim unit values at the 1% extremes. Ad valorem equivalents are imputed using the Census' estimated collected tariffs. Each observation is a TSUS 5-digit import, averaged across all years. Source: US Census Annual Import Data Bank (IDB).

Figure A.11: 1980s American Imports: Heatmap of Unit Values, Ad Valorem Equivalents, and Specificity

Notes: This figure shows the relationship between unit values, ad valorem equivalents, and specificity in American imports 1980-1985. Ad valorem equivalents are imputed using the Census' estimated collected tariffs. Each observation is a TSUS 5-digit import, averaged across all years. Orange bins have a higher share of specific tariffs, blue bins have a higher share of ad valorem tariffs, and more opaque bins have more imports (in dollars). Source: US Census Annual Import Data Bank (IDB).

Table A.1: Schematic of Moving from Label to Goods

Panel A. Wool Flannels

Year	Wool Flannels < $0.50	Wool Flannels $0.50–1.00	Wool Flannels $1.00–2.00	Wool Flannels $2.00–4.00	Wool Flannels > $4.00
1789	All other goods				
1790	All other goods				
1791					
1794					
1816	Wool Manufactures				
1824	Wool Manufactures Over 33 1/3 cents a square yard and all Woolen Flannels				
1828	Wool Mfg: < $0.50	$0.50–1.00	$1.00–2.00	$2.00–4.00	>$4.00
1832	Flannels				
1842	Flannels, not cotton				

Panel B. Cotton Flannels

Year	Cotton Flannel, uncolored < $0.20	Cotton Flannel, uncolored $0.20–0.25	Cotton Flannel, colored < $0.25	Cotton Flannel, uncolored $0.25–0.30	Cotton Flannel, colored $0.25–0.30	Cotton Flannel, $0.30–0.35	Cotton Flannel, ≥ $0.35
1789	All other goods						
1790	All other goods						
1791	Cotton Manufactures						
1794	Cotton Manufactures						
1816	Cotton Manufactures < $0.25			Cotton Manufactures			
1824	Cotton Manufactures < $0.30					Cotton Manufactures	
1828	Cotton Cloth < $0.35						
1832	Flannels						
1842	Cotton Mfg uncolored < $0.20	Cotton Mfg other	Cotton Mfg colored < $0.30	Cotton Mfg other	Cotton Mfg colored < $0.30	Cotton Mfg other	Cotton Mfg other

Notes: This table shows how we created twelve flannel goods from the various relevant labels that appeared in the *Tariff Acts of the United States*, 1789–1842. The labels shown are broadly the text that appears in the acts themselves (though we compressed the language, for space). See Figure A.1 for examples of the labels that appeared in 1828 related to manufactures of wool. We manually linked the text of the acts in order to create *intersections*: each good has a unique path of tariffs. Every label links to at least one good in every year, and every good links to no more than one tariff rate. The horizontal solid lines imply that the good is associated with a new label, while the horizontal dashed lines imply that the old label (and tariff) continues to apply. All prices refer to dollars per square yard. We only include years in which at least one flannel good is implicated.

Table A.2: 1980s American Imports: Specificity and Price Dispersion Across Time

	(1)	(2)
Price Dispersion Over Time	-0.032	0.0035
	(0.011)	(0.0099)
Mean Outcome	0.20	0.20
Number of Goods	3908	3908
Number of Observations	3917	3917
4-Digit FEs		✓
Sample	Pooled	Pooled

Notes: This table shows the relationship between price dispersion across time in American imports 1980-1985 and tariff specificity. Each observation is a TSUS 5-digit import, and the outcome is if it has a specific tariff. Price dispersion is calculated as the dispersion of the log (within-year-average) price of the good across time. Column 2 additionally includes fixed effects for each 4-digit TSUS code. All regressions are clustered by 5-digit TSUS code. Source: US Census Annual Import Data Bank (IDB).

Table A.3: Tariff Regime and Reported Total Values

	(1)	(2)	(3)	(4)
Post × Specific in 1842	-0.24	-0.24	0.035	0.035
	(0.15)	(0.15)	(0.071)	(0.071)
Include Wool		✓		✓
Full Years Only			✓	✓
Number of Goods	17	18	17	18
Number of Observations	133	141	84	89

Notes: This table shows the effect of switching from specific to ad valorem tariffs on reported total values. We use the initially-specific goods listed in the 1849 Annual Report of the Secretary of the Treasury for 1843–1849, and compare them to the goods without specific tariffs for which unit values were reported (two types of coffee, two types of tea, and wool). Because pre-1846 wools of different values had different tariffs, the odd columns exclude them from the control group. Columns 3 and 4 show results only for the import statistics with 12 months of reporting (dropping 1843 and two reports for 1847, one for each tariff act). The change in the ad valorem equivalent for each good-year is calculated as its ad valorem tariff under the 1846 act minus the ad valorem equivalent that the good would have faced under the 1842 tariff act. All regressions include good and year fixed effects, and standard errors clustered by good. Source: *Commerce and Navigation of the United States* and *Annual Reports of the Secretary of the Treasury on the State of the Finances.*

Table A.4: Singleton Labels Predict Specific Tariffs, Robustness to Potential Industrial Policy

Panel A. Non-Manufactured Goods

	(1)	(2)	(3)	(4)
Singleton Label	0.065	0.082	0.10	0.053
	(0.013)	(0.013)	(0.014)	(0.046)
Mean Outcome	0.12	0.12	0.12	0.49
Number of Goods	581	581	581	489
Number of Observations	10459	10459	10459	489
Specification	Year FE	Year × SITC FE	Year & Good FE	Only 1842

Panel B. Goods Without Splits By Price

	(1)	(2)	(3)	(4)
Singleton Label	0.077	0.082	0.11	0.24
	(0.010)	(0.0094)	(0.011)	(0.034)
Mean Outcome	0.11	0.11	0.11	0.48
Number of Goods	899	899	899	801
Number of Observations	17297	17297	17297	801
Specification	Year FE	Year × SITC FE	Year & Good FE	Only 1842

Panel C. Industries Not Mentioned in Hamilton (1791)

	(1)	(2)	(3)	(4)
Singleton Label	0.031	0.072	0.080	-0.017
	(0.014)	(0.012)	(0.013)	(0.047)
Mean Outcome	0.11	0.11	0.11	0.42
Number of Goods	541	541	541	453
Number of Observations	10092	10092	10092	453
Specification	Year FE	Year × SITC FE	Year & Good FE	Only 1842

Notes: This table shows the relationship between the aggregation in the tariff code and if a good has a specific tariff. The sample in Panel A is only goods that are not manufactured. Panel B drops all goods that ever have a different tariff as a function of their price (such as unmanufactured wool in 1842, which has a different tariff depending on if it is valued at above or below seven cents a pound). Panel C drops all goods whose industry was mentioned in Hamilton (1791) *Report on Manufactures*. Each good is included only in the years in which it is covered by a tariff and it has either a specific or ad valorem tariff. Each observation is a good-by-year. The outcome for all regressions is if the good has a specific tariff, and the independent variable is if the label for the good is a singleton (only covers one good). All regressions control for year. Column 2 additionally controls for year interacted with three digit SITC classifications. Column 3 controls for good fixed effects. Column 4 only includes 1842. All regressions cluster by good. Source: *Tariff Acts of the United States* and *Commerce and Navigation of the United States*.

Table A.5: Homogeneity Predicts Specific Tariffs, Robustness to Potential Industrial Policy

Panel A. Non-Manufactured Goods

	(1)	(2)	(3)
Homogeneous	0.17	0.16	0.44
	(0.010)	(0.042)	(0.041)
Mean Outcome	0.12	0.12	0.49
Number of Goods	581	581	489
Number of Observations	10459	10459	489
Specification	Year FE	Year × SITC FE	Only 1842

Panel B. Goods Without Splits By Price

	(1)	(2)	(3)
Homogeneous	0.15	0.16	0.33
	(0.0095)	(0.041)	(0.036)
Mean Outcome	0.11	0.11	0.48
Number of Goods	899	899	801
Number of Observations	17297	17297	801
Specification	Year FE	Year × SITC FE	Only 1842

Panel C. Industries Not Mentioned in Hamilton (1791)

	(1)	(2)	(3)
Homogeneous	0.18	0.17	0.51
	(0.011)	(0.053)	(0.041)
Mean Outcome	0.11	0.11	0.42
Number of Goods	541	541	453
Number of Observations	10092	10092	453
Specification	Year FE	Year × SITC FE	Only 1842

Notes: This table shows the relationship between price dispersion across varieties of a good and if it has a specific tariff. The sample in Panel A is only goods that are not manufactured. Panel B drops all goods that ever have a different tariff as a function of their price (such as unmanufactured wool in 1842, which has a different tariff depending on if it is valued at above or below seven cents a pound). Panel C drops all goods whose industry was mentioned in Hamilton (1791) *Report on Manufactures*. Each good is included only in the years in which it is covered by a tariff and it has either a specific or ad valorem tariff. Each observation is a good-by-year. The outcome for all regressions is if the good has a specific tariff, and the independent variable is if the good is mentioned as a commodity by the *Executive Documents* of the House of Representatives. All regressions control for year. Column 2 additionally controls for year interacted with two digit SITC classifications. Column 3 only includes 1842. All regressions cluster by good. Source: *Tariff Acts of the United States*.

Table A.6: Modern Customs House Capacity Predicts More Complex Tariffs and More Specific Tariffs, Only For Broadly-Traded Goods

Panel A. Tariff Complexity

	ln(Number of Distinct Tariff Rates)		Share of HS 6-digit Codes with Variation in Tariffs	
	(1)	(2)	(3)	(4)
Overall Export Misreporting Index	-0.35 (0.085)		-0.018 (0.0086)	
Overall Trade Misreporting Index		-0.40 (0.093)		-0.019 (0.0094)
Mean Outcome	2.85	2.85	0.13	0.13
Number of Observations	125	125	125	125

Panel B. Specific Tariffs

	Share of Tariff Lines with Specific Tariffs		Average of HS 6-digit Share with Specific Tariffs	
	(1)	(2)	(3)	(4)
Overall Export Misreporting Index	-0.020 (0.0081)		-0.017 (0.0078)	
Overall Trade Misreporting Index		-0.020 (0.0093)		-0.018 (0.0090)
Mean Outcome	0.043	0.043	0.034	0.034
Number of Observations	125	125	125	125

Notes: This table shows the relationship between modern Customs house capacity and the tariff code. Customs house capacity is measured with a (normalized) index of each country's agreement with its trading partners, following Farhad et al. (2024), where the first row considers only export and the second row considers all trade flows. For all results, we only consider goods that are imported by *all* of the 20 (non-OPEC) countries with the worst scores on the exporting measure. Panel A considers measures of the complexity of the tariff code: the (ln) number of distinct tariff rates, and the share of HS 6-digit codes with variation in tariffs. Panel B considers the specificity of the tariff code. The outcome in Columns 1 and 2 is the share of tariff lines that are specific. Columns 3 and 4 take the unweighted average share of tariff lines within each HS-6 that are specific, and then take the average across all HS-6 codes. Each observation is a country in its most recently available data, and we consider MFN tariff rates. Robust standard errors reported in parentheses. Source: WTO Integrated Database.

Table A.7: Narrow Labels Predict Specific Tariffs

Panel A. $\leq$ 5 goods in Label

	(1)	(2)	(3)	(4)
$\leq$ 5 goods in Label	0.078	0.080	0.082	0.11
	(0.0072)	(0.0071)	(0.0083)	(0.025)
Mean Outcome	0.11	0.11	0.11	0.49
Number of Goods	1692	1692	1692	1588
Number of Observations	34639	34639	34639	1588
Specification	Year FE	Year $\times$ SITC FE	Year & Good FE	Only 1842

Panel B. $\leq$ 10 goods in Label

	(1)	(2)	(3)	(4)
$\leq$ 10 goods in Label	0.074	0.075	0.066	0.035
	(0.0062)	(0.0055)	(0.0067)	(0.026)
Mean Outcome	0.11	0.11	0.11	0.49
Number of Goods	1692	1692	1692	1588
Number of Observations	34639	34639	34639	1588
Specification	Year FE	Year $\times$ SITC FE	Year & Good FE	Only 1842

Notes: This table shows the relationship between the aggregation in the tariff code and if a good has a specific tariff. Panel A considers labels covering 5 or fewer goods, and Panel B considers labels covering 10 or fewer goods. Each good is included only in the years in which it is covered by a tariff and it has either a specific or ad valorem tariff. Each observation is a good-by-year. The outcome for all regressions is if the good has a specific tariff, and the independent variable is if the label for the good is relatively disaggregated, as defined in the panel heading. All regressions control for year. Column 2 additionally controls for year interacted with three-digit SITC classifications. Column 3 controls for good fixed effects. Column 4 only includes 1842. All regressions cluster by good. Source: *Tariff Acts of the United States* and *Commerce and Navigation of the United States.*

Table A.8: Homogeneity Predicts Specific Tariffs, Alternative Definition of Homogeneous

	(1)	(2)	(3)
Homogeneous	0.074	0.10	0.12
	(0.0068)	(0.020)	(0.027)
Mean Outcome	0.11	0.11	0.49
Number of Goods	1692	1692	1588
Number of Observations	34639	34639	1588
Specification	Year FE	Year × SITC FE	Only 1842

Notes: This table shows the relationship between price dispersion across varieties of a good and if it has a specific tariff. Each good is included only in the years in which it is covered by a tariff and it has either a specific or ad valorem tariff. Each observation is a good-by-year. The outcome for all regressions is if the good has a specific tariff, and the independent variable is if the good (in the modern era) is sold on an organized exchange or has a reference price, following the conservative classification of Rauch (1999). All regressions control for year. Column 2 additionally controls for year interacted with two-digit SITC classifications. Column 3 only includes 1842. All regressions cluster by good. Source: *Tariff Acts of the United States.*

Table A.9: Singleton & Homogeneous Labels Predict Specific Tariffs

	(1)	(2)	(3)
Singleton Label	0.023	0.17	0.11
	(0.012)	(0.033)	(0.032)
Homogeneous	0.11	0.060	0.27
	(0.011)	(0.034)	(0.044)
Homogeneous & Singleton Label	0.00019	0.26	-0.064
	(0.019)	(0.053)	(0.066)
Mean Outcome	0.11	0.11	0.49
Number of Goods	1692	1692	1588
Number of Observations	34639	34639	1588
Specification	Year FE	Year $\times$ SITC FE	Only 1842

Notes: This table shows the relationship between the aggregation in the tariff code, price dispersion across varieties of a good, and if a good has a specific tariff. Each good is included only in the years in which it is covered by a tariff and it has either a specific or ad valorem tariff. Each observation is a good-by-year. The outcome for all regressions is if the good has a specific tariff, and the independent variables are if the label for the good is a singleton (only covers one good), if the good is listed as a commodity in the *Executive Documents* of the House of Representatives, and their interaction. All regressions control for year. Column 2 additionally controls for year interacted with two-digit SITC classifications. Column 3 only includes 1842. All regressions cluster by good. Source: *Tariff Acts of the United States.*

Table A.10: Singleton Labels Predict Specific Tariffs, Robustness to Specification

Panel A. Controlling for 1842 Ad Valorem Equivalent

	(1)	(2)	(3)	(4)
Singleton Label	0.063	0.067	0.14	0.19
	(0.012)	(0.014)	(0.015)	(0.032)
Mean Outcome	0.15	0.15	0.15	0.71
Number of Goods	763	763	763	734
Number of Observations	15751	15751	15751	734
Specification	Year FE	Year × SITC FE	Year & Good FE	Only 1842

Panel B. Weighting by SITC

	(1)	(2)	(3)	(4)
Singleton Label	0.13	0.082	0.14	0.34
	(0.017)	(0.0088)	(0.017)	(0.046)
Mean Outcome	0.084	0.084	0.084	0.46
Number of Goods	1692	1692	1692	1588
Number of Observations	34639	34639	34639	1588
Specification	Year FE	Year × SITC FE	Year & Good FE	Only 1842

Notes: This table shows the relationship between the aggregation in the tariff code and if a good has a specific tariff. Panel A additionally controls for the 1842 ad valorem equivalent, for the goods reported in the import statistics. Panel B weights by the inverse of the number of goods within each three-digit SITC classification. Each good is included only in the years in which it is covered by a tariff and it has either a specific or ad valorem tariff. Each observation is a good-by-year. The outcome for all regressions is if the good has a specific tariff, and the independent variable is if the label for the good is a singleton (only covers one good). All regressions control for year. Column 2 additionally controls for year interacted with three-digit SITC classifications. Column 3 controls for good fixed effects. Column 4 only includes 1842. All regressions cluster by good. Source: *Tariff Acts of the United States* and *Commerce and Navigation of the United States*.

Table A.11: Homogeneity Predicts Specific Tariffs, Robustness to Specification

Panel A. Controlling for 1842 Ad Valorem Equivalent

	(1)	(2)	(3)
Homogeneous	0.11	0.17	0.29
	(0.011)	(0.066)	(0.025)
Mean Outcome	0.15	0.15	0.71
Number of Goods	763	763	734
Number of Observations	15751	15751	734
Specification	Year FE	Year × SITC FE	Only 1842

Panel B. Weighting by SITC

	(1)	(2)	(3)
Homogeneous	0.072	0.17	0.38
	(0.014)	(0.047)	(0.056)
Mean Outcome	0.084	0.084	0.46
Number of Goods	1692	1692	1588
Number of Observations	34639	34639	1588
Specification	Year FE	Year × SITC FE	Only 1842

Notes: This table shows the relationship between price dispersion across varieties of a good and if it has a specific tariff. Each good is included only in the years in which it is covered by a tariff and it has either a specific or ad valorem tariff. Each observation is a good-by-year. The outcome for all regressions is if the good has a specific tariff, and the independent variable is if the good is listed as a commodity in the *Executive Documents* of the House of Representatives. Panel A additionally controls for the 1842 ad valorem equivalent, for the goods reported in the import statistics. Panel B weights by the inverse of the number of goods within each three-digit SITC classification. All regressions control for year. Column 2 additionally controls for year interacted with two digit SITC classifications. Column 3 only includes 1842. All regressions cluster by good. Source: *Tariff Acts of the United States.*

Table A.12: Singleton Labels Predict Specific Tariffs, Robustness to Sample

Panel A. Goods That Are Never Free

	(1)	(2)	(3)	(4)
Singleton Label	0.040	0.075	0.088	0.18
	(0.012)	(0.012)	(0.013)	(0.032)
Mean Outcome	0.12	0.12	0.12	0.51
Number of Goods	1310	1310	1310	1310
Number of Observations	28820	28820	28820	1310
Specification	Year FE	Year × SITC FE	Year & Good FE	Only 1842

Panel B. Major Acts Only

	(1)	(2)	(3)	(4)
Singleton Label	0.085	0.17	0.18	0.14
	(0.019)	(0.018)	(0.018)	(0.028)
Mean Outcome	0.25	0.25	0.25	0.49
Number of Goods	1692	1692	1692	1588
Number of Observations	14273	14273	14273	1588
Specification	Year FE	Year × SITC FE	Year & Good FE	Only 1842

Notes: This table shows the relationship between the aggregation in the tariff code and if a good has a specific tariff. Panel A drops all goods that are ever free. Panel B only includes the major acts (in which at least half of the goods are covered: 1789, 1790, 1792, 1794, 1816, 1824, 1832, and 1842). Each good is included only in the years in which it is covered by a tariff and it has either a specific or ad valorem tariff. Each observation is a good-by-year. The outcome for all regressions is if the good has a specific tariff, and the independent variable is if the label for the good is a singleton (only covers one good). All regressions control for year. Column 2 additionally controls for year interacted with three-digit SITC classifications. Column 3 controls for good fixed effects. Column 4 only includes 1842. All regressions cluster by good. Source: *Tariff Acts of the United States.*

Table A.13: Homogeneity Predicts Specific Tariffs, Robustness to Sample

Panel A. Goods That Are Never Free

	(1)	(2)	(3)
Homogeneous	0.16	0.21	0.42
	(0.011)	(0.065)	(0.029)
Mean Outcome	0.12	0.12	0.51
Number of Goods	1310	1310	1310
Number of Observations	28820	28820	1310
Specification	Year FE	Year × SITC FE	Only 1842

Panel B. Major Acts Only

	(1)	(2)	(3)
Homogeneous	0.16	0.17	0.27
	(0.017)	(0.040)	(0.032)
Mean Outcome	0.25	0.25	0.49
Number of Goods	1692	1692	1588
Number of Observations	14273	14273	1588
Specification	Year FE	Year × SITC FE	Only 1842

Notes: This table shows the relationship between price dispersion across varieties of a good and if it has a specific tariff. Panel A drops all goods that are ever free. Panel B only includes the major acts (in which at least half of the goods are covered: 1789, 1790, 1792, 1794, 1816, 1824, 1832, and 1842). Each good is included only in the years in which it is covered by a tariff and it has either a specific or ad valorem tariff. Each observation is a good-by-year. The outcome for all regressions is if the good has a specific tariff, and the independent variable is if the good is listed as a commodity in the *Executive Documents* of the House of Representatives. All regressions control for year. Column 2 additionally controls for year interacted with two-digit SITC classifications. Column 3 only includes 1842. All regressions cluster by good. Source: *Tariff Acts of the United States*.

Table A.14: Singleton Labels Predict Specific Tariffs, Robustness to Definition of a Good

Panel A. Strict Deduplication

	(1)	(2)	(3)	(4)
Singleton Label	0.054	0.079	0.092	0.15
	(0.0096)	(0.0098)	(0.011)	(0.028)
Mean Outcome	0.12	0.12	0.12	0.49
Number of Goods	1616	1616	1616	1545
Number of Observations	33505	33505	33505	1545
Specification	Year FE	Year × SITC FE	Year & Good FE	Only 1842

Panel B. Extremely Strict Deduplication

	(1)	(2)	(3)	(4)
Singleton Label	0.058	0.081	0.10	0.17
	(0.010)	(0.010)	(0.012)	(0.031)
Mean Outcome	0.12	0.12	0.12	0.54
Number of Goods	1169	1169	1169	1125
Number of Observations	24226	24226	24226	1125
Specification	Year FE	Year × SITC FE	Year & Good FE	Only 1842

Notes: This table shows the relationship between the aggregation in the tariff code and if a good has a specific tariff. Our definition of a good is imports which are physically distinct than all other imports with the same path of tariffs. This necessitates some judgment calls. In Panel A, we are stricter than in our baseline specification, dropping all plausibly similar objects. In Panel B, we pick one random good from each path of tariffs, dropping all others regardless of physical similarity. Each good is included only in the years in which it is covered by a tariff and it has either a specific or ad valorem tariff. Each observation is a good-by-year. The outcome for all regressions is if the good has a specific tariff, and the independent variable is if the label for the good is a singleton (only covers one good). All regressions control for year. Column 2 additionally controls for year interacted with three-digit SITC classifications. Column 3 controls for good fixed effects. Column 4 only includes 1842. All regressions cluster by good. Source: *Tariff Acts of the United States*.

Table A.15: Homogeneity Predicts Specific Tariffs, Robustness to Definition of a Good

Panel A. Strict Deduplication

	(1)	(2)	(3)
Homogeneous	0.12	0.19	0.29
	(0.0093)	(0.043)	(0.032)
Mean Outcome	0.12	0.12	0.49
Number of Goods	1616	1616	1545
Number of Observations	33505	33505	1545
Specification	Year FE	Year × SITC FE	Only 1842

Panel B. Extremely Strict Deduplication

	(1)	(2)	(3)
Homogeneous	0.13	0.20	0.28
	(0.0095)	(0.049)	(0.033)
Mean Outcome	0.12	0.12	0.54
Number of Goods	1169	1169	1125
Number of Observations	24226	24226	1125
Specification	Year FE	Year × SITC FE	Only 1842

Notes: This table shows the relationship between price dispersion across varieties of a good and if it has a specific tariff. Our definition of a good is imports which are physically distinct than all other imports with the same path of tariffs. This necessitates some judgment calls. In Panel A, we are stricter than in our baseline specification, dropping all plausibly similar objects. In Panel B, we pick one random good from each path of tariffs, dropping all others regardless of physical similarity. Each good is included only in the years in which it is covered by a tariff and it has either a specific or ad valorem tariff. Each observation is a good-by-year. The outcome for all regressions is if the good has a specific tariff, and the independent variable is if the good is listed as a commodity in the *Executive Documents* of the House of Representatives. All regressions control for year. Column 2 additionally controls for year interacted with two-digit SITC classifications. Column 3 only includes 1842. All regressions cluster by good. Source: *Tariff Acts of the United States*.

Table A.16: Lower Unit Values Predict Higher Ad Valorem Tariffs (Weighted)

	(1)	(2)	(3)
Unit Value (Standardized)	-2.28	-1.13	-1.88
	(0.54)	(0.65)	(0.55)
Mean Outcome	30.077	29.402	29.775
Number of Imports	854	405	1163
Number of Observations	1069	605	1674
Tariff Regime	1842	1846	Both

Notes: This table shows the relationship between unit values and ad valorem duties. The outcome in all regressions is the ad valorem rate for the import after 1846. The sample is the imports with specific tariffs in the 1842 tariff act (which all have ad valorem tariffs after the 1846 Walker Tariff, the rate being analyzed). All regressions additionally control for each import's 1842 ad valorem equivalent, and we weight by the import's share of the total value of imports in each year. Column 1 includes only observations from years where the 1842 tariff applied, Column 2 only from the years where the 1846 tariff applied, and Column 3 includes all observations. Results clustered by import. Each observation is a row in the import tables of *Commerce and Navigation of the United States*. Additional source: *Tariff Acts of the United States*.

Table A.17: Modern Customs House Capacity Predicts More Complex Tariffs and More Specific Tariffs, Robustness

	ln(Number of Distinct Tariff Rates)	Share of HS 6-digit Codes with Variation in Tariffs	Share of Tariff Lines with Specific Tariffs	Average of HS 6-digit Share with Specific Tariffs
	(1)	(2)	(3)	(4)
Bureaucracy Quality	0.53	0.015	0.057	0.054
	(0.15)	(0.0068)	(0.026)	(0.026)
Mean Outcome	3.05	0.078	0.038	0.035
Number of Observations	99	99	99	99

Notes: This table shows the relationship between modern Customs house capacity and the tariff code. Customs house capacity is measured with a (normalized) index of each country's Bureaucracy Quality index from the International Country Risk Guide. Columns 1 and 2 consider measures of the complexity of the tariff code: the (ln) number of distinct tariff rates, and the share of HS 6-digit codes with variation in tariffs. Columns 3 and 4 consider the specificity of the tariff code. The outcome in Column 3 is the share of tariff lines that are specific. Column 4 takes the unweighted average share of tariff lines within each HS-6 that are specific, and then take the average across all HS-6 codes. Each observation is a country in its most recently available data, and we consider MFN tariff rates. Robust standard errors reported in parentheses. Source: WTO Integrated Database and International Country Risk Guide.

Table A.18: Modern AVEs Predict Specificity, Robustness

Panel A. Ad Valorem Equivalents Below 2.5%

	Specific Tariff		
	(1)	(2)	(3)
AVE $<$ 2.5	0.085	0.085	0.062
	(0.006)	(0.006)	(0.006)
Homogeneous		0.0068	0.0053
		(0.001)	(0.001)
AVE $<$ 2.5 $\times$			0.042
Homogeneous			(0.009)
Mean Outcome	0.017	0.017	0.017
Number of Observations	801016	801016	801016

Panel B. Ad Valorem Equivalents Below 10%

	Specific Tariff		
	(1)	(2)	(3)
AVE $<$ 10	0.0038	0.0031	0.0050
	(0.001)	(0.001)	(0.001)
Homogeneous		0.0076	0.0096
		(0.001)	(0.001)
AVE $<$ 10 $\times$			-0.0044
Homogeneous			(0.002)
Mean Outcome	0.017	0.017	0.017
Number of Observations	801016	801016	801016

Notes: This table shows the relationship between modern tariff line price variation, ad valorem equivalents, and specific tariffs. Panel A considers the tariff burden calculated as an indicator for the ad valorem equivalent being below 2.5% and Panel B uses an indicator for the ad valorem equivalent being below 10%. Each observation is a tariff line, and the outcome is if the tariff is specific. Homogeneous goods are those that are sold on an organized exchange or have a reference price, following the conservative classification of Rauch (1999). All regressions include importer fixed effects, and standard errors are clustered by HS 6-digit code. We use the most recently available data for each country. Source: WTO Integrated Database and UN Trade Analysis Information System.

Table A.19: Modern Homogeneous Goods Predict Specificity, Robustness

	Specific Tariff				
	(1)	(2)	(3)	(4)	(5)
Homogeneous	0.0073	0.0065	0.0049	0.0077	0.0023
	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)
AVE $<$ 5		0.044	0.033		
		(0.003)	(0.004)		
AVE $<$ 5 $\times$ Homogeneous			0.020		
			(0.005)		
Non-Grouped				0.011	0.0046
				(0.001)	(0.001)
Homogeneous $\times$ Non-Grouped					0.015
					(0.003)
Mean Outcome	0.017	0.017	0.017	0.017	0.017
Number of Observations	801016	801016	801016	801016	801016

Notes: This table shows the relationship between modern tariff line price variation and specific tariffs. Each observation is a tariff line, and the outcome is if the tariff is specific. Homogeneous goods are those that are sold on an organized exchange or have a reference price, following the liberal classification of Rauch (1999). Non-grouped goods are those that do not share the same tariff with all of the other tariff lines within the same HS 4-digit code. All regressions include importer fixed effects, and standard errors are clustered by HS 6-digit code. We use the most recently available data for each country. Source: WTO Integrated Database and UN Trade Analysis Information System.

Table A.20: Modern Low AVEs and Homogeneous Goods Predict Specificity, Robustness Only Including New Imports

Panel A. Ad Valorem Equivalents

	Specific Tariff		
	(1)	(2)	(3)
AVE < 5	0.013	0.013	0.0022
	(0.003)	(0.003)	(0.002)
Homogeneous		0.0042	0.0018
		(0.001)	(0.001)
AVE < 5 × Homogeneous			0.023
			(0.005)
Mean Outcome	0.010	0.010	0.010
Number of Observations	316079	316079	316079

Panel B. Homogeneity

	Specific Tariff			
	(1)	(2)	(3)	(4)
Homogeneous	0.0044		0.0049	0.00075
	(0.0009)		(0.0009)	(0.0003)
Non-Grouped		0.0037	0.0044	0.00054
		(0.0007)	(0.0008)	(0.0004)
Homogeneous × Non-Grouped				0.011
				(0.002)
Mean Outcome	0.010	0.010	0.010	0.010
Number of Observations	316079	316079	316079	316079

Notes: This table shows modern predictors of specific tariffs. Each panel includes coefficients from three predictors, one of which is always if the good is homogeneous. Panel A also considers the tariff burden, calculated as an indicator for the ad valorem equivalent being below 5% and Panel B considers if the good is non-grouped. Each panel also shows the interaction of the two variables. Each observation is a tariff line, and the outcome is if the tariff is specific. Homogeneous goods are those that are sold on an organized exchange or have a reference price, following the conservative classification of Rauch (1999). Non-grouped goods are those that do not share the same tariff with all of the other tariff lines within the same HS 4-digit code. All regressions include importer fixed effects, and standard errors are clustered by HS 6-digit code. We use the most recently available data for each country, and limit the sample to goods in SITC Revision 2 three-digit sectors that did not have a tariff before 1846. Source: WTO Integrated Database and UN Trade Analysis Information System.

Table A.21: 1980s American Imports: Drivers of Specificity, Only Non-Concentrated Products

Panel A. Unit Values

	(1)	(2)	(3)	(4)
ln(Unit Values)	-0.027	-0.011	-0.029	-0.015
	(0.0022)	(0.0031)	(0.0022)	(0.0048)
Mean Outcome	0.19	0.19	0.19	0.19
Number of Goods	3571	3571	3571	3571
Number of Observations	13717	13717	2973	2973
4-Digit FEs		✓		✓
Sample	Annual	Annual	Pooled	Pooled

Panel B. Ad Valorem Equivalents

	(1)	(2)	(3)	(4)
ln(Ad Valorem Equivalent)	-0.16	-0.11	-0.16	-0.13
	(0.0068)	(0.0087)	(0.0074)	(0.013)
Mean Outcome	0.16	0.16	0.16	0.16
Number of Goods	4152	4152	4152	4152
Number of Observations	16287	16287	3507	3507
4-Digit FEs		✓		✓
Sample	Annual	Annual	Pooled	Pooled

Panel C. Price Dispersion Across Exporters

	(1)	(2)	(3)	(4)
Price Dispersion	-0.13	-0.012	-0.15	-0.027
	(0.0095)	(0.0059)	(0.013)	(0.017)
Mean Outcome	0.19	0.19	0.19	0.19
Number of Goods	3519	3519	3519	3519
Number of Observations	13369	13369	2972	2972
4-Digit FEs		✓		✓
Sample	Annual	Annual	Pooled	Pooled

Notes: The outcome in this table is if the import has a specific tariff. Each observation is a TSUS 5-digit import. We only include products where the HHI by origin is less than 0.5, as described in the text. Columns 1 and 2 include each good by year separately, and Columns 3 and 4 average across years within goods. Columns 2 and 4 additionally include fixed effects for each 4-digit TSUS code. All regressions are clustered by 5-digit TSUS code. Panel A considers the unit values of the goods (the value divided by the quantity). Panel B considers the ad valorem equivalents (the tariff revenue divided by the value). Panel C considers the dispersion of unit values across shipments, as described in the text. Source: US Census Annual Import Data Bank (IDB).