Where Do the Data Come From? Endogenous Classification in Administrative Data*

Matthew Grant†

May 2023

Abstract

Classification systems group heterogeneous objects such as products into categories called "codes" and define what level of policy, such as tariffs, will be applied to each group. I develop a theory of endogenous classification and show that accounting for the motives underlying the design of classification systems has important implications for empirical research. I build a model in which the choice of which things to group together reflects the objectives of a policymaker. A finer classification better targets policy but is harder to design and enforce. The degree to which heterogeneous objects are grouped together will vary across codes within a classification system and may be systematically related to policies and attributes of interest to both the econometrician and the policymaker. Taking the classification of U.S. imports as a leading example, I show that its design is consistent with this theory. Product codes vary greatly in their size and specificity – codes are more subdivided when tariffs are high, and when codes are more subdivided goods within them are more similar. Heterogeneity in product attributes within codes leads to large biases when estimating parameters such as demand and supply elasticities. I show that the amount of heterogeneity and therefore the size of this bias is correlated with tariffs as expected under a theory of endogenous classification, and correcting it changes our understanding of important empirical relationships such as the one between tariffs and demand elasticities.

*I thank Kyle Bagwell, Mostafa Beshkar, Renee Bowen, Swati Dhingra, Gene Grossman, Kyle Handley, Giovanni Maggi, Eduardo Morales, Ezra Oberfield, Ralph Ossa, Steve Redding, Paulo Somaini, Isaac Sorkin, Meredith Startz, and Richard Startz for helpful comments, and gratefully acknowledge the hospitality of the International Economics Section at Princeton University while working on an early version of this paper.

†Dartmouth College. Email: Matthew.W.Grant@Dartmouth.edu

1 Introduction

Classification systems group objects into "codes" according to common features, such as product type, industry, or occupation. A primary purpose of these systems is to define the application of policy. For instance, the definition of codes may determine which tariffs apply to which products or which workers are covered under collective bargaining agreements. However, they also govern the collection and transmission of government data that are commonly used by researchers – we often observe data on international trade at the product code level and wages at the industry level. While such systems aggregate objects that are heterogenous in their actual characteristics – for instance, a single code in the trade data contains products with many different barcodes – economists rarely consider why classifications are designed as they are, or what the implications might be. To the extent that we do, it is often viewed simply as a practical data cleaning challenge, such as how to concord data sets in which definitions of codes change over time. In this paper, I introduce a theory of endogenous classification, and show empirically that the gap between how a classification system is designed by a policy maker and the way it is typically treated by the econometrician can matter a great deal.

The classifications we see *could* have been defined differently. Why do we observe the aggregation that we do? I begin by showing that a classification system can be thought of as a tool for targeting policy. I model the decision of a policymaker who simultaneously sets policy and chooses how to group together objects into codes that define how the policy will be applied. The model implies that the extent of aggregation will vary within a single classification system, in ways that are systematically related to both policy and the underlying characteristics of the objects being classified. I take the classification of traded goods in the United States as a leading example, and show that the design and application of the system is consistent with this theory of classification. This, in turn, means that many empirical estimates that rely on data reported at the code level will not only be biased, but that the extent of this bias will covary with policy. Accounting for the endogeneity of the classification system substantially changes the conclusions we reach about important parameters used to estimate the gains from trade, and reverses the sign of the empirical relationship between import elasticities and tariffs, restoring the relationship predicted by theory.

This paper is motivated by the observation that trade data are highly aggregated, and throughout the paper I will take trade as an important example of a broader set of insights about classification, policy, and government data. The Harmonized Tariff Schedule (HTS) is used to classify products imported into the United States and determines which tariff will be applied. Even at the finest widely available level of aggregation, imports into the U.S. are measured at the level of roughly twenty thousand 10-digit HTS categories. For example, all men's cotton shirts fall under a single code. As a consumer, it is clear that products within this category are far from uniform in price, style, and so on. However, what is visible to the researcher are measures aggregated to the level of the code: in 2015, the U.S. imported roughly 16 million kilograms of such shirts at a total cost of about $1 billion, for a weighted average price of $66.39 per kg. This type of code-level weighted average is the most common sort of data used in empirical trade research. While this research typically takes codes as given, they are in fact far from uniform or constant. HTS 8-digit codes,

which define the application of tariffs, vary wildly in size, and the classification changes frequently – nearly half of codes changed at least once over the period 1989-2004 (Bernard, Jensen, Redding, Schott (2009)).

To help explain these facts, I develop a model of endogenous classification. While there is a large literature devoted to modeling policy choices as a function of a policy maker's objective, this literature has not previously considered the system that defines the mapping from policies to objects as a choice variable. Decisions about how to define the targeting of policy are a core function of government and lawmaking. We can understand a classification system as the outcome of an optimization problem of a policy maker (who I refer to as the classifier), who simultaneously divides up a set of objects (e.g. traded products) into codes and chooses the level of a policy to apply uniformly within each code.[1] The policy maker may choose more or less broad categories in order to trade off the benefits of more finely targeted policy against the costs of implementing a more nuanced system. Because both the value of targeting and the effective costs of creating more homogeneous codes may vary in different parts of the object space, the coarseness of classification will vary along with the classifier's objective. In the context of the HTS, this means that some codes contain a relatively homogeneous set of products, while others are more heterogeneous.

I use several key insights to impose order on the problem. First, I obtain tractability by treating the classifier's payout on each classified object as quadratic, which can be thought of as a second order-approximation to a general objective. Second, I transform the problem to show that maximizing the objective is equivalent to minimizing the losses due to mistargeting that arise from setting policy at the code level rather than treating every object separately, subject to the costs of defining and enforcing the classification. I show that the cost of policy mistargeting is related to the heterogeneity of objects within each code, which is a weighted function of the covariances across "properties" of each object that the policymaker may care about.

To microfound the costs of classification, I distinguish between these "properties", which the policymaker cares for policy, and "characteristics", which the classifier may or may not care about per se but which are verifiable features that can be used to define and enforce groups in practice. Since the classifier has to use definitions in terms of characteristics to control the way objects are grouped in terms of their properties, the local relationship between characteristics and properties will determine the difficulty (and hence costliness) of designing the classification. For instance, the classifier may care about the elasticity of demand when setting tariffs, but since this is not readily verifiable or enforceable in court, has to use characteristics such as the horsepower of an engine or the thickness of a sheet of metal to define codes in the classification. To the extent that these verifiable characteristics don't correspond perfectly to the properties of true interest, it will be difficult and costly for the classifier to create homogenous codes that reduce policy mistargeting.

In equilibrium under the chosen classification, within-code heterogeneity will be smaller in parts of the object space where policy mistargeting is particularly costly to the classifier and larger in the

[1] Although I focus on implications for measurement in this paper, the insight also applies to systems that define the application of policy but are not used in empirical research. For instance, in the U.S., income tax differentiates between types of earnings (e.g., salary vs capital gains) and expenditures (e.g., regular expenditure vs. mortgage interest or charitable giving). The U.S. Clean Air Act discriminates across types of sources (e.g., cars vs factories) and regions (e.g., everywhere else vs non-attainment regions). Choices about how to define the set of objects over which aspects of these policies apply are analogous to the model I lay out below.

parts of the object space where describing the classification is more difficult. For example, variation in within-code elasticities of import demand will be particularly costly for the classifier in HTS codes with low average elasticities of import demand. Policies are set according to the average level properties across objects in a code, so a low average elasticity in the code implies a high tariff, and consequently amplifies the costs of mistargeting.

Next, I derive some predictions of this theory of classification that can be taken to data in an international trade setting, and show that these insights apply to the HTS classification system. The model predicts that, in the absence of a policy targeting motive, there is no reason to create additional codes, while when there is policy to be applied, the gains from additional subdivision will be increasing in the level of the policy and the number or value of objects. Consistent with these predictions, I show that the United States almost never subdivides codes when the applied tariff is zero, while the number of subdivisions is increasing in the tariff and the size of trade flows, both in the cross-section and over time. The theory also predicts that there will be variation in the extent of within-code heterogeneity across different parts of the classification system. Using trade microdata that allow for finer disaggregation of products than the HTS, I show that there is substantial variation in the degree of heterogeneity within each code in terms of prices and import shares. There is less within-code heterogeneity in parts of the HTS where the U.S. government uses more codes to divide the product space. Consistent with the theory, this subdivision is finer and within-code heterogeneity is lower where tariffs are higher, both across codes and within codes over time. These patterns strongly suggest that these codes are a conscious choice of the U.S. government in service of government objectives.

Why does it matter that the classification is designed with policy objectives in mind? In the final section of the paper, I show how failing to account for the endogeneity of the classification system when using classified data can substantially distort or even overturn the conclusions we reach from empirical analysis. I examine a particular source of bias related to data aggregation, which arises from taking a non-linear function of data averaged at the code level, and provide a formula for correcting it to a second order. This bias arises in a specific case that is crucial in empirical trade: estimating import elasticities from trade data using average weighted import prices and shares at the HS code level. When I implement a lower-bound correction for this bias using trade microdata, I find that the elasticities of import demand and foreign export supply are substantially downward biased – the median corrected estimates are nearly twice as large as the uncorrected ones. This difference has quantitatively important implications – for instance, using the method of Arkolakis, Costinot, and Rodriguez-Clare (2012), the implied gains from trade are 58% lower once this bias is accounted for. Critically, the size of the bias also varies in systematic ways across products, and is negatively correlated with the level of tariffs. The empirical trade literature has been puzzled by a positive correlation between elasticities and tariffs across products, which is contrary to most theoretical models of tariff-setting. When the bias due to endogenous classification is corrected, the sign flips and we obtain the relationship predicted by theory.

The motivation for and empirical work in this paper builds on a small literature in trade that has noted how changes in classification have been used to manipulate policy and has documented frequent changes in the classification system. Gowa and Hicks (2018) observe that in a series of bilateral U.S.

agreements, the U.S. split its tariff lines to reduce free-riding on via MFN. Similarly, Gulotty (2018) shows that tariff lines were subdivided in order to balance a trade agreement between the U.S. and France in the early 19th century. Tavares (2006) finds that reclassification can be a "sneaky" way of raising tariffs when tariffs are bound (by moving goods from low-tariff to high-tariff lines) and finds that reclassification is correlated with lobbying. Schott (2003), Schott (2004), and Hallak and Schott (2011) explore how goods classified under the same code actually vary substantially across source countries. Bernard, Redding, Jensen, and Schott (2009) and Pierce and Schott (2012) document that classifications change frequently over time; the changes in classification suggest that classifications are not perfect descriptions of goods: if multiple codes can be derived from a single code when it is subdivided or multiple codes can be merged together into a single one, then the trade codes must aggregate together multiple different goods.

My main contribution is to explain theoretically how we can think of a classification system as an endogenous choice of a policy maker, and what this tells us about why we observe the classification that we do. The spirit of my approach is most similar to that of the literature on endogenously incomplete contracts (see Dye (1985); Battigalli and Maggi (2002)). This literature contains some analogies to the theory in this paper, showing why contracts are more detailed regarding states of the world that are more important to the parties of the contract. I also contribute to a longstanding literature on tariff-setting objectives, including (among many others) Baldwin (1987), Grossman and Helpman (1994, 1995), Bombardini (2008), and Ossa (2011), and empirical investigations in Goldberg and Maggi (1999), Gawande and Bandyopadhyay (2000), and McCalman (2004). While this literature models policy choices as a function of a policymaker's objective, it has not previously considered the system that defines the mapping from policies to objects.

In dealing with biases in empirical estimation that arise from classification systems, this paper shares some applications with work on bias from aggregation or mismeasurement. Aggregation bias has been shown to be important in the case of trade elasticities by Imbs and Mejean (2015), who find lower elasticites at the industry level than at more disaggregate levels of classification, and in price adjustment of imports to exchange rate shocks in Imbs, Mumtaz, Ravn, and Rey (2005). Understanding the extent of the bias I document is related to the concept of "aggregation factors" in the aggregation bias literature. By providing a theory of the source of the aggregation itself – the classification system – I show not only why this bias arises but why the extent of bias is likely to be correlated with the parameters of interest in ways that further confound empirical analysis. I show that the bias can be corrected to a second order using only some limited moments describing the underlying heterogeneity in a way that is related to the work of Chesher (1991) on errors-in-variables.

Finally, I make use of techniques and estimates from a long literature focused on estimating trade elasticities, which originated with Feenstra (1994). Other works include Broda and Weinstein (2006), Broda, Greenfield, and Weinstein (2017), Broda, Limao, and Weinstein (2008), Soderberry (2015), Hottman, Redding and Weinstein (2016), Soderberry (2018), and Redding and Weinstein (2018). Recent work estimating consumption elasticities has been trending towards using scanner bar code data, in which every product is a different observation. Assuming that barcodes yield observations at the level of the data generating process, this reduces or eliminates the sort of bias I identify in this paper (see, e.g. Broda and Weinstein (2008) for an early use of bar code data in this

literature, although there are many other examples including papers previously listed). However, this data is generally limited in scope and does not cover goods that comprise a large fraction of trade. As a result, the majority of the empirical trade literature continues to rely on estimates based on HTS-classified data.

2 The HTS classification

I take the Harmonized Tariff Schedule (HTS) as a leading example of the design and implications of classification systems. In this section, I begin by introducing the design of the HTS and exploring some of the empirical patterns in its application that motivate the need for a model of endogenous classification.

2.1 Institutional context of the HTS

The HTS is a set of codes used to classify imports into the United States. Codes are divided into broad groups called "sections", which are the coarsest level of classification. Each section contains 2-digit codes, called "chapters". In turn, chapters are divided into one or more 4-digit "headings", 6-digit "subheadings", 8-digit "tariff lines", and 10-digit "statistical reporting numbers".

The first six digits of all codes (i.e. sections, chapters, headings, and subheadings) are called the Harmonized System (HS) and are shared by nearly all countries.[2] The HS codes are administered by the World Customs Organization (WCO), an international organization that is jointly managed by its members. The broad objective of the WCO is trade facilitation, and it has a number of activities beyond managing the HS codes. One objective of the HS system is to facilitate classification of traded goods, so that an exporter who knows the classification its good falls under in one country will be able to classify the good in a different country. The HS codes also make it easier to compare trade flows across countries.

The last four digits of all HTS codes are unilateral choices of the U.S. government, and are managed by the U.S. International Trade Commission (U.S. ITC). The U.S. ITC changes the set of tariff lines following changes in trade policy in order to implement the policy changes. For example, as a consequence of the trade policies of the Trump Administration, nearly 200 pages of additional classifications and their descriptions were added to the HTS.[3] The 8-digit code determines the tariff treatment of the good, while 10-digit codes permits tracking of trade flows in greater detail within the 8-digit tariff line.

The classification of goods in the HTS is hierarchical. Goods are first assigned to the section that fits best; next, they are assigned to the best-fitting chapter within that section, and so forth. Thus it is possible to classify any traded product, regardless of whether there is a code that fits the good in question well or not.[4] Changes to the HTS are sometimes interpreted in the literature as reflecting

[2] With the exception of Section 22, which is Chapters 98 and 99.

[3] Similarly, the U.S. ITC releases revised editions of the HTS as the policies change; a typical year might see 2 or fewer revised editions; 2018 saw 16 such revisions, 2019 saw 20 revisions, and there were 28 revisions in 2020. This was a direct consequence of the frequent changes to trade policy during the Trump Administration

[4] In fact, it's possible to classify goods that do not exist. Were the reader to come into possession of a Martian heat-ray gun from H.G. Wells's *The War of the Worlds*, it would be classified under 9301.90.90.90.

changes in the underlying set of products being described (see, e.g., Broda and Weinstein (2006)). While changes may indeed be motivated by the arrival of new product varieties, it is important to note that this is not mechanically true, since hierarchical classification means all new goods can be fit into the existing set of codes. Instead, splitting or adding codes is necessarily a conscious decision by the U.S. ITC that it is worthwhile to alter the classification to better capture the current set of varieties.

A related point is that codes do not generally correspond to unique product varieties, although they are often treated that way in practice in the empirical trade literature for lack of a better approach. In fact, multiple real-world products or product varieties are almost always aggregated into a single code. The level at which one defines a "true" product and might ideally want to observe data may depend on the application – the barcode level might be a good approximation in many cases – but the HTS code level is essentially always an aggregation of these underlying objects. It is therefore sensible to talk about heterogeneity within a code, where heterogeneity means differences in the characteristics of the underlying products or varieties that are being aggregated together. For instance, in the case of men's cotton shirts presented in the introduction, it is clear that there is variation in the actual price of different types of shirts within the code, as well as in their color, style, quality, and so on.

2.2 Changes in trade classifications over time

Both the HTS and the HS classifications undergo changes. Changes to the HS codes occur infrequently at semi-regular intervals. Since their introduction in 1988, the HS codes have changed six times: in 1996, 2002, 2007, 2012, 2017, and 2022. In contrast, changes to the unilateral portions of the HTS occur frequently and can happen at any time. For example, Pierce and Schott (2012) find that in the 15 year period from 1989 to 2004, nearly 45% of all 10-digit HTS codes changed, and these codes encompassed nearly 60% of all U.S. imports in 2004. These changes include merging multiple codes into one, splitting one code into multiple codes, and dividing and re-combining multiple codes into multiple new codes.

There are a large number of codes at the finer levels. As shown in Figure 1, the number of HTS 8-digit codes has been increasing while the number of HS 6-digit codes has stayed largely consistent over time (although there are many reorganizations within a roughly constant number of 6-digit codes). At the start of 1989, the HTS included 8,552 8-digit codes, while the HS included 4,954 6-digit codes. By the end of 2016, this had increased to 12,583 8-digit codes and 5,316 6-digit codes.

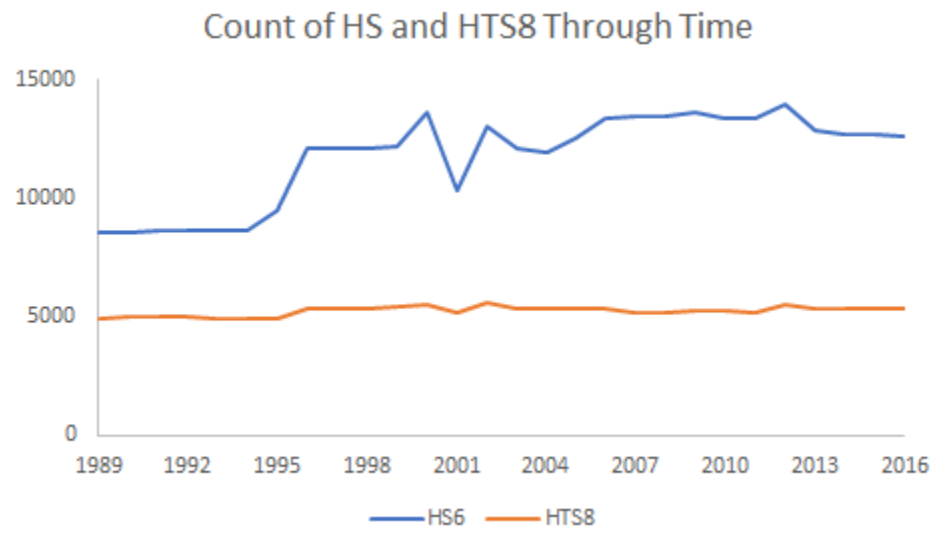

Figure 1: HTS 8-digit and HS-6 digit codes over time

2.3 Subdivision of HS 6-digit codes

Because the number of 6-digit codes has stayed relatively constant, the increase in the number of 8-digit codes is driven by U.S. government choices about how much to subdivide the internationally-set codes.

One might think that subdivision is targeted at maintaining codes of roughly equal size. In fact, HTS 8-digit codes vary widely in the value of trade flows within them. The average coefficient of variation for import values across HTS 8-digit codes within a year is 11.5 from 1989-2016– i.e. within a given year the standard deviation of import values across codes is 11.5 times the mean import value.

There is also substantial variation in the extent of heterogeneity across varieties of goods within a single HTS 8-digit code. To see this, I use import data disaggregated by the district of entry and district of unlading pair.[5] I observe the total value of imports and the weighted average price of goods within each HTS8, arriving via different districts of entry and terminating in different districts of unlading. Using this data, I find that the coefficients of variation for within-code variation in price and expenditure share[6] are 87.6 and 1,147, respectively. This implies that some codes have relatively little within-code heterogeneity in the varieties of goods they contain while other codes have much more.[7]

[5] This disaggregation is made public by the U.S. Census and broadly disseminated by Schott (2008).

[6] This expenditure share is the share of expenditure in a category on goods which enter through a particular district of entry and particular district of unlading.

[7] Although these numbers seem large, they are consistent with prior investigations of large within-code dispersion in price in some codes. For example the 1995 GAO report "Unit Values Vary Widely for Identically Classified Commodities" investigated dispersion of average prices within a code for fax machines, because the GAO was worried about fraud. Instead, it found that the same code encompassed fax machine parts, finished fax machines, and commercial telegraphs used for communicating with stock exchanges, and that there was truly a 5 order-of-magnitude difference in the price of these different goods.

Instead, the extent of subdivision is clearly associated with policy decisions. Table 1 shows variation in the extent to which 6-digit codes are subdivided into 8-digit tariff-line codes over the period 1989-2017.[8] On average, there are 2.01 8-digit codes per 6-digit code, but this masks substantial variation: the most-divided 6-digit code nests 62 8-digit codes, while many only have one. Furthermore, the extent of subdivision is related to tariffs (using measures of tariffs from Feenstra, Romalis, and Schott (2002) and Besedes, Lake, and Kohl (2020)). When there is no tariff[9] on any part of a 6-digit code – which is the case for 12,034 code-year pairs – it is almost never subdivided, while when there is a tariff on at least part of a 6-digit code, it is more likely to be divided.[10] This is consistent with the fact that tariffs are legally defined at the 8-digit level; i.e. tariffs may differ across HTS 8-digit codes but are by definition the same across products within one.[11]

HTS8 per HS6	Mean	SD	Min	Max	Count
All	2.02	2.50	1	62	144,432
No tariff	1.00	0.0637	1	2	12,034
Any tariff	2.11	2.59	1	62	130,076

Note that the second and third rows do not add up to the first because tariff information is missing for some lines: I omit a HS6-year pair if any of the HTS8 subcodes is missing tariff information in rows 2 and 3.

Table 1: Subdivision of 6-digit codes into 8-digit codes, 1989-2017

The design of the HTS defies simple explanations – the classification does not correspond mechanically to homogenous products, the arrival of new product varieties, or division into bins of a consistent size. In contrast, it is clearly related to policy choice, both in the sense that it legally defines the application of trade policies and that the extent of subdivision is empirically correlated with those policies. The institutional structure of the HTS and these patterns suggest the need for a model of intentional classification choices, which I turn to in the following section. The model will suggest a set of more precise predictions about the design and application of classification systems, which I will then examine empirically in application to the HTS in Section 4.

3 A theory of endogenous classification

In this section, I model the problem of a classifier who decides how to group a set of objects into codes and how to set policy based on those groupings. I use general notation, as the theory applies

[8] This figure excludes Section 22, which overrides other sections of the classification. Within Section 22, HTS 10-digit codes can map to 6-digit codes anywhere in the classification.

[9] I.e. no Column 1 / MFN tariff and no Column 2 tariff.

[10] An alternative interpretation of Table 1 might be that 6-digit codes that contain more subcodes may be mechanically more likely to have a tariff on at least one subcode, if for instance the probability of any tariff were randomly and uniformly distributed across HTS8s. Two supplemental facts suggest this is not the case. First, Table 4 (in Section 4 of the paper) shows that changes in subdivision are also related to changes in the *level* of the tariff. Second, if the relationship were mechanical, we would expect many 6-digit codes to have multiple subcodes with zero tariff. However, this almost never happens: only 0.4% of all HS 6-digit code-year pairs have multiple subcodes with zero tariffs. These rare examples are likely due to cases where 8-digit codes are also being used to implement other policies such as rules of origin or the Generalized System of Preferences.

[11] This is true for Column 1 and 2 tariffs which are the primary tariff instruments of the U.S. government, but temporary trade barriers and other special policies may differ within an 8-digit code. Furthermore, other policies (beyond the Column 1 and Column 2 tariffs) such as rules of origin, special rates of duty, etc. can also be set at the HTS 8-digit code.

to a wide range of classification systems. To fix ideas, however, one can think of the case of U.S. imports, in which the classifier is the U.S. government, objects are imported goods, and the codes are those of the HTS.

This section proceeds in four parts. First, I provide a formal definition of a classification and set up the problem of the classifier. Second, I show that when policy is set on a group of objects rather than separately for every object, the costs of policy mistargeting will depend on the variances and covariances across properties of objects within the group. Third, I solve the problem of a classifier simultaneously grouping objects and setting policy based on those groups in a simplified setting. And fourth, I show that this solution generalizes to encompass additional aspects of classification systems which are not included in the baseline model. Throughout, I relegate proofs and derivations to the Appendix.

3.1 Problem and general setting

A classification is a mapping from a set of objects to groups called codes. Each object is assigned to only one code, and a code can encompass a set of objects. The purpose of a classification system is to implement policy.[12] Codes define the set of objects that will be treated together: objects in the same code will be subject to the same policy. Thus, the overall problem for the classifier is to maximize its payout from objects in the set $\mathbb{X}$ by simultaneously choosing a set of codes, indexed by i, and the policy applied to each code. Formally, the classifier chooses a partition of the set of objects, $\mathbb{P}$, and the policy γ_i that is applied to each code $i \in \mathbb{P}$.[13] I will refer to the vector of the level of the policy across codes by $\vec{\gamma} \equiv (\gamma_1, \gamma_2, \ldots)$. The classifier chooses the classification and sets policy to maximize the sum of payouts from all codes less the costs of the classification:

$$\max_{\mathbb{P},\ \{\gamma_i\}_{i\in\mathbb{P}}} \left[\sum_{i\in\mathbb{P}} \left(\sum_{x\in i} \phi\left(\vec{\gamma}, \mathbb{P}, x\right) \right) - C\left(\mathbb{P}\right) \right] \tag{1}$$

The payout from each code i is the sum of payouts from each object in the code. In turn, the payout from each object x, denoted $\phi\left(\vec{\gamma}, \mathbb{P}, x\right)$, depends on the policy applied to the object and the attributes of the object itself, in addition to the policies applied to other objects and the attributes of those objects. The classifier also incurs costs of classification ($C\left(\mathbb{P}\right)$), which may depend on the sets of objects within each code or other aspects of the partition such as the number of codes. Consequently, this expression permits the possibility that the classifier's payout from any given object may depend on policies applied to other objects (either within that code or outside of it). However, I assume that the costs of classification are separable from the applied policies.

Objects are equivalent to a bundle of "characteristics". Characteristics capture all the "fundamental" attributes of an object, such as form, function, color, density, size, elasticities of supply and demand (if these elasticities are exogenous), other fundamentals of supply and demand (such as productivity and taste shifters in a CES supply and demand system), and so on. I assume there are

[12] Later, I will discuss how classification can be used to collate information, as in the ten-digit statistical reporting categories of the HTS. This is isomorphic to the problem of a classification used to implement policy.

[13] In the Appendix, I show that the theory extends readily to a setting with multiple policies.

finitely many characteristics numbering K and indexed by k. For tractability, I assume that every characteristic k can be expressed as a continuous real-valued variable between $z_{k,min}$ and $z_{k,max}$ where there is no requirement that either of these bounds be finite. Consequently, objects lie in Euclidean characteristic space of arbitrary dimension. I will denote a given bundle of characteristics using the vector $\mathbf{z} = (z_1, z_2, \ldots, z_K)$. The entire set of objects $\mathbb{X}$ in characteristic space is continuously distributed according to the pdf $f(\mathbf{z})$, and I am implicitly assuming that there is a measure of objects in order for $f(\mathbf{z})$ to be continuously distributed.[14] Throughout, I assume the classifier has perfect information about the distribution of characteristics. The partition maps objects to codes in terms of their characteristics, so that I can denote the code for a particular bundle of characteristics $\mathbf{z}$ by the function $i(\mathbf{z})$. The distribution of objects in code i is $f_i(\mathbf{z})$, where $f_i(\mathbf{z}) = f(\mathbf{z})$ if $\mathbf{z} \in i$ and $f_i(\mathbf{z}) = 0$ otherwise.[15] Thus, I can write the classifier's payout from objects in code i as $\int_{\mathbf{z}} \phi(\gamma(\mathbf{z}), \mathbf{z}) f_i(\mathbf{z}) dz_1 \cdots dz_K$, where policy $\gamma(\mathbf{z})$ is applied to an object with characteristics $\mathbf{z}$.

3.2 The classifier's payout and the costs of mistargeted policy

If there were no costs of classification, the classifier would simply set a policy for each object. Implicitly, this is the way the literature has approached the problem of optimal policy setting. For example, the tax literature describes the optimal tax on a specific good as a function of features of that good, such as the elasticity of demand. Goods with different features should face different tax rates. Non-trivial classification arises, therefore, because there is some cost of treating each object separately that leads the policymaker to group them together into codes. However, when policy is set at the code level, the classifier experiences some loss from applying a different policy to objects in the code than would be optimally applied if each were treated separately, i.e. a cost of policy mistargeting. This trade-off between the cost of classification and the cost of policy mistargeting is what drives the choice of classification. In this section, I characterize policy mistargeting while holding the classification constant. In the main text, I examine a setting in which payouts are separable across objects (i.e. the payouts to a given object do not depend on policies applied to other objects). This setting captures all of the important intuition, and in this setting I establish Proposition 1, which relates the classification to the cost of policy mistargeting, in this setting. Afterwards, I will (briefly) explain how the main intuition extents to a general setting in which payouts are not separable. Full details for this more general setting are provided in the Appendix.

3.2.1 Mistargeting when payouts are separable across objects

In order to describe how the costs of mistargeting are related to the distribution of characteristics of objects in i, I assume that the classifier's payout is quadratic in policies and characteristics.[16] This

[14] This distribution is of the "types" of object and is assumed to be exogenous to policy choices of the classifier. Some measure of the number of objects might respond to the policy (e.g. trade flows may respond to tariffs). However, these flows can be predicted in terms of the fundamentals – $\mathbf{z}$ – and the endogenous policy choice, and thus are indirectly a components of $\phi(\cdot)$.

[15] This function exists because the set of i are elements in a partition – i.e. every $\mathbf{z}$ maps to exactly one i. To give an example of such a function, consider the case in which $\mathbf{z}$ is a scalar and codes are sequential intervals of the characteristic; in this case $i(\mathbf{z})$ is a step function.

[16] When applying this theory to the HTS at the end of this section of the paper, I will use a second-order approximation to capture a more general form of the objective functions while still applying these results.

assumption greatly simplifies the problem: I will show in Proposition 1 that under this assumption the costs of policy mistargeting are proportional to a weighted average of the covariances across characteristics within a code. When payouts are separable across objects, the payout on an object with characteristics $\mathbf{z}$ can be expressed as $\phi(\gamma, \mathbf{z})$, and I show in the Appendix that this quadratic payout simplifies to

$$\phi(\gamma, \mathbf{z}) = \hat{\phi}(\mathbf{z}) + \phi_\gamma \gamma + \vec{\phi}_{\gamma \mathbf{z}} \cdot \mathbf{z}\gamma + \frac{1}{2}\phi_{\gamma\gamma}\gamma^2 \tag{2}$$

The term $\hat{\phi}(\mathbf{z})$ is a function of the objects characteristics and, while it affects the level of the classifier's payout, is invariant to the policy and so does not affect the optimal policy or mistargeting. The terms $\phi_\gamma$and $\phi_{\gamma\gamma}$ are constants, while $\vec{\phi}_{\gamma\mathbf{z}} = (\phi_{\gamma z_1}, \ldots, \phi_{\gamma z_K})$ is a vector of constants. I assume that payouts are strictly concave in policy (i.e. $\phi_{\gamma\gamma} < 0$).

If the classifier could set policy separately on every object, which I refer to as perfect targeting, the optimal policy would be $\gamma^*(\mathbf{z})$, and this would yield a payout (ignoring classification costs) of $\int_{\mathbf{z}} \Phi(\gamma^*(\mathbf{z}), \mathbf{z}) f_i(\mathbf{z}) dz_1 \cdots dz_K$ over the set of objects in code i. In contrast, when the classifier sets a uniform level of policy to all objects in code i, the optimal choice of which I denote γ_i^*, the payout (again ignoring classification costs) is $\int_{\mathbf{z}} \phi(\gamma_i^*, \mathbf{z}) f_i(\mathbf{z}) dz_1 \cdots dz_K$.[17] Consequently, the cost of mistargeting, $\Delta\Phi_i$, will be

$$\Delta\Phi_i = \int_{\mathbf{z}} \left[\phi(\gamma^*(\mathbf{z}), \mathbf{z}) - \phi(\gamma_i^*, \mathbf{z})\right] f_i(\mathbf{z}) dz_1 \cdots dz_K \tag{3}$$

which is the difference between the payout from policy with perfect targeting and the payout from code-level policy.

I next turn to evaluating the cost of policy mistargeting for the classifier. It follows immediately from the quadratic payout that the optimal policy under perfect targeting for an object with characteristics $\mathbf{z}$ is

$$\gamma^*(\mathbf{z}) = \frac{\phi_\gamma + \vec{\phi}_{\gamma\mathbf{z}} \cdot \mathbf{z}}{-\phi_{\gamma\gamma}} \tag{4}$$

while the optimal code-level policy is

$$\gamma_i^* = \frac{\phi_\gamma + \vec{\phi}_{\gamma\mathbf{z}} \cdot \mathbb{E}[\mathbf{z}|\mathbf{z} \in i]}{-\phi_{\gamma\gamma}} \tag{5}$$

where I use $\mathbb{E}[\mathbf{z}|\mathbf{z} \in i]$ to denote the weighted average of characteristics of objects falling within the code. Thus, under perfect targeting the classifier takes the characteristics of each object into account, while under code-level policy the chosen policy only reflects the weighted average characteristics.

As a result, the classifier's payout is higher under perfect targeting. The payout for for any given

[17] Conditioning the code on $\mathbf{z}$ is redundant because I integrate with respect to f_i, which is zero for bundles of characteristics that do not fall into code i. I adopt this notation to better distinguish between the notation for uniform policy and perfect targeting.

object under perfect targeting is

$$\phi\left(\gamma^{*}\left(\mathbf{z}\right),\mathbf{z}\right)=\hat{\phi}\left(\mathbf{z}\right)+\frac{\left(\phi_{\gamma}+\vec{\phi}_{\gamma\mathbf{z}}\cdot\mathbf{z}\right)^{2}}{-2\phi_{\gamma\gamma}} \tag{6}$$

While under code-level policy the payout for the same object is

$$\phi\left(\gamma_{i}^{*},\mathbf{z}\right)=\hat{\phi}\left(\mathbf{z}\right)+\frac{\left(\phi_{\gamma}+\vec{\phi}_{\gamma\mathbf{z}}\cdot\mathbf{z}\right)^{2}}{-2\phi_{\gamma\gamma}}+\frac{\left(\vec{\phi}_{\gamma\mathbf{z}}\cdot\left[\mathbf{z}-\mathbb{E}\left[\mathbf{z}|\mathbf{z}\in i\right]\right]\right)^{2}}{2\phi_{\gamma\gamma}} \tag{7}$$

This payout is the same as under perfect targeting except for the 3rd term. This term is weakly negative because of the assumption that the objective is strictly concave in the policy. This term will be strictly negative if the policy under perfect targeting is different from the code-level policy, i.e. if there is at least one characteristic k for which $\phi_{\gamma\mathbf{z}_{k}}\neq 0$ and where the object has a different value of characteristic k than the code-level average. In fact, the difference in policy under perfect targeting from code-level policy is one measure of the distance of a given object with characteristics $\mathbf{z}$ from the code level average, and as this distance increases the mistargeting increases and the payout decreases. Thus, for codes with greater heterogeneity, there is more heterogeneity in object-level optimal policy, and therefore more mistargeting.

In what follows, I leave the difference in payouts in terms of characteristics (rather than expressing it as a difference in levels of policy). By taking the difference between Equations (6) and (7) and integrating over the set of $\mathbf{z}$ in code i, I immediately obtain Proposition 1, which provides an expression for the costs of policy mistargeting in code i.

Proposition 1:

For a given code i, the cost of policy mistargeting is

$$\Delta\Phi_{i}=F_{i}\sum_{k'=1}^{K}\sum_{k=1}^{K}L_{kk'}\sigma_{ikk'}^{2}$$

where

$$L_{kk'}=\frac{\phi_{\gamma\mathbf{z}_{k}}\phi_{\gamma\mathbf{z}_{k'}}}{-2\phi_{\gamma\gamma}}$$

This proposition says that the cost of policy mistargeting is a weighted sum of the covariances of characteristics across objects within a code. Intuitively, if there were no heterogeneity in characteristics of objects within a code, then the code level policy and the policy with perfect targeting would necessarily be the same for every object, and so there would be no cost of mistargeting. Thus, mistargeting is related to within-code heterogeneity, and under the assumption that payouts are quadratic, the heterogeneity that matters is the covariances of characteristics across objects within a code. Under a more general payout function, the classifier would care about many more moments of heterogeneity (skewness, kurtosis, etc.) and the problem would quickly become intractable, but the idea that within-code heterogeneity is responsible for policy mistargeting is perfectly general. The simplicity of Proposition 1 (and the analogous result when payouts are not separable) keeps the

problem tractable without sacrificing intuition.

Furthermore, within-code heterogeneity as captured by the covariances in characteristics translates into a cost of policy mistargeting through the weights $L_{kk'}$. These weights are related to how the optimal policy under perfect targeting responds to the characteristics in question (i.e. the cross derivatives of the payout between the optimal policy and the characteristic). Thus, some of these weights could be zero. In particular, if for some characteristic k, $\phi_{\gamma \mathbf{z}_k} = 0$, then for all k', $L_{kk'}$ will be zero and so none of the covariances involving k will matter for the classifier's payout from policy. Consequently, the characteristic will have no influence on the classification. For example, the color of an object may not matter in setting an optimal tax. I refer to the subset of characteristics for which $\phi_{\gamma \mathbf{z}_k} \neq 0$ as "properties relevant to the optimal policy", or "properties". And although some weights might be negative, the weighted sum is always weakly positive, which follows from the fact that covariance matrices are positive semi-definite.

For convenience, I will index the covariances of properties by m, and denote their total number as M (note that if I number by $\hat{K} \leq K$, then $M = \frac{\hat{K}^2+\hat{K}}{2}$). The costs of mistargeting are therefore:

$$\Delta\Phi_i = F_i \sum_{m=1}^{M} L_m \sigma_{im}^2 \tag{8}$$

where if m indexes the covariance between characteristics $k \neq k'$, then $L_m = \frac{\phi_{\gamma \mathbf{z}_k} \phi_{\gamma \mathbf{z}_{k'}}}{-\phi_{\gamma\gamma}}$, while if $k = k'$, then $L_m = \frac{\phi_{\gamma \mathbf{z}_k} \phi_{\gamma \mathbf{z}_{k'}}}{-2\phi^i_{\gamma\gamma}}$.[18]

3.2.2 Mistargeting when payouts are not separable across objects

The idea of Proposition 1 – that when payouts are quadratic, the cost of policy mistargeting is a weighted sum of covariances of characteristics across objects within a code – still holds true even when payouts are not separable across objects, for instance if there are cross-elasticities in demand across products. However, the weights on these covariances will now reflect additional terms which capture how policies on one good affect the classifier's payout on other objects.

Whereas in the separable case, policy on an object is chosen to maximize payouts from the object itself, now the choice of the level of policy also reflects how the level of policy affects the payouts from other goods. However, both of these effects only vary based on an object's characteristics (in a linear way). Consequently, when objects are grouped together into codes, the choice of policy will reflect weighted average levels of characteristics across objects in the code, and the mistargeting of policy will reflect deviations of an object's characteristics from the weighted average characteristics in the code. This is the same as in the separable case and so the form of the expression for the mistargeting of policy will be the same.

I provide full details and a derivation in the Appendix.

[18] This is because $\sigma^2_{ikk'} = \sigma^2_{ik'k}$, so that covariances appear twice in the expression for $\Delta\Phi_i$ in Proposition 1 while variances only appear once.

3.3 Codes and the optimal classification

Having characterized the cost of policy mistargeting conditional on the classification, I now consider the optimal choice of the classification itself. As described in Section 3.1, the classifier's problem is to choose a partition of the set of objects, $\mathbb{P}$, and the policy γ_i that is applied to each code $i \in \mathbb{P}$. This is optimally done to maximize a payout that is decreasing in policy mistargeting as specified in Section 3.2, and in the cost of classification itself. If the partition involves choosing a discrete number of codes, then classification is a combinatorial optimization problem. To make the problem more tractable while preserving the generality of the objective function, I therefore make three further assumptions. The first concerns the form of the cost of classification, while the second and third make the problem differentiable.

First, I assume the costs of classification take the form of a fixed cost, C, of adding additional codes. This implies that the total cost of administering a classification system is increasing in the fineness of classification. This cost could be thought of as a cost of defining or enforcing a code. [19]

Second, I assume codes take the form of hyperrectangles in characteristic space, so that the classification problem is a choice of how large each code is in each dimension. If there were only one characteristic, each code would be an interval along a line; with two characteristics, codes are rectangles in characteristic space; and so forth. I use the vector $\mathbf{D}_i$ to denote the distances (or width) of code i in every characteristic. Note that the measure of objects and covariances across properties within the code can be written as functions of these distances and the location of the code in space. E.g., if code i has center (in characteristic space) given by the vector $\mathbf{q}_i$, then the mass of objects in the code is given by

$$F_i = \int_{q_K - \frac{D_{iK}}{2}}^{q_K + \frac{D_{iK}}{2}} \cdots \int_{q_1 - \frac{D_{i2}}{2}}^{q_2 + \frac{D_{i2}}{2}} \int_{q_1 - \frac{D_{i1}}{2}}^{q_1 + \frac{D_{i1}}{2}} f(\mathbf{z})\, dz_1 dz_2 \cdots dz_K$$

with analogous expressions for the means and covariances in properties across characteristics within the code.

Finally, I relax the classifier's problem by treating all codes as infinitesimal; this is equivalent to assuming there are a continuum of "small" codes (by which I mean each code is measure zero in every dimension, so that such a code has length $D_1 dz_1$ in dimension 1).[20] In effect, I am allowing the classifier to choose a non-integer number of codes, and am permitting the classifier to separately manipulate every part of a code in each dimension. The second and third of these assumptions mean I can abstract from considerations about the geometry of codes (i.e. ensuring codes cover the entire characteristic space without overlapping so that they form a partition of the characteristic space) and from the restriction that the classifier choose an integer number of codes.

[19] The assumption that this cost is the same across codes could be relaxed – it could be indexed by code, C_i, which implies costs could be functions of things like the size of a code (i.e. it is more difficult to describe a narrow code than a broad one), as long as the costs are separable across codes and continuous in code space. However, I assume costs are the same across codes going forward.

[20] This is analogous to models which assume a continuum of firms, workers, or goods, often with a goal of relaxing integer constraints.

3.3.1 A one-dimensional example

I start with a simplified problem to build intuition, in which there is a single characteristic – which I assume matters to the classifier for setting policy, and is therefore also a "property". While working with a single dimension in characteristic space, I will temporarily suppress k subscripts and convert vectors to scalars.

I permit the classifier to use a continuum of "small" codes, $i \in [0, N]$, where N is the total measure of codes. Code i has location in characteristic space given by $z(i)$.[21] When I describe a code as small, I mean that it can be treated as an infinitesimal fraction di of a code with measure 1 with the same location and width. Suppose there were a large code (centered around $z(i)$) with width D_i, a cost of mistargeting $F_i L \sigma_i^2$ – where F_i is the mass of objects in the code, L is the cost of within-code heterogeneity, and σ_i^2 is the variance in property i – and classification cost C. Then when the code is small it has cost of mistargeting $F_i L \sigma_i^2 di$ and classification cost $C di$.

Thus, the classifier's payout can be expressed as

$$\int_0^N \left(F_i L \sigma_i^2 + C\right) di$$

and the classifier has the constraint that these codes partition the characteristic space, which requires $D_i > 0$ and that these codes line up end to end to full cover the characteristic space

$$z(i) = z_{min} + \int_0^n D_i di$$

Maximizing the classifier's objective is equivalent to minimizing the sum of classification costs and policy mistargeting costs such that the set of codes partitions the characteristic space. Thus I can write the classifier's problem as

$$\min_{N, \{D_i\}_{\forall\, i \in [0,N]}} \int_0^N \left(F_i L \sigma_i^2 + C\right) di$$

s.t.

$$z(i) = z_{min} + \int_0^i D_n dn$$

$$D_i > 0$$

Using the constraint, I can use a change of variables ($\frac{\partial z}{\partial i} = D_i$) to index codes by their location in

[21] There is no distinction between the left and right endpoints of i because the code has measure 0 in the characteristic, so the endpoints must coincide.

characteristic space, z, instead of number i. When I do so, the classifier's problem can be re-written

$$\min_{\{D_z\}_{\forall\, z\in[z_{min},z_{max}]}} \int_{z_{min}}^{z_{max}} \left(\frac{F_z L\sigma_z^2 + C}{D_z}\right) dz$$

s.t.

$$D_z > 0$$

and in this problem the constraint that the codes partition the characteristic space is always satisfied (it has been fully captured by the change of variables). Consequently the problem is separable and I can characterize optimal policy by the first order condition with respect to the choice of the size of each code, D_z. It is also useful to define the elasticity between $F_i\sigma_i^2$ (the total variation in the code) and the width of the code in characteristic space, $\eta_i = \frac{D_i}{F_i\sigma_i^2} \cdot \frac{\partial}{\partial D_i}\left(F_i\sigma_i^2\right)$. Note that this elasticity is an equilibrium relationship and is assigned the value it would attain given the equilibrium width of all codes in the classification. Using the first order condition and this elasticity, I can express within-code variance in code i as:[22]

$$\sigma_i^2 = \frac{1}{L} \cdot \frac{C}{\eta_i - 1} \cdot \frac{1}{F_i}$$

This result says that the optimal within-code variance in the property, σ_i^2, will be a decreasing function of the marginal cost of policy mistargeting per unit measure of objects, L, the measure of objects in the code, F_i, and the strength of the local relationship between characteristic space and the total variation in the code, captured by the elasticity η_i. The elasticity captures the effectiveness of adding additional codes at reducing the within-code variance. Where this elasticity is large in equilibrium, adding additional codes is very effective at reducing the cost of policy mistargeting, and the classifier will choose a classification that yields comparatively less within-code variance in the property. Where this elasticity is small in equilibrium, divisions are not very effective at reducing the cost of policy mistargeting and the equilibrium classification will have comparatively more within-code variance in the property. The chosen within-code variance will be increasing in classification costs, C. The term $\frac{C}{\eta_i - 1}$ can be thought of as a measure of effective classification costs.

3.3.2 General problem

I now extend the intuition of the simplified case with only one characteristic to a general setting with arbitrarily many properties and characteristics. As in the simplified case, I index codes by $i \in [0, N]$. I can write the location of code i in characteristic space by the vector $\mathbf{z}(i)$.[23] As in the single dimension case, when I say that a code is "small", I mean that if the large code (with the same center) and distances $\mathbf{D}_i$ has a cost of mistargeting $\sum_{m=1}^{M} F_i L_m \sigma_i^2$ and classification cost C, then when the code is small it has cost of mistargeting $\sum_{m=1}^{M} F_i L_m \sigma_{im}^2 di$ and classification cost Cdi.

[22] I present the result changed back into code number for consistency with earlier work.

[23] There is no distinction between the left and right endpoints of i in any dimension because the code has measure 0 in every characteristic, so the endpoints must coincide.

Thus, the classifier's payout can be expressed as

$$\int_0^N \left(\sum_{m=1}^{M} F_i L_m \sigma_{im}^2 + C \right) di$$

and the classifier has the same constraint that these codes partition the characteristic space. However, unlike in the single dimension case, it is hard to describe the location of codes such that they form a partition.[24] Thus, in the many dimensional case I will write the constraint as

$$\bigcup_{\forall\, i} \{\mathbf{z} \,|\, \mathbf{z} \in i\} = \mathbf{Z}$$
$$\{\mathbf{z} \,|\, \mathbf{z} \in i\} \bigcap \{\mathbf{z} \,|\, \mathbf{z} \in i'\} = \emptyset \text{ if } i \neq i'$$
$$D_{ik} > 0 \;\forall\, i, k$$

which says that codes cover the entire characteristic space (denoted $\mathbf{Z}$), that different codes in the set never intersect, and no code is empty.

Again, maximizing the classifier's objective is equivalent to minimizing the sum of classification costs and policy mistargeting costs such that the set of codes partitions the characteristic space. Thus I can write the classifier's problem as

$$\min_{N, \{D_{ik}\}_{\forall\, i,k}} \int_0^N \left(\sum_{m=1}^{M} F_i L_m \sigma_{im}^2 + C \right) di$$

s.t.

$$\bigcup_{\forall\, i} \{\mathbf{z} \,|\, \mathbf{z} \in i\} = \mathbf{Z}$$
$$\{\mathbf{z} \,|\, \mathbf{z} \in i\} \bigcap \{\mathbf{z} \,|\, \mathbf{z} \in i'\} = \emptyset \text{ if } i \neq i'$$
$$D_{ik} > 0 \;\forall\, i, k$$

As in the single-dimensional case, I will use a change of variables to express the problem in characteristic space, where it is separable. However, it is no longer possible to take this change of variables directly from the constraint. Instead, I use the identity that the cost of a code is equal to the cost per unit hypervolume of the code integrated over the characteristic space falling within that code:

$$\sum_{m=1}^{M} F_i L_m \sigma_{im}^2 + C = \int_{\mathbf{z} \in i} \left(\frac{\sum_{m=1}^{M} F_i L_m \sigma_{im}^2 + C}{\prod_{k=1}^{K} D_{ik}} \right) dz_1 z_2 \cdots dz_K$$

and the first and second constraints (every $\mathbf{z}$ falls into exactly one code) to write the sum across codes as the integral across characteristic space (and to define the function $i(\mathbf{z})$ which maps a point

[24] The particularly tricky part is to describe how a change in the size of a positive measure of codes will cause the rest of the codes to adjust to that the set of codes continues to partition the characteristic space.

in characteristic space to the code encompassing that point).

$$\int_0^N \left(\sum_{m=1}^{M} F_i L_m \sigma_{im}^2 + C \right) di = \int_{z_{K,min}}^{z_{K,max}} \cdots \int_{z_{2,min}}^{z_{2,max}} \int_{z_{1,min}}^{z_{1,max}} \left(\frac{\sum_{m=1}^{M} F_{i(\mathbf{z})} L_m \sigma_{i(\mathbf{z})m}^2 + C}{\prod_{k=1}^{K} D_{i(\mathbf{z})k}} \right) dz_1 dz_2 \cdots dz_K$$

Then, because every code is small in every characteristic, there is a bijection between a location in code space and a location in characteristic space.[25] This means that $i(\mathbf{z})$ is invertible and I can write D_{ik} as $D_{\mathbf{z}k}$. This completes the change of variables and I can write the classifier's problem as

$$\min_{\{D_{\mathbf{z}k}\}_{\forall k,\mathbf{z}}} \int_{z_{K,min}}^{z_{K,max}} \cdots \int_{z_{2,min}}^{z_{2,max}} \int_{z_{1,min}}^{z_{1,max}} \left(\frac{\sum_{m=1}^{M} F_{\mathbf{z}} L_m \sigma_{\mathbf{z}m}^2 + C}{\prod_{k=1}^{K} D_{\mathbf{z}k}} \right) dz_1 dz_2 \cdots dz_K$$

s.t.

$$D_{\mathbf{z}k} > 0 \ \forall \, \mathbf{z}, k$$

In this formulation, the problem is separable and I can characterize optimal policy by the first order condition with respect to the choice of the size of each code in every dimension, $D_{\mathbf{z}k}$. It is also useful to define the elasticity between $F_i \sigma_{im}^2$ and the width of the code in characteristic k space, $\eta_{ikm} = \frac{D_{ik}}{F_i \sigma_{im}^2} \cdot \frac{\partial}{\partial D_{ik}} \left(F_i \sigma_{im}^2 \right)$. This elasticity is an equilibrium relationship and is assigned the value it would attain given the equilibrium width of all codes in the classification. Using these elasticities and first-order conditions with respect to every distance in characteristic space at all points, I can obtain an expression related to the optimal within-code covariance in every characteristic. I present this result as Proposition 2.

Proposition 2:

If $\mathbf{H}_i$ has rank M, then the optimal within-code covariances in code i will satisfy

$$\vec{\Theta}_i = \frac{1}{F_i} \left(\mathbf{H}_i \right)_L^{-1} \mathbf{L}^{-1} C \vec{1}$$

In this result, $\vec{\Theta}_i$ is a vector of variances and covariances in code i (i.e. $\vec{\Theta}_{im} = \sigma_{im}^2$), $\mathbf{L}$ is a diagonal matrix of the classifier's losses from variance (i.e. $\mathbf{L}_{mm} = L_m$ and $\mathbf{L}_{mn} = 0$ for $m \neq n$), $\mathbf{H}_i$ is a (rectangular) matrix of elasticities between attributes and covariances (such that $\mathbf{H}_{ikm} = \eta_{ikm} - 1$), $\left(\mathbf{H}_i \right)_L^{-1}$ is the left inverse of $\mathbf{H}_i$, and $\vec{1}$ is vector of ones.

The result in Proposition 2 is a straightforward extension of the one-dimensional result of the prior section. There are two main differences: first, this is a setting with multiple characteristics, and second, here the classification is written in terms of characteristics, while the targeted moments are covariances in properties (which number $M = \frac{\hat{K}^2 + \hat{K}}{2}$ where $\hat{K}$ is the number of properties). Consequently, for $\mathbf{H}_i$ to have rank M, there must be more characteristics than properties (keeping in mind that the properties are a subset of characteristics), and that enough of the characteristics

[25] This is consistent with the well-known result that there is a bijection between any positive measure of real numbers and K-dimensional Euclidean space.

must have linearly independent relationships to the covariances of interest.[26]

3.4 Extensions

In this section, I provide intuition for how the theory can be extended in two directions relevant to real world applications. First, I explain why classifications are not written in terms of what would seem like natural characteristics, such as the political weights of domestic producers of objects or elasticities of demand. And second, I explain how a classification used to collate information is consistent with the theory presented here.

In the theory, the frequency of divisions in a particular characteristic should be increasing in the elasticity between distance in the characteristic space and covariances across properties. From this, it would be natural to infer that the properties themselves (which are a subset of the characteristics) should be frequently used to describe classifications. For example, what would be a better way to determine within-code variance in elasticity of demand than describing the code in terms of elasticities of demand? However, this does not happen empirically. Instead, HTS codes tend to be described in terms of things which are unlikely to be of direct interest to a government. For instance, goods are classified according to their power output ("Of an output not exceeding 75 kVA" vs "Of an output exceeding 75 kVA but not exceeding 375 kVA" vs "Of an output exceeding 375 kVA but not exceeding 750 kVA" vs "Of an output exceeding 750 kVA"), their source ("Animal Products" vs "Vegetable Products" vs "Mineral Products" vs ...), or many other features.

This is consistent with the idea that a classification system must be described in terms of verifiable characteristics. By this, I mean that in order to assign two objects to separate codes, the classifier must be able to describe the set of objects that fall into each code in terms of attributes which can be readily verified by, e.g., a customs inspector or a court.[27] As long as there are sufficient numbers of verifiable characteristics with proper relationships to the covariances in properties, Proposition 2 still holds; however, now $\mathbf{H}_i$ only corresponds to verifiable characteristics. In practice, properties of interest to the classifier like elasticities and political weights are unlikely to be verifiable, and so we observe other characteristics used to define classifications instead.

Second, some classification systems are not associated with a particular policy like tariffs. It seems likely that many of the classifications are used to collate information. One example is the 10-digit HTS codes called "statistical reporting categories". Others are the industry classifications used by the U.S. government, such as the North American Industry Classification System (NAICS) .[28] In

[26] If $\mathbf{H}_i$ did not have rank M, it would not mean that $\vec{\Theta}_i$ were under-determined: all elements can be written as functions of only the location of the code in space, the distances in each characteristic, and the distribution $f_i(\mathbf{z})$. However, the "missing" equations are differential equations stemming from the shared distribution which relate all of the covariances to each other. These differential equations are messy, and also requires knowledge of $f_i(\mathbf{z})$ for all codes. Since this distribution is not observable in many settings (e.g. data are only reported at the code level), solutions presented in terms of $f_i(\mathbf{z})$ would not allow taking this theory to the data.

[27] In the case of traded goods, this is critical. Because different codes carry different tariffs, importers will try to enter their goods under lower-tariff classifications if there is ambiguity in the classification. In consequence, there is monitoring at ports of entry to ensure goods have been properly declared and there are frequent court cases disputing the classification of traded goods.In this setting, clear descriptions are key to enforcing the system. For popular reporting on this subject, see NPR's "Planet Money" Episode 501 on the classification of traded goods, "When The Supreme Court Decided Tomatoes Were Vegetables." (NPR December 26, 2013)

[28] Consistent with this story, during the change between the Standard Industry Classification (SIC) codes and the NAICS codes, there was significant public comment and lobbying from users of industrial classifications for information

the Appendix I show how collating information using a classification is isomorphic to implementing a policy. The classifier has full information and wishes to share at least some part of it. To do so, it must describe which objects a particular piece of data encompasses. This is very similar to the problem faced when setting policy, and the model already presented neatly nests these incentives. In particular, in an information setting, properties will be features of objects that the classifier wishes to share information about. The classification determines the level at which information to be shared is aggregated into codes. The covariances in the properties across objects within the code reflect the precision of the communicated information. For instance, the U.S. Government reports the weighted average unit value of all goods that fall within a particular 10-digit HTS code in a particular year. The costs of allowing greater within-code heterogeneity will be related to the changing value of communicating the information as it becomes less precise. In this way, the model extends readily to this case.

4 The relationship between tariffs and the HTS

In the previous section, I derived the optimal classification from the perspective of the classifier. Now, I turn to deriving some predictions from this theory about what we should expect classifications to look like and how they will relate to policy. I begin by making some general theoretical predictions, and then make them more specific in the case of the HTS and tariff policy. Then, I examine these predictions in the data, and show that the relationships between subdivision in the HTS, heterogeneity across products within codes, and tariffs follows what would be expected under a standard tariff-setting policy objective from the literature.

4.1 Relationships between classification and policy

Section 3 describes how a classifier will choose the way objects are grouped into codes and the policy that is set for each code. Now, I describe how we expect the number of codes, the extent of within-code heterogeneity, and the level of policy to be related under this optimal classification. Importantly, these predictions cannot take the form of comparative statics with respect to policy across codes, because neither the policy nor the classification is exogenous – they are endogenously and jointly determined.

Instead, I impose order on the problem in the following way. I take two sets of objects and hold them fixed, so that the properties of objects within each set are invariant to the classification. I then make comparisons across the two sets, and ask how the *average* policy in each set relates to the extent to which the classifier wants to subdivide objects into codes *within* each set. This allows us to ask how the level of policy is related to the extent of subdivision in an "all else equal" sense, by holding the the sets of objects across which comparisons are being made fixed.

When the classifier uses a larger number of codes to subdivide a given set of objects, the average heterogeneity in properties within each code must decrease. Holding classification costs fixed, the extent to which the classifier chooses to do this will covary with the properties that determine the

purposes (e.g. market research for industry) related to the change.

costs of policy mistargeting. The question is whether this will also covary systematically with the level of policy chosen, which depends on how the cost of policy mistargeting (as captured by the vector of weights $\mathbf{L}$) is correlated with the level of policy. Thus the key question is whether the vector of weights $\mathbf{L}$ is larger when the properties imply a higher average policy.

Consider a fixed set of objects. Without loss of generality I can normalize all properties so that the optimal policy is weakly increasing in all of the properties, e.g. by defining a new property which is the inverse of the old one. [29] Suppose that for some property k'' all of the costs of within code covariances, $L_{kk'}$, are (weakly) increasing with respect to k'', with at least one strictly increasing – this implies that, holding the levels of other properties and the costs of classification fixed, the extent of subdivision is increasing in k''.[30] Then, comparing across two sets of objects, when the level of $k''$is higher, both the level of the average policy and the number of subdivisions into codes will increase under the optimal classification.[31]

Intuitively, this means that we should expect the coarseness of the classification to be correlated with policy if the properties that drive higher levels of policy also make mistargeting – driven by greater heterogeneity in properties within a code – more costly. This will not necessarily be true for every property and every policy. It will be the case when the variation in optimal policy across objects and the concavity of the classifier's payout with respect to the policy increase (or decrease) systematically in the property of interest. Otherwise, the relationships between policy and within-code heterogeneity is theoretically ambiguous, and will depend on the relative magnitudes and average levels of characteristics within the codes.

4.2 The theoretical relationship between tariffs and the HTS

I now apply this general insight to the case of tariffs as a specific policy that is targeted using the HTS. I show that, under standard assumptions from the trade policy literature about the drivers of optimal tariff policy, we should expect a greater degree of subdivision in the classification to be associated with lower within-code heterogeneity and higher average tariff levels.

I consider an ad-valorem tariff on imported goods, which maximizes a government objective that is a weighted welfare function that places a higher weight on firm profits than other components of welfare. I adopt a partial equilibrium framework in both production and consumption, and assume that home and foreign varieties are perfect substitutes, that both production and import demand

[29] In other words, the payout from policy is increasing in every property k, $\phi_{\gamma z_k} > 0$, which also implies that the cost of within code covariances are positive, i.e. $L_{kk'} > 0$ for every k and k'.

[30] Formally, the derivative of $L_{kk'}$ with respect to $z_{k''}$ (adjusted to the form of an elasticity) is given by

$$\frac{z_{k''}}{L_{kk'}}\frac{\partial L_{kk'}}{\partial z_{k''}} = \left[\frac{z_{k''}}{\phi_{\gamma z_k}}\phi_{\gamma z_k z_{k''}} + \frac{z_{k''}}{\phi_{\gamma z_{k'}}}\phi_{\gamma z_{k'} z_{k''}} + \frac{z_{k''}}{-\phi_{\gamma\gamma}}\phi_{\gamma\gamma z_{k''}}\right]$$

Therefore $L_{kk'}$ is (weakly) increasing with respect to k'' if the elasticity of $\phi_{\gamma z_k}$ and $\phi_{\gamma z_{k'}}$ with respect to $z_{k''}$ are greater than the elasticity of $\phi_{\gamma\gamma}$ with respect to $z_{k''}$.

Note that if the payout were truly quadratic so that $\vec{\phi}_{\gamma\mathbf{z}}$ and $\phi_{\gamma\gamma}$ were constant and independent of the $\{z_k\}$, then this derivative is always zero. In practice, most payouts are not quadratic, even though we can think of a quadratic payout as a local approximation to a more general payout (in the style of, e.g., Harberger Triangles). Therefore, the reason that the the vector of weights $\mathbf{L}$ would differ across sets of objects is that we are taking a local approximation to the objective within each set.

[31] Another force pushing in the same direction is $\phi_{\gamma\gamma}$ – a smaller $\phi_{\gamma\gamma}$ drives both a higher average level of policy and greater costs of all dimensions heterogeneity.

areconstant elasticity functions of price, and that the country is small.[32] I take the form of the policymaker's payout from a longstanding literature on tariff-setting objectives (e.g. Baldwin (1987), Grossman and Helpman (1994, 1995), and Goldberg and Maggi (1999), Gawande and Bandyopadhyay (2000), McCalman (2004)), and keep the economic setting similar to the classic approach of Baldwin (1987) and Grossman and Helpman (1994). Full details are in the Appendix.

The formula for the optimal ad-valorem tariff (denoted t^*) under these assumptions is:[33]

$$t^* = \frac{\lambda y_0}{\epsilon_p^M M_0 \left(1 + \epsilon_p^M\right) - \lambda y_o \epsilon_p^y}$$

where λ is the additional weight (beyond welfare) put on domestic producer profits in the objective function, y_0 is the domestic supply at the world price, M_0 are imports at the world price, and ϵ_p^y and ϵ_p^M are the price elasticities of domestic supply and import demand. The properties of traded goods in this framework are y_0, λ, ϵ_p^y, and p^w.[34]

The costs of policy mistargeting under the government objective are also increasing in these properties. The key intuition is that these objects all amplify each other in the optimal tariff formula, e.g. the effects of a high domestic output at the world price will be higher when the political weight λ is larger. In other words, all of the cross derivatives of the optimal tariff in these properties are positive. This means the cost of mistargeting will increase in the square of the variance in the tariffs under perfect targeting if the second derivative of the objective held fixed, in a way that is analogous to the logic of Harberger Triangles.[35] Formally, I show in the Appendix that y_0, λ, ϵ_p^y, and p^w) are properties, derive the weights $\mathbf{L}$ on within-code heterogeneity, and show that all are (weakly) increasing in the properties.

This application of the general theory to a trade setting leads to three predictions that we can investigate empirically in the trade data. First, in parts of the HTS where there is more subdivision, there will be less within-code heterogeneity in the properties that drive tariffs. This follows from Proposition 2: heterogeneity in within-code properties is the source of mistargeting and the driving force behind subdivision. Second, where average tariffs are high, there will be greater dispersion in the optimal tariffs that would be set under perfect good-by-good targeting in the absence of the classification.[36] And third, there will be a positive correlation between the average level of the tariff and the degree of subdivision, all else equal.

[32] I could alternatively adopt the trade talks format of Grossman and Helpman (1995) or a large country setting unilateral policy without affecting the overall result; however, these alternative frameworks introduce many additional parameters that would complicate the exposition.

[33] This expression is slightly different than the one used in the literature. The reason for the difference is that this expression is in terms of fundamentals while the standard expression is an equilibrium relationship.

[34] Note that M_0 and ϵ_p^M are not properties because they only enter the payout via the second derivative with respect to the tariff – their cross derivatives of the objective with respect to the tariff are zero.

[35] The second derivative of the objective with respect to the tariff, $\frac{1}{\left(\epsilon_p^M\right)^{-1}} - \lambda\left(\frac{y_0}{M_0}\right)\epsilon_p^y$, is decreasing in the properties, but (largely) in a first-order sense. Consequently, the increased variance in the tariffs under perfect targeting, which is increasing quadratically in the properties, will win out, and the cost of mistargeting will be higher when tariffs are high.

[36] This follows directly from the optimal tariff formula and the fact that all of the cross derivatives of the optimal tariff in the properties are positive.

4.3 Investigating classification and tariffs in the HTS

In this section, I will investigate the predictions of the theory empirically in the context of trade data organized under the HTS system. To implement this empirically, I will take an HS 6-digit code in a given year as a set of objects with some distribution of properties. Then, I will make comparisons across these 6-digit codes – either cross-sectionally or over time – and show how sub-division into HTS 8-digit codes, and heterogeneity in policy and properties across and within those 8-digit codes vary.

Prediction 1 – More subdivision of the HS6 into HTS8 codes is associated with more heterogeneity in properties within codes

The first implication of the theory I investigate is that greater subdivision of an HS 6-digit code reduces heterogeneity in the properties of subsidiary HTS 8-digit codes. The price of imported goods is a property whose heterogeneity I can observe directly. I will also look at heterogeneity in the expenditure share on goods within an HTS 8-digit code as this should be closely correlated with the world price of the good.

In order to look at heterogeneity within 8-digit codes, I require information about what is happening below the 8-digit level. I do this by using the import data disaggregated by the district of entry and district of unlading pair that was introduced in Section 2. I observe the total value of imports and the weighted average price of goods within each HTS8, arriving via different districts of entry and terminating in different districts of unlading.[37] Heterogeneity across districts within an HTS8 will be an underestimate of true heterogeneity across products, but all else equal should be expected to be correlated with it. Using this data, I perform a variance decomposition at the HS 6-digit level: variance within the HS 6-digit code must either lie across the subsidiary HTS 8-digit codes or be within them. The greater the share of variance across the HTS 8-digit codes, the less that is within them, and consequently the more uniform the objects within these codes are.

Table 2 shows that this within-code heterogeneity at the HTS 8-digit level is systematically related to how finely the parent HS 6-digit code has been subdivided. The dependent variables are the share of the variance in the weighted-average price within the HS 6-digit code which is across HTS 8-digit codes and the share of the variance in expenditure share within the HS 6-digit code which is across HTS 8-digit codes. Greater subdivision of a HS 6-digit code (i.e. more HTS8s within the HS6) increases the share of variance in prices and expenditure shares across HTS 8-digit codes. This holds true overall in the data, comparing across codes in Columns (1) and (3). Columns (2) and (4) add HS6 and year fixed effects, showing that the same pattern also holds in changes over time within an HS6 – increases in the subdivision of a given 6-digit code over time are negatively correlated with the amount of variation within subordinate 8-digit codes.

[37] An alternative would be to look at differences across HTS 10-digit codes within an HTS 8-digit code. However, because 10-digit codes are also an endogenous choice of the U.S. governemnt, observed heterogeneity in subdivisions from HTS8 to HTS10 is also likely to be correlated with characteristics of interest. While the district definitions may also be endogenous, they are shared across all goods, and thus differences in heterogeneity across goods are not driven by the government's decision about how to subdivide 8-digit codes.

	(1) Share of Var. Price across HTS8	(2) Share of Var. Price across HTS8	(3) Share of Var. Exp. Share across HTS8	(4) Share of Var. Exp. Share across HTS8
Number of HTS8 in HS6	0.0562***	0.0348***	0.0678***	0.0382***
	(0.00161)	(0.00448)	(0.00197)	(0.00447)
Observations	129,719	129,578	129,719	129,578
Dep. var. mean	0.103	0.1023	0.125	0.125
Dep. var. SD	0.204	0.204	0.221	0.221
FEs		HS6, year		HS6, year
Adj. R2	0.301	0.786	0.371	0.817
Clustering	HS6	HS6	HS6	HS6

Note: *** significant at the 1% level, ** significant at the 5% level, * significant at the 10% level

Table 2: Within-code variance correlated with subdivision

Prediction 2 – Higher average tariffs in the HS6 is associated with more dispersion in tariffs across subsidiary HTS8 codes

In Section 4.2 I established that that in the case of tariffs, a higher value of any of the properties implies greater heterogeneity in the good-by-good optimal tariff. While we do not observe good-by-good optimal tariffs – by definition, tariffs are applied at the HTS 8-digit code level and not the good level – we should expect there to be more dispersion in HTS8 tariffs when there is more dispersion in the (unobserved) optimal good level tariffs for a given set of goods. Therefore, we can investigate this prediction by looking at the relationship between the average tariff at the HS6 level and the dispersion in tariffs across subsidiary HTS8s within that 6-digit code.

Table 3 shows they are positively related, as predicted by the theory Columns (1) shows that this relationship holds overall in the data. Column (2) shows that this is not driven mechanically by a larger number of codes – it holds even conditional on the number of HTS 8-digit codes within the HS6. Columns (3) and (4) add HS6 and year fixed effects, showing that the same relationship holds within HS 6-digit codes as they change over time – i.e. when the average tariff increases, so does the variation in tariffs across subsidiary codes.

	(1) St.Dev. MFN Tariff in HS6	(2) St.Dev. MFN Tariff in HS6	(3) St.Dev. MFN Tariff in HS6	(4) St.Dev. MFN Tariff in HS6
Avg. MFN Tariff in HS6	0.192***	0.172***	0.337***	0.335***
	(0.0161)	(0.0155)	(0.0715)	(0.0713)
Count of HTS8 in HS6		0.00401***		0.00366***
		(0.000203)		(0.000586)
Observations	129,271	129,271	129,113	129,113
Dep. var. mean	0.00639	0.00639	0.00639	0.00639
Dep. var. SD	0.0274	0.0274	0.0274	0.0274
FEs			HS6, year	HS6, year
Adj. R2	0.154	0.237	0.870	0.871
Clustering	HS6	HS6	HS6	HS6

Note: *** significant at the 1% level, ** significant at the 5% level, * significant at the 10% level.

Table 3: Tariffs correlated with subdivision of HS6 codes into HTS8 codes

Prediction 3 – HS6s with higher average tariffs are subdivided into more HTS8 codes

The most substantial implication of the theory derived in Section 4.2 is that higher tariffs will be associated with finer subdivision across different parts of the HTS classification. I investigate this prediction in Table 4 by looking at the relationship between average tariffs at the HS6 level and the number of subsidiary HTS8 codes within that 6-digit code.

Column (1) of Table 4 shows that, across the data as a whole, higher average tariffs are associated with a HS 6-digit codes beings split into a larger number of 8-digit codes. One concern might be that tariffs are correlated with import value, and that codes with a larger volume of imports might be mechanically subdivided more for reasons other than those in the model presented. Column (2) shows that the positive relationship between tariffs and subdivision holds strongly even conditional on the value of trade in the HS6. Columns (3) and (4) add HS6 and year fixed effects to these specifications, looking only at what happens as tariffs change within codes over time relative to other codes within a given year. The coefficients are substantially smaller, as would be expected given that the fixed effects soak up much of the variation that the theory predicts, but the relationship still holds strongly.

	(1) No. HTS8 within HS6	(2) No. HTS8 within HS6	(3) No. HTS8 within HS6	(4) No. HTS8 within HS6
Avg. MFN Tariff in HS6	5.03***	5.08***	0.545***	0.546***
	(0.299)	(0.300)	(0.201)	(0.201)
Imports in HS6 ($US trillions)		40.4		4.82
		(18.0)		(3.30)
Observations	129,271	129,271	129,113	129,113
Dep. var. mean	1.88	1.88	1.88	1.88
Dep. var. SD	1.99	1.99	1.99	1.99
FEs			HS6	HS6
Adj. R2	0.0201	0.0237	0.981	0.981
Clustering	HS6	HS6	HS6	HS6

Note: *** significant at the 1% level, ** significant at the 5% level, * significant at the 10% level.

Table 4: Tariffs correlated with subdivision of HS6 codes into HTS8 codes

5 Endogenous classification bias

Under the theory of endogenous classification, the extent of subdivision, the degree of heterogeneity across objects within codes, and the level of policy may be systematically related. I have shown that this is the case in practice in the trade classification, where the heterogeneity across product varieties within a code is correlated with the level of the tariff applied to that code. While the primary purpose of the HTS classification is to target the application of policy, economists use this data (or similar data from countries other than the U.S.) for essentially all empirical analysis of international trade. In this section, I show how the fact that underlying heterogeneity in products is aggregated up to the code level in data can bias important empirical estimates, and how this bias is systematically correlated with policy. This influences relationships that an econometrician might examine in the data such as that between tariffs and elasticities.

This section proceeds in three parts. First, I lay out the problem and describe the bias, which follows immediately from Jensen's inequality when taking non-linear functions of aggregated data. I also calculate the bias to a second-order approximation. Second, I examine an application to import elasticities estimated using the method of Feenstra (1994), an approach that is widely used in the trade literature. I explain how to correct the bias using observed within-code variation in prices and shares. And third, I show that this bias is endogenous to the classification system and affects the estimated relationship between MFN tariffs and elasticities.

Throughout this section I relegate derivations to the Appendix.

5.1 Non-linear functions of aggregated data yields bias

There are multiple ways in which data aggregation can lead to estimation biases. The particular one that I focus on arises when taking non-linear functions of weighted averages. I show how this bias can be estimated to a second order, as this is useful in settings in which perfectly disaggregate data cannot be observed but moments of the distribution plausibly can be observed or estimated

indirectly.

Suppose there are a set of objects that have been assigned to some finite number of codes. I denote observations at the object level with subscript j, outcomes at the code level with the subscript i, and I use J_i to denote the set of objects j falling into code i. The true model at the level of the objects, j is

$$f(y_j) = \sum_{n=1}^{N} \beta_n g_n\left(x_j^n\right) + \epsilon_j$$

where $f(\cdot)$ and $\{h_n(\cdot)\}_{n=1}^{N}$ are functions (I consider a setting with multiple right-hand side variables and functions which are indexed by n). [38]

However, the econometrician does not observe the outcomes at the j level, and instead observes outcomes at the code i level $y_i = \mathbb{E}[y_j | j \in J_i]$ and $x_i^n = \mathbb{E}\left[x_j^n | j \in J_i\right] \ \forall n$ as is common in classified data. Suppose instead the econometrician runs the specification

$$f(y_i) = \sum_{n=1}^{N} \tilde{\beta}_n g_n(x_i^n) + \tilde{\epsilon}_i$$

then the estimates $\tilde{\beta}_n$ will be biased if either or both of the functions $f(\cdot)$ and $g_n(\cdot)$ are non-linear, following the logic of Jensen's inequality. In order to derive the bias, I impose an additional assumption, which is that the true errors are also independent of the regressors used by the econometrician.[39]

This bias will depend both on the local non-linearity of the functions in question and the within-code variance for all of the left hand side and right hand side variables. I can calculate the bias to a second order approximation. I use σ_{ni}^2 to denote the within-code variance in x_n for code i and σ_{yi}^2 to denote the within-code variance in y for code i. I use $\tilde{\mathbf{X}}$ to denote a matrix of $g_n(x_i^n)$, while $\mathbf{G}$ is a matrix of $\frac{1}{2} g''(x_i^n)\sigma_{ni}^2$ and $\mathbf{F}$ is a vector of $\frac{1}{2} f''(y_i)\sigma_{yi}^2$. The bias is

$$\mathbb{E}\left[\tilde{\beta} | \tilde{\mathbf{X}}\right] - \beta = \left(\tilde{\mathbf{X}}'\tilde{\mathbf{X}}\right)^{-1} \tilde{\mathbf{X}}'(\mathbf{G}\beta - \mathbf{F}) \tag{9}$$

The sign of this bias is ambiguous, and the impact on each element of β depends on the relative size and signs of the corresponding elements of $\mathbf{G}\beta$ and $\mathbf{F}$. This implies the sign of the bias will depend on whether the average creates a larger bias in the outcome or the regressor, and whether the relevant functions are convex or concave. Also, all else equal, these terms will be larger the greater the within-code variances and the more curvature in the relevant functions.

[38] For simplicity, I assume that the standard conditions hold such that the estimate of β_n will be consistent and unbiased. If they do not, then the type of bias I identify here arising from aggregation may interact with other sources of bias in complicated ways.

[39] Formally, I assume $\mathbb{E}\left[\epsilon | \tilde{\mathbf{X}}\right] = 0$,where ϵ is vector of the true errors and $\tilde{\mathbf{X}}$ is a matrix of the function applied to within-code averages, $g_n\left(x_i^n\right)$.

5.2 Application to trade elasticities

In this section, I explain how to apply this insight to estimating elasticities of supply and demand across varieties of a good calculated following the method of Feenstra (1994). First, I explain why this method suffers from bias of the type I identify in the prior subsection. Second, I show how to correct this bias using the microdata introduced in Section 2. I present corrected elasticities and show that this correction matters substantively for estimates of the gains from trade.

The price elasticity of import demand defines how much the quantity of a good imported into a particular country will increase in response to change in the cost of that good. These elasticities are key parameters for answering essentially all questions in empirical trade, such as measuring the gains from trade and the welfare costs of trade policy, among many others. Although the type of bias I highlight would arise under essentially any approach to estimating elasticities (due to the log-log form), I focus on trade elasticities calculated using the Feenstra (1994) method because it is widely used and well-understood.[40]

In the Feenstra (1994) method, trade is assumed to be in the style of Armington, so that each country supplies a unique variety of a given good.[41] Then, it assumes CES supply and demand of product varieties to obtain an estimating equation to jointly estimate elasticities of demand (σ_i) and supply (ω_i)

$$\left(\Delta^k P_{ict}\right)^2 = \theta_{i1}\left(\Delta^k S_{ict}\right)^2 + \theta_{i2}\left(\Delta^k S_{ict}\right)\left(\Delta^k P_{ict}\right) + \epsilon_{ict} \tag{10}$$

where ict subscripts denote good i from country c at time t, P denotes log price, S denotes log share, Δ^k denotes the time and reference-country difference in a variable, and ϵ_{ict} is a mean zero error. The coefficients θ_{i1} and θ_{i2} are functions of the elasticity of supply and demand given by $\theta_{i1} \equiv \frac{\omega_i}{(\sigma_i-1)(1+\omega_i)}$ and $\theta_{i2} \equiv \frac{\omega_i\sigma_i-2\omega_i-1}{(\sigma_i-1)(1+\omega_i)}$). Feenstra (1994) estimates this equation using country dummies as instrumental variables, and then inverts the θ_{i1} and θ_{i2} to obtain σ_i and ω_i.

The method treats a source country-HTS8 code pair as a variety, and so applies the model to country-code level average prices and shares. To consider the bias that will arise if there are in fact multiple product varieties aggregated within a country-code pair, I keep all of the original assumptions except that I allow there to be some set of varieties that fall within each code, and assume that the relationships posited by Feenstra (1994) hold true at the variety level. Formally, I suppose that there are some varieties j that fall into code i and that there is CES utility over the ijc varieties. This implies CES demands over the ij varieties, each of which is uniquely produced in some country c. Thus the true estimating equation is

$$\left(\Delta^k P_{ijct}\right)^2 = \theta_{i1}\left(\Delta^k S_{ijct}\right)^2 + \theta_{i2}\left(\Delta^k S_{ijct}\right)\left(\Delta^k P_{ijct}\right) + \epsilon_{ijct} \tag{11}$$

where the coefficients θ_{i1} and θ_{i2} are functions of the elasticity of supply and demand as previously

[40] It is also well-suited to answering questions about endogenous classification for some of the same reasons that it is so popular as a method for estimating elasticities to begin with. In particular, it does not require data from multiple countries (which cannot be easily constructed at disaggregate levels, precisely because tariff-line level classifications are unilateral choices and not internationally concorded), and it does not require instruments, which is critical when examining thousands of traded goods).

[41] Or, as Feenstra (1994) notes, this is analogous to a continuum of goods which are all sold at the same price, as in this case there is an isomorphism between taste and measure of variety.

defined.

When the true structural relationships are those reflected in equation (11), but the estimating equation (10) is used instead, the estimates of the coefficients θ_{i1} and θ_{i2} (and consequently the parameters of interest σ_i and ω_i) will be biased in the way described by equation (9), because the estimating equation takes non-linear functions of variety-level objects aggregated to the country-code level.[42] Note that this bias will arise even when the true parameters σ and ω are assumed to be uniform across all varieties within each country-code group, which distinguishes this from what is most commonly thought of as "aggregation bias".

Fortunately, the second-order approach used to calculate the bias can also be used to correct these estimates if the second moments of the joint distribution of p_{ijct} and s_{ijct} for every source c and time t are available. The Feenstra (1994) method would yield unbiased estimates if it were applied to correctly constructed price and consumption indices instead of prices and expenditure shares averaged at the code level. The correction I propose is roughly analogous to taking second-order approximations to these indices. In particular, by taking second order approximations of the "true" estimation equations and taking averages within a code, I can obtain an analogue of the unbiased estimation equation.

To obtain corrected estimates, I estimate the following instead of Equation (10):

$$Y_{ict} = \theta_{i1} X_{1ict} + \theta_{i2} X_{2ict} + \upsilon_{ict} \tag{12}$$

where θ_{i1} and θ_{i2} are defined as before, υ_{ict} is a mean zero shock, and Y_{ict}, X_{1ict}, and X_{2ict} are corrected versions of the $\Delta^k S_{ict}$ and $\Delta^k P_{ict}$ that appear in Equation (10). The full expressions for these terms are quite complicated, and are shown in full in the Appendix.[43] Because they involve terms that require knowing the within-code variances and covariances in expenditure and quantity, implementing this correction in practice requires some empirical measure of within-code heterogeneity. To get these terms, I use the variance across districts of entry and unlading.

When I estimate the corrected system of equations,[44] I obtain significantly higher elasticities of substitution and smaller inverse elasticities of foreign export supply (implying larger elasticities of foreign export supply) than if I apply the same Feenstra (1994) technique on HS8-level data

[42] I show formally in the Appendix that taking double-differences does not solve this problem unless there are precisely parallel time-trends in the bias across country pairs. Generically, this will not be true.

[43] There are two main issues. First, Equation (11) involves nonlinear transformations of time and reference country differences of variables. This implies that not only does the correction involve the joint distribution of p_{ijct} and s_{ijct}, but also the correlation of these variables with reference country variables and the autocorrelations in these variables. To remove these nuisance terms and simplify the resulting expressions, I make assumptions about the form of the supply and demand shocks and take a third reference country difference. This greatly simplifies the empirical implementation but with sufficient data is not strictly necessary. And second, Equation (11) references prices. In the data, "prices" (or "unit values") are actually quantity-weighted average prices. However, the observed average expenditures are not quantity weighted. Consequently, when I take averages of a second order approximation, to obtain an expression in terms of the observed quantity-weighted average price I would need to put different weights in the expression than I would use to obtain average shares. Due to the log form, the Feenstra (1994) estimating equation is isomorphic to an expression in terms of aggregate expenditure and aggregate quantity. I thus use this isomorph to get around the trouble of using observed prices.

[44] In practice, I employ the limited-information maximum likelihood (LIML) approach of Soderberry (2015) to estimate these relationships; this also involves including a constant to correct for measurement error and inverse-variance weighting at the supplying country level. In the event that this approach fails to produce economically reasonable estimates of σ_i and ω_i, then I combine the LIML approach of Soderberry (2015) with a grid search following Broda and Weinstein (2006).

without running this correction. I summarize these results in Table 5.[45] Notably, the bias does not always have the same sign – sometimes the corrected elasticity is larger than the original and sometimes it is smaller – and so the average change understates the average absolute value of the change. The difference between the corrected and uncorrected elasticities has quantitatively important implications. If I use the median elasticity and apply the formula of Arkolakis, Costinot, and Rodriguez-Clare (2012) to calculate the implied gains from trade, I find that they are 58% lower using the corrected elasticity.[46]

Estimated σ	25th Percentile	Median	75th Percentile
Uncorrected	1.4	2.0	4.9
Corrected	1.9	3.9	11
Difference	-0.75	1.0	7.7
Abs. value of diff.	0.91	3.6	19
Estimated ω	25th Percentile	Median	75th Percentile
Uncorrected	0.093	0.57	2.5
Corrected	0.052	0.28	0.97
Difference	-1.7	-0.11	0.36
Abs. value of diff.	0.21	0.87	3.3

Table 5: Comparison of corrected and uncorrected elasticity estimates

In addition to yielding substantially different elasticity estimates, the corrected data also fit the Feenstra (1994) model better. Using the correction, I am able to obtain viable estimates for more HTS8s. Without the correction, I do not obtain viable estimates for 2,464 HTS8s (out of a total of 40,428), while with the correction I do not obtain viable estimates for only 1,594 HTS8s; reducing the set of unviable HTS8s by over a third relative to the uncorrected data. Additionally, the data yield viable estimates with fewer constraints. I first estimate the model without any constraints, and if I obtain viable estimates ($\sigma \geq 1$ and $\omega \geq 0$), then I adopt the resulting numbers. If not, I run constrained GMM as in Soderbery (2015) and if I obtain viable estimates I stop. Finally, if not I use a grid search procedure as in Broda and Weinstein (2006). With my correction, many more of the parameters are estimated without constraints relative to the uncorrected estimates. The results are summarized in Table 6.

Method	Unconstrained	Constrained	Grid search	Total estimates
Uncorrected	0.397	0.161	0.442	37,964
Corrected	0.467	0.156	0.376	38,834

Table 6: Comparison of estimation techniques for corrected and uncorrected data

[45] Note that estimates are actually implemented for each HTS8-version – the HS6 changes over time, and I take each HTS8 within a given version of the HS6 to be a separate code to avoid complications related to concordances in the HS6.

[46] If I use the method of Ossa (2015) to take into account variation in the estimated elasticities and the correlation between trade elasticities and import share, I find that gains from trade with the corrected elasticities are roughly one quarter of the gains from trade from uncorrected ones. The difference is because the gains from trade are a convex function of the elasticity, so that the function of the average elasticity is less than the average of the function. Also, the share of expenditure on foreign goods and the elasticity are complements in the gains from trade formula. Taking into account both of these forces (and their importants to different degrees for the corrected and uncorrected elasticities) is the reason for the change in result.

5.3 The degree of bias is endogenous

Bias in the elasticity estimation arises from the fact that there is heterogeneity across true product varieties within an HTS code. I showed in Sections 2 and 4 that the degree of heterogeneity varies substantially in different parts of the HTS, suggesting that the extent of the bias will also vary. Further following the logic of Section 4, the amount of within-code heterogeneity is correlated with the level of the tariff, and so we should expect bias in the elasticity estimates to covary systematically with tariffs as well.

This is precisely what the data show. In Table 7, I show that the difference between the corrected and uncorrected elasticity is negatively (and significantly) correlated with variation in subdivision, i.e. that the bias is smaller in HS 6-digit codes that are more subdivided. In these regressions, the dependent variable is the normalize bias in the elasticities of substitution (σ) and inverse foreign export supply (ω), defined as the difference between the corrected and uncorrected value divided by the average of the corrected and uncorrected value.[47] The right-hand side variable of interest is the number of HTS 8-digit codes within the relevant HS 6-digit code (which has been de-meaned and standardized to aid in the interpretation of the results).

The first two columns examine bias in the elasticity of demand. This bias is positive on average and the impact of greater subdivision at the HS 6-digit level is negative; thus, greater subdivision implies less bias. Furthermore, subdivision explains a large share of the bias – a one standard deviation increase in subdivision explains roughly a third to a half of the variation in the bias. Column (2) adds HS 4-digit code and year fixed effects, to control for potential differences in the effectiveness of subdivision to reduce within-code heterogeneity in different parts of the schedule.[48]

Columns (3) and (4) examine bias in the inverse foreign export supply elasticity. This bias is negative on average and the impact of greater subdivision at the HS 6-digit level is positive; thus, greater subdivision implies less bias. Here, the point estimate is not statistically significant, but a one standard deviation change in classification reduces the bias by a fifth to a quarter. In column (4) I include HS 4-digit code and year fixed effects to control for differences in the effectiveness of subdivision to reduce within-code heterogeneity in different parts of the schedule.

[47] I adopt this approach because some of the estimates can be quite large; as demand and supply approach being perfectly elastic, there is little economic difference but a substantial re-weighting of observations in the regression; in contrast the normalized measure gives a more equal weighting to all observations.

[48] I cannot used fixed effects at a finer level than the HS 4-digit code because the variation on the right-hand side is at the HS 6-digit level and there is no time variation in the outcome.

	(1) Bias in σ (Fraction)	(2) Bias in σ (Fraction)	(3) Bias in ω (Fraction)	(4) Bias in ω (Fraction)
Count of HTS8 in	-0.415***	-0.577***	0.288	0.3854
HS6 (standardized)	(0.154)	(0.210)	(0.201)	(0.257)
Observations	206,525	206,525	206,525	206,525
Dep. var. mean	0.373	0.373	-0.308	-0.308
Dep. var. SD	1.12	1.12	1.51	1.51
FEs		HS4, year		HS4, year
Clustering	HS6	HS6	HS6	HS6
Adj. R2	0.0003	0.133	0.0001	0.136

Note: *** significant at the 1% level, ** significant at the 5% level, * significant at the 10% level

Table 7: Correlation between tariffs and bias

The fact that bias in elasticity estimation is correlated with the level of tariffs suggests that our understanding of the relationships between tariffs and elasticities could be affected by endogenous classification bias. Essentially all theories of tariff setting suggest that the relationship between tariffs and elasticities should be negative (see, e.g., Baldwin (1987); Grossman and Helpman (1994, 1995)).[49] In Table 8, I investigate whether this theoretical relationship comes through in the data, and whether it is affected by the correction for classification bias in the elasticity estimates.

The first two columns of Table 8 relate the MFN tariffs to the corrected elasticities. In Column (1), I regress MFN tariffs against the corrected demand elasticity alone. In Column (2), I also include the inverse foreign export supply elasticities in order to account for any potential terms-of-trade motive for tariffs.[50] I also regress MFN tariffs against the uncorrected demand elasticity alone (Column (3)) and against both the uncorrected demand and inverse foreign export supply elasticities (Column (4)). Since I do not observe import penetration ratios and political weights at the HTS 8-digit level, I control for these missing variables with HS 6-digit code fixed effects in all specifications. This approach should pick up the part of these omitted variables which is correlated across all HTS 8-digit codes within the same HS 6-digit code.

When using the corrected elasticities (Columns (1) and (2)), the results are consistent with the theory and statistically significant. If I regress MFN tariffs on demand elasticities alone (Column (1)), I obtain a negative relationship which is significant at the 1% level. Thus, if the MFN tariffs are the politically optimal ones, they have the relationship with the elasticities as suggested by the theory. If I regress MFN tariffs on demand elasticities and inverse foreign export supply elasticities

[49] An important caveat is that this prediction does not hold in the case that tariffs are motivated solely by terms-of-trade manipulation, and for industries that are not politically organized in Grossman and Helpman (1995) and examined empirically in Goldberg and Maggi (1999). However, Goldberg and Maggi (1999) also find that most industries are organized.

[50] It is not clear whether the MFN tariffs reflect full elimination of terms-of-trade manipulation or not; Ossa (2014) finds that most but not all terms-of-trade manipulation has been removed from applied tariffs. Theoretically (see e.g. Grossman and Helpman (1995)), the MFN tariffs should be between the politically optimal tariffs from an efficient trade agreement and the Nash tariffs of a trade war. The politically optimal tariffs should in equilibrium be negatively correlated with the elasticity of import demand (in this setting the analogue is the elasticity of substitution across varieties) and should also be positively correlated with other variables like political weights and inverse import penetration ratios. The Nash tariffs should incorporate the same relationships as for the politically optimal tariff, but should also be positively correlated with the inverse foreign export supply elasticity.

(Column (2)), I obtain a negative relationship to the import demand elasticities and a positive relationship to the inverse foreign export supply elasticities. Again, these are the relationship with the elasticities as suggested by the theory.[51]

In contrast, when using the uncorrected elasticities (Columns (3) and (4)), the results do not support the theory. If I regress MFN tariffs on uncorrected demand elasticities alone (Column (3)), I obtain a positive relationship. Thus, if the MFN tariffs are the politically optimal ones, they have the opposite relationship with the elasticities as would have been suggested by the theory. If I regress MFN tariffs on uncorrected demand elasticities and inverse foreign export supply elasticities (Column (4)), I obtain a positive relationship to the import demand elasticities and a positive relationship to the inverse foreign export supply elasticities. Again, the sign on the import demand elasticity is wrong, and in this specification although the inverse foreign supply elasticity has the right sign, it is not statistically significant.[52]

Thus, correcting the elasticities changes the correlation between U.S. Column 1 tariffs and import demand elasticities from a positive relationship to a negative one, so correcting the elasticities restores the relationship between tariff and elasticity suggested by the theory. Prior work has been surprised by a positive correlation between tariffs and elasticities. Most notably, Kee, Nicita, and Olarreaga (2008) find that tariffs are positively correlated with elasticity across a large sample of countries. This relationship is attributed to omitted variables – in particular, that political weights might be positively correlated with elasticity, and they conclude that this is an important direction for study in lobbying models. However, the results in Table 8 suggest a simpler explanation: endogenous bias arising from different degrees of within-code heterogeneity in parts of the classification where tariffs are high relative to parts of the classification where tariffs are lower.[53][54]

[51] In this specification, the fixed effect also needs to control for an additional omitted variable: potentially negotiations will have reduced tariffs differentially on different goods. I need the fixed effects to control for any such differential reductions.

[52] In the Appendix, I extend this analysis to the Column 2 tariffs. If we interpret the Column 2 tariffs as the Nash or unilateral tariffs as in Ossa (2014), this extension has the advantage of a clearer relationship to the elasticities, but on the other hand Column 2 tariffs aren't as well measured as MFN tariffs. The results are very similar in spirit to the result in the main text. In Table 8, I also constrain the sample to be the same for all 4 columns. Some observations have corrected elasticities but not uncorrected ones, or vice-versa. By restricting the sample to only observations which have both corrected and uncorrected elasticities, I avoid the possibility that differences in sample composition drive the change in the sign of the correlation with import demand elasticity. In the Appendix, I show that this is not necessary for my result: I obtain the same change in sign if I include all potential observations in both regressions.

[53] An important difference is that Kee, Nicita, and Olarreaga (2008) work at the HS 6-digit level, while I work at the HTS 8-digit level. I believe these results are consistent: the objectives of the WCO (which manages the 6-digit codes) and regression results for the U.S. classifications (available on request) suggest that the WCO creates finer 6-digit codes in parts of the good space where countries tend to have a finer level of classification. This would explain why there are similar impacts at the 6-digit level, and this tends to be true across many countries.

[54] In general, data is not available at a fine enough level to control directly for other forces that might affect the relationship between tariffs and elasticities (such as residual terms of trade manipulation which has not been eliminated through international agreements or the political weight placed on particular industries). Instead, I include fixed effects at the HS6 and year levels. It is possible that remaining variation in omitted variables across HTS8s within an HS6 could contribute to the positive correlation between tariffs and the uncorrected elasticities. Since these factors are present in both the regressions with the corrected and the uncorrected elasticities, the results still suggest that the correction substantially changes our understanding of the relationship between tariffs and elasticities

	(1) MFN tariff	(2) MFN tariff	(3) MFN tariff	(4) MFN tariff
Corrected	-2.66e-8***	-2.66e-8***		
Sigma	(7.22e-9)	(7.22e-9)		
Corrected		1.34e-7***		
Omega		(5.58e-9)		
Uncorrected			2.88e-7	2.88e-7
Sigma			(2.11e-7)	(2.11e-7)
Uncorrected				4.42e-6
Omega				(7.52e-6)
Observations	205,674	205,674	205,674	205,674
Dep. var. mean	0.0490	0.0490	0.0490	0.0490
Dep. var. SD	0.0676	0.0676	0.0676	0.0676
FEs	HS6, year	HS6, year	HS6, year	HS6, year
Adj. R2	0.570	0.570	0.570	0.570
Clustering	HS6	HS6	HS6	HS6

Note: *** significant at the 1% level, ** significant at the 5% level, * significant at the 10% level

Table 8: Correlation between tariffs and elasticities

6 Conclusion

Classifications have always been a nuisance for empirical research. Many researchers know the headache of concording data organized by multiple different classifications or correcting data for changes through time to create a panel. However, in wrestling with the practicalities we may miss the deeper issues related to classification and endogenous bias. While it would be easier if the econometrician could treat data organized under such systems as objective reflections of reality – or at least constant, exogenous ones – this is not consistent with the way classifications appear to work and change over time in practice. As James Scott writes in *Seeing Like A State*, "These state simplifications... did not successfully represent the actual activity of the society they depicted, nor were they intended to; they represented only that slice of it that interested the official observer."

I ask why classification systems in general, and the HTS in particular, are designed in the way that they are. I argue that classifications are used to implement policy, and the design of a classification system trades off the benefits of better targeted policy with the greater costs of designing and enforcing the system. Economists often think about policies as the choice of a policymaker with a particular objective function. Although decisions about how to target those policies are a key function of government and lawmaking, the economics literature has not previously considered the system that defines the mapping from policies to objects as a choice variable. I bridge that gap in this paper.

I show that the degree to which heterogeneity in individual objects is aggregated into codes will vary, and may often be correlated with the policy the classification system is designed to implement. For parts of economic activity that "interest the official observer", great attention may be paid to fine distinctions. For those that do not, perhaps because they will not be taxed anyway, very different things may be lumped together indiscriminately.

Given a standard tariff setting objective, I show that imported product classifications should be more disaggregate where tariffs are higher. Digging into the details of the HTS classification system, I show that, consistent with the theory, where tariffs are high, codes are more subdivided and the imports within them are more homogeneous. Within-code heterogeneity introduces substantial bias in our estimates of trade elasticities. Since within-code heterogeneity is inversely related to tariffs, this causes tariffs to be inversely correlated with aggregation. In fact, the bias changes the sign of the correlation between tariffs and elasticities, and correcting the elasticities restores the relationship predicted by theory.

Although I focus on the HTS, there are many classification systems that are important to empirical research in economics. What policy objectives drives these other classification systems, such as industrial classifications, and how might this affect empirical work using those classifications?

Understanding the motives underlying the design of classification systems may also point toward other ways that we can learn from them. For instance, many classifications including the HTS change substantially over time. What information, if any, can be extracted from these changes? Or, can we learn something about how policymakers' ability to describe or enforce differential policies across groups has changed from the use of finer classifications? Future research should consider what purpose a classification was originally designed to serve, and how that may affect its design.

Bibliography

"A Shirt, A Meat Grinder And The Book Of Everything." NPR Planet Money Episode 501. Dec 6, 2013.

Arkolakis, Costas, Arnaud Costinot, and Andrés Rodríguez-Clare (2012), "New Trade Models, Same Old Gains?," *The American Economic Review* 102(1): 94-130.

Baldwin, Richard (1987), "Politically realistic objective functions and trade policy PROFs and tariffs," *Economics Letters* 24(3): 287-290.

Battigalli, Pierpaolo and Giovanni Maggi (2002), "Rigidity, Discretion and the Costs of Writing Contracts," *The American Economic Review* 92(4): 798-817.

Bernard, Andrew B., J. Bradford Jensen, Stephen J. Redding, Peter K. Schott (2009), "The Margins of U.S. Trade (Long Version)," mimeo.

Besedes, Tibor, James Lake, and Tristan Kohl (2020), "Phase Out Tariffs, Phase In Trade?," *Journal of International Economics* 127: 103385.

Bombardini, Matilde (2008), "Firm heterogeneity and lobby participation," *Journal of International Economics* 75(2): 329-348.

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

Broda, Christian and David E. Weinstein (2008), "Understanding International Price Differences Using Barcode Data," mimeo.

Broda, Christian, Joshua Greenfield, and David E. Weinstein (2017), "From Groundnuts to Globalization: A Structural Estimate of Trade and Growth," *Research in Economics* 71(4): 759-783.

Broda, Christian, Nuno Limao, and David E. Weinstein (2008), "Optimal Tariffs and Market Power: The Evidence," *The American Economic Review* 98(5): 2032-2065.

Chesher, Andrew (1991), "The effect of measurement error," *Biometrika* 78(3): 451–462

Dye, Ronald A. (1985), "Costly Contract Contingencies," *International Economic Review* 26(1): 233-250.

Feenstra, Robert C. (1994), "New Product Varieties and the Measurement of International Prices," *The American Economic Review* 81(4): 157-177.

Feenstra, Robert C., John Romalis, and Peter K. Schott (2002), "U.S. Imports, Exports and Tariff Data, 1989-2001," mimeo.

Gawande, Kishore and Usree Bandyopadhyay (2000), "Is Protection for Sale? Evidence on the Grossman-Helpman Theory of Endogenous Protection," *The Review of Economics and Statistics* 82(1): 139–152.

Goldberg, Pinelopi Koujianou and Giovanni Maggi (1999), "Protection for Sale: An Empirical Investigation," *The American Economic Review* 89(5): 1135-1155.

Gowa, Joanne and Raymond Hicks (2018), ""Big" Treaties, Small Effects: The RTAA Agreements", *World Politics* 70(2): 165-193.

Government Accountability Office (1995), "U.S. Imports: Unit Values Vary Widely for Identically Classified Commodities," GAO Publication No. GGD-95-90, Washington, D.C.: U.S. Government Printing Office..

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

Grossman, Gene M. and Elhanan Helpman (1995), "Trade Wars and Trade Talks," *Journal of Political Economy* 103(4): 675-708.

Gulotty, Robert (2018), "Mismeasured Cooperation: The Trans-Atlantic Origins of Reciprocity in Market Access," mimeo.

Hallak, Juan Carlos and Peter K. Schott (2011), "Estimating Cross-Country Differences in Product Quality," *The Quarterly Journal of Economics* 126(1): 417–474.

Hottman, Colin J., Stephen J. Redding and David E. Weinstein (2016), "Quantifying the Sources of Firm Heterogeneity," *The Quarterly Journal of Economics* 131(3): 1291-1364.

Imbs, Jean and Isabelle Mejean (2015), "Elasticity Optimism," *American Economics Journal: Macroeconomics* 7(3): 43-83.

Imbs, Jean, Haroon Mumtaz, Morten O. Ravn, and Helene Rey (2005), "PPP Strikes Back: Aggregation And the Real Exchange Rate," *The Quarterly Journal of Economics* 120(1): 1-43.

Ludema, Rodney D. and Anna Maria Mayda (2013), "Do Terms-of-Trade Effects Matter for Trade Agreements? Theory and Evidence from WTO Countries" *The Quarterly Journal of Economics* 128 (4): 1837-1893.

Ludema, Rodney D., Anna Maria Mayda, Miaojie Yu and Zhi Yu (2018), "Endogenous Trade Policy in a Global Value Chain: Evidence from Chinese Micro-level Processing Trade," mimeo.

McCalman, Phillip (2004), "Protection for Sale and Trade Liberalization: an Empirical Investigation," *Review of International Economics* 12(1): 81-94.

Kee, Hiau Looi, Alessandro Nicita, and Marcelo Olarreaga (2008), "Import Demand Elasticities and Trade Distortions," *The Review of Economics and Statistics* 90(4): 666-682.

Ossa, Ralph (2014), "Trade Wars and Trade Talks with Data," *The American Economic Review* 104(12): 4104–4146.

Ossa, Ralph (2011), "A 'New Trade' Theory of GATT/WTO Negotiations," *The Journal of Political Economy* 119(1): 122-152.

Pierce, Justin R. and Peter K. Schott (2012), "Concording U.S. Harmonized System Codes over Time" *Journal of Official Statistics* 28(1): 53-68.

Redding, Stephen J. and David E. Weinstein (2018), "Accounting for Trade Patterns," mimeo.

Schott, Peter K. (2003), "One Size Fits All? Heckscher-Ohlin Specialization in Global Production" *The American Economic Review* 93(2): 686-708.

Schott, Peter K. (2004), "Across-Product versus Within-Product Specialization in International Trade," *The Quarterly Journal of Economics* 119(2): 647-678.

Schott, Peter K. (2008), "The Relative Sophistication of Chinese Exports," *Economic Policy* 53:5-49.

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

Soderberry (2015), "Estimating Import Supply and Demand Elasticities: Analysis and Implications." *Journal of International Economics* 96: 1-17 .

Soderberry (2018), "Trade Elasticities, Heterogeneity, and Optimal Tariffs," *Journal of International Economics* 114: 44-62.

Tavares, Samia Costa (2006), "The political economy of the European customs classification," *Public Choice* 129: 107–130.

Appendix

Proofs and derivations from Section 3

The classifier's payout when payouts are separable across objects

If the payout is quadratic across objects in characteristics and policy, then

$$\phi\left(\gamma,\mathbf{z}\right)=\sum_{k=1}^{K}\left(\phi_{z_k}z_k+\frac{1}{2}\phi_{z_kz_k}z_k^2+\sum_{k'\neq k}\phi_{z_{k'}z_k}z_{k'}z_k\right)+\phi_\gamma\gamma+\vec{\phi}_{\gamma\mathbf{z}}\cdot\mathbf{z}\gamma+\frac{1}{2}\phi_{\gamma\gamma}\gamma^2$$

If I define

$$\hat{\phi}\left(\mathbf{z}\right)\equiv\sum_{k=1}^{K}\left(\phi_{z_k}z_k+\frac{1}{2}\phi_{z_kz_k}z_k^2+\sum_{k'\neq k}\phi_{z_{k'}z_k}z_{k'}z_k\right)$$

then I immediately obtain the form for the payout presented in the text.

Proposition 1: For a given code i, the cost of policy mistargeting is

$$\Delta\Phi_i = F_i \sum_{k'=1}^{K} \sum_{k=1}^{K} L_{kk'} \sigma^2_{ikk'}$$

where

$$L_{kk'} = \frac{\phi_{\gamma \mathbf{z}_k} \phi_{\gamma \mathbf{z}_{k'}}}{-2\phi_{\gamma\gamma}}$$

Proof of Proposition 1: I start with Equation (3) from the text and substitute from Equations (6) and (7) the payouts for each object under optimal uniform policy and optimal policy with perfect targeting

$$\begin{aligned}
\Delta\Phi_i &= \int_{\mathbf{z}} \left[\phi\left(\gamma^*\left(\mathbf{z}\right), \mathbf{z}\right) - \phi\left(\hat{\gamma}^*\left(i\left(\mathbf{z}\right)\right), \mathbf{z}\right)\right] f_i\left(\mathbf{z}\right) dz_1 \cdots dz_K \\
&= \int_{\mathbf{z}} \left[\frac{\left(\vec{\phi}_{\gamma \mathbf{z}} \cdot \left[\mathbf{z} - \mathbb{E}\left[\mathbf{z} | \mathbf{z} \in i\right]\right]\right)^2}{-2\phi_{\gamma\gamma}}\right] f_i\left(\mathbf{z}\right) dz_1 \cdots dz_K \\
&\equiv F_i \sum_{k'=1}^{K} \sum_{k=1}^{K} L_{kk'} \sigma^2_{ikk'}
\end{aligned}$$

where the terms $L_{kk'}$ are defined as in the statement of the Proposition.

Mistargeting when payouts are not separable across objects

As in the separable case, I assume a quadratic payout function. I will focus attention on the payout for some object $\mathbf{z}$ with level of policy γ; this payout will have the same terms as in the separable case[55] along with additional terms which capture the impact of policies applied to other objects. These additional terms will capture the first and second order effects of the policy applied to other objects and a set of cross-terms between the policies on other objects and their characteristics. To understand these terms, consider two additional (arbitrary) objects whose characteristics and level of policy I denote with a $'$ and $''$ superscripts. The object $\mathbf{z}'$ will contribute terms $\phi_{\gamma'}\gamma'$, $\frac{1}{2}\phi_{\gamma'\gamma'}\left(\gamma'\right)^2$, and $\vec{\phi}_{\gamma'\mathbf{z}'}\gamma' \cdot \mathbf{z}'$ to the payout from $\mathbf{z}$ by itself. Furthermore, in conjunction with object $\mathbf{z}$, it will contribute $\vec{\phi}_{\gamma\mathbf{z}'}\gamma \cdot \mathbf{z}'$, $\vec{\phi}_{\gamma'\mathbf{z}}\gamma' \cdot \mathbf{z}$, and $\phi_{\gamma\gamma'}\gamma\gamma'$. to this payout. And finally, in conjunction with object $\mathbf{z}''$ it will contribute $\vec{\phi}_{\gamma'\mathbf{z}''}\gamma' \cdot \mathbf{z}''$ and $\phi_{\gamma'\gamma''}\gamma'\gamma''$ to this payout.[56] Thus, I integrate over the set of all $\mathbf{z}'$ and $\mathbf{z}''$to obtain the impacts of policies applied to other objects and obtain the payout for object $\mathbf{z}$:

[55] Technically, the first term now reflects the characteristics of all objects and not just the object's own characteristics, but as it does not affect the payout from policy it is not important to what follows.

[56] Note that cross-derivatives between γ'' and $\mathbf{z}'$ are redundant with cross-derivatives between γ' and $\mathbf{z}''$.

$$\phi\left(\gamma,\vec{\gamma}_{-\mathbf{z}},\mathbf{z},\vec{\mathbf{z}}_{-\mathbf{z}}\right)=\hat{\phi}\left(\vec{\mathbf{z}}\right)+\tilde{\phi}_{\gamma}\gamma+\vec{\phi}_{\gamma\mathbf{z}}\cdot\mathbf{z}\gamma+\frac{1}{2}\phi_{\gamma\gamma}\gamma^{2}+\ldots$$
$$\tilde{\phi}_{\mu}\mu_{\vec{\gamma}}+\tilde{\phi}_{\mu^{2}}\mu_{\vec{\gamma}}^{2}+\frac{1}{2}\tilde{\phi}_{\sigma^{2}}\sigma_{\vec{\gamma}}^{2}+\sum_{k=1}^{K}\tilde{\phi}_{\gamma' z_{k}'}\mathrm{Cov}\left(\vec{\gamma},\vec{z}_{k}\right)$$

In this expression, $\mu_{\vec{\gamma}}$ is the average policy applied to all objects, $\sigma_{\vec{\gamma}}^{2}$ is the variance of the policy applied to all objects, and $\mathrm{Cov}\left(\vec{\gamma},\vec{z}_{k}\right)$ is the covariance across all objects between the policy on an object and the level of characteristic. As in the separable case, $\tilde{\phi}\left(\vec{\mathbf{z}}\right)$ affects the level of the classifier's payout but is invariant to the policy and so does not affect the optimal policy or classification (now, however, it is a function of the characteristics of all objects). Similarly, the terms $\tilde{\phi}_{\gamma}$and $\phi_{\gamma\gamma}$ are constants, while $\tilde{\phi}_{\gamma\mathbf{z}}$ is a vector of constants, which are defined by

$$\begin{aligned}
\tilde{\phi}_{\gamma} =&\phi_{\gamma}+F\vec{\phi}_{\gamma\mathbf{z}'}\cdot\mathbb{E}\left[\mathbf{z}'\right]+F\phi_{\gamma\gamma'}\mu_{\gamma}\\
\tilde{\phi}_{\mu} =&F\left(\phi_{\gamma'}+\vec{\phi}_{\gamma'\mathbf{z}}\cdot\mathbf{z}+\vec{\phi}_{\gamma'\mathbf{z}''}\cdot\mathbb{E}\left[\mathbf{z}''\right]+\vec{\phi}_{\gamma'\mathbf{z}}\cdot\mathbb{E}\left[\mathbf{z}\right]\right)\\
\tilde{\phi}_{\mu^{2}} =&F\left(\phi_{\gamma'\gamma''}+\phi_{\gamma'\gamma'}\right)\\
\tilde{\phi}_{\sigma^{2}} =&F\phi_{\gamma'\gamma'}\\
\tilde{\phi}_{\gamma' z_{k}'} =&F\vec{\phi}_{\gamma' z_{k}'}
\end{aligned}$$

For this setting, I will think about optimal policies set jointly on an arbitrary set of $\mathbf{z}$ in code i; there are also a set of objects in $\mathbf{z}$ which are not in code i (denoted $-i$); these objects are grouped in codes which I take as exogenous[57] but the code-level policies for these objects are chosen jointly. If I integrate the object level payouts across all $\mathbf{z}$ (in both i and $-i$) I obtain the objective of the classifier (following Equation (2) in the text)

$$\int_{\mathbf{z}\in\{i\cup-i\}}\phi\left(\gamma,\vec{\gamma}_{-\mathbf{z}},\mathbf{z},\vec{\mathbf{z}}_{-\mathbf{z}}\right)f\left(\mathbf{z}\right)dz_{1}\cdots dz_{K}=F\left[A+B\mu_{\vec{\gamma}}+\sum_{k=1}^{K}C_{k}\mathrm{Cov}\left(\vec{\gamma},\vec{z}_{k}\right)+\frac{1}{2}D\sigma_{\vec{\gamma}}^{2}+\frac{1}{2}E\mu_{\vec{\gamma}}^{2}\right]$$

where as usual F denotes the measure of objects and where constants A, B, $\{C_{k}\}_{k=1}^{K}$, D, and E are defined

$$\begin{aligned}
A&=\mathbb{E}\left[\tilde{\phi}\left(\vec{\mathbf{z}}\right)\right]\\
B&=\tilde{\phi}_{\gamma}+\tilde{\phi}_{\mu}+\mathbf{C}\cdot\mathbb{E}\left[\mathbf{z}\right]\\
C&=\phi_{\gamma z_{k}}+\tilde{\phi}_{\gamma' z_{k}'}\\
D&=\phi_{\gamma\gamma}+\tilde{\phi}_{\sigma^{2}}\\
E&=2\tilde{\phi}_{\mu^{2}}+\phi_{\gamma\gamma}+\tilde{\phi}_{\sigma^{2}}
\end{aligned}$$

[57] These codes are of arbitrary size, and they include the possibility of perfectly disaggregated codes which would permit perfectly targeted policy.

So that payouts are concave in every policy choice, I assume that $D < 0$.

The classifier will choose $\{\gamma(\mathbf{z})\}_{\mathbf{z} \in i \cup -i}$ to maximize this objective. For each such $\mathbf{z}$, the FOC for optimal policy under perfect targeting implies

$$\gamma^*(\mathbf{z}, \vec{\gamma}_{-\mathbf{z}}, \vec{\mathbf{z}}_{-\mathbf{z}}) = \frac{B + \mathbf{C} \cdot (\mathbf{z} - \mathbb{E}[\mathbf{z}]) + (E - D)\mu_{\vec{\gamma}}}{-D}$$

while under code-level policy the FOC for optimal policy implies

$$\gamma_i^* = \frac{B + \mathbf{C} \cdot (\mathbb{E}[\mathbf{z} | \mathbf{z} \in i] - \mathbb{E}[\mathbf{z}]) + (E - D)\mu_{\vec{\gamma}}}{-D}$$

Using these two results, it follows immediately that $\gamma_i = \mathbb{E}[\gamma^*(\mathbf{z}, \vec{\gamma}_{-\mathbf{z}}, \vec{\mathbf{z}}_{-\mathbf{z}}) | \mathbf{z} \in i]$: the classification does not affect the average level of policy across all objects within a code. Consequently, I can solve for the average level of policy when objects in i are subject to perfect targeting via

$$\begin{aligned}
\mu_{\vec{\gamma}} &= \frac{F_i}{F}\mathbb{E}[\gamma^*(\mathbf{z}, \vec{\gamma}_{-\mathbf{z}}, \vec{\mathbf{z}}_{-\mathbf{z}}) | z \in i] + \frac{F_{-i}}{F}\mathbb{E}[\gamma^*(\mathbf{z}) | z \in -i] \\
&= \frac{B + (E - D)\mu_{\vec{\gamma}}}{-D} \\
\mu_{\vec{\gamma}} &= \frac{B}{E}
\end{aligned}$$

And in turn, this yields optimal levels of policy under both perfect targeting and code-level policy as

$$\begin{aligned}
\gamma^*(\mathbf{z}, \vec{\gamma}_{-\mathbf{z}}, \vec{\mathbf{z}}_{-\mathbf{z}}) &= B\left(\frac{1}{-D} + \frac{1}{E}\right) + \frac{\mathbf{C} \cdot (\mathbf{z} - \mathbb{E}[\mathbf{z}])}{-D} \\
\gamma_i^* &= B\left(\frac{1}{-D} + \frac{1}{E}\right) + \frac{\mathbf{C} \cdot (\mathbf{z} - \mathbb{E}[\mathbf{z}])}{-D}
\end{aligned}$$

With these optimal policies in hand, it is possible to calculate the difference in the classifier's payouts under these different levels of policy. Notably, the classification does not affect the average level of policy across all objects within the code; thus the difference in payouts comes from the difference in the variance of policy and the difference in covariance with characteristics. I calculate these terms as

$$\begin{aligned}
\mathrm{Var}[\gamma_i^*(\mathbf{z}) | \mathbf{z} \in i] &= 0 \\
\mathrm{Var}[\gamma^*(\mathbf{z}, \vec{\gamma}_{-\mathbf{z}}, \vec{\mathbf{z}}_{-\mathbf{z}}) | \mathbf{z} \in i] &= \frac{\sum_{k'=1}^{K}\sum_{k=1}^{K} C_{k'}C_k\sigma^2_{ik'k}}{2D^2} \\
\mathrm{Cov}[\gamma_i^*(\mathbf{z}), \vec{z}_k | \mathbf{z} \in i] &= 0 \\
\mathrm{Cov}[\gamma^*(\mathbf{z}, \vec{\gamma}_{-\mathbf{z}}, \vec{\mathbf{z}}_{-\mathbf{z}}), \vec{z}_k | \mathbf{z} \in i] &= \frac{\sum_{k'=1}^{K} C_{k'}\sigma^2_{ik'k}}{-D}
\end{aligned}$$

where $\sigma^2_{ik'k}$ denotes the covariance between $\vec{z}_{k'}$ and $\vec{z}_k$ in code i.

An overall covariance (or, in the special case, variance) of some outcomes p and q can be decom-

posed into the covariances within subgroups as

$$F\left(\text{Cov}[p,q]+\mu_p\mu_q\right)=F_i\left(\text{Cov}[p,q|s\in i]+\mu_{p,i}\mu_{q,i}\right)+F_{-i}\left(\text{Cov}[p,q|s\in -i]+\mu_{p,-i}\mu_{q,-i}\right)$$

And consequently, the differences in variance in policy and covariance between policy and characteristics will be (which I denote using a delta)

$$\begin{aligned}\Delta\text{Cov}\left(\vec{\gamma},\vec{z}_k\right)&=\frac{F_i}{F}\frac{\sum_{k'=1}^K\sum_{k=1}^K C_{k'}C_k\sigma^2_{ik'k}}{2D^2}\\ \Delta\sigma^2_{\vec{\gamma}}&=\frac{F_i}{F}\frac{\sum_{k'=1}^K C_{k'}\text{Cov}\sigma^2_{ik'k}}{-D}\end{aligned}$$

so that the classifier's payout under perfect targeting is larger by

$$\begin{aligned}&=F\left[D\cdot\frac{F_i}{F}\frac{\sum_{k'=1}^K\sum_{k=1}^K C_{k'}C_k\sigma^2_{ik'k}}{2D^2}+\sum_{k=1}^K C_k\frac{F_i}{F}\frac{\sum_{k'}C_{k'}\sigma^2_{ik'k}}{-D}\right]\\ &=F_i\frac{\sum_{k'=1}^K\sum_{k=1}^K C_{k'}C_k\sigma^2_{ik'k}}{-2D}\end{aligned}$$

this is an expression with exactly the same form as Proposition 1, except now $L_{k'k}=\frac{C_{k'}C_k}{-2D}$. To put this in terms of the fundamentals in the payout,

$$L_{k'k}=\frac{\left(\phi_{\gamma z_{k'}}+F\phi_{\gamma' z'_{k'}}\right)\left(\phi_{\gamma z_k}+F\phi_{\gamma' z'_k}\right)}{-2\left(\phi_{\gamma\gamma}+F\phi_{\gamma'\gamma'}\right)}$$

i.e. mistargeting reflects the curvature in the payout from own policy, the curvature in the payout from cross-object policy, and how characteristics affect the marginal payout of own policy and cross-object policy.

Simplified setting

In this section, I derive the result for the simplified setting with a single characteristic and a single property presented in the text.

The solution follows the steps presented in the text. Once the problem is separable (i.e. presented with all choices and payouts in terms of characteristic space), then I can take the FOC with respect to D_z and obtain

$$\begin{aligned}0&=\frac{\partial}{\partial D_z}\left(\frac{F_zL\sigma_z^2+C}{D_z}\right)\\ &=-\frac{F_zL\sigma_z^2+C}{D_z^2}+\eta_z\frac{F_zL\sigma_z^2}{D_z^2}\\ \sigma_z^2&=\frac{1}{L}\cdot\frac{C}{\eta_z-1}\cdot\frac{1}{F_z}\end{aligned}$$

Then, using the function $i(z)$, I can transform the result back to code space

$$\sigma_i^2 = \frac{1}{L} \cdot \frac{C}{\eta_i - 1} \cdot \frac{1}{F_i}$$

General setting

These results are for a general setting with arbitrarily many characteristics and properties (so long as the characteristics are sufficiently informative about the properties).

Proposition 2: If $\mathbf{H}_i$ has rank M, then the optimal within-code covariances in code i will satisfy

$$\vec{\Theta}_i = \frac{1}{F_i} (\mathbf{H}_i)_L^{-1} \mathbf{L}^{-1} C\vec{1}$$

Proof: I follow the steps presented in the text and take the FOC with respect to every characteristic. Then the K first-order conditions imply

$$F_i \mathbf{H}_i \mathbf{L} \vec{\Theta}_i = C\vec{1}$$

And if $\mathbf{H}_i$ has rank at least M, then its left inverse exists, so that

$$\vec{\Theta}_i = \frac{1}{F_i} (\mathbf{H}_i)_L^{-1} \mathbf{L}^{-1} C\vec{1}$$

which establishes the proof.

Extensions

Multiple policies: Suppose a classification is used to implement multiple policies. I show that this will yield identical results, except now the equilibrium classification will reflect the combined costs of mistargeting in all of the policies.

I will focus on the set of policies implemented by a given classification in equilibrium.[58] I will denote this set of policies with the vector $\vec{\gamma}$, which I index with p and WLOG has P elements. I can then repeat the steps used to derive Proposition 1. I then obtain an analogue of Proposition 1: For a given code i, the cost of policy mistargeting is

$$\Delta\Omega_i = F_i \sum_{k'=1}^{K} \sum_{k=1}^{K} L_{kk'} \sigma_{ikk'}^2$$

where

$$L_{kk'} = \sum_{p=1}^{P} \frac{\phi^i_{z_k \gamma_p} \phi^i_{z_{k'} \gamma_p}}{-2\phi^i_{\gamma_p \gamma_p}}$$

[58] One could imagine that the set of policies to be implemented by a given classification is a choice of the classifier, for example, the U.S. government could adopt one classification for tariffs and one classification for rules of origin, or a combined classification for both policies. This is a combinatorial problem and hard to solve. However, by focusing on the policies implemented by the classification in equilibrium, I can avoid this issue.

Thus, it is clear that the formulas and proofs for the optimal classification will be exactly the same, except that now the cost of policy mistargeting reflects the cost of mistargeting for all policies in $\vec{\gamma}$.

Collating information: Suppose a classification is used to collect information. In particular, for code i, the classifier reports information $\mathbf{y}$, and this information is understood to reflect all parts of the characteristic space in code i. The classifier has a payout function $\psi(\mathbf{y}, \mathbf{z})$ for information $\mathbf{y}$ given about objects with characteristics $\mathbf{z}$.

I again assume a quadratic payout.[59] It follows immediately from the first order conditions that the optimal policy in code i is $\mathbf{y} = \mathbb{E}[\mathbf{z}|\mathbf{z} \in i]$, while under point-by-point classification, the optimal policy is $\mathbf{y} = \mathbf{z}$, assuming more accurate information is always valuable. Furthermore, it is possible to follow the steps outlined in the Multiple Policies extension, above, to see that the analogue to Proposition 1 in this setting: For a given code i, the cost of imperfect collation of information is

$$\Delta\Psi_i = F_i \sum_{k'=1}^{K} \sum_{k=1}^{K} L_{kk'} \sigma^2_{ikk'}$$

where

$$L_{kk'} = \sum_{p=1}^{K} \frac{\psi_{z_k y_p} \psi_{z_{k'} y_p}}{-2\psi_{y_p y_p}}$$

Thus, it is clear that, apart from the change in the weights in Proposition 1, the formulas and proofs for the optimal classification will be exactly the same.

Proofs and derivations from Section 4

The theoretical relationship between tariffs and the HTS

The jumping off point is for a small country which cannot influence the world price of any good. The economic framework has an outside good produced 1-for-1 from only labor (and the country has a sufficient labor endowment such that the good is produced in any equilibrium). Each good g is produced from sector specific capital and labor via a CRS production function. Utility is quasilinear in the outside good and separable across the various goods g. Domestic and foreign output of every good are perfect substitutes. Furthermore, the production and utility functions are such that the domestic output of good g has a constant own-price elasticity ϵ^y_{pg} and demand for imports of good g has a constant own-price elasticity ϵ^M_{pg}. For simplicity, I assume that the allocation of sector specific capital for every good g is such that home exports the outside good and imports all goods g under any equilibrium.

[59] This could be understood as a second order approximation to a more general payout).

The government objective Ω is a weighted welfare function as described in the text

$$\begin{aligned}\Omega &= Y_0 + \sum_g (\lambda_g + 1)\,\mathrm{PS}_g + \mathrm{CS}_g + \mathrm{Rev}_g \\ &= Y_0 + \sum_g \Omega_g\end{aligned}$$

where PS_g denotes producer surplus (capital quasi-rents), CS_g denotes consumer surplus, and Rev_g denotes tariff revenue. Y_0 denotes labor income and Ω_g denotes the payout from good g.

I will think of Ω_g as a function of the tariff, the imports and output at the world price, price elasticities of domestic output and import demand, and the political weight. In particular, I will approximate Ω_g to a second order around the tariff $t = 0$, M_0 and y_0 (the average imports and output across all g at the average world price), the average world price p^w, and the average elasticities of import demand and domestic output ϵ_p^M and ϵ_p^y across all g. The key terms in the approximation (from the perspective of policy mis-targeting) all evaluated at the points described above are

$$\begin{aligned}\frac{d\Omega}{dt} &= \lambda y_0 p^w \\ \frac{d^2\Omega}{dt^2} &= \left(\lambda y_o \epsilon_p^y - \epsilon_p^M M_0 \left(1 + \epsilon_p^M\right)\right) p^w \\ \frac{d^2\Omega}{dt dy_0} &= \lambda p^w \\ \frac{d^2\Omega}{dt d\lambda} &= y_0 p^w \\ \frac{d^2\Omega}{dt d\epsilon_p^y} &= \lambda y_0 p^w \ln\left(p^w\right) \\ \frac{d^2\Omega}{dt dp^w} &= \lambda y_0 \left(1 + \epsilon_p^y\right) \\ \frac{d^2\Omega}{dt dM_0} &= 0 \\ \frac{d^2\Omega}{dt d\left(\epsilon_p^M\right)^{-1}} &= 0\end{aligned}$$

First, this implies the average tariff in the code will be

$$t_{ave} = \frac{\lambda y_0}{\epsilon_p^M M_0 \left(1 + \epsilon_p^M\right) - \lambda y_o \epsilon_p^y}$$

This tariff is increasing in λ, y_0, σ^{-1}, M_0^{-1}, and ϵ_p^y.

Second, the properties are y_0, λ, ϵ_p^y, and p^w. These are the terms with non-zero cross derivatives of the objective and tariff. (Note that σ^{-1} and M_0^{-1} only affect the optimal tariff via the second derivative with respect to the tariff, while p^w is a property even though it doesn't affect the optimal tariff since it has a non-zero cross derivative). I can calculate the cost of mistargeting of the four properties

$$
\begin{aligned}
L_{y_0 y_0} &= \frac{\lambda^2 p^w}{2\left(\epsilon_p^M M_0\left(1+\epsilon_p^M\right) - \lambda y_o \epsilon_p^y\right)} \\
L_{\lambda\lambda} &= \frac{y_0^2 p^w}{2\left(\epsilon_p^M M_0\left(1+\epsilon_p^M\right) - \lambda y_o \epsilon_p^y\right)} \\
L_{\epsilon_p^y \epsilon_p^y} &= \frac{\left(\lambda y_0 \ln\left(p^w\right)\right)^2 p^w}{2\left(\epsilon_p^M M_0\left(1+\epsilon_p^M\right) - \lambda y_o \epsilon_p^y\right)} \\
L_{p^w p^w} &= \frac{\left(\lambda y_0\left(1+\epsilon_p^y\right)\right)^2}{2\left(\epsilon_p^M M_0\left(1+\epsilon_p^M\right) - \lambda y_o \epsilon_p^y\right) p^w} \\
L_{y_0 \lambda} &= \frac{\lambda y_0 p^w}{\epsilon_p^M M_0\left(1+\epsilon_p^M\right) - \lambda y_o \epsilon_p^y} \\
L_{y_0 \epsilon_p^y} &= \frac{\lambda^2 p^w y_0 \ln\left(p^w\right)}{\epsilon_p^M M_0\left(1+\epsilon_p^M\right) - \lambda y_o \epsilon_p^y} \\
L_{y_0 p^w} &= \frac{\lambda^2 y_0\left(1+\epsilon_p^y\right)}{\epsilon_p^M M_0\left(1+\epsilon_p^M\right) - \lambda y_o \epsilon_p^y} \\
L_{\lambda \epsilon_p^y} &= \frac{\lambda y_0^2 p^w \ln\left(p^w\right)}{\epsilon_p^M M_0\left(1+\epsilon_p^M\right) - \lambda y_o \epsilon_p^y} \\
L_{\lambda p^w} &= \frac{\lambda y_0^2\left(1+\omega^{-1}\right)}{\epsilon_p^M M_0\left(1+\epsilon_p^M\right) - \lambda y_o \epsilon_p^y} \\
L_{\epsilon_p^y p^w} &= \frac{\left(\lambda y_0\right)^2 \ln\left(p^w\right)\left(1+\epsilon_p^y\right)}{\epsilon_p^M M_0\left(1+\epsilon_p^M\right) - \lambda y_o \epsilon_p^y}
\end{aligned}
$$

As can be immediately verified, all of these objects are weakly increasing in the properties λ, y_0, ϵ_p^y, and also the factors which increase the tariff only via the second order condition M_0^{-1}, and ϵ_p^M.

Derivations and supplementary tables from Section 5

Derivation of the bias:

In the text, I present an expression for a second-order approximation to the bias which arises when estimation is nonlinear and the econometrician uses functions of weighted averages as opposed to weighted averages of the function. If $\mathbb{E}\left[\epsilon|\tilde{\mathbf{X}}\right] = 0$, then this bias is

$$
\mathbb{E}\left[\tilde{\beta}|\tilde{\mathbf{X}}\right] - \beta = \left(\tilde{\mathbf{X}}'\tilde{\mathbf{X}}\right)^{-1}\tilde{\mathbf{X}}'\left(\mathbf{G}\beta - \mathbf{F}\right)
$$

where $\tilde{\mathbf{Y}}$ is a matrix of $f\left(y_i\right)$, $\tilde{\mathbf{X}}$ is a matrix of $g_n\left(x_i^n\right)$, $\mathbf{G}$ is a matrix of $\frac{1}{2}g''\left(x_i^n\right)\sigma_{ni}^2$ and $\mathbf{F}$ is a vector of $\frac{1}{2}f''\left(y_i\right)\sigma_{yi}^2$ (note that I use σ_{ni}^2 to denote the within-code variance in x_n for code i and σ_{yi}^2 to denote the within-code variance in y for code i.

I start with the standard expression for $\tilde{\beta}$ and take the conditional expectation:

$$\tilde{\beta} = \left(\tilde{\mathbf{X}}'\tilde{\mathbf{X}}\right)^{-1}\tilde{\mathbf{X}}'\tilde{\mathbf{Y}}$$
$$\mathbb{E}\left[\tilde{\beta}|\tilde{\mathbf{X}}\right] = \left(\tilde{\mathbf{X}}'\tilde{\mathbf{X}}\right)^{-1}\tilde{\mathbf{X}}'\mathbb{E}\left[\tilde{\mathbf{Y}}|\tilde{\mathbf{X}}\right]$$

I then simplify $\mathbb{E}\left[\tilde{\mathbf{Y}}|\tilde{\mathbf{X}}\right]$ through a local second-order approximation. In particular, for observation i within code j, I can take a second order expansion for $f(\cdot)$ around $y_i = \mathbb{E}[y_j|j \in J_i]$ and then take expectations:

$$f(y_j) \approx f(y_i) + (y_j - y_i) f'(y_i) + \frac{1}{2}(y_j - y_i)^2 f''(y_i)$$
$$\mathbb{E}[f(y_j)|j \in J_i] \approx f(y_i) + \frac{1}{2}f''(y_i)\sigma_{yi}^2$$
$$f(y_i) \approx \mathbb{E}[f(y_j)|j \in J_i] - \frac{1}{2}f''(y_i)\sigma_{yi}^2$$

Furthermore, returning to the true model presented in the text, I have that

$$f(y_j) = \sum_{n=1}^{N} \beta_n g_n\left(x_j^n\right) + \epsilon_j$$

and by taking a second-order expansion around $x_i^n = \mathbb{E}[g_n(x_j)|j \in J_i]$ I obtain

$$h_n\left(x_j^n\right) \approx g_n\left(x_i^n\right) + \left(x_j^n - x_i^n\right) g_n'\left(x_i^n\right) + \frac{1}{2}\left(x_j^n - x_i^n\right)^2 g_n''\left(x_i^n\right)$$
$$\mathbb{E}[h_n(x_j)|j \in J_i] \approx g_n\left(x_i^n\right) + \frac{1}{2}g_n''\left(x_i^n\right)\sigma_{ni}^2$$

so that

$$f(y_i) \approx \sum_{n=1}^{N} \beta_n \left[g_n\left(x_i^n\right) + \frac{1}{2}g_n''\left(x_i^n\right)\sigma_{ni}^2\right] - \frac{1}{2}f''(y_i)\sigma_{yi}^2 + \mathbb{E}[\epsilon_j|j \in J_i]$$

which implies

$$\mathbb{E}\left[\tilde{\mathbf{Y}}|\tilde{\mathbf{X}}\right] = \mathbf{G}\beta - \mathbf{F}$$

which completes the derivation.

Bias in a difference and double-difference setting:

In the text, I assert that the bias arising from applying non-linear estimated to weighted averages will not be corrected by taking either simple differences or differences-in-differences. I show that this is the case by deriving the bias under both strategies; it is then clear that the bias is generically not 0 in these settings.

The true model can be expressed as

$$\Delta_t f\left(y_{jt}\right) = \sum_{n=1}^{N} \beta_n \Delta_t g_n\left(x_{jt}^n\right) + \epsilon_{jt}$$

where I define $\Delta_t N_t \equiv N_t - N_{t-1}$, and this is perfectly analogous to the true model presented in the derivation without first differences.

I define $\Delta_t \tilde{\mathbf{Y}}$ as a matrix of $\Delta_t f\left(y_{it}\right)$, and $\Delta_t \tilde{\mathbf{X}}$ as a matrix of $\Delta_t g_n\left(x_{it}^n\right)$. I then use the same first-order expansions as in the derivation without first differences at each point t (and make analogous definitions $y_{it} = \mathbb{E}\left[y_{jt} | j \in J_i\right]$ and $x_{it}^n = \mathbb{E}\left[g_n\left(x_{jt}\right) | j \in J_i\right]$) and take expectations as before:

$$f\left(y_{it}\right) \approx \mathbb{E}\left[f\left(y_{jt}\right) | j \in J_i\right] - \frac{1}{2} f''\left(y_{it}\right) \sigma_{yit}^2$$
$$\mathbb{E}\left[g_n\left(x_j\right) | j \in J_i\right] \approx g_n\left(x_i^n\right) + \frac{1}{2} g_n''\left(x_i^n\right) \sigma_{ni}^2$$

and as before (but now taking into account that variables have been first-differenced) and taking conditional expectations[60] (and note I assume $\mathbb{E}\left[\epsilon_{jt} | \Delta_t \tilde{\mathbf{X}}\right] = 0$ which is analogous to the assumption made earlier)

$$\hat{\beta} = \left(\Delta_t \tilde{\mathbf{X}}' \Delta_t \tilde{\mathbf{X}}\right)^{-1} \Delta_t \tilde{\mathbf{X}}' \Delta_t \tilde{\mathbf{Y}}$$
$$\begin{aligned} \mathbb{E}\left[\hat{\beta} | \Delta_t \tilde{\mathbf{X}}\right] &= \left(\Delta_t \tilde{\mathbf{X}}' \Delta_t \tilde{\mathbf{X}}\right)^{-1} \Delta_t \tilde{\mathbf{X}}' \mathbb{E}\left[\Delta_t \tilde{\mathbf{Y}} | \Delta_t \tilde{\mathbf{X}}\right] \\ &= \left(\Delta_t \tilde{\mathbf{X}}' \Delta_t \tilde{\mathbf{X}}\right)^{-1} \Delta_t \tilde{\mathbf{X}}' \left(\Delta_t \mathbf{G} \beta - \Delta_t \mathbf{F}\right) \end{aligned}$$

where $\Delta_t \mathbf{G}$ is a matrix of $\Delta_t \left[\frac{1}{2} g''\left(x_{it}^n\right) \sigma_{nit}^2\right]$ and $\Delta_t \mathbf{F}$ is a vector of $\Delta_t \left[\frac{1}{2} f''\left(y_{it}\right) \sigma_{yit}^2\right]$. It is clear from this expression that neither $\Delta_t \mathbf{G}$ nor $\Delta_t \mathbf{F}$ will generally be zero.

For double differences, I adopt perfectly analogous notation (for which I skip definitions are they're obvious) and all the analogous assumptions. The true model can be expressed as

$$\Delta_k \Delta_t f\left(y_{jtk}\right) = \sum_{n=1}^{N} \beta_n \Delta_k \Delta_t g_n\left(x_{jtk}^n\right) + \epsilon_{jtk}$$

and by using the same expansions I can obtain

$$\mathbb{E}\left[\hat{\beta} | \Delta_k \Delta_t \tilde{\mathbf{X}}\right] = \left(\Delta_k \Delta_t \tilde{\mathbf{X}}' \Delta_k \Delta_t \tilde{\mathbf{X}}\right)^{-1} \Delta_k \Delta_t \tilde{\mathbf{X}}' \left(\Delta_k \Delta_t \mathbf{G} \beta - \Delta_k \Delta_t \mathbf{F}\right)$$

Again, it is clear that this bias need not be zero.

[60] Also note I use $\hat{\beta}$ to denote the biased estimate in differences to distinguish it from $\tilde{\beta}$ which is the biased estimate without differences.

Correction of bias in method of Feenstra (1994):

In the text, I discussion to the Feenstra method (introduced in Equation (12) in the text). In this section of the appendix, I provide formulas for Y_{ict}, X_{1ict}, and X_{2ict} and I present the derivation of this correction (and why I have adopted this approach).

First, the corrected estimating equation is

$$Y_{ict} = \theta_1 X_{1ict} + \theta_2 X_{2ict} + \upsilon_{ijct}$$

where

$$\begin{aligned}
Y_{ict} \equiv & \Delta^{k'}\left[\left(\Delta_k P_{ict}\right)^2\right] - \Delta^{k'}\left[\Delta_k P_{ict}\Delta_k\left(\mathrm{CV}\left(e_{ijct}\right)\right)^2\right]\ldots \\
& + \Delta^{k'}\left[\Delta_k P_{ict}\Delta_k \mathrm{CV}\left(q_{ijct}\right)^2\right]\ldots \\
& + \Delta^{k'}\left[\left(\mathrm{CV}\left(e_{ijct}\right)\right)^2 + \left(\mathrm{CV}\left(e_{ijkt}\right)\right)^2 + \left(\mathrm{CV}\left(q_{ijct}\right)\right)^2 + \left(\mathrm{CV}\left(q_{ijkt}\right)\right)^2\right]\ldots \\
& - 2\Delta^{k'}\left[\frac{\mathrm{Cov}\left(\Omega_{it}e_{ijct}, \Omega_{it}q_{ijct}\right)}{e_{ict}q_{ict}} + \frac{\mathrm{Cov}\left(\Omega_{it}e_{ijkt}, \Omega_{it}q_{ijkt}\right)}{e_{ikt}q_{ikt}}\right] \\
X_{1ict} \equiv & \Delta^{k'}\left[\left(\Delta_k E_{ict}\right)^2\right] - \Delta^{k'}\left[\Delta_k E_{ict}\Delta_k\left(\mathrm{CV}\left(e_{ijct}\right)\right)^2\right]\ldots \\
& + \Delta^{k'}\left[\left(\mathrm{CV}\left(e_{ijct}\right)\right)^2 + \left(\mathrm{CV}\left(e_{ijkt}\right)\right)^2\right] \\
X_{2ict} \equiv & \Delta^{k'}\left[\left(\Delta_k E_{ict}\right)\left(\Delta_k P_{ict}\right)\right]\ldots \\
& - \Delta^{k'}\left[\left(\frac{\Delta_k E_{ict}}{2} - \frac{\Delta_k P_{ict}}{2}\right)\Delta_k\left(\mathrm{CV}\left(e_{ijct}\right)\right)^2\right]\ldots \\
& + \Delta^{k'}\left[\left(\mathrm{CV}\left(e_{ijct}\right)\right)^2 + \left(\mathrm{CV}\left(e_{ijkt}\right)\right)^2\right]\ldots \\
& + \Delta^{k'}\left[\frac{\Delta_k E_{ict}}{2}\Delta_k\left(\mathrm{CV}\left(q_{ijct}\right)\right)^2\right]\ldots \\
& - \Delta^{k'}\left[\frac{\mathrm{Cov}\left(\Omega_{it}e_{ijct}, \Omega_{it}q_{ijct}\right)}{e_{ict}q_{ict}} + \frac{\mathrm{Cov}\left(\Omega_{it}e_{ijkt}, \Omega_{it}q_{ijkt}\right)}{e_{ikt}q_{ikt}}\right]
\end{aligned}$$

Second, I present the derivation of the estimating equation above. My assumptions are largely the same as in Feenstra (1994); however, I put greater structure on the supply and demand shocks across varieties within a code and across time. I also introduce some additional notation for my setting. I describe all of these assumptions and notation below.

I start by defining notation. Following notation introduced in the body of the paper, I denote a good with ij subscripts and a code by an i subscript. I denote by Ω_{it} the measure of goods ij in code i at time i. I use c subscripts to denote variables for supplying country c, and t subscripts to denote variables at time t. Thus, for good ij sourced from country c at time t, I denote expenditure by e_{ijct}, the quantity sourced by q_{ijct}, and the price by p_{ijct}. I denote the aggregate expenditure on good ij in time t regardless of country of origin by e_{ijt}. Throughout, a capitalized variable denotes the log of a variable, a Δ_k denotes a reference country difference, and a $\Delta^{k'}$ denotes a time and reference country difference, i.e. $\Delta_k E_{ct} \equiv E_{ct} - E_{kt}$ and $\Delta^k E_{ct} \equiv \left(\ln e_{ct} - \ln e_{c(t-1)}\right) - \left(\ln e_{kt} - \ln e_{k(t-1)}\right)$.

I assume the economic structure from Feenstra (1994): a reduced form constant elasticity export

supply curve and a constant elasticity of substitution utility function across varieties. As in Feenstra (1994), the economic structure follows Armington (1969): each country is assumed to perfectly competitively supply a unique variety of the good. Following these assumptions, price and expenditure for a given variety are jointly determined by the demand and supply equations

$$e_{ijct} = b_{ijct}\left(\frac{p_{ijct}}{\phi_{ijt}}\right)^{1-\sigma_i} e_{ijt}$$
$$e_{ijct} = \exp\left(-\frac{a_{ijct}}{\omega_i}\right) p_{ijct}^{\frac{\omega_i+1}{\omega_i}}$$

where σ_i denotes the elasticity of substitution across varieties, ω_i denotes the inverse foreign export supply elasticity, b_{ijct} is a demand shock, and a_{ijct} is a production cost shock (alternatively, $\exp\left(-\frac{a_{ijct}}{\omega_i}\right)$ is an inverse productivity shock). The exponential form of the export supply shock and multiplicative form of the demand shock are adopted to conform to the notation of Feenstra (1994).

I next turn to the supply and demand shocks, $\exp\left(-\frac{a_{ijct}}{\omega_i}\right)$ and b_{ijct}. First, I decompose these shocks (in logs) into a country-good average across time, a code-supplier-time shock, a common good-time shock shared by all suppliers, and an idiosyncratic good-supplier-time shock, i.e.

$$a_{ijct} = \hat{a}_{ijc} + \hat{a}_{ijt} + \hat{a}_{ijct}$$
$$B_{ijct} = \hat{B}_{ijc} + \hat{B}_{ijt} + \hat{B}_{ijct}$$

Furthermore, I assume that the $\hat{a}_{ijt}$ and $\hat{B}_{ict}$ are shared across suppliers, and the $\hat{a}_{ijct}$ and the $\hat{B}_{ijct}$ are supplier-good-time idiosyncratic shocks. I normalize the time-dependent terms so that they are mean zero – any average across time of the time dependent components can be absorbed by the $\hat{a}_{ijc}$ and $\hat{B}_{ijc}$. I further assume that the $\hat{a}_{ijt}$, $\hat{a}_{ijct}$, $\hat{B}_{ijt}$, and $\hat{B}_{ijct}$ are mutually uncorrelated (including across sources, i.e. $\hat{a}_{ijct}$ is uncorrelated with $\hat{a}_{ijc't}$ for $c' \neq c$).[61] Finally, I assume heterogeneous variances in the $\hat{a}_{ijct}$ and $\hat{B}_{ijct}$ across source countries such that the ratio of these variances is generically different.

These assumptions on the supply and demand shocks are consistent with the assumptions in Feenstra (1994), but the assumption that the components of the supply and demand shocks are mutually uncorrelated is somewhat stronger. In particular, Feenstra (1994) assumes the residuals of the supply and demand shocks after taking time and reference differences are uncorrelated with each other. I maintain this assumption, but add the stronger assumption that these residuals are also uncorrelated with the parts of the supply and demand shocks which are removed when taking time and reference differences. More formally, the first part of the assumption in Feenstra (1994) is that $\mathbb{E}\left[\left(\Delta^k a_{ijct}\right)\left(\Delta^k B_{ijct}\right)\right] = 0$. This continues to hold in my framework – $\Delta^k a_{ijct} = \hat{a}_{ijct}$ and $\Delta^k B_{ijct} = \hat{B}_{ijct}$. Furthermore, $\hat{a}_{ijct}$ and $\hat{B}_{ijct}$ are uncorrelated mean 0 shocks – so that the expectation of their product is zero. The second part of the assumption in Feenstra (1994) is heteroskedasticity – that there exist two source countries c and c' such that $\frac{\mathbb{V}\left[\Delta^k a_{ijct}\right]}{\mathbb{V}\left[\Delta^k a_{ijct}\right]} \neq \frac{\mathbb{V}\left[\Delta^k a_{ijc't}\right]}{\mathbb{V}\left[\Delta^k a_{ijc't}\right]}$;

[61] Note that a_{ijct} and $a_{ijc't}$ are still be correlated with each other for $c \neq c'$ due to the shared $\hat{a}_{ijt}$ term).

my assumption on the variances of the $\hat{a}_{ijct}$ and $\hat{B}_{ijct}$ across countries ensure that this condition will be satisfied.

Finally, my approach requires that there are at least five countries, where two of those countries share at least three different sets of shared consecutive years with the remaining source countries. After taking second reference differences, this implies 3 remaining sources, and after taking time differences, two observations for each source country. Three source countries is the minimum to identify three parameters of interest: a supply and demand elasticity plus a constant. And two time periods for each source is the minimum to have variances and covariances across time in price and quantity, which are the dependent and independent variables in the regression. Effectively, this method requires one additional source country (with data in the necessary time periods) relative to Feenstra (1994).

With my assumptions in hand, I turn to deriving the estimating equation itself. I start by multiplying both the demand and supply equations by the measure of varieties in code i at time t, Ω_{it}, and take logs to obtain

$$\tilde{E}_{ijct} = B_{ijct} + (1-\sigma_i)(P_{ijct} - \Phi_{ijt}) + \tilde{E}_{ijt}$$
$$\tilde{E}_{ijct} = -\frac{a_{ijct}}{\omega_i} + \left(\frac{\omega_i + 1}{\omega_i}\right) P_{ijct} + \ln(\Omega_{it})$$

where I define $\tilde{E}_{ijct} \equiv \ln(\Omega_{it} e_{ijct})$ and $\tilde{E}_{ijt} \equiv \ln(\Omega_{it} e_{ijt})$ as variety adjusted expenditures). These equations are in terms of the "true" variety ij, but this is not a substantive departure from Feenstra (1994) – this is simply a question of data, not theory. However, there are two true departures from Feenstra (1994) here, although neither is substantive. Feenstra (1994) uses expenditure shares and does not adjust by the measure of varieties; however, in time and reference country differences both adjustments drop out, and so these equations are exactly equivalent to the analogous equations in Feenstra (1994).[62]

I then take reference country differences of both equations to remove the log of the price index and variety-adjusted expenditure to obtain

$$\Delta_k \tilde{E}_{ijct} = \Delta_k B_{ijct} + (1-\sigma_i)\Delta_k P_{ijct}$$
$$\Delta_k \tilde{E}_{ijct} = -\frac{\Delta_k a_{ijct}}{\omega_i} + \left(\frac{\omega_i + 1}{\omega_i}\right) \Delta_k P_{ijct}$$

At this point, Feenstra (1994) also takes a time difference of both equations; I omit this step for now, as doing so would bury the time difference inside non-linear functions, complicating the second-order approximations.[63]

[62] To put this another way, if all of the necessary terms were observed, I could adopt either this demand equation or the Feenstra (1994) one and, by following the Feenstra (1994) method the rest of the way get exactly the same regressors and regressands for any dataset. But making these changes makes the second-order approximations somewhat simpler to follow, which is hwy I adopt them.

[63] In particular, it would make the approximations depend on the serial correlations of price and quantity at the source level. In general, there is no reason to expect these serial correlations to be the same across source countries, and so the second reference difference strategy I employ would not eliminate these terms. This then would be a problem for settings with no microdata, as governments are unlikely to target these moments.This difficulty could be overcome by taking a second time difference in addition to a second reference difference; this would remove the

With the exception of the skipped time differences, I continue to follow Feenstra (1994) after the step I have skipped. I manipulate both equations to move the price and expenditure terms onto the left-hand side, and then I multiply both equations. By doing so, I obtain

$$(\Delta_k P_{ijct})^2 - \theta_1 \left(\Delta_k \tilde{E}_{ijct}\right)^2 - \theta_2 \left(\Delta_k P_{ijct}\right) \left(\Delta_k \tilde{E}_{ijct}\right) = u_{ijct}$$

where I define

$$\begin{aligned} u_{ijct} &\equiv -\frac{1}{\omega_i} \left(\Delta_k a_{ijct}\right) \left(\Delta_k B_{ijct}\right) \\ \theta_1 &\equiv \frac{\omega_i}{\left(\sigma_i - 1\right)\left(\omega_i + 1\right)} \\ \theta_2 &\equiv \frac{\left(\sigma_i - 1\right)\omega_i - \left(\omega_i + 1\right)}{\left(\sigma_i - 1\right)\left(\omega_i + 1\right)} \end{aligned}$$

Although Feenstra (1994) defines θ_1 and θ_2 in terms of ρ instead of ω, it is straightforward to show that my definitions are equivalent to his by plugging in the definition of ρ in terms of σ and ω. I then follow Feenstra (1994) in moving the $\left(\Delta_k \tilde{E}_{ijct}\right)^2$ and $\left(\Delta_k P_{ijct}\right) \left(\Delta_k \tilde{E}_{ijct}\right)$ on to the righthand side. Finally, I take a time difference and a second reference country difference to obtain

$$\Delta^{k'} \left(\Delta_k P_{ijct}\right)^2 = \theta_1 \Delta^{k'} \left(\Delta_k \tilde{E}_{ijct}\right)^2 + \theta_2 \Delta^{k'} \left[\left(\Delta_k P_{ijct}\right) \left(\Delta_k \tilde{E}_{ijct}\right)\right] + \upsilon_{ijct}$$

The Feenstra (1994) method relies on the average of the error being mean zero across time for a given source country. The last task (even if a rather straightforward one) is for me to show that this holds here as well, i.e. $\mathbb{E}_t \left[\upsilon_{ijct}\right] = 0$, where I place a t subscript on the expectation to denote an expectation across time (with the ijc given). Under my assumption on the decomposition of the shocks,

$$u_{ijct} = -\frac{1}{\omega_i} \left(\Delta_k \hat{a}_{ijc} + \Delta_k \hat{a}_{ijct}\right) \left(\Delta_k \hat{B}_{ijc} + \Delta_k \hat{B}_{ijct}\right)$$

This term is not mean zero because of the $\Delta_k \hat{a}_{ijc}$ and $\Delta_k \hat{B}_{ijc}$ terms. However, the additional time and second reference differences of this expression will eliminate the $\Delta_k \hat{a}_{ijc}$ and $\Delta_k \hat{B}_{ijc}$ terms (since these are constant over time and uncorrelated with the $\Delta_k \hat{a}_{ijct}$ and $\Delta_k \hat{B}_{ijct}$), yielding an error which is mean zero for each source country.

$$\upsilon_{ijct} = \Delta^{k'} u_{ijt} = -\frac{1}{\omega_i} \Delta^{k'} \left[\left(\Delta_k \hat{a}_{ijct}\right) \left(\Delta_k \hat{B}_{ijct}\right)\right]$$

Next, I construct second order approximations to the nonlinear terms in the estimating equation: $\left(\Delta_k P_{ijct}\right)^2$, $\left(\Delta_k \tilde{E}_{ijct}\right)^2$, and $\left(\Delta_k P_{ijct}\right) \left(\Delta_k \tilde{E}_{ijct}\right)$. I will then take these expressions in time and second reference differences to obtain the estimating equation (note the approximation need not take into account the time and second reference differences because the estimating equation is linear in these terms).

One complication is that the observed prices are quantity weighted while expenditures are variety-

additional covariances at the cost of reducing the amount of data available to the estimator.

weighted. This mean that if I take second order expansions around weighted average prices and variety weighted expenditures, when I then average across the varieties j within a given code i at least one of the first-order terms will not drop out (depending on the weights in the average). This problem could be handled directly at the cost of more complicated approximations, but it is simpler to instead use the identity $p_{ijct} \equiv \frac{e_{ijct}}{q_{ijct}}$. This permits expansions around variety-weighted expenditure and variety-weighted quantity, so that all the first order terms drop out when taking a variety-weighted average.

Following this logic, I take second order approximations to $\left(\Delta_k \tilde{E}_{ijct}\right)^2$, $\left(\Delta_k \tilde{Q}_{ijct}\right)^2$ (where I define $\tilde{Q}_{ijct} \equiv \Delta_t \ln\left(\Omega_{it} q_{ijct}\right)$), and $\left(\Delta_k \tilde{E}_{ijct}\right)\left(\Delta_k \tilde{Q}_{ijct}\right)$ around $e_{ijct} = \frac{e_{ict}}{\Omega_{it}}$ and $q_{ijct} = \frac{q_{ict}}{\Omega_{it}}$ (where e_{ict} and q_{ict} the aggregate expenditure and quantity for country c in code i). I start by averaging $\left(\Delta_k \tilde{E}_{ijct}\right)^2$ across all j in i for source c at time t, which I denote by putting a j subscript on the expectation,

$$\begin{aligned}\mathbb{E}_j\left[\left(\Delta_k \tilde{E}_{ijct}\right)^2\right] \approx{}& \left(\Delta_k E_{ict}\right)^2 - \Delta_k E_{ict} \Delta_k \left(\mathrm{CV}\left(e_{ijct}\right)\right)^2 \ldots \\ &+ \left(\mathrm{CV}\left(e_{ijct}\right)\right)^2 + \left(\mathrm{CV}\left(e_{ijkt}\right)\right)^2 - \frac{\mathrm{Cov}\left(\Omega_{it} e_{ijct}, \Omega_{it} e_{ijkt}\right)}{e_{ict} e_{ikt}}\end{aligned}$$

where $\mathrm{CV}_j\left(e_{ijct}\right)$ denotes the coefficient of variation of e_{ijct} across the j within code i for country c at time t.

And for the average of $\left(\Delta_k \tilde{Q}_{ijct}\right)^2$, I obtain

$$\begin{aligned}\mathbb{E}_j\left[\left(\Delta_k \tilde{Q}_{ijct}\right)^2\right] \approx{}& \left(\Delta_k Q_{ict}\right)^2 - \Delta_k Q_{ict} \Delta_k \left(\mathrm{CV}\left(q_{ijct}\right)\right)^2 \ldots \\ &+ \left(\mathrm{CV}\left(q_{ijct}\right)\right)^2 + \left(\mathrm{CV}\left(q_{ijkt}\right)\right)^2 - \frac{\mathrm{Cov}\left(\Omega_{it} q_{ijct}, \Omega_{it} q_{ijkt}\right)}{q_{ict} q_{ikt}}\end{aligned}$$

where $\mathrm{CV}_j\left(q_{ijct}\right)$ denotes the coefficient of variation of q_{ijct} across the j within code i for country c at time t.

And finally, for the average of $\left(\Delta_k \tilde{E}_{ijct}\right)\left(\Delta_k \tilde{Q}_{ijct}\right)$, I obtain

$$\begin{aligned}\mathbb{E}_j\left[\left(\Delta_k \tilde{E}_{ijct}\right)\left(\Delta_k \tilde{Q}_{ijct}\right)\right] \approx{}& \left(\Delta_k E_{ict}\right)\left(\Delta_k Q_{ict}\right) - \frac{\Delta_k E_{ict}}{2} \Delta_k \left(\mathrm{CV}_j\left(q_{ijct}\right)\right)^2 \ldots \\ &- \frac{\Delta_k Q_{ict}}{2} \Delta_k \left(\mathrm{CV}\left(e_{ijct}\right)\right)^2 \ldots \\ &+ \left[\frac{\mathrm{Cov}\left(\Omega_{it} e_{ijct}, \Omega_{it} q_{ijct}\right)}{e_{ict} q_{ict}} - \frac{\mathrm{Cov}\left(\Omega_{it} e_{ijct}, \Omega_{it} q_{ijkt}\right)}{e_{ict} q_{ikt}}\right] \ldots \\ &- \left[\frac{\mathrm{Cov}\left(\Omega_{it} e_{ijkt}, \Omega_{it} q_{ijct}\right)}{e_{ikt} q_{ict}} - \frac{\mathrm{Cov}\left(\Omega_{it} e_{ijkt}, \Omega_{it} q_{ijkt}\right)}{e_{ikt} q_{ikt}}\right]\end{aligned}$$

The next task is to transform the nonlinear terms in the estimating equation in terms of the second order expansions using the identity $p_{ijct} \equiv \frac{e_{ijct}}{q_{ijct}}$. One of the terms is trivial, as I provide the

expression for $\left(\Delta_k \tilde{E}_{ijct}\right)^2$ above. Turning next to $\left(\Delta_k P_{ijct}\right)^2$

$$\begin{aligned}
\left(\Delta_k P_{ijct}\right)^2 &= \left(\Delta_k \tilde{E}_{ijct}\right)^2 - 2\left(\Delta_k \tilde{E}_{ijct}\right)\left(\Delta_k \tilde{Q}_{ijct}\right) + \left(\Delta_k \tilde{Q}_{ijct}\right)^2 \\
\mathbb{E}_j\left[\left(\Delta_k P_{ijct}\right)^2\right] &\approx \left(\Delta_k P_{ict}\right)^2 - \Delta_k P_{ict}\Delta_k\left(\mathrm{CV}\left(e_{ijct}\right)\right)^2 + \Delta_k P_{ict}\Delta_k\left(\mathrm{CV}_j\left(q_{ijct}\right)\right)^2 \ldots \\
&\quad + \left(\mathrm{CV}\left(e_{ijct}\right)\right)^2 + \left(\mathrm{CV}\left(e_{ijkt}\right)\right)^2 + \left(\mathrm{CV}\left(q_{ijct}\right)\right)^2 + \left(\mathrm{CV}\left(q_{ijkt}\right)\right)^2 \ldots \\
&\quad - \frac{\mathrm{Cov}\left(\Omega_{it}e_{ijct}, \Omega_{it}e_{ijkt}\right)}{e_{ict}e_{ikt}} - \frac{\mathrm{Cov}\left(\Omega_{it}q_{ijct}, \Omega_{it}q_{ijkt}\right)}{q_{ict}q_{ikt}} \ldots \\
&\quad - 2\left[\frac{\mathrm{Cov}\left(\Omega_{it}e_{ijct}, \Omega_{it}q_{ijct}\right)}{e_{ict}q_{ict}} - \frac{\mathrm{Cov}\left(\Omega_{it}e_{ijct}, \Omega_{it}q_{ijkt}\right)}{e_{ict}q_{ikt}}\right] \ldots \\
&\quad + 2\left[\frac{\mathrm{Cov}\left(\Omega_{it}e_{ijkt}, \Omega_{it}q_{ijct}\right)}{e_{ikt}q_{ict}} - \frac{\mathrm{Cov}\left(\Omega_{it}e_{ijkt}, \Omega_{it}q_{ijkt}\right)}{e_{ikt}q_{ikt}}\right]
\end{aligned}$$

And finally,

$$\begin{aligned}
\left(\Delta_k P_{ijct}\right)\left(\Delta_k \tilde{E}_{ijct}\right) &= \left(\Delta_k \tilde{E}_{ijct}\right)^2 - \left(\Delta_k \tilde{E}_{ijct}\right)\left(\Delta_k \tilde{Q}_{ijct}\right) \\
\mathbb{E}_j\left[\left(\Delta_k P_{ijct}\right)\left(\Delta_k \tilde{E}_{ijct}\right)\right] &\approx \left(\Delta_k E_{ict}\right)\left(\Delta_k P_{ict}\right) - \left(\frac{\Delta_k E_{ict}}{2} - \frac{\Delta_k P_{ict}}{2}\right)\Delta_k\left(\mathrm{CV}\left(e_{ijct}\right)\right)^2 \ldots \\
&\quad + \left(\mathrm{CV}\left(e_{ijct}\right)\right)^2 + \left(\mathrm{CV}\left(e_{ijkt}\right)\right)^2 + \frac{\Delta_k E_{ict}}{2}\Delta_k\left(\mathrm{CV}_j\left(q_{ijct}\right)\right)^2 \ldots \\
&\quad - \left[\frac{\mathrm{Cov}\left(\Omega_{it}e_{ijct}, \Omega_{it}q_{ijct}\right)}{e_{ict}q_{ict}} - \frac{\mathrm{Cov}\left(\Omega_{it}e_{ijct}, \Omega_{it}q_{ijkt}\right)}{e_{ict}q_{ikt}}\right] \ldots \\
&\quad + \left[\frac{\mathrm{Cov}\left(\Omega_{it}e_{ijkt}, \Omega_{it}q_{ijct}\right)}{e_{ikt}q_{ict}} - \frac{\mathrm{Cov}\left(\Omega_{it}e_{ijkt}, \Omega_{it}q_{ijkt}\right)}{e_{ikt}q_{ikt}}\right] \ldots \\
&\quad - \frac{\mathrm{Cov}\left(\Omega_{it}e_{ijct}, \Omega_{it}e_{ijkt}\right)}{e_{ict}e_{ikt}}
\end{aligned}$$

The next step is to show that the "nuisance" covariance terms – those between source country and reference country variables which are not plausibly targeted by the classifier – all (approximately) cancel out in time and second reference country differences. I assume that for all c and j, e_{ijct} is reasonably close to $\frac{e_{ict}}{\Omega_{it}}$ and q_{ijct} is reasonably close to $\frac{q_{ict}}{\Omega_{it}}$. Under these assumptions, I can approximate for variables (in an abuse of notation to keep expressions general) $x \neq y,\ x, y \in \{e_{ct}, e_{kt}, q_{ct}, q_{kt}\}$, so that $x_{ij} \in \{e_{ijct}, e_{ijkt}, q_{ijct}, q_{ijkt}\}$ and $x_i \in \{e_{ict}, e_{ikt}, q_{ict}, q_{ikt}\}$

$$\begin{aligned}
\frac{\mathrm{Cov}\left(\Omega_{it}x_{ij}, \Omega_{it}y_{ij}\right)}{x_i y_i} &= \mathbb{E}\left[\left(\frac{\Omega_{it}x_{ij} - x_i}{x_i}\right)\left(\frac{\Omega_{it}y_{ij} - y_i}{y_i}\right)\right] \\
&\approx \mathbb{E}\left[\ln\left(x_{ij}\right)\ln\left(y_{ij}\right)\right]
\end{aligned}$$

Turning to the equilibrium values of price and expenditure (as a function of supply and demand

shocks and the price index),

$$\ln\left(\Omega_{it}e_{ijct}\right) = \frac{1-\sigma_i}{\sigma_i\omega_i+1}a_{ijct} + \frac{\omega_i+1}{\sigma_i\omega_i+1}\left(B_{ijct}+\tilde{E}_{ijt}\right) + \frac{\left(\omega_i+1\right)\left(\sigma_i-1\right)}{\sigma_i\omega_i+1}\Phi_{ijt}$$
$$\ln\left(\Omega_{it}q_{ijct}\right) = \frac{-\sigma_i}{\sigma_i\omega_i+1}a_{ijct} + \frac{1}{\sigma_i\omega_i+1}\left(B_{ijct}+E_{ijt}\right) + \frac{\sigma_i-1}{\sigma_i\omega_i+1}\Phi_{ijt}$$

Following the assumed decompositions of a_{ijct} and B_{ijct} and the independence of the different components of the shocks from each other, I find for the four covariances of interest (note that $\frac{\text{Cov}\left(\Omega_{it}e_{ijct},\Omega_{it}q_{ijkt}\right)}{e_{ict}q_{ikt}} \approx \frac{\text{Cov}\left(\Omega_{it}e_{ijkt},\Omega_{it}q_{ijct}\right)}{e_{ikt}q_{ict}}$ so I only provide expressions for one of these covariances instead of the same expression twice).

$$\begin{aligned}
\frac{\text{Cov}\left(\Omega_{it}e_{ijct},\Omega_{it}e_{ijkt}\right)}{e_{ict}e_{ikt}} &\approx \left(\frac{1-\sigma_i}{\sigma_i\omega_i+1}\right)^2\left(\mathbb{E}_j\left[\hat{a}_{ijc}\right]\mathbb{E}_j\left[\hat{a}_{ijk}\right]+\mathbb{V}_j\left[\hat{a}_{ijt}\right]\right)\ldots \\
&+ \left(\frac{\omega_i+1}{\sigma_i\omega_i+1}\right)^2\left(\mathbb{E}_j\left[\hat{B}_{ijc}\right]\mathbb{E}_j\left[\hat{B}_{ijk}\right]+\mathbb{V}_j\left[\hat{B}_{ijt}+\tilde{E}_{ijt}\right]\right)\ldots \\
&+ \left(\frac{\left(\omega_i+1\right)\left(\sigma_i-1\right)}{\sigma_i\omega_i+1}\right)^2\mathbb{V}_j\left[\Phi_{ijt}\right] \\
\frac{\text{Cov}\left(\Omega_{it}q_{ijct},\Omega_{it}q_{ijkt}\right)}{q_{ict}q_{ikt}} &\approx \left(\frac{\sigma_i}{\sigma_i\omega_i+1}\right)^2\left(\mathbb{E}_j\left[\hat{a}_{ijc}\right]\mathbb{E}_j\left[\hat{a}_{ijk}\right]+\mathbb{V}_j\left[\hat{a}_{ijt}\right]\right)\ldots \\
&+ \left(\frac{1}{\sigma_i\omega_i+1}\right)^2\left(\mathbb{E}_j\left[\hat{B}_{ijc}\right]\mathbb{E}_j\left[\hat{B}_{ijk}\right]+\mathbb{V}_j\left[\hat{B}_{ijt}+\tilde{E}_{ijt}\right]\right)\ldots \\
&+ \left(\frac{\sigma_i-1}{\sigma_i\omega_i+1}\right)^2\mathbb{V}_j\left[\Phi_{ijt}\right] \\
\frac{\text{Cov}\left(\Omega_{it}e_{ijct},\Omega_{it}q_{ijkt}\right)}{e_{ict}q_{ikt}} &\approx \left(\frac{1-\sigma_i}{\sigma_i\omega_i+1}\right)\left(\frac{\sigma_i}{\sigma_i\omega_i+1}\right)\left(\mathbb{E}_j\left[\hat{a}_{ijc}\right]\mathbb{E}_j\left[\hat{a}_{ijk}\right]+\mathbb{V}_j\left[\hat{a}_{ijt}\right]\right)\ldots \\
&+ \left(\frac{\omega_i+1}{\sigma_i\omega_i+1}\right)\left(\frac{1}{\sigma_i\omega_i+1}\right)\left(\mathbb{E}_j\left[\hat{B}_{ijc}\right]\mathbb{E}_j\left[\hat{B}_{ijk}\right]+\mathbb{V}_j\left[\hat{B}_{ijt}+\tilde{E}_{ijt}\right]\right)\ldots \\
&+ \left(\omega_i+1\right)\left(\frac{\sigma_i-1}{\sigma_i\omega_i+1}\right)^2\mathbb{V}_j\left[\Phi_{ijt}\right]
\end{aligned}$$

As can be seen, all of the expressions are either time invariant ($\mathbb{E}_j\left[\hat{a}_{ijc}\right]\mathbb{E}_j\left[\hat{a}_{ijk}\right]$ and $\mathbb{E}_j\left[\hat{B}_{ijc}\right]\mathbb{E}_j\left[\hat{B}_{ijk}\right]$) or source country invariant ($\mathbb{V}_j\left[\hat{a}_{ijt}\right]$ and $\mathbb{V}_j\left[\hat{B}_{ijt}+\tilde{E}_{ijt}\right]$); thus all of these covariances (approximately) drop out with second reference country and time differences.

Note that the $\frac{\text{Cov}\left(\Omega_{it}e_{ijct},\Omega_{it}q_{ijct}\right)}{e_{ict}q_{ict}}$ and $\frac{\text{Cov}\left(\Omega_{it}e_{ijkt},\Omega_{it}q_{ijkt}\right)}{e_{ikt}q_{ikt}}$ terms need not cancel out in second differences. Both e_{ijct} and q_{ijct} are functions of the $\hat{a}_{ijct}$ and $\hat{B}_{ijct}$ – and these terms might have difference variances across goods j within code i for different source countries. Consequently, the $\frac{\text{Cov}\left(\Omega_{it}e_{ijct},\Omega_{it}q_{ijct}\right)}{e_{ict}q_{ict}}$ and $\frac{\text{Cov}\left(\Omega_{it}e_{ijkt},\Omega_{it}q_{ijkt}\right)}{e_{ikt}q_{ikt}}$ terms are likely to vary by source-reference pair.

By combining all of these terms, I obtain the estimating equation presented at the beginning of this subsection and in the text.

	(1) MFN tariff	(2) MFN tariff	(3) MFN tariff	(4) MFN tariff
Corrected	-1.88e-8***	-1.88e-8***		
Sigma	(6.53e-9)	(6.53e-9)		
Corrected		1.35e-7***		
Omega		(5.30e-9)		
Uncorrected			2.73e-7	2.73e-7
Sigma			(2.13e-7)	(2.13e-7)
Uncorrected				4.37e-6
Omega				(7.52e-6)
Observations	210,090	210,090	207,777	207,777
Dep. var. mean	0.0493	0.0493	0.0491	0.0491
Dep. var. SD	0.0741	0.0741	0.06767	0.0677
FEs	HS6, year	HS6, year	HS6, year	HS6, year
Adj. R2	0.484	0.484	0.568	0.5678
Clustering	HS6	HS6	HS6	HS6

Note: *** significant at the 1% level, ** significant at the 5% level, * significant at the 10% level

Table 9: Correlation between tariffs and elasticities

Supplementary tables for the correlation between tariffs and elasticities

In the text, I examine the correlations between the corrected and uncorrected elasticities and tariffs in Table 8. In this section, I consider two adjustments to my approach, and I find similar results.

First, in Table 8 in the main body of the paper, I constrain the sample to be the same for all 4 columns. Some observations have corrected elasticities but not uncorrected ones, or vice-versa. By restricting the sample to only observations which have both corrected and uncorrected elasticities, I avoid the possibility that differences in sample composition drive the change in the sign of the correlation with import demand elasticity. In Appendix Table 9, I show that this is not necessary for my result: I obtain the same change in sign if I include all potential observations in both regressions.

And second, in Table 8 in the body of the paper I use MFN tariffs as the dependent variable. This has some advantages: MFN tariffs are broadly applied and consequently well measured. However, their mapping to theory is also more complicated because they partly, but not necessarily entirely, reflect trade negotiations as discussed in the body of the paper. In contrast, the Column 2 tariffs do not reflect trade negotiations and so are easily to interpreted as unilateral tariffs. However, these tariffs are applied to a tiny fraction of trade – for many goods no imports are subject to these tariffs – and so they are much more difficult to measure. In Appendix Table 10, I show that the flavor of Table 8 in the body of the paper is preserved when I use Column 2 tariffs. Again there is a sign flip for the elasticity of import demand, although now the relationship with the corrected sigma and omega is no longer significant (while the relationship with the uncorrected sigma is the opposite of the sign predicted by the theory and highly significant).

	(1)	(2)
	Col2 tariff	Col2 tariff
Corrected	-4.15e-8	
Sigma	(3.95e-8)	
Corrected	1.07e-8	
Omega	(4.47e-8)	
Uncorrected		1.48e06***
Sigma		(4.36e-7)
Uncorrected		5.46e-5*
Omega		(3.02e-5)
Observations	203,926	203,926
Dep. var. mean	0.353	0.353
Dep. var. SD	0.268	0.268
FEs	HS6, year	HS6, year
Adj. R2	0.570	0.570
Clustering	HS6	HS6

Note: *** significant at the 1% level, ** significant at the 5% level, * significant at the 10% level

Table 10: Correlation between Column 2 tariffs and elasticities