

Quantitative Comparison of Ad-Valorem Equivalents for Non-Tariff Trade Measures: WTO Versus TRAINS

Sionegael Ikeme; Amanda M. Countryman; Dale T.
Manning; Diane Charlton

Accessibility Disclaimer:

For a more accessible version of this document, please submit an accessibility request form through the Montana State University Library website.

AAEA INVITED PAPER OPEN ACCESS

Quantitative Comparison of Ad-Valorem Equivalents for Non-Tariff Trade Measures: WTO Versus TRAINS

Sionegael Ikeme¹  | Amanda M. Countryman¹  | Dale T. Manning² | Diane Charlton³

¹Department of Agricultural and Resource Economics, Colorado State University, Fort Collins, Colorado, USA | ²Department of Agricultural and Resource Economics and Baker School of Public Policy, University of Tennessee, Knoxville, Tennessee, USA | ³Department of Agricultural Economics and Economics, Montana State University, Bozeman, Montana, USA

Correspondence: Sionegael Ikeme (sionegael.ikeme@colostate.edu)

Received: 29 March 2025 | **Revised:** 20 October 2025 | **Accepted:** 28 October 2025

Funding: This work was supported by U.S. Department of Agriculture, 2023-67023-39108.

ABSTRACT

Non-tariff measures (NTMs) play a growing role in trade policy. Economists quantify NTMs by calculating their Ad Valorem Equivalents (AVE), typically employing one of two global NTM datasets: TRAINS from UNCTAD or WTO notifications. While TRAINS data better measure NTMs, their limited coverage over countries and time is inadequate for many types of analyses, and thus researchers often resort to using WTO data, which is not as accurate as the TRAINS but has better coverage. We compare the two datasets to assess whether AVEs from WTO notifications align with those from TRAINS, given similar country and time coverage. We find significant AVE discrepancies between datasets, sometimes reversing signs, warranting caution when using WTO notifications in the absence of TRAINS data to construct NTMs.

1 | Introduction

Economists are increasingly studying non-tariff measures (NTMs) as crucial elements of international trade policy. While the use of traditional tariffs declined following World War II, the implementation of NTMs has surged (Orefice 2017; Niu et al. 2020). Recognizing their growing significance to global trade, numerous studies have estimated the effects of NTMs not only on trade but also on outcomes such as supply chains (Raimondi et al. 2023), labor (Porto 2018), and the environment (Mao et al. 2023). However, measuring NTMs and their impacts can be challenging. Economists typically quantify the effects of NTMs using Ad Valorem Equivalents (AVEs). An AVE is the tariff rate that would have an equivalent effect on imported quantities as an NTM, making heterogeneous NTMs more comparable (Disdier and Fugazza 2020). To estimate AVEs, the literature has mostly relied on either the Trade Analysis Information System (TRAINS) database from the United Nations Conference on Trade and Development (UNCTAD) or World Trade Organization (WTO) notifications from the Integrated Trade Intelligence Portal (I-TIP). TRAINS

data provide more accurate information on NTMs compared to WTO notifications (de Melo and Nicita 2018; UNCTAD 2018; Rau and Vogt 2017) but have more limited temporal and country coverage (Grübler and Reiter 2021). TRAINS provides a detailed snapshot of NTMs implemented by select countries, with UNCTAD actively compiling data from various sources, including domestic regulations, customs agencies, and other relevant bodies (de Melo and Nicita 2018). In contrast, WTO notification data rely on a more passive approach, based on self-reported notifications of newly implemented NTMs, and do not consistently record when NTMs are withdrawn. This results in the omission of pre-existing measures, the overreporting of withdrawn measures, and the introduction of potential self-notification bias. Despite these limitations, WTO data offer significantly broader country and time coverage than TRAINS, due to the lower cost of data collection. Many studies rely on TRAINS data, but its limited coverage often prohibits estimating AVEs over extended periods (Beghin et al. 2015; Bratt 2017; Cadot and Gourdon 2016; Cadot et al. 2018). Since UNCTAD collects data over successive years for few countries, it is difficult to obtain time-varying AVEs from TRAINS. To date, Niu et al. (2018)

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2025 The Author(s). *Applied Economic Perspectives and Policy* published by Wiley Periodicals LLC on behalf of Agricultural & Applied Economics Association.

is one of the only studies to our knowledge that estimates time-varying AVEs using the TRAINS database, while only including three-year intervals for a limited number of countries. On the other hand, WTO notification data have been used in many studies (Ghodsi et al. 2017; Ghodsi and Stehrer 2022; Grübler and Reiter 2021; Adarov and Ghodsi 2023) due to their broader country and time coverage. This expanded coverage not only facilitates the estimation of time-varying AVEs (Adarov and Ghodsi 2023) but also allows for analysis of a greater number of countries. Since AVEs are estimated based on cross-country variation, the limited number of countries in the TRAINS data can introduce a different source of bias. Despite its common use, it remains unclear whether using WTO data produces similar AVE estimates to the TRAINS dataset, conditional on similar time and country coverage. This research helps fill this gap in the literature.

In this research, we compare the two primary global NTM databases and assess how similar AVEs are when derived from WTO notifications and TRAINS, examining only countries and years where the databases have comparable coverage. Substantial differences in AVEs derived from the two datasets would suggest that WTO data are not a good substitute for TRAINS data. Since the data collection method for TRAINS is more objective than that of the WTO notifications and does not include NTMs indefinitely, even after they expire, we suggest that TRAINS is the preferred dataset if the datasets are not comparable. To assess this, we conduct a statistical analysis to test the null hypothesis that the two datasets produce equivalent AVEs. Rejecting this null would suggest that the WTO data are not an adequate substitute for TRAINS. We also examine whether differences in AVEs vary by importing country, product, and time, which could provide insights into sources of measurement error since some countries might be more likely to correctly notify the WTO of NTMs. Finally, we explore whether observed characteristics of the importer, exporter, and product can explain the differences in AVEs. Specifically, we assess whether the importance of imported goods for both the importer and exporter, the GDP of both countries, and the overall NTM intensity imposed by the importer can explain the magnitude of AVE differences. Assessing the sensitivity of AVE estimates to data sources reveals whether WTO notification data can be substituted for TRAINS data in years and countries for which TRAINS data do not exist. This would suggest that the WTO can provide an alternative to using TRAINS and allow for the estimation of AVEs across more countries and over longer time periods. Alternatively, if estimated AVEs differ between the two datasets, it can serve as a warning of potential biases or inconsistencies inherent in the WTO data.

To estimate AVEs, we employ a trade quantity-based approach proposed by Kee and Nicita (2022). This method enables us to derive bilateral AVEs at the yearly Harmonized System (HS) 6 digit product level from 2013 to 2016 while addressing the endogeneity of NTMs and tariffs using the NTMs and tariffs imposed by neighboring importers as instrumental variables (IVs). We concentrate on nine South American countries because of their geographic proximity and relatively good coverage in both datasets. In addition, we focus on two types of NTMs: Technical Barriers to Trade (TBT) and Sanitary and Phytosanitary (SPS) measures. We prioritize these measures because they are the most frequently used types of NTMs (Ghodsi et al. 2017) and are particularly dominant in agri-food sectors (Raimondi

et al. 2023). We find that while some AVEs yield relatively similar values, this is not the case for all estimates at the product-importer-exporter-year level. In fact, there are instances where the differences in AVEs are substantial, sometimes even resulting in opposite signs. This is particularly concerning since it implies that the estimated effect of NTMs on trade flows can be reversed in some cases depending on the data source used.

While previous studies have explored the qualitative differences between the two NTM databases (Rau and Vogt 2017; de Melo and Nicita 2018), to our knowledge, no research has quantitatively compared the resulting AVE estimates. This study bridges that gap by analyzing the differences in the estimated AVEs derived from both sources for similar time and country coverage. This paper contributes to the literature by offering insights into meaningful differences between the datasets used to estimate AVEs for NTMs. It can provide valuable guidance for future analyses, especially those exploring the effects of trade policy on international trade and other economic outcomes. We find that WTO and TRAINS data do not yield similar AVEs even when examining countries and time periods where both datasets have high coverage. This highlights the need for improved data quality and collection methods to study NTMs. Future data collection should focus on both accuracy and coverage. Furthermore, our findings suggest that measurement errors in WTO notification data likely lead to significant miscalculations of AVEs, underscoring the importance of using extreme caution when using the WTO data. This issue is particularly salient in the agricultural and food sector, where accurate NTM data are essential for evaluating the complex and heterogeneous effects of SPS and TBT measures. The agri-food sector is highly regulated due to its sensitivity to health, safety, and quality concerns. NTMs can be a barrier to export markets for some trade partners while they might act as a catalyst for quality upgrading or improve consumer confidence for other trade partners, thus improving their export market access. Unlike tariffs, the effects of NTMs are less predictable and often asymmetric. Given the uneven distribution of compliance costs, especially among small and resource-constrained producers, detailed and reliable NTM data are critical to capture these nuanced effects on export market access. Better data not only enhance the accuracy of AVE estimation but also improve our understanding of how NTMs shape trade patterns, value chain participation, investment, and broader development outcomes in the agri-food sector. Before delving into the quantitative analysis, we provide background on NTM data and their use in trade policy. Then, we qualitatively compare the two databases and outline how the literature has quantified NTMs through AVEs. Following this, we describe our methods and provide an overview of the data used to estimate AVEs in this study. We then present results and conclude with a discussion of key findings and policy implications.

2 | Background

2.1 | Current NTM Data

There are multiple sources of NTM data, differing in collection methods and temporal and country coverage. Rau and Vogt (2017) provide a comprehensive overview of inventories, legislative reviews, notifications, surveys, import refusals, and other data sources. Among these, national legislation inventories

emerge as the most reliable. These inventories meticulously catalog national legislation, allowing measures to be classified into different types of NTMs using the MAST classification system.¹ Notably, the TRAINS database, jointly developed by UNCTAD and the International Trade Centre (ITC), stands out as a primary resource for researchers investigating NTM effects. This database meticulously gathers and classifies data for NTMs from all legal texts issued by countries, including bilateral and plurilateral agreements. The process involves active data collection from multiple sources, which vary depending on the type of NTM and are subject to independent regulatory review by UNCTAD.² The TRAINS database provides HS6 level NTM data for 103 countries, detailing the implementing country, the affected countries, descriptions of the regulation and product, and the dates of NTM implementation and withdrawal. While data collection began in 1994, a major overhaul in 2006 introduced a new methodology, so the most reliable NTM records are available only from 2008 (UNCTAD 2018). Key strengths of the database include its traceability, cross-country comparability, transparency, and open accessibility. However, the inconsistent frequency of data collection across countries presents challenges in reliably estimating AVEs over time. This limitation is particularly restrictive in panel analyses, where only countries with multiple years of data can be effectively included.

Another valuable source of NTM data is WTO notifications. As WTO members, countries are required to report the implementation of NTMs to enhance the transparency and predictability of trade policies. The current WTO notification dataset covers NTMs for 140 member countries since 1979, providing information at the HS6 product code level on the type of NTM, regulation details, notification date, and, in some cases, the withdrawal year. While this dataset covers more countries and periods than TRAINS,³ its passive data collection relies on self-reported submissions, which introduce bias. Countries do not always notify the WTO when an NTM is implemented, or in a timely manner, causing discrepancies between actual implementation and notification years (de Melo and Nicita 2018). In addition, WTO members are only required to notify the WTO of NTMs when they may have a nonmarginal impact on other trading countries or when the regulation is not based on “relevant international standards” (UNCTAD 2018). This results in delays, omissions, and self-notification bias since there is no external verification process. Additionally, some self-reported NTMs are missing HS product codes. We address this by retrieving missing codes using product information as in Ghodsi et al. (2015). Another limitation is the inconsistent recording of NTM withdrawals, particularly for SPS and TBT measures, where no withdrawal records exist. Furthermore, the dataset sometimes lacks implementation dates, requiring the use of notification dates as a proxy. In addition to direct notifications, the WTO dataset includes reverse notifications, where WTO members report NTMs imposed by other members (Rau and Vogt 2017). These reverse notifications fall under the Special Trade Concerns (STCs) database. While combining STC data with direct WTO notifications could help account for under-reporting by certain countries, matching STC cases with actual notifications is challenging (Ghodsi et al. 2015). These limitations likely introduce measurement errors in AVE estimation. Consequently, this database is generally less favored compared to TRAINS data for studies where country and temporal coverage in the TRAINS are adequate.

Business surveys and complaint portals serve as other possible sources of NTM data and offer insights from the perspective of businesses facing trade barriers. Examples include the ITC NTM surveys and the EU market access database. However, survey data tend to be country-specific and often represent potentially biased reports from firms affected by NTMs. Consequently, survey results tend to overlook NTMs that have minimal perceived impact (Rau and Vogt 2017). Hence, such data must be interpreted with caution, considering the inherent biases and limitations associated with the perspective of surveyed businesses. We focus exclusively on NTM data from the UNCTAD TRAINS and WTO notification databases for this analysis. They are the most commonly used sources for estimating AVEs, and their broad coverage and relatively reliable information make them arguably the most suitable databases for AVE estimation.

2.2 | Global NTM Data Limitations

Despite extensive information on NTMs, these databases share common limitations for quantitative analysis. These challenges stem not only from the way NTMs are collected but also from the nature of NTMs themselves. de Melo and Nicita (2018) highlight six key limitations to consider when using databases such as TRAINS or WTO notifications. Limitations include limited comprehensiveness, lack of precision, data dimensionality, time dimension, product-specific requirements, and endogeneity.

Comprehensiveness refers to both omissions and double-counting. As discussed previously, there are multiple sources for NTM databases, making data collection challenging. Double-counting can arise when identical NTMs from different sources refer to the same regulation. In addition, cross-country comparisons are difficult because data availability and collection efforts vary across countries. Notification-based data also face challenges, as reporting requirements are not always strictly followed, and overlapping measures may be counted twice.

The lack of precision stems from the qualitative nature of NTM data, making it difficult to determine the stringency of trade regulations. Data often fail to specify how regulation impacts trade costs or the economic sectors it affects. Additionally, the same regulation can have varying effects across countries, with trade costs differing by industry and exporting country. To address this challenge, economists use NTM data to infer stringency by estimating AVEs.

Dimensionality poses challenges in disentangling the econometric effects of individual NTMs on products due to the heterogeneous nature of measures classified as NTMs, particularly at the HS6 level. Collinearity among NTMs is a common issue, as multiple NTMs are often implemented on the same product, making it difficult to isolate the impact of any single measure. When multiple NTMs are applied simultaneously, their effects may overlap, creating statistical challenges in identifying their distinct contributions. As a result, analyses often aggregate NTMs into broader categories, such as SPS and TBT, to improve the reliability of AVE estimations.

Issues related to time dimensions stem from data collection. While NTM databases often include implementation

or notification dates, this information is not always precise enough for time-series or event-study types of analyses. In notification-based datasets, measures that have existed unchanged for a long time are generally not reported, as countries are only required to notify newly implemented NTMs or modifications to existing policy. As a result, long-standing historical NTMs may be missing from the data. Similarly, in the case of the UNCTAD TRAINS database, data is not collected continuously for all countries across years. This irregularity makes it particularly challenging to perform time-series analyses.

Product-specific requirements refer to NTMs that affect products at the HS6 level, covering approximately 5000 traded goods. However, NTMs do not impact all products uniformly. Certain categories, such as food and chemicals, are more heavily regulated and more likely to be subject to NTMs than raw materials, for instance. This variation requires caution when estimating the effects of NTMs, ensuring that the analysis is sufficiently disaggregated to capture these differences. However, the more disaggregated the analysis, the more complex it becomes, posing additional methodological challenges.

Finally, endogeneity presents a major challenge in assessing the causal effects of NTMs on trade flows and prices, with reverse causality being a key concern. Countries may impose stricter NTMs on sectors that are economically or politically significant, leading to simultaneity problems. This can be an especially important issue when estimating AVEs, where we would like to isolate and quantify the true impact of NTMs. Often, analysis employs IVs to address endogeneity (Bratt 2017; Kee and Nicita 2022; Adarov and Ghodsi 2023).

Due to the various challenges arising from data collection and the inherent nature of NTMs, quantifying their effects is not straightforward. In particular, inconsistencies in data collection may lead to discrepancies between sources.

2.3 | Quantification of NTM Effects

To understand how the effects of NTMs are quantified, we first examine their impact on trade. When an importing country implements an NTM on a product from an exporting country, the exporter may face additional costs to comply with the regulation. These can take the form of higher fixed and/or variable costs, and the burden of compliance may vary across exporting countries. Those with strong infrastructure and regulatory alignment can likely adapt more easily, while those lacking such capacity face significantly higher compliance costs. However, NTMs can also play a constructive role in market development. In particular, some SPS and TBTs help reduce asymmetric information by establishing uniform standards, reassuring consumers about product safety and quality, and ultimately stimulating demand (Xiong and Beghin 2014; Beghin et al. 2015; Cadot et al. 2018; Gourdon et al. 2020). Beyond their regulatory role, NTMs can sometimes act as demand catalysts, particularly when firms respond by upgrading product quality to meet more stringent requirements. In this context, NTMs might drive innovation, lead exporters to upgrade quality and processes, and enable countries to access higher-value markets where compliance becomes

a signal of competitiveness and reliability (Aghion 2018). Under competitive pressure, firms can even sometimes “escape competition” by meeting higher standards and upgrading product quality through innovation (Daron et al. 2006; Aghion et al. 2005, 2009). Thus, NTMs can have two opposing effects: compliance costs (represented by an inward product supply shift) and demand catalysts (represented by an outward shift in product demand). [Supporting Information](#): Figure 6 from Section II in the appendix illustrates these mechanisms.

In the context of agricultural goods, the two opposing effects of NTMs, acting both as catalysts and as barriers to trade, remain clearly observable. Curzi et al. (2015) discuss how standards in food can be both a barrier to trade and a catalyst, especially in terms of modernizing supply chains and stimulating product upgrading. Swinnen (2016) additionally shows that both trade-enhancing and trade-restricting effects can coexist in the agricultural and food sectors. The paper emphasizes that the impact of food standards is highly context-dependent. While NTMs can create new market opportunities for compliant producers, they may also exclude others, particularly in agriculture, where producers vary widely in their ability to meet compliance requirements. Santeramo and Lamonaca (2019), through a meta-analysis, confirm this heterogeneity in the effects of NTMs on agricultural trade outcomes. In light of this, our work employs a methodological approach that captures both the barrier and catalyst roles of NTMs.

Although various methods exist to quantify the effects of NTMs, the econometric approach enables us to estimate the price and/or quantity changes induced by their presence. Within this framework, two main methodologies emerge: the price-based approach, which captures only compliance costs, and the quantity-based approach, which accounts for both compliance costs and market catalyst effects. The quantity-based method examines how NTMs affect trade volumes and translates these effects into trade costs using elasticities of substitution. This approach identifies instances where import quantities rise (indicating that the demand creation effect outweighs compliance costs, leading to negative AVEs) and cases where they decline, reflecting the opposite effect. In contrast, the price-based approach focuses on how NTMs influence prices, typically expecting an increase when compliance costs rise (Cadot and Gourdon 2016). Both approaches rely on gravity models, which regress bilateral trade flows on importer-exporter characteristics and trade barriers, including tariffs and NTMs. Past studies have employed both quantity-based methods (Kee et al. 2009; Essaji 2008; Beghin et al. 2015; Ghodsi and Stehrer 2022) and price-based methods (Cadot and Gourdon 2016; Ing et al. 2016; Cadot et al. 2018; Kravchenko et al. 2022) to estimate AVEs, with the choice often depending on data availability and model specifications. Cadot et al. (2018) compare these approaches, highlighting key differences in their measured effects. If we narrow the literature to econometric papers that exclusively quantify NTMs of agri-food products, Xiong and Beghin (2014) Arita et al. (2015), and Mao et al. (2023) adopt quantity-based approaches, while Fiankor (2023) employs a price-based approach. These studies use either TRAINS (Fiankor 2023) or WTO notification data (Mao et al. 2023), as well as other sources specific to the type of NTM or countries of interest. For instance, (Arita et al. 2015) uses the USTR's 2014 Report on Sanitary

and Phytosanitary Measures and the European Commission's Market Access Database among other sources.

While both the price- and quantity-based approaches offer valuable insights, we adopt the quantity-based method in this paper, as it captures both positive and negative AVEs. Negative AVEs are particularly relevant for SPS and TBT measures (Gourdon et al. 2020). Among quantity-based methods, we focus on bilateral AVE estimation, which accounts for AVE variation across importing and exporting countries at the product-year level. This approach is consistent with an Armington perspective, where goods are differentiated by country of origin. Indeed, compliance costs for NTMs vary across countries, even under uniform regulations (Bratt 2017; Kee and Nicita 2022), with exceptions for SPS measures that target specific exporters. To capture this variation, we follow Kee and Nicita (2022), modeling imports as a function of gravity variables, tariffs, and NTMs. A key advantage of this method is its simultaneous estimation of import elasticities, eliminating the need for pre-estimated values from studies like Bratt (2017).

3 | Methods

To estimate AVEs, we use a gravity model to assess trade effects at the importer-exporter-year level for products disaggregated to the HS6 product code from 2013 to 2016. We compare AVEs derived from two NTM data sources by estimating the gravity model twice: once with data from WTO notifications and once with data from TRAINS. In addition, we focus on nine South American importers and 82 exporters, selected based on data availability across multiple years in both databases. In this section, we first address potential endogeneity in estimating tariff and NTM effects on trade, and then we present the gravity framework. Finally, we explain how the estimated coefficients are used to calculate AVEs.

3.1 | Accounting for Endogeneity

The gravity model estimates the impact of NTMs on import quantities by regressing trade flows on NTMs. This approach quantifies the average effect of an additional NTM on the volume of imports while controlling for gravity characteristics and other trade barriers. However, we must address the issue of endogeneity of trade volume and prices inherent in both NTMs and tariffs (Kee et al. 2009; Bratt 2017; Kee and Nicita 2022). Reverse causality is a concern since the importance of a traded good might influence the implementation of trade policies. Specifically, factors such as import penetration in industries can drive protectionist policies (Trefler 1993; Lee and Swagel 1997). A common approach to address endogeneity of trade policies is to instrument for the importing country's NTMs using the NTMs imposed by neighboring countries (Kee et al. 2009; Cadot et al. 2018; Bratt 2017; Kee and Nicita 2022). Following this approach, we adopt a two-stage regression framework to correct for potential bias. In the first step, we regress an importer's NTM implementation on the mean implementation of its three closest neighbors.⁴ This regression is performed separately for each year in our sample since the endogeneity problem persists over time, which is a concern that has also been acknowledged

in cross-sectional studies (e.g., Kee et al. (2009); Kee and Nicita (2022)). In the second stage, we use the predicted NTMs from the first stage in the gravity equation. We apply the same instrumental variable strategy for tariffs since tariffs have the same endogeneity concerns. Specifically, we estimate the number of NTMs (Equation 1) and the level of tariffs (Equation 2) imposed by country i on country j for product k at time t as a function of the average of NTMs and tariffs of three neighboring countries of i , noted as h , for each year t . In addition, following Kee and Nicita (2022), we introduce importer and exporter heterogeneity in AVEs by interacting the NTM and tariff variables with importer market share ($share_{kit}$) and exporter market share ($share_{kjt}$). Since these interaction terms are endogenous, they also require valid instruments. To address this, we construct instruments by interacting the NTMs and tariffs of the three geographically closest neighboring countries with the corresponding market shares, as specified in Equations (3–6).

$$NTM_{ijkt}^n = Z_{ijt} \hat{\alpha}_t^1 + \hat{\gamma}_t^1 \overline{NTM}_{hjk}^n + \epsilon_{ijkt} \quad (1)$$

$$t_{ijkt} = Z_{ijt} \hat{\alpha}_t^2 + \hat{\gamma}_t^2 \bar{t}_{hjk} + \epsilon_{ijkt} \quad (2)$$

$$share_{kit} NTM_{ijkt}^n = Z_{ijt} \hat{\alpha}_t^3 + \hat{\gamma}_t^3 share_{kit} \overline{NTM}_{hjk}^n + \epsilon_{ijkt} \quad (3)$$

$$share_{kjt} NTM_{ijkt}^n = Z_{ijt} \hat{\alpha}_t^4 + \hat{\gamma}_t^4 share_{kjt} \overline{NTM}_{hjk}^n + \epsilon_{ijkt} \quad (4)$$

$$share_{kit} t_{ijkt} = Z_{ijt} \hat{\alpha}_t^5 + \hat{\gamma}_t^5 share_{kit} \bar{t}_{hjk} + \epsilon_{ijkt} \quad (5)$$

$$share_{kjt} t_{ijkt} = Z_{ijt} \hat{\alpha}_t^6 + \hat{\gamma}_t^6 share_{kjt} \bar{t}_{hjk} + \epsilon_{ijkt} \quad (6)$$

where the term $\overline{NTM}_{hjk}^n$ refers to the average count of non-tariff measures (NTMs) of type n (where $n \in \{SPS, TBT, \text{other NTMs}\}$) imposed by the three neighboring countries h of i on imports from j for good k at year t . Z_{ijt} represents gravity variables, including distance, contiguity, landlocked status, common language, regional trade agreements (RTA), and colonial relationships, as well as the logarithm of GDP for both the importer and exporter. $\bar{t}_{hjk}$ denotes the average applied tariff imposed by the 3 neighboring countries h of i on imports from j for good k at year t . The variables $share_{kit}$ and $share_{kjt}$ denote the importer's (i) and exporter's share (j) in the world market for product k , respectively. We apply Equation (1) to all types of NTMs considered in our study, including SPS measures, TBTs, and other types of NTMs.⁵ We then use the estimated NTMs and tariffs at the product-importer-exporter-year level to assess their impact on trade. However, employing NTMs from other countries as IVs introduces a potential concern: it may exacerbate the measurement error inherent in WTO notification data. This approach risks amplifying existing biases, as NTMs from neighboring countries, which may themselves be subject to inaccuracies, are used to predict NTMs imposed by others.

3.2 | Quantity Based Gravity Model

Once we obtain the predicted NTMs and tariffs from our first stage regressions, we include them in the second stage gravity

model, where we express bilateral import value as a function of bilateral gravity characteristics, NTMs, and tariffs.⁶ We derive the following gravity equation for each good k and year t , capturing trade flows from importer i to exporter j .

$$\ln Q_{ijkt} = \beta_{kt} + (\beta_{kt}^t + \beta_{k1t}^t \text{share}_{kit} + \beta_{k2t}^t \text{share}_{kjt}) t_{ijkt} + \sum_n (\beta_{kt}^{NTM^n} + \beta_{k1t}^{NTM^n} \text{share}_{kit} + \beta_{k2t}^{NTM^n} \text{share}_{kjt}) NTM_{ijkt}^n + Z_{ijkt} \beta + \varepsilon_{ijkt} \quad (7)$$

In the gravity equation, Q_{ijkt} represents the value of imports of product k by importer i from exporter j at year t . The variables share_{kit} and share_{kjt} denote the importer's and exporter's share in the world market for product k , respectively. The term t_{ijkt} refers to the predicted applied tariff imposed by importer i on imports of product k from exporter j (with its corresponding coefficient β^t), while NTM_{ijkt}^n indicates the predicted count of non-tariff measures imposed by importer i on imports of product k from exporter j (with its corresponding coefficient β^{NTM^n}). The NTM effect is estimated at the importer-exporter level by interacting the NTM count of type n with the market share of the importer and exporter for that good in global trade. Z_{ijkt} consists of all gravity variables introduced earlier. We also adjust some of the gravity variables by GDP to account for multilateral resistance, a concept introduced by Anderson and Van Wincoop (2003), following the approach of Bratt (2017); Baier and Bergstrand (2009). We provide more details in [Supporting Information: Section III](#) in the appendix. To account for zero trade flows, we employ a Pseudo-Poisson Maximum Likelihood (PPML) specification. Following Kee and Nicita (2022), we estimate this model separately for each year and at the HS6 digit product level. We denote the linear combinations of estimated coefficients as:

$$\beta_{ijkt}^t \equiv \beta_{kt}^t + \beta_{k1t}^t \text{share}_{kit} + \beta_{k2t}^t \text{share}_{kjt} \quad (8)$$

$$\beta_{ijkt}^{NTM^n} \equiv \beta_{kt}^{NTM^n} + \beta_{k1t}^{NTM^n} \text{share}_{kit} + \beta_{k2t}^{NTM^n} \text{share}_{kjt} \quad (9)$$

We use the coefficients β_{ijkt}^t and $\beta_{ijkt}^{NTM^n}$ to obtain AVEs at the product-importer-exporter-year level. Finally, we cluster standard errors at the importer level to address heteroscedasticity in the standard errors of the gravity model (Fell and Duver 2024).

3.3 | Estimation of AVEs

To derive the AVEs, we follow Kee and Nicita (2022), where we determine the marginal effects of NTMs on quantity traded from Equation (9). From the gravity Equation (7), we know that the proportionate change in import quantity due to an additional NTM from m to $m+1$ is:

$$\ln(Q_{ijkt} | NTM = m+1) - \ln(Q_{ijkt} | NTM = m) = \hat{\beta}_{ijkt}^{NTM} \quad (10)$$

$$\frac{Q_{ijkt} | NTM = m+1}{Q_{ijkt} | NTM = m} = \exp(\hat{\beta}_{ijkt}^{NTM}) \quad (11)$$

$$\frac{Q_{ijkt} | NTM = m+1 - Q_{ijkt} | NTM = m}{Q_{ijkt} | NTM = m} = \exp(\hat{\beta}_{ijkt}^{NTM}) - 1 \quad (12)$$

The proportionate change in quantity imported due to a 1% point increase in an import tariff is:

$$\frac{\partial \ln Q_{ijkt}}{\partial t_{ijkt}} = \hat{\beta}_{ijkt}^t \quad (13)$$

The AVE of an NTM on product k in importing country i from exporting country j measures the Ad-Valorem tariff that induces the same proportionate change in quantity imported as the presence of the NTM (Kee and Nicita 2022). Thus, we compute the $AVE_{ijkt}^{NTM^n}$ for each NTM type n using the following expression. Following Kee and Nicita (2022), we have:

$$AVE_{ijkt}^{NTM^n} = \frac{\exp(\hat{\beta}_{ijkt}^{NTM^n}) - 1}{\hat{\beta}_{ijkt}^t} \quad (14)$$

This represents the mean AVE for an additional NTM on product k imposed by importer i on exporter j . Using Equation (14), we can obtain the effect of the total number of NTMs imposed by country i by multiplying the NTM coefficient in Equation (15) by the total count of NTMs to obtain the total AVE for all NTMs imposed as follows (Kravchenko et al. 2022; Fell and Duver 2024).

$$AVE_{ijkt}^{NTM^n} = \frac{\exp(\hat{\beta}_{ijkt}^{NTM^n} NTM_{ijkt}^n) - 1}{\hat{\beta}_{ijkt}^t} \quad (15)$$

4 | Data

4.1 | Trade and Tariff Data

This study uses a dataset covering trade flows between 2013 and 2016 for 9 South American importing countries and 82 exporting countries. Following this approach, we adopt a two-stage regression framework to correct for potential bias. The data are disaggregated to the HS6 digit product level, following HS 4 revisions.⁷ Trade values (in U.S. dollars) are obtained from the World Integrated Trade Solution (WITS) database, while gravity variables (log of distance, common language, contiguity, colonial ties, how landlocked importer and exporter are) come from the Center for Prospective Studies and International Information (CEPII) (Conte et al. 2022), and regional trade agreement (RTA) data are sourced from Blöthner and Larch (2022). GDP values are taken from the World Bank. For tariff data, we use Most-Favored-Nation (MFN) tariffs from the World Integrated Trade Solution (WITS) database. After merging all data by product-importer-exporter-year, the final dataset forms a balanced panel that is then subset by year and product for gravity analysis. The dataset, which includes 9 importing countries, 82 exporting countries, and 5152 products over 4 years (2013–2016),⁸ is then merged with the NTM data. Summary statistics for the variables used in this analysis are provided in the [Supporting Information: appendix Section IV-II](#).

4.2 | NTM Data

We use the UNCTAD TRAINS database and WTO notification data available through the I-TIP. Since the TRAINS data are reported at the HS6 digit level, our analysis is conducted at this product aggregation level. Because I-TIP data contains missing HS codes, we rely on the dataset provided by Ghodsi et al. (2017), which is derived from I-TIP data at the HS6 digit level. Both datasets provide information on NTM types (e.g., SPS, TBT), the count of applied NTMs, the implementing country (the importing country), the affected country (the exporting country), relevant regulations, and affected product details. WTO notifications require certain assumptions and adjustments to fit our estimation equation, so we follow the recommendations of Ghodsi et al. (2015) to transform the WTO data. First, the WTO notification data has missing product classifications. To address this, Ghodsi et al. (2017) retrieved missing HS codes by cross-referencing product descriptions with other datasets. However, this likely introduces some error in product identification. Second, since the WTO does not record when NTMs are withdrawn, we assume that NTMs, once implemented, remain in place indefinitely as in Ghodsi et al. (2017) and Ghodsi and Stehrer (2022). This assumption enables us to construct a cumulative count of NTMs per year. However, it also imposes an upward bias, overstating the number of active NTMs over time, at least among NTMs that were ever recorded in the dataset. Finally, since the WTO notification database only includes notified NTMs, we assume that all notified NTMs represent the total number of NTMs implemented, although, in reality, some measures may exist without formal notification. Additionally, the WTO might be notified of certain NTMs with a delay. After processing the NTM data, we merge each NTM dataset with bilateral trade and country characteristics to create a bilateral panel dataset at the HS6 digit level. This dataset includes the count of NTMs imposed by an importer on an exporter. Since our analysis focuses on SPS and TBT measures, all other NTMs are grouped into a single aggregated category. To maintain consistency across analyses, we use TRAINS data for non-SPS and non-TBT NTMs in both regressions (the gravity model with TRAINS data and with WTO notification data). This approach ensures that the only variation in estimated AVEs is due to differences in data sources for the SPS and TBT measures, allowing for a more direct comparison. [Supporting Information: Table 3](#) in Section IV-III of the appendix presents the summary statistics for the observed count of NTMs in the finalized dataset. The average number of applied NTMs does not seem to vary between the two datasets, but the third quantile reveals that the WTO data for SPS measures contain more zeros than the TRAINS data, as shown in [Tables 4 and 5](#), which report the percentage of observations with an NTM count of zero. Additionally, the maximum number of NTMs is considerably higher in the WTO data for both SPS and TBT cases, which might be due to unreported withdrawals of NTMs in the WTO.

We also compare NTM coverage across HS6 codes and differences in NTM counts between the two datasets, specifically by looking at the variation in NTM coverage at the HS6 digit level and differences in the total number of NTMs imposed in TRAINS and WTO data. [Supporting Information: Figures 7 and 8](#) in appendix Section IV-IV display the number of distinct HS codes affected by NTMs over the years for each importing

country. We observe discrepancies in the number of HS6 codes covered by at least one NTM and note that these discrepancies vary both across importers and NTM types. Another perspective is provided in [Supporting Information: Figures 9 and 10](#), which, instead of focusing on HS coverage, represent the total number of NTMs. Since we account for NTM stringency through total NTM counts, this metric is equally important. Interestingly, for some countries, even when the number of distinct HS codes covered is similar across datasets, the total NTM count still differs. These observed discrepancies likely stem from measurement errors in WTO notification data. These figures highlight two main factors contributing to differences in AVE estimates between the datasets: either the absence of certain HS6 codes from NTM coverage or discrepancies in the total number of NTMs applied to a given HS code. A wide gap in NTM coverage could contribute to significant differences in AVE estimates. However, if the impact of NTMs on trade is relatively small, the resulting difference in AVEs could be marginal, so estimating AVEs in each dataset remains essential. Finally, to compare NTM coverage across HS sections, [Supporting Information: appendix Section IV-V](#) provides a detailed comparison of NTM coverage across each HS section, with additional details on HS sections also given in [Supporting Information: appendix Section IV-VI](#).

5 | Results

5.1 | Differences in AVE Estimates

We obtained gravity coefficients from the 20,608 regressions (5152 products $\times$ 4 years), which are summarized in the [Supporting Information: appendix Section V-I](#). We then estimated AVEs at the product-importer-exporter-year level, yielding 15,023,232 AVEs for each dataset. To meaningfully compare AVE estimates between the two datasets, we match each AVE from the TRAINS dataset to its equivalent from the WTO dataset at the HS6 digit product-importer-exporter-year, and NTM type level. [Tables 10 and 11](#) in the [Supporting Information: appendix Section V-III](#) present summary statistics for the AVE estimates obtained for SPS and TBT measures across both datasets. Given that we estimate multiple regressions at the product-year level, we present the results for the obtained AVEs while also removing extreme values.⁹

We begin with a global comparison of AVE values due to convergence issues in the gravity regressions.¹⁰ [Tables 1 and 2](#) compare AVEs from TRAINS and WTO notifications across all estimates, showing the proportion of AVEs in TRAINS relative to their corresponding values in WTO. The two tables classify observations into three groups: AVEs equal to 0, NA (due to lack of convergence in regression estimation), or nonzero/non-NA across both NTM datasets. Comparing these categories helps assess dataset alignment and identify initial discrepancies. For both SPS and TBT, we find that AVEs are zero in both datasets for roughly 22%–24% of observations, reflecting cases where NTMs are recorded as zero in both datasets.¹¹ However, NTMs are recorded as zero more frequently than the proportion of zero AVEs (see [Supporting Information: appendix Section IV-III](#) for summary statistics of the NTM variables). This arises from convergence issues in the gravity estimation, which often occur when there is little

or no variation in the NTM variables (e.g., when NTMs are uniformly recorded as zero across all importers). As a result, AVEs are frequently missing (NA) in both datasets, as shown in Table 3. This table illustrates how the presence of NTMs is reflected in estimated AVEs across datasets and NTM types by reporting the share of observations in different NTM value categories and their corresponding AVEs. For both SPS and TBT, a large share of observations have NTMs equal to zero while AVEs are recorded as NA. This suggests that the missing values are not problematic, since they likely correspond to cases where the NTM count is zero and the actual AVE would also be zero.

A notable discrepancy arises when AVEs based on WTO data are nonzero while those from TRAINS are zero (occurring in 2% of SPS observations and 13% for TBT). Since AVEs are the product of the NTM gravity coefficient and count, they are zero when the NTM count is null. This suggests that these cases occur when NTMs are recorded in the WTO dataset but not in TRAINS. A likely explanation is that the WTO dataset assumes NTMs are never withdrawn, meaning some measures persist in its records even when absent from TRAINS. Conversely, in 9%–12% of SPS and TBT cases, TRAINS produces nonzero AVEs while WTO produces AVEs of zero. This indicates that some NTMs may be missing from the WTO dataset but are captured in TRAINS, possibly due to

incomplete WTO notifications by importers or the exclusion of NTMs implemented before 1995, which are not included in our WTO data. Finally, in 8% of SPS and 26% of TBT observations, both datasets yield nonzero AVEs. Within this subset for SPS measures, 4% have the same sign for AVEs, while 4% have opposite signs. For TBT measures, 14% of the AVEs share the same sign across both datasets, suggesting a consistent directional effect on trade. However, in 12% of TBT cases, AVEs have opposite signs. This noteworthy share of AVEs with opposite signs raises concerns, as it suggests conflicting interpretations of the impact of NTMs on trade. Such divergence highlights inconsistencies in the NTM data, which could significantly influence policy analysis.

The subsequent analysis focuses on AVEs that are neither NA nor equal to zero in both datasets. Specifically, we examine 20% of the original 15,023,232 SPS AVE observations (2.45% + 8.25% + 9.36% from Table 1) and 52% of the TBT observations (13.73% + 26.03% + 12.94% from Table 1). Because we drop some of the extreme AVE values as discussed previously, the number of observations is reduced for both SPS (5.18%) and TBT (27.7%). We choose a more conservative cutoff to be cautious not to base our findings on discrepancies in AVEs across datasets caused by outliers.

To quantify the differences in AVE estimates between TRAINS and WTO notification data, we plot the density distribution of the percentage point difference in AVEs (computed as TRAINS AVE minus WTO notification AVE for each product-importer-exporter-year observation), as shown in Figure 1. In the figure, SPS differences are shown in purple and TBT in pink. The two distributions exhibit distinct patterns: TBT is relatively symmetric around zero with differences dispersed on both sides, and the discrepancies are generally more substantial. In contrast, SPS is slightly skewed, with a higher concentration of positive differences and less dispersion. This pattern may be explained by the fact that TBT measures include more negative AVE values than SPS measures and that TBT AVEs are generally larger in magnitude across the two datasets (see [Supporting Information: Appendix section V-III](#)). From Table 4, the mean AVE difference is 0.57% points for SPS measures (median 0.00%) and 3.52% points for TBT measures (median of 0.01%). For both types of NTMs, the differences in AVEs appear broadly centered around zero. The interquartile range also indicates limited variation, with the first quartile between −0.35% and −1.60% points and the third quartile between 1.04% and 2.26% points. Although the difference may appear small, it remains substantial when viewed relative to the overall mean and median AVE values of our obtained estimates. When expressed relative to the average

TABLE 1 | Comparison of WTO and TRAINS AVE estimates for SPS measures (0, NA, and ≠0).

SPS AVEs		TRAINS		
		0	NA	≠ 0
WTO	0	22.92%	3.52%	9.36%
	NA	31.98%	7.68%	13.32%
	≠ 0	2.45%	0.51%	8.25%

TABLE 2 | Comparison of WTO and TRAINS AVE estimates for TBT measures (0, NA, and ≠0).

TBT AVEs		TRAINS		
		0	NA	≠ 0
WTO	0	24.18%	5.15%	12.94%
	NA	1.74%	4.44%	8.63%
	≠ 0	13.73%	3.17%	26.02%

TABLE 3 | Comparison of WTO and TRAINS AVEs and corresponding NTM count values.

Data	NTM	NTM = 0, AVE = NA	NTM = 0, AVE = 0	NTM = 0, AVE ≠ 0	NTM > 0, AVE = NA	NTM > 0, AVE = 0	NTM > 0, AVE ≠ 0
TRAINS	SPS	8.43%	57.30%	0.00%	3.28%	0.05%	30.93%
	TBT	8.56%	39.54%	0.00%	4.20%	0.12%	47.59%
WTO	SPS	44.68%	35.76%	0.00%	8.30%	0.04%	11.21%
	TBT	5.84%	42.08%	0.00%	17.51%	0.12%	34.44%

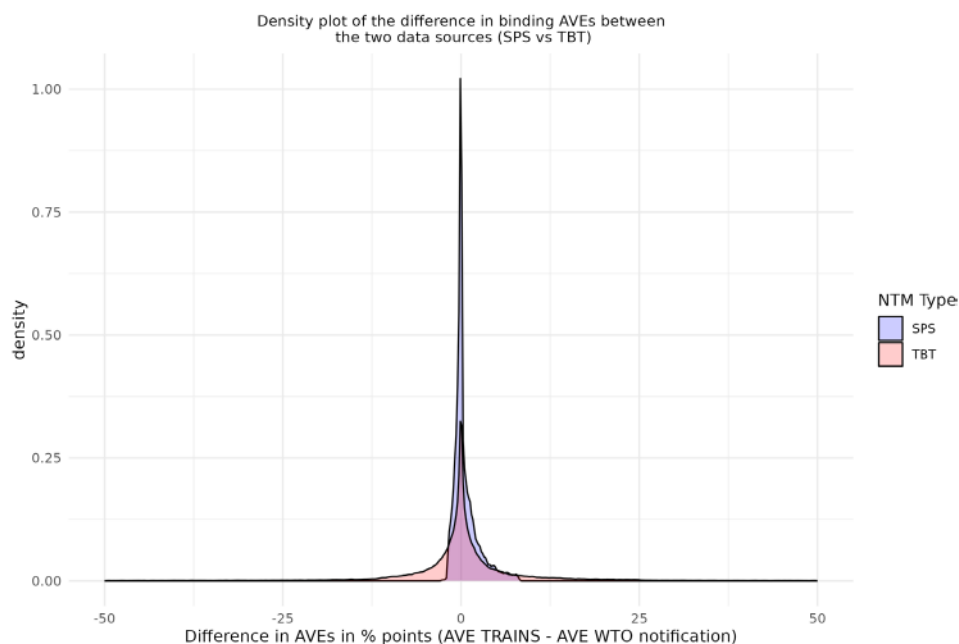


FIGURE 1 | Density plot of the difference in AVEs for both SPS and TBT measures (in percentage points).

TABLE 4 | Summary statistics for the difference in AVEs between WTO notifications and TRAINS.

Diff in AVE	Mean	Median	SD	Q25	Q75	Min	Max
SPS	0.57	0.00	1.72	−0.35	1.04	−2.80	8.22
TBT	3.52	0.01	24.68	−1.60	2.26	−62.36	325.64

Note: The summary statistics are calculated from the obtained AVEs, excluding observations where AVEs are null or NA in both datasets, and dropping extreme values.

of the AVE values, the observed differences are sizable. For SPS, the average difference corresponds to 88% of the mean TRAINS binding AVE (0.57/0.65), and for TBT, approx. 86% of the mean TRAINS binding AVE (3.52/4.09).

For robustness, we test alternative cutoff thresholds for excluding extreme AVEs. The results based on these alternative values, presented in [Supporting Information: Figure 12](#) in appendix Section V-IV, indicate that for TBT, changes in the cutoff do not noticeably affect the observed distribution of AVE differences between the two datasets. However, for SPS, we find that with smaller cutoffs (i.e., when more observations are retained in our sample), the distribution becomes less skewed around zero, and the number of cases with larger differences increases. The cutoff chosen for the results presented in this section is therefore more conservative, yet we still observe considerable differences when comparing the mean and median AVE values.

To further analyze these differences and capture their heterogeneity across the four cases presented in Tables 1 and 2, we examine the distribution of AVE differences by: (i) AVEs with the same sign in both datasets, (ii) AVEs with opposite signs, (iii) cases where AVEs are zero in TRAINS but nonzero in WTO notifications, and (iv) cases where AVEs are nonzero in TRAINS but zero in WTO notifications. Figure 2 displays

these four distributions, with SPS shown in purple and TBT in pink. The percentages above each panel indicate the share of observations relative to the total sample that fall into each case for both SPS and TBT. For example, among the SPS sample, 9.4% of the AVEs fall into case (i). We observe that for both SPS and TBT, the distribution of differences varies by case. In case (i), where the AVEs have the same sign across the two datasets, the distribution is tight around zero, indicating that the differences are generally small in magnitude. By contrast, in case (ii), where the AVEs obtained from the two datasets have opposite signs, the distribution is much more spread out, reflecting larger discrepancies, especially for TBT. Cases (iii) and (iv) arise when an NTM is recorded in one dataset but not in the other. Although the resulting differences are often centered near zero, a nontrivial share of observations still displays sizeable gaps. Such discrepancies may reflect bias, either from WTO notifications over-recording NTMs absent in TRAINS or from omissions of NTMs that appear in TRAINS. For SPS, case (iv) alone accounts for a large share of comparisons (80.5%), while for TBT, cases (iii) and (iv) together cover a substantial share, around 63% of observations in each case. Finally, the TBT distributions are largely symmetric, indicating that discrepancies arise regardless of whether the estimated AVE is positive (a trade barrier) or negative (a trade catalyst). This suggests that the differences are not systematically biased in one direction. By contrast, SPS AVEs are

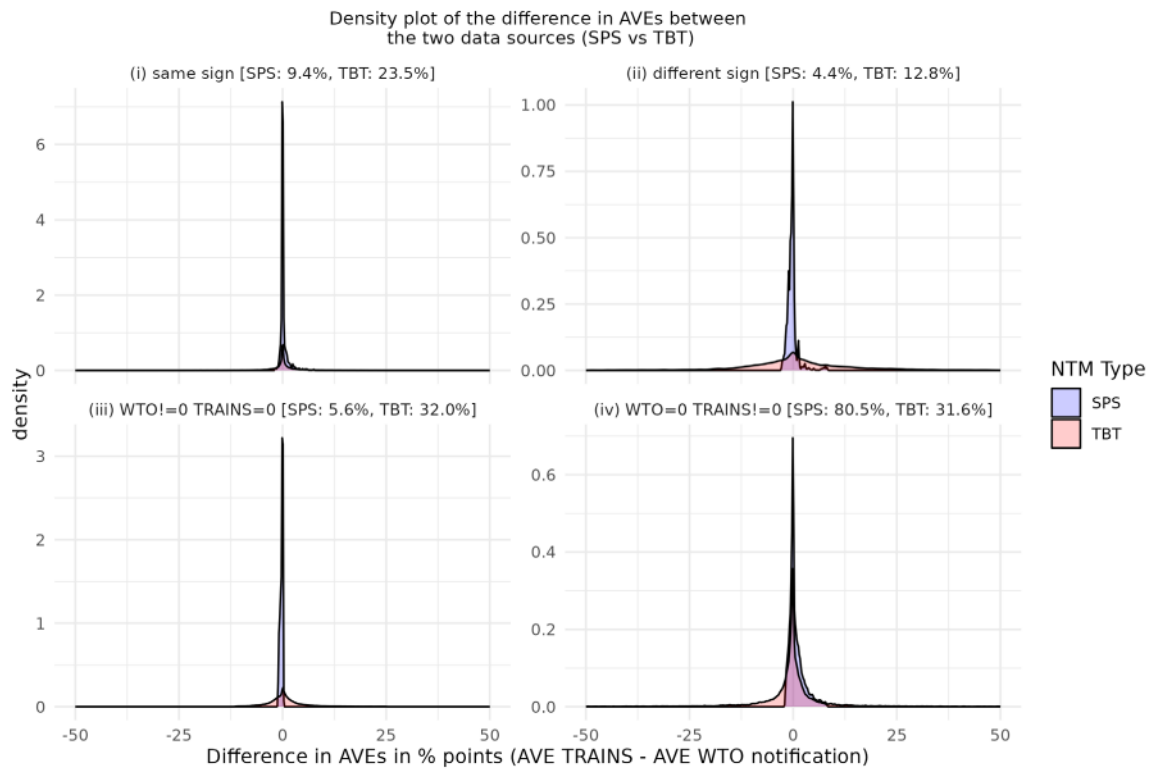


FIGURE 2 | Density plot of the difference in AVEs for both SPS and TBT measures (in percentage points).

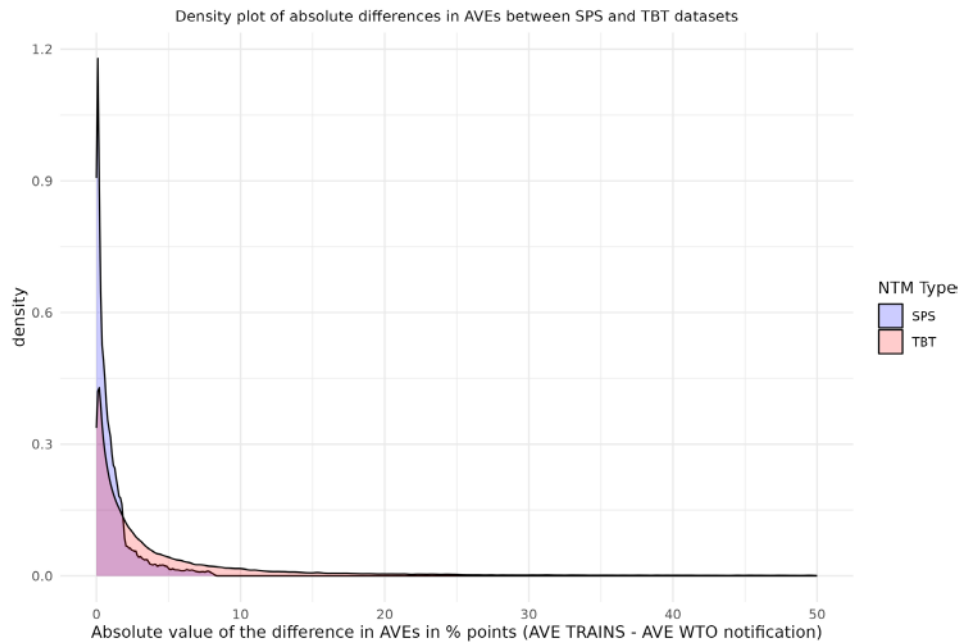


FIGURE 3 | Density plot of the absolute value difference in AVEs for both SPS and TBT measures (in percentage points).

TABLE 5 | Summary statistics for the absolute value of the difference in AVEs between WTO notifications and TRAINS (using truncated data).

Diff in AVE	Mean	Median	SD	Q25	Q75	Min	Max
SPS	1.09	0.58	1.45	0.16	1.38	0.00	8.22
TBT	7.83	1.90	23.67	0.55	5.78	0.00	325.64

predominantly positive, which makes the distribution of differences slightly right-skewed in case (iv).

From the above results, we find that although the differences in AVEs can be substantial for some observations, the overall distribution remains centered around zero. This could be because we allow for both positive and negative differences (i.e., whether TRAINS AVEs are greater or lower than WTO notification AVEs), and these offsetting effects pull the distribution toward zero. To better capture the extent of divergence, we also examine the absolute differences in AVEs between TRAINS and WTO notification data, which reflect only the magnitude of discrepancies regardless of direction. Figure 3 and Table 5 present the distribution and summary statistics for these absolute differences. On average, discrepancies amount to 0.58% points for SPS and 7.83% points for TBT. The medians are smaller (1.66 for SPS and 1.90 for TBT), which may seem modest at first. However, when compared with the average binding AVEs reported in Supporting Information: Table 11 (less than 1 for SPS and -1 to 4 for TBT), these median differences are not negligible. Moreover, the third-quartile values of 1.38 for SPS and 5.78 for TBT show that for a sizable share of observations, the discrepancies across the two datasets are economically meaningful, especially for TBT.

Finally, we examine whether differences in estimated AVEs across datasets vary by importing country, affected product, and year of NTM implementation. The results are presented in Supporting Information: appendix Section V-VI. This expanded analysis shows that while the distribution of AVE differences varies only slightly across importers and years, greater heterogeneity emerges when examined by HS section, where differences in AVEs appear more diverse across types of products.

Focusing on agricultural goods (HS sections 1–4), we report the absolute differences in AVEs regressed on HS section dummies in Supporting Information: Table 12 in appendix Section V-V. We find that the discrepancies are consistently larger for TBT measures than for SPS measures. On average, HS sections covering agricultural goods show AVE differences of about 1% point for SPS and 7% points for TBT between the two datasets. To illustrate discrepancies in AVE estimates for specific agricultural products, Supporting Information: Figure 13 in Appendix V-V presents the case of SPS measures imposed by Brazil on Argentina for a selection of HS6 digit agricultural goods. The comparison reveals that while some products (such as cane sugar) show highly consistent AVEs across both datasets, others (such as grapes or tomatoes) exhibit substantial divergence. These product-specific cases underscore important inconsistencies in AVE estimation methodologies between the two sources, emphasizing the need for greater alignment in NTM effect quantification.

5.2 | Explaining the Difference in AVE Estimates

To explore the observed bias and detect systematic errors, we regress the absolute value of the difference in AVEs on various factors that may contribute to measurement errors in the WTO data as follows:

$$|AVE_{ijkt}^{n, TRAINS} - AVE_{ijkt}^{n, WTO}| = importImp_{ikl} + Share_{ikl} + Share_{jkl} + GDPpercapita_{it} + GDPpercapita_{jt} + totNTM_{it}^n + AgProd_{kt} + \tau_t + \epsilon_{ijkt} \quad (16)$$

Explanatory variables include the importance of the imported product in total imports for the importing country ($importImp_{ikl}$), measured as the ratio of the good's import value to total imports in a given country; the importer's importance in the world market using its global market share for that specific good ($Share_{ikl}$); exporter's global market share ($Share_{jkl}$); GDP per capita of both the importing ($GDPpercapita_{it}$) and exporting countries ($GDPpercapita_{jt}$); and total number of SPS or TBT measures implemented by the importer in a given year, using TRAINS data ($totNTM_{it}^n$), which helps to assess whether countries that impose a high number of NTMs overall are more prone to reporting inconsistencies in the WTO notification data. We include $AgProd_{kt}$, a dummy variable indicating whether the product covered by the NTM is agricultural, with agricultural goods defined as HS sections 1–4. We control for time fixed effects (τ_t) to account for the assumption of cumulative implementation of NTMs over time in WTO notifications, which would overcount previous NTMs in the WTO as time progresses. We estimate the model using clustered standard errors at the product-importer-year level, restricting the sample to AVEs that are different from zero or NA in TRAINS or WTO notifications. The regression results are presented in Table 6. We first find that generally, fixed effects are statistically significant and tend to decline slightly over time. For SPS measures, importer market share is statistically significant at the 95% confidence level, with a negative coefficient. As the importer market share increases, the difference in AVEs tends to decrease. Exporter market share is also negative and significant at the 90% confidence level, indicating that larger exporters are likewise associated with smaller AVE differences. This may suggest that when an importing country has a larger presence in the global market, or when its trading partners have substantial market shares, discrepancies in reporting or notification of NTMs are less frequent. However, this pattern should not be attributed solely to misreporting, since other sources of measurement error, such as differences in how NTMs are recorded or interpreted across datasets, may also contribute. In the case of TBT measures, the coefficient on import importance is positive and significant at the 95% confidence level, indicating that when a good represents a larger share of the importer's import basket, the discrepancies in AVEs tend to be greater. Nevertheless, this analysis remains exploratory. The relatively low adjusted R -squared values indicate that the model explains only a small portion of the observed variation in AVE differences. This implies that additional, unobserved factors likely contribute to the discrepancies between the two datasets.

6 | Conclusion and Policy Implications

Accurate and comprehensive NTM data are essential for understanding the role of NTMs in global trade. Unlike tariffs, which are subject to stricter WTO oversight, NTMs are less

TABLE 6 | Regression results: Difference in AVE estimates for SPS and TBT measures.

	SPS	TBT
	abs(diff_AVE_SPS)	abs(diff_AVE_TBT)
Constant	1.537*** (0.1064)	7.104** (1.208)
Year2014	−0.3575*** (0.0188)	−1.315** (0.1911)
Year2015	−0.2933** (0.0228)	−1.469*** (0.0646)
Year2016	−0.2170*** (0.0124)	−1.111** (0.0940)
tot_SPS	-2.25×10^{-8} (1e−5)	—
Reporter_GDP_per_cap	-1.20×10^{-5} (1.1e−5)	0.0002 (8.41e−5)
Partner_GDP_per_cap	-2.83×10^{-7} (1e−5)	1.49×10^{-6} (1e−5)
import_importance	0.2803 (0.2766)	4.902* (1.343)
import_share	−0.0238* (0.0053)	−0.0960 (0.1321)
export_share	−0.0048. (0.0020)	0.0132 (0.0104)
Ag_Prod	0.0349 (0.0837)	−0.2141 (0.7457)
tot_TBT	—	-3.03×10^{-8} (1e−5)
S.E.: Clustered	By: Repo. and Prod. and Year	By: Repo. and Prod. and Year
Observations	778,565	4,176,296
R ²	0.01236	0.00201
Adjusted R ²	0.01235	0.00200

Note: Coefficient values and their standard errors (in parentheses) for the regressions in Equation (16). In the first column, the dependent variable is the difference in AVE for the SPS type, and in the second column, the dependent variable is the difference in AVE for the TBT type. Significance levels are denoted as follows: “***” 0.001, “**” 0.01, “*” 0.05, “.” 0.1.

transparent, harder to monitor, and increasingly used as substitutes for tariffs (Niu et al. 2018, 2020). Although NTMs are imposed unilaterally by importers, they can have discriminatory effects that disproportionately impact certain exporters. This makes transparency and reliable data even more crucial. Broad country coverage is important, since AVE estimates depend on cross-country variation. Insufficient coverage therefore increases the risk of introducing bias into policy analysis. Since NTMs vary significantly across countries due to differences in regulatory frameworks and enforcement practices, a dataset with extensive country coverage is necessary for more reliable policy analysis and research.

Time coverage is also important for identifying the impact of major trade events, such as the implementation of trade agreements or trade wars. For example, Chen et al. (2022) highlight the impact of NTMs in the United States–China trade war and find that NTMs accounted for nearly 50% of the overall decline in Chinese imports from the United States. However, their study infers the effects of NTMs by attributing trade impacts to unexplained trade costs rather than directly using NTM data. The lack of comprehensive NTM data and subsequent challenges in measuring AVEs introduce additional measurement errors by attributing impacts of changes in trade values to NTMs that might actually arise from other sources.

Finally, this study is particularly relevant for research on the agricultural and food sector, where SPS and TBT measures often have a significant impact on trade flows (Gourdon et al. 2020;

Disdier and Fugazza 2020). This relevance arises not only from the frequent application of NTMs in the sector but also from the substantial compliance burden they often impose (Santeramo and Lamonaca 2019). Meeting such standards typically requires significant adjustments to farming practices, the implementation of traceability mechanisms, and the acquisition of certifications. Moreover, the effects of NTMs tend to be asymmetric: while some exporters benefit by leveraging compliance to differentiate their products and gain access to premium markets, others, particularly smaller or less-resourced producers, may face disproportionate challenges or even be excluded from trade entirely.

The findings of this research demonstrate that data coverage is crucial, and while the TRAINS database has more reliable methods than WTO notifications for the data it collects, TRAINS does not always provide sufficient country and time coverage. To address this, we assess whether WTO data can produce AVE estimates similar to those derived from TRAINS when conditioned on the same country and time coverage. Our results indicate that at the HS6 digit product-importer-exporter-year level, AVE estimates frequently do not align across datasets. Among nonzero AVEs, the mean absolute difference between TRAINS and WTO estimates is roughly 1%–8% points for both SPS and TBT measures, which is substantial when considering the overall AVE distribution in our analysis. Additionally, we identify many cases where AVEs are zero in TRAINS but nonzero in WTO, and vice versa, as well as instances where the estimated AVEs switch sign across datasets. Such discrepancies suggest

that the direction of the effect may depend on the dataset used, raising concerns about inconsistent policy interpretations and potential bias in AVE estimates. In sum, our findings suggest that the assumptions required when using WTO notification data introduce substantial measurement errors in recorded NTMs and ultimately bias AVE estimates. Notably, the magnitude of these differences varies depending on the product, NTM type, importing country, and year. Moreover, for SPS measures, discrepancies in AVE estimates tend to decrease with the importer's share in the global market for that good. Although we attempt to identify conditions under which AVE estimates from both datasets align more closely, this remains challenging due to the multiple sources of measurement error inherent in WTO notification data, which are difficult to fully disentangle.

A limitation of this study is country coverage, as we only include nine importing countries. The choice to focus on a relatively small set of countries was driven by the limited overlap of countries covered in both data sets over successive years and their geographic proximity, which was important for our instrumental variable approach. However, this limits external validation and comes at the expense of generalizability across a broader set of countries. As more data become available across a wider variety of countries and for more years, future research might expand our analysis to test external validity while preserving the instrumental variable approach to address concerns of endogenous trade policies. In addition, further research could investigate how differences in AVEs between the two datasets vary across more disaggregated NTM types beyond SPS and TBT, as well as across products at a finer level of disaggregation. Such analyses would help reveal more nuanced patterns of heterogeneity in NTM effects.

Building on our findings, a key policy implication is to highlight the need for improved data collection and reporting of NTMs. Strengthening both the consistency and reliability of available data is essential for producing more robust and comparable AVE estimates. Currently, limitations in coverage across countries, products, and years, as well as variation in measurement errors across sources, narrow the scope of analysis and may introduce bias into empirical analysis. These constraints may also hinder policymakers' ability to base decisions on a clear and accurate understanding of NTM impacts. A more comprehensive and accurate dataset is essential not only for understanding the trade effects of NTMs but also for evaluating their broader economic and welfare implications. Given the challenges with NTM data that we explore in this analysis, some studies resort to estimating AVEs of NTMs without using direct NTM data, particularly in the services sectors (Fontagné et al. 2011; Guillin 2013). They infer AVEs from all unexplained variation in trade costs. However, this approach relies on the assumption that any trade value not explained by the gravity model or other known trade regulations, such as tariffs, is solely attributed to the effects of NTMs. This assumption can be problematic, as it overlooks other potential factors that influence trade.

The findings of this study highlight key challenges when using different NTM data sources to assess their trade effects, underscoring the importance of data accuracy when estimating AVEs. Specifically, we show that although the WTO notification data has better coverage than TRAINS, inaccurate reporting in the

WTO can lead to significant bias in AVE estimations. Given the central role of NTMs in trade policy, this work raises awareness of data limitations and encourages a more informed use of these datasets. Identifying measurement issues provides a foundation for future improvements in NTM data, contributing to more accurate and reliable trade policy analysis.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Endnotes

- ¹ The Multi-Agency Support Team group (MAST), established under UNCTAD, has been instrumental in shaping the International Classification of NTMs. This classification system is extensively used to categorize and grasp the diverse forms of NTMs impacting global trade.
- ² For instance, SPS-related information can be gathered from ministries of agriculture, standardization agencies, and sometimes even from WTO notifications (UNCTAD 2018).
- ³ **Supporting Information:** appendix Section I provides a comparison of data coverage between WTO notification data and TRAINS data.
- ⁴ Neighboring countries are defined based on the geographical distance between their capitals.
- ⁵ NTMs other than SPS and TBT are categorized in one variable, but, as this is not the focus of our study, we will mostly focus on the results for SPS and TBT.
- ⁶ We use import trade value as a proxy for trade quantity following the literature (Ferrantino 2006), since quantity data are often noisy or incomplete (Kee et al. 2009; Bratt 2017).
- ⁷ The importing and exporting countries used in this analysis can be found in the **Supporting Information:** appendix section IV-I.
- ⁸ This is equivalent to 15,023,232 observations. We have $4 \times 5152 \times 9 \times 81$ observations. Although there are 82 distinct exporters, we exclude observations where the importer is the same as the exporter.
- ⁹ Specifically, we drop AVE values below the 10th percentile and above the 90th percentile, as some estimates reach extreme magnitudes. This issue was also noted by Bratt (2017), who similarly excluded extreme AVEs from their analysis, attributing these outliers to the estimation process, which relies on a combination of six different coefficients to compute AVEs. To check the robustness of our results to the choice of cutoff, we present the results using alternative thresholds in the **Supporting Information:** appendix Section V-IV. These show that the reported results are best interpreted as a lower bound, while the main conclusions remain unchanged across different cutoffs.
- ¹⁰ Details on convergence issues are provided in the appendix Section V-II.
- ¹¹ The AVE takes the value of exactly zero when the NTM associated with a specific product–importer–exporter–time observation is null. However, if the NTM has only a negligible effect, the AVE will instead be close to zero, reflecting a coefficient in the gravity regression that is itself near zero.

References

- Adarov, A., and M. Ghodsi. 2023. "Heterogeneous Effects of Nontariff Measures on Cross-Border Investments: Bilateral Firm-Level Analysis." *Review of International Economics* 31: 158–179.
- Aghion, P., N. Bloom, R. Blundell, R. Griffith, and P. Howitt. 2005. "Competition and Innovation: An Inverted-U Relationship." *Quarterly Journal of Economics* 120: 701–728.

- Aghion, P., R. Blundell, R. Griffith, P. Howitt, and S. Prantl. 2009. "The Effects of Entry on Incumbent Innovation and Productivity." *Review of Economics and Statistics* 91: 20–32.
- Aghion, P. 2018. "Innovation and Growth From a Schumpeterian Perspective." *Revue D'économie Politique* 128: 693–711.
- Anderson, J. E., and E. Van Wincoop. 2003. "Gravity With Gravitas: A Solution to the Border Puzzle." *American Economic Review* 93: 170–192.
- Arita, S., L. Mitchell, and J. Beckman. 2015. "Estimating the Effects of Selected Sanitary and Phytosanitary Measures and Technical Barriers to Trade on U.S.-EU Agricultural Trade." Economic Research Report No. ERR-199, U.S. Department of Agriculture, Economic Research Service.
- Baier, S. L., and J. H. Bergstrand. 2009. "Bonus Vetus OLS: A Simple Method for Approximating International Trade-Cost Effects Using the Gravity Equation." *Journal of International Economics* 77: 77–85.
- Beghin, J. C., M. Maertens, and J. Swinnen. 2015. "Nontariff Measures and Standards in Trade and Global Value Chains." *Annual Review of Resource Economics* 7: 425–450.
- Blöthner, T., and M. Larch. 2022. "Economic Determinants of Regional Trade Agreements Revisited Using Machine Learning." *Empirical Economics* 63: 1771–1807.
- Bratt, M. 2017. "Estimating the Bilateral Impact of Nontariff Measures on Trade." *Review of International Economics* 25: 1105–1129.
- Cadot, O., and J. Gourdon. 2016. "Non-Tariff Measures, Preferential Trade Agreements, and Prices: New Evidence." *Review of World Economics* 152: 227–249.
- Cadot, O., J. Gourdon, and F. van Tongeren. 2018. "Estimating Ad Valorem Equivalents of Non-Tariff Measures: Combining Price-Based and Quantity-Based Approaches." OECD Trade Policy Papers No. 215, OECD Publishing, Paris.
- Chen, T., C. T. Hsieh, and Z. M. Song. 2022. "Non-Tariff Trade Barriers in the U.S.-China Trade War." NBER Working Paper No. 30318, National Bureau of Economic Research.
- Conte, M., P. Cotterlaz, and T. Mayer. 2022. "The CEPII Gravity Database." CEPII Working Paper No. 2022–05, CEPII (Centre d'Études Prospectives et d'Informations Internationales), Paris.
- Curzi, D., V. Raimondi, and A. Olper. 2015. "Quality Upgrading, Competition and Trade Policy: Evidence From the Agri-Food Sector." *European Review of Agricultural Economics* 42: 239–267.
- Daron, A., P. Aghion, and F. Zilibotti. 2006. "Distance to Frontier, Selection, and Economic Growth." *Journal of the European Economic Association* 4: 37–74.
- de Melo, J., and A. Nicita. 2018. "Non-Tariff Measures: Data and Quantitative Tools of Analysis." FERDI Working Paper No. P218, FERDI, Clermont-Ferrand, France.
- Disdier, A. C., and M. Fugazza. 2020. *A Practical Guide to the Economic Analysis of Non-Tariff Measures*. United Nations.
- Essaji, A. 2008. "Technical Regulations and Specialization in International Trade." *Journal of International Economics* 76: 166–176.
- Fell, J., and A. Duver. 2024. "Non-Tariff Measures: A Methodology for the Quantification of Bilateral Trade Effects of Policy Measures at a Product Level." *Applied Economics* 56: 4374–4388.
- Ferrantino, M. J. 2006. "Quantifying the Trade and Economic Effects of Non-Tariff Measures." OECD Trade Policy Papers No. 28, OECD Publishing, Paris.
- Fiankor, D. D. D. 2023. "Estimating Ad Valorem Equivalents of Non-Tariff Measures in Swiss Agriculture." *Agroscopie Science* 156: 1–35.
- Fontagné, L., A. Guillin, and C. Mitaritonna. 2011. "Estimations of Tariff Equivalents for the Services Sectors."
- Ghodsi, M., J. Grübler, O. Reiter, and R. Stehrer. 2017. "The Evolution of Non-Tariff Measures and Their Diverse Effects on Trade." Working Paper, Wiiw Research Report.
- Ghodsi, M., O. Reiter, and R. Stehrer. 2015. "Compilation of a Database for Non-Tariff Measures From the WTO Integrated Trade Intelligence Portal (WTO I-TIP)." Working Paper/Project Report No. PRONTO/wiiw, Vienna.
- Ghodsi, M., and R. Stehrer. 2022. "Trade Policy and Global Value Chains: Tariffs Versus Non-Tariff Measures." *Review of World Economics* 158: 887–916.
- Gourdon, J., S. Stone, and F. van Tongeren. 2020. "Non-Tariff Measures in Agriculture." OECD Food, Agriculture and Fisheries Papers No. 147, OECD Publishing, Paris.
- Grübler, J., and O. Reiter. 2021. "Characterising Non-Tariff Trade Policy." *Economic Analysis and Policy* 71: 138–163.
- Guillin, A. 2013. "Assessment of Tariff Equivalents for Services Considering the Zero Flows." *World Trade Review* 12: 549–575.
- Ing, L. Y., S. F. de Cordoba, and O. Cadot. 2016. "Non-Tariff Measures in ASEAN." In *Economic Research Institute for ASEAN and East Asia and United Nations Conference on Trade and Development*.
- Kee, H. L., A. Nicita, and M. Olarreaga. 2009. "Estimating Trade Restrictiveness Indices." *Economic Journal* 119: 172–199.
- Kee, H. L., and A. Nicita. 2022. "Trade Fraud and Non-Tariff Measures." *Journal of International Economics* 139: 103682.
- Kravchenko, A., A. Strutt, C. Utoktham, and Y. Duval. 2022. "New Price-Based Bilateral Ad-Valorem Equivalent Estimates of Non-Tariff Measures." *Journal of Global Economic Analysis* 7.
- Lee, J. W., and P. Swagel. 1997. "Trade Barriers and Trade Flows Across Countries and Industries." *Review of Economics and Statistics* 79: 372–382.
- Mao, R., Y. Liu, and X. Wang. 2023. "Economic and Environmental Impacts of Agricultural Non-Tariff Measures: Evidence Based on Ad Valorem Equivalent Estimates." *International Food and Agribusiness Management Review* 26: 379–396.
- Niu, Z., C. Liu, S. Gunessee, and C. Milner. 2018. "Non-Tariff and Overall Protection: Evidence Across Countries and Over Time." *Review of World Economics* 154: 675–703.
- Niu, Z., C. Milner, S. Gunessee, and C. Liu. 2020. "Are Nontariff Measures and Tariffs Substitutes? Some Panel Data Evidence." *Review of International Economics* 28: 408–428.
- Orefice, G. 2017. "Non-Tariff Measures, Specific Trade Concerns and Tariff Reduction." *World Economy* 40: 1807–1835.
- Porto, G. G. 2018. "Labor Market Effects of Non-Tariff Measures in Latin America." UNCTAD, pp. 195.
- Raimondi, V., A. Piriou, J. Swinnen, and A. Olper. 2023. "Impact of Global Value Chains on Tariffs and Non-Tariff Measures in Agriculture and Food." *Food Policy* 118: 102469.
- Rau, M. L., and A. Vogt. 2017. "Data Concepts and Sources of Non-Tariff Measures (NTMs): An Exploratory Analysis." Working paper No. ETSG-Paper 267, ETSG/World Trade Institute, Bern.
- Santeramo, F. G., and E. Lamonaca. 2019. "The Effects of Non-Tariff Measures on Agri-Food Trade: A Review and Meta-Analysis of Empirical Evidence." *Journal of Agricultural Economics* 70: 595–617.
- Swinnen, J. 2016. "Economics and Politics of Food Standards, Trade, and Development#." *Agricultural Economics* 47: 7–19.

Trefler, D. 1993. "Trade Liberalization and the Theory of Endogenous Protection: An Econometric Study of US Import Policy." *Journal of Political Economy* 101: 138–160.

UNCTAD. 2018. "UNCTAD TRAINS: The Global Database on Non-Tariff Measures User Guide (2017, Version 2)."

Xiong, B., and J. Beghin. 2014. "Disentangling Demand-Enhancing and Trade-Cost Effects of Maximum Residue Regulations." *Economic Inquiry* 52: 1190–1203.

Supporting Information

Additional supporting information can be found online in the Supporting Information section. **Data S1:** aepp70039-sup-0001-Appendix.pdf.