What Matters for Agricultural Trade?
Assessing the Role of Trade Deal Provisions using Machine Learning *

Stepan Gordeev Jeremy Jelliffe Dongin Kim Sandro Steinbach

Abstract

This paper applies machine learning methods to identify provisions in preferential trade agreements (PTAs) that enhance agricultural trade and to determine the socio-economic and political factors influencing their inclusion. Using a theory-consistent three-way gravity model, we employ Lasso regularized regression to identify PTA provisions impacting agricultural trade. The Lasso estimates highlight the relevance of competition policies, export taxes, intellectual property rights, capital movement, state enterprises, and technical trade barriers. Next, we employ Random Forests to determine the economic, political, social, and geographic factors driving the inclusion of these PTA provisions. The results emphasize the importance of contagion, governance quality, energy use, and geographic proximity.

Keywords: Preferential trade agreements, non-tariff provisions, machine learning, agricultural trade, Lasso regularized regression, Random Forests

JEL Codes: F14, Q17

---

* Stepan Gordeev, Corresponding Author, Department of Agricultural and Resource Economics, University of Connecticut, email: stepan.gordeev@uconn.edu; Jeremy Jelliffe, Economic Research Service, United States Department of Agriculture, email: jeremy.jelliffe@usda.gov; Dongin Kim, Department of Agricultural and Resource Economics, University of Connecticut, email: dong.i.kim@uconn.edu; and Sandro Steinbach, Department of Agribusiness and Applied Economics, North Dakota State University, email: sandro.steinbach@ndsu.edu. We acknowledge the excellent research assistance provided by Deexith Reddy and thank participants of the 2022 AAEA annual meeting and the 2023 IATRC summer symposium for their comments on an earlier version of this paper. This work was supported by the U.S. Department of Agriculture, National Institute of Food and Agriculture through the Agriculture and Food Research Initiative Award 2022-67023-36399. The findings and conclusions in this work are those of the authors and should not be construed to represent any official USDA or U.S. Government determination or policy.

1. Introduction

Multilateral economic integration through the World Trade Organization (WTO) has largely given way to bilateral or regional trade agreements in recent decades (Limão 2016). As of June 2023, the WTO recognizes 355 active preferential trade agreements (PTAs) with 583 notifications from WTO members, covering goods, services, and accessions separately (WTO 2023). These agreements have evolved from focusing primarily on reducing import tariffs to addressing non-tariff barriers and harmonizing behind-the-border policies. This design shift is evident in the provisions found in modern PTAs, which now encompass areas as diverse as technical standards, public procurement regulation, or intellectual property rights (Dür, Baccini and Elsig 2014; Hofmann, Osnago and Ruta 2017, 2019; Mattoo, Rocha and Ruta 2020). Modern trade agreements also increasingly include agricultural trade in their purview, treating it either alongside other sectors or in specialized provisions like sanitary measures and rules of origin (Thompson-Lipponen and Greenville 2019). The transition toward bilateral and regional trade agreements and its implications for agricultural trade have garnered significant research interest (see, e.g., Grant and Lambert 2008; Disdier, Fontagné and Mimouni 2008; Santeramo and Lamonaca 2019; Duvaleix et al. 2021; He 2022). However, due to the complex and multidimensional nature of modern PTAs, estimating the trade implications of different provisions and understanding the determinants of PTA design are daunting empirical challenges.

The economic literature offers various explanations for why countries enter into bilateral or regional trade agreements. Previous studies focused on competition among major trading nations, mobilization of interest groups, and market characteristics (Baldwin 1993; Chase 2008; Dür 2007; Limão 2016). Regarding the agricultural sector, there is only limited research highlighting federations and market power as influential factors in a country's decision to engage in PTAs (Ruppel, Boadu and Peterson 1991; McCalla 1992; Josling et al. 2010). Most empirical studies omit the agricultural and food sector entirely and fail to explain the inclusion of additional provisions beyond tariff reductions and those covered under the WTO mandate (Dür, Baccini and Elsig 2014; Falvey and Foster-McGregor 2018; Lee, Rocha and Ruta 2021). In contrast, a significant body of literature has been examining the relationship between PTAs and agricultural trade, exploring aspects such as trade creation and diversion (Grant and Lambert 2008; Sheldon, Chow and McGuire 2018;

He 2022), sanitary and phytosanitary measures (SPS) and rules of origin (Josling 2006; Disdier, Fontagné and Mimouni 2008; Duvaleix et al. 2021), and economic welfare implications (Alston et al. 1997; Martin 2018). Recent research has gone beyond studying the overall impact of PTAs and looked at the economic implications of specific provisions (Rodrik 2018; Breinlich et al. 2022; Kim and Steinbach 2023). However, including numerous PTA provisions in empirical studies poses a challenge due to pervasive collinearity concerns (Mattoo, Rocha and Ruta 2020). Conventional statistical methods, such as gravity regressions with binary indicators for individual provisions, fail to address these identification issues, casting doubts on the conventional wisdom regarding the implications of PTAs (Breinlich et al. 2022; Kim and Steinbach 2023).

This paper leverages modern machine learning (ML) techniques, combined with high-dimensional economic data, to identify PTA provisions that promote agricultural trade and to examine the socio-economic and political factors that determine the inclusion of such provisions. To evaluate the impact of PTA provisions on agricultural trade, we extend the theory-consistent three-way gravity model using a Lasso regularized regression approach (Yotov et al. 2016; Breinlich et al. 2022). This approach addresses empirical challenges associated with overfitting and multicollinearity, a particular issue in earlier studies focusing on the trade implications of a single or few provisions simultaneously (Scoppola, Raimondi and Olper 2018; Duvaleix et al. 2021; He 2022). To identify the PTA provisions that enhance agricultural trade, we employ repeated simulations and apply the plug-in Poisson pseudo-maximum likelihood (PML) estimator for regularization regression, as proposed by Breinlich et al. (2022). Through this analysis, we identify nine provisions related to competition policy, export taxes, intellectual property rights (IPR), movement of capital, state enterprises, and technical barriers to trade (TBT) that enhance agricultural trade. A subsequent post-Lasso gravity estimation reveals that six PTA provisions exhibit a statistically significant impact on agricultural trade. More specifically, we find that provisions on competition policies increase agricultural trade by 40.6 percent, while including those on export taxes and state enterprises leads to an increase in agricultural trade by 23.3 percent and 139.6 percent, respectively. In contrast, PTAs with provisions covering geographical indicators (GIs) reduce agricultural trade by 36.5 percent, likely driven by considerable heterogeneity across product groups.

We proceed by constructing Random Forest (RF) models (Breiman 2001) for each of the identified

PTA provisions important for agricultural trade to determine the significant economic, political, social, and geographic factors that influence their inclusion in modern trade agreements. The presence of a specific provision promoting agricultural trade within an agreement can be influenced by various factors, including domestic economic and political conditions, existing trade relationships, and shared cultural or institutional characteristics (Baier and Bergstrand 2004; Baccini and Dür 2012; Gamso and Grosse 2021; Bergstrand, Egger and Larch 2016). We assemble a dataset of hundreds of factors observed at the individual country and country-pair levels. Traditional econometric techniques often struggle to account for the non-linear and interactive effects of these factors, leading previous studies to focus on individual mechanisms one at a time (Raess, Dür and Sari 2018; Kucik 2012; Lechner 2016; Raimondi et al. 2023). To address this limitation, we employ RF models to predict the inclusion of provisions fostering agricultural trade within PTAs. RFs excel in handling high-dimensional and interacted predictors (Ziegler and König 2014). We utilize a robust variable importance measure to determine the critical determinants of PTA provisions related to agriculture (Altmann et al. 2010). Our analysis reveals that contagion, wherein a country negotiating a provision into its trade agreement prompts other countries competing for the same market to follow suit, is a key driver of all PTA provisions. These findings extend the results of earlier studies on PTA formation (Baldwin 1993; Egger and Larch 2008; Baldwin and Jaimovich 2012). Several other factors consistently emerge as important determinants across multiple PTA provisions relevant to agricultural trade, including the governance quality of the involved countries, their energy consumption, and various geographic measures. These results highlight areas where future work should focus on identifying the mechanisms through which these factors drive the ability—and political will—of trading partners to include agriculture-relevant provisions in their trade agreements.

Our paper contributes to the existing literature on the agricultural trade implications of PTA provisions and the factors that influence the inclusion of those provisions in modern trade agreements. We make two distinct contributions in this regard. First, we build upon the extensive body of literature that utilizes the three-way gravity framework to estimate the effects of individual PTA provisions on agricultural trade (e.g., Disdier, Fontagné and Cadot 2014; Scoppola, Raimondi and Olper 2018; Huysmans and Swinnen 2019; Duvaleix et al. 2021). We extend this framework by incorporating ML techniques for feature selection. Specifically, we employ a plug-in Lasso regularized regression approach to address multicollinearity concerns recognized in earlier studies as a challenge for model identification (Breinlich et al. 2022; Kim and Steinbach 2023). Through this approach, we identify the PTA provisions that have a statistically significant and economically meaningful impact on agricultural trade. Our findings indicate that provisions related to anti-discriminatory policies and the movement of capital effectively facilitate agricultural trade, while those associated with intellectual property rights hinder it.

Second, we leverage ML methods to address a significant challenge in evaluating the determinants of PTA provisions. The design of PTAs involves numerous potential determinants that may have non-linear and interacting effects, making it infeasible to study them using standard econometric techniques (Varian 2014; Mullainathan and Spiess 2017). To overcome this limitation, we adopt the RF approach, which allows us to effectively analyze high-dimensional data and simultaneously assess multiple potential mechanisms behind including PTA provisions that promote agricultural trade. Using two complementary modern ML techniques in this way can provide valuable insights for policymakers. By applying Lasso and RF methods, our research identifies PTA provisions that promote agricultural trade and explores the socioeconomic, political, and geographic factors that influence the inclusion of these provisions. These findings contribute to a better understanding of the factors driving agricultural trade and offer potential avenues for addressing global challenges like hunger and climate change through trade integration policies.

The remainder of this paper is organized as follows. Section 2 discusses the evolving nature of PTAs. Section 3 describes the two ML methods and data used. Section 4 presents results on identifying binding PTA provisions and their effect on agricultural trade. Section 5 presents results on the economic, political, social, and geographic determinants of binding PTA provisions. Section 6 concludes.

2. Background

Since the early 1990s, governments have increasingly pursued bilateral and regional trade agreements to enhance economic and political integration among participating countries beyond the multilateral commitments of the WTO (Hofmann, Osnago and Ruta 2017). The number of active

PTAs has steadily increased over the past two decades, reaching 357 as of June 2023. As more countries have joined these agreements, the landscape of preferential trade has become more complex, leading to changes in their content and configuration (Mattoo, Rocha and Ruta 2020). As the PTA regime evolves towards a more integrated economy, the agreements have expanded to include various market access concessions surpassing the WTO commitments of individual countries. In fact, over 90 percent of PTAs have reduced tariff rates below the most favored nation rate of the WTO, with many aiming for eventual duty-free status (Thompson-Lipponen and Greenville 2019). Furthermore, modern PTAs have incorporated norms for non-tariff measures to complement the tariff reduction policies. Panel (A) of Figure 1 shows the number of those non-tariff policy areas covered in PTAs over time. Until the late 1990s, most PTAs covered fewer than ten policy areas. However, a significant shift occurred then, with most new PTAs now encompassing between 10 and 20 policy areas. Particularly after the 2010s, there has been a trend toward PTAs covering more than 20 policy areas. This trend highlights the broadening scope and depth of trade and policy issues these agreements address. Panel (B) of the same figure reveals a consistent increase in the average number of PTA provisions over time, particularly since the mid-1990s. This pattern demonstrates the expanding range of commitments and obligations that PTAs entail. Overall, PTAs have evolved into comprehensive agreements beyond traditional tariff reductions (Hofmann, Osnago and Ruta 2019). They now encompass various policy areas and non-tariff measures, reflecting a deeper integration of economies and a broader range of commitments among participating countries.

Most PTAs contain various provisions relevant to agricultural trade. SPS, TBT, and trade remedies provisions started to be included in PTAs in the mid-1990s (Josling 2006). While SPS and TBT measures can be employed in a protectionist manner to impede trade flows (Peterson et al. 2013; Murina and Nicita 2017), they can also facilitate agricultural trade by providing assurance to consumers about the safety of products, particularly when information on traded products is incomplete (Disdier, Fontagné and Mimouni 2008). Furthermore, these measures are being utilized to enhance transparency and implement equivalence-based approaches, potentially contributing to smoother trade flows (Disdier, Fontagné and Cadot 2014). The elimination of import tariffs has heightened the relevance of trade remedies, including those pertaining to the agricultural sector. Among the trade remedies utilized, safeguard duties are the most frequently observed, followed by anti-dumping and countervailing duties. While these measures are often perceived as protective measures for domestic agricultural producers, some studies have found them to be an ineffective policy instrument due to the trade diversion effects they can cause (Carter and Gunning-Trant 2010; Carter and Steinbach 2018). Modern PTAs have expanded into policy areas beyond the scope of the WTO, encompassing domains such as IPR, agricultural research and development (R&D), and export measures.

There has been an increasing relevance of agriculture-related IPR provisions (Thompson-Lipponen and Greenville 2019). In the last decade, most agreements that have come into force include provisions for protecting plant varieties or reference geographical indications (GI). Particularly, commitments to protect specific agricultural products are prevalent, primarily used by the European Union (EU). The impact of these provisions within and outside the EU member states has been the subject of several studies, yet the empirical evidence regarding their trade effects remains inconclusive (Moschini, Menapace and Pick 2008; Huysmans and Swinnen 2019; Duvaleix et al. 2021; Curzi and Huysmans 2022). About one-third of the active PTAs incorporate provisions for agricultural research, development, and training (Hofmann, Osnago and Ruta 2017). These explicit references and potential implicit inclusion in broader provisions highlight the significance of international cooperation in agriculture. Covered policy areas include cooperation on vocational training and research within the framework of standard agricultural policies. Lastly, regarding export restrictions and other measures, Fulponi, Shearer and Almeida (2011) found that most PTAs include chapters prohibiting quantitative export restrictions, except for reasons falling under Article XI of the General Agreement on Tariffs and Trade (GATT). However, a few bilateral and regional trade agreements exempt certain agricultural products from the prohibition of export restrictions, implying that those may play an important role in agricultural trade (Thompson-Lipponen and Greenville 2019).

3. Empirical Methods and Data

3.1 Identifying Binding PTA Provisions using Plug-In Lasso Regularized Regression

To identify PTA provisions that promote agricultural trade, we rely on a panel data gravity model of international trade that is consistent with economic theory (Baier and Bergstrand 2007; Yotov et al. 2016; Sun and Reed 2010). Our modeling approach is multiplicative, as we represent the expected trade flows as an exponential function of relevant covariates, along with three sets of high-dimensional fixed effects. These fixed effects account for multilateral trade resistances and unobserved time-invariant trade costs (Anderson 2011; Fally 2015). Within the three-way gravity framework, we tackle the empirical challenges of dealing with numerous correlated covariates and an abundance of zero observations simultaneously (Silva and Tenreyro 2006; Correia, Guimarães and Zylkin 2020; Weidner and Zylkin 2021). Specifically, we introduce a variable selection algorithm to the gravity model, which allows us to identify PTA provisions with a non-zero impact on agricultural trade (Breinlich et al. 2022). We start with an empirical model that assesses the relationship between trade flows X_{ijt} and PTA provisions τ_{ijt}:

$$\mu_{ijt} := \mathrm{E}\left(X_{ijt}|\tau'_{ijt}, \alpha_{it}, \gamma_{jt}, \delta_{ij}\right) = \exp\left(\tau'_{ijt}\beta' + \alpha_{it} + \gamma_{jt} + \delta_{ij}\right) \ , \tag{1}$$

where i, j, and t denote the exporter, importer, and year, respectively. X_{ijt} indicates agricultural exports from country i to country j in year t and τ'_{ijt} denotes the vector of PTA provisions, which includes each provision included in an enforced bilateral or regional trade agreement.[1] We account for the multilateral trade resistances with the high-dimensional fixed effects α_{it} and γ_{jt}. In addition, we include time-invariant exporter-importer fixed effects δ_{ij}, which account for unobserved trade costs potentially correlated with the PTA provisions.

Traditional econometric methods could fail to identify the parameters of interest in Equation 1 accurately due to overfitting and multicollinearity (Mattoo, Rocha and Ruta 2020; Breinlich et al. 2022; Kim and Steinbach 2023). This empirical challenge arises due to the large number of PTA provisions likely correlated with each other. In addition to the many explanatory variables, only a subset of those PTA provisions will have a binding (non-zero) effect on trade flows (Breinlich et al. 2022). To resolve this empirical challenge, we adopt a two-stage approach, where we rely on an ML variable selection algorithm to identify binding PTA provisions in the first stage and estimate the

[1] We use the first lag of the PTA provision indicators to account for the potential endogeneity caused by different enforcement dates in a given year following Baier and Bergstrand (2007) and Anderson and Yotov (2010).

impact of those provisions on agricultural trade in the second stage. This approach circumvents potential econometric and research design challenges as the data patterns determine the model specification through repetitive simulations with varying covariate sets. More specifically, we rely on a plug-in Lasso regularized regression approach, which allows us to specify the regression model to be consistent with the gravity model of international trade (Breinlich et al. 2022). We amend the minimization problem that defines the three-way gravity by adding a penalization term that purges PTA provisions with coefficients equal to zero:

$$\left(\hat{\alpha}, \hat{\gamma}, \hat{\delta}, \hat{\beta}\right) := \arg \min_{\alpha,\gamma,\delta,\beta} \frac{1}{n} \left(\sum_{i,j,t} \left(\mu_{ijt} - X_{ijt} \ln \mu_{ijt} \right) \right) + \frac{1}{n} \sum_{l=1}^{m} \lambda \hat{\phi}_l |\beta_l| \, , \tag{2}$$

where the notation is the same as in Equation 1, apart from n, which indicates the number of observations. The first part of Equation 2 represents the standard Poisson Pseudo Maximum Likelihood (PML) minimization problem using the pseudo-likelihood function, while the second part is the Lasso penalty term, which consists of two tuning parameters, $\lambda \geq 0$ and $\hat{\phi}_l \geq 0$. Refining the model iteratively across PTA provisions, the tuning parameters shrink the β coefficients to zero. The Lasso penalty, denoted by λ, determines the extent of penalization applied to all regressors. By adjusting the lambda value, researchers can control the rate at which the penalization occurs. In the plug-in Lasso methods, we use a diagonal matrix $\hat{\phi}_l$ to account for regressor-specific penalty weights in addition to the standard Lasso penalty λ outlined by Belloni et al. (2016). The regressor-specific penalty allows for the customization of penalty weights for individual regressors, taking into account heteroskedasticity and within-cluster correlation. Consequently, the plug-in method outperforms the regular Lasso by effectively addressing estimation bias (Belloni et al. 2012).[2] Because the high-dimensional fixed effects have a structural meaning in the three-way gravity framework, we do not penalize them in the plug-in Lasso. This implies that for any given β, α, γ, and δ, we estimate the model by solving the standard Poisson PML minimization problem.

[2] The plug-in method is parsimonious in selecting the variables, resulting in superior performance over cross-validation approaches in finite samples (Breinlich et al. 2022). Therefore, the post-Lasso estimates have a "near-oracle" property, which implies that they identify the correct model if the sample is sufficiently large (Belloni et al. 2012).

Based on the exponential mean form, the Poisson PML estimator addresses potential bias from censored observations and enables the inclusion of zero trade flows in the estimation equation (Silva and Tenreyro 2006). To handle statistical separation and convergence issues caused by high-dimensional fixed effects, we employ a robust modified version of the iteratively re-weighted least-squares algorithm (Correia, Guimarães and Zylkin 2020). This recent computational innovation ensures consistent estimation and asymptotic distribution of the three-way fixed effects model using the three-way PPML estimator (Weidner and Zylkin 2021). Lastly, following standard practice in the related literature, we cluster all standard errors at the exporter-importer level (Cameron and Miller 2015).

3.2 Identifying the Determinants of PTA Provisions using Random Forests

Once the plug-in Lasso regularized regression has selected the binding PTA provisions for agricultural trade outcomes, we use a different ML method to identify the important economic, social, geographic, and political factors that determine whether these provisions are included in a PTA between a pair of countries.

Random Forests — The objective for each PTA provision is to construct a model that predicts whether a given PTA between a country pair includes the provision in question or not, using various observable characteristics of the country pair as predictors. Because the outcome variable is binary (is the provision included in the PTA or not), the prediction task boils down to a binary classification problem. We employ the RF algorithm for this task (Breiman 2001). RF is a supervised machine learning algorithm that grows many classification trees (hence forest), with each tree using a random bootstrap sample of the dataset (hence random). Each classification tree is constructed node by node. At each node, the algorithm randomly selects a subset of predictor variables to consider and searches among them for the best predictor variable and the best way to split the tree into two branches on the values of this variable. For continuous variables, a threshold value is picked, with observations whose value of this variable is below the threshold going to one branch and observations whose value is above going to another. For categorical variables, a partition of categories into two branches is picked instead. The variable and the value/partition to split on are chosen to optimize the goodness of fit measure. Traditionally, the fit is assessed with the Gini measure of misclassification error, but we instead utilize the area under the receiver operating characteristic curve (AUC-ROC) as it achieves better discriminating ability, especially with imbalanced data (Biau and Scornet 2016; Ling, Huang and Zhang 2003). This modification is appropriate because, for most provisions, the number of PTAs that do not contain the provision far exceeds the number of PTAs that do. The tree keeps branching in this way until each terminal node contains a small pre-specified number of observations out of the tree's training sample. The value of the outcome variable for training observations in each terminal node (leaf) determines the tree's prediction of this branch. To obtain a prediction for a new observation, it is "dropped" down the tree, taking branches depending on the values of its predictors until it reaches the terminal node that provides the prediction. The prediction of the entire Random Forest is then simply a majority vote among the constituent trees' predictions.

The forest has three free parameters: the number of trees, the number of predictor variables to draw as splitting candidates at each node, and the terminal number of observations per node, at which point nodes are not split further. We pick a large enough number of trees so that further increases do not improve prediction error. We tune the other two parameters using k-fold cross-validation (with $k = 3$) for each provision's forest individually (Bischl et al. 2021). The conventional measure of predictive performance we employ for tuning is the out-of-bag (OOB) misclassification error: the algorithm comes up with a prediction for each PTA (does the PTA contain the provision in question or not), only using trees for which the PTA is "out-of-bag" (i.e., the PTA is not present in the tree's bootstrap sample), and then computes the share of PTAs for which the provision was classified incorrectly (Ziegler and König 2014).

The RF algorithm has several advantages over the traditional Ordinary Least Squares (OLS) regression and its derivatives. First, it performs well when dealing with high-dimensional data in which the number of predictors is on the same order of magnitude as the number of observations (Schonlau and Zou 2020). This feature allows us to consider many potential PTA determinants simultaneously. Second, RF methods adapt on the fly to the non-linearities and interactions present in the data, which is critical because pre-specifying a flexible structure of interactions and higher order terms between hundreds of potential determinants (as OLS would require) would not be feasible (Ziegler and König 2014).

Variable Importance Measure — Another advantage of RF methods is the existence of well-developed variable importance measures that allow us to identify the country-pair characteristics with the highest predictive power of including a particular PTA provision. The Mean Decrease Accuracy (MDA) is the most commonly used variable importance measure (Ziegler and König 2014). For each predictor, it shuffles the vector of its values (breaking any real association between the predictor and the outcome variable) and computes the drop in the forest's predictive performance. If the drop is large, the predictor is critical for the forest's predictive performance and thus identifies an important PTA provision. The method, however, lacks a notion of statistical significance and suffers from certain mechanical biases (Strobl et al. 2007). Therefore, we employ the permutation importance method developed by Altmann et al. (2010) that builds on the MDA measure. It randomly permutes the outcome vector, re-constructing the RF and re-computing predictors' MDAs on each permutation. The procedure produces a null MDA distribution for each PTA provision and determinant. The determinant's original MDA can be compared with the null distribution to calculate its p-value. The p-value provides a corrected measure of variable importance that is easily comparable across provisions and, at the same time, corrects for MDA's biases.

Missing Values — Another major advantage of RF methods is that they allow for missing values in the data without resorting to the two standard solutions of either dropping entire rows with some missing values (which would remove a large portion of our data that combines hundreds of variables from different sources) or imputing missing observations using non-missing data (which would create spurious variable importance results) (Scheffer 2002). To avoid these issues, we adopt the on-the-fly-imputation method proposed by Ishwaran et al. (2008) and Tang and Ishwaran (2017), which permits maximizing the data used without biasing the variable importance measures. At each tree node, this procedure discards missing values when looking for the optimal predictor and value to split on. But once the optimal split is found, each missing value is temporarily replaced with a randomly drawn non-missing one to determine which branch the observation is sent down. Therefore, missing values are neither discarded nor imputed (despite the method's name), preserving the interpretability of the variable importance measure.

3.3 Data

PTA Provisions — We rely on the Deep Trade Agreements (DTAs) database for the content and evolution of bilateral and regional trade agreements (Mattoo, Rocha and Ruta 2020). The database covers eighteen policy areas in 283 trade agreements notified to the WTO, classified by legal experts. Table 1 presents the number of PTA provisions and their stated objectives. The table also provides the average number of essential provisions and their standard deviation categorized by policy area for all mapped agreements. Out of the 937 PTA provisions, we rely on 305 essential PTA provisions for the Lasso regression following Breinlich et al. (2022). It is important to note that the DTAs database covers PTAs notified to the WTO between 1958 and 2017. More specifically, the dataset is limited to agreements in effect as of December 2017, excluding any trade agreements that have since expired. To deal with this data limitation, we exclude observations associated with expired trade agreements from the plug-in Lasso analysis. To identify those trade agreements, we rely on the Design of Trade Agreements (DESTA) database by Dür, Baccini and Elsig (2014). This database contains information on all 356 agreements notified to the WTO.

Agricultural Trade Flows — We obtained product-level export data from UN Comtrade (2023). Specifically, our analysis covers all agricultural trade flows between 1968 and 2017. Agricultural trade is classified under the Standard International Trade Classification (SITC) codes 00–09. The final balanced panel dataset includes 213 exporters and 260 importers.

PTA Determinants — Various factors can drive the inclusion of specific PTA provisions. First, the PTA formation literature has found several factors relevant to whether countries sign a PTA, and these factors may also prove important for influencing the PTA content (Baccini 2019). Second, most PTA provisions discuss a narrow topic of international trade governance, motivating the inclusion of measures relating to that topic. To identify the potential relevance of these PTA determinants, we assemble a large set of observable country-pair-year characteristics driven by these possible inclusion motives.

Consistent with earlier studies on PTA formation highlighting the importance of aggregate economic characteristics like development level and similarity, we include several aggregate economic measures as potential PTA provision determinants (Baier and Bergstrand 2004; Bergstrand, Egger and Larch 2016). To differentiate countries by the extent of agriculture, we include sectoral shares of agriculture, manufacturing, and services in the gross domestic product (GDP). Because many provisions deal specifically with intellectual property, labor regulation, and environment protection, we include several measures on the labor market, innovation, and natural resources. All of these measures are obtained from *Penn World Tables* and the *World Development Indicators* (Feenstra, Inklaar and Timmer 2015; The World Bank 2023*b*). We also include several measures of geographic, institutional, and cultural proximity from *CEPII Gravity* (Conte, Cotterlaz and Mayer 2022), *CEPII Language* (Melitz and Toubal 2014), *GeoDist* (Mayer and Zignago 2011), and *UNCTADstat* (United Nations Conference on Trade and Development 2023). We include these measures because physical proximity is an important determinant of PTA formation, while institutional or cultural similarity may be relevant to harmonizing regulations (Baier and Bergstrand 2004; Bergstrand, Egger and Larch 2016).

Various trade-related variables, such as bilateral trade imbalance, foreign direct investment flows, and intra-industry trade, are also relevant for PTA formation and depth (Gaulier and Zignago 2010; Grossman and Helpman 1995; Kucik 2012; Facchini, Silva and Willmann 2021; Chase 2008; Baccini, Dür and Elsig 2018; Gamso and Grosse 2021). We include these and many other measures of trade flows, their sectoral breakdown, tariff levels, shipping costs, and FDI flows collected and constructed from *UN Comtrade* (UN Comtrade 2023), *WITS* (The World Bank 2023*a*), *BACI* (Gaulier and Zignago 2010), *CEPII Gravity* (Conte, Cotterlaz and Mayer 2022), and *IMF CDIS* (International Monetary Fund 2023). Several studies have also shown domestic political concerns to be relevant for PTA formation and design (Mansfield and Milner 2012; Baccini and Urpelainen 2014; Raess, Dür and Sari 2018; Lechner 2016). Therefore, we include several measures of domestic political leaning, competitiveness, and governance quality from the *Database of Political Institutions* (Cruz, Keefer and Scartascini 2021), the *Worldwide Governance Indicators* (Kaufmann, Kraay and Mastruzzi 2010), and the *World Development Indicators* (The World Bank 2023*b*).

Lastly, PTA contagion is a critical factor behind PTA formation (Baldwin 1993; Egger and Larch 2008). This literature posits that a country has a strong incentive to sign a PTA with another country if the latter is an important export market and that many other nations competing for this market have already signed agreements with it. To exploit the richness of our data, we extend this idea to the question of contagion in provisions by constructing a provision-level analog of the contagion index in Baldwin and Jaimovich (2012). This approach allows us to evaluate whether there are spillovers in the inclusion of agriculture-relevant provisions in PTAs.

Some of the measures discussed above are observed at the country-pair-year level (like bilateral trade imbalances or the contagion index), while others are defined at the country-year level (like GDP or the rule of law index). Because the outcome variable (presence of a provision in a PTA signed in a certain year) is at the country-pair-year level, we have to aggregate the latter country-year variables. We construct two aggregates for all such numerical variables: the mean and the difference of the values of both countries (either in levels or logs, where appropriate). For all categorical (including binary) variables, we construct two different aggregates: an indicator of whether the values for both countries are the same and a dummy variable encoding the combination of the countries' values. Counting all composite variables, we include 291 potential determinants in the RF analysis. They are listed in Table A.1.

4. Impact of PTA Provisions on Agricultural Trade

Table 2 summarizes estimates for the impact of PTA provisions on agricultural trade. Column (1) shows the average treatment effect of bilateral and regional trade integration. We used an indicator variable to represent the enforcement of such an agreement between the trading partners. We find that PTAs increase agricultural trade by 35.9 percent on average.[3] Column (2) presents the results of the plug-in Lasso regression, which identifies the PTA provisions with non-zero trade effects, while Column (3) displays the post-Lasso regression results. For the post-Lasso regression, we estimate the impact of the selected PTA provisions in the three-way gravity model. Lastly, we include the PTA dummy variable in the post-Lasso regression presented in Column (4). This specification accounts for potential trade policy effects on trade flows not captured by the selected PTA provisions.

The plug-in Lasso regression identified nine provisions related to competition policy, export taxes, IPR, movement of capital, state enterprises, and TBT. Subsequently, the post-Lasso gravity esti-

[3] The coefficient estimate β can be transformed into the elasticity form by $\exp(\beta) - 1 = \exp(0.307) - 1 = 0.359$.

mation revealed that six of these provisions impacted agricultural trade at conventional levels of statistical significance. The observed trade effects of the selected PTA provisions can be rationalized from the related literature. For example, competition policy, export taxes, and state enterprises are at the core of preferential trading in agriculture because of transparency and non-discriminatory principles (Buccirossi, Marette and Schiavina 2002; Anderson, Martin and Valenzuela 2006; Martin and Anderson 2012). Our analysis identifies three PTA provisions under this scheme that positively impact agricultural trade. We find that Competition Policy provision 23, which promotes transparency between trading partners, leads to a 26.6 percent increase in agricultural trade. In addition, Export Taxes provision 18 and State Enterprises provision 43 are associated with a 36.9 percent and 141.8 percent increase in trade flows, respectively.[4]

In contrast, two PTA provisions falling under the IPR policy area are found to have a negative impact on agricultural trade. Specifically, IPR provision 58, which promotes trade in GIs by bypassing certain registration steps if the products meet the required standards, is found to have adverse trade effects. This provision leads to a decrease in agricultural trade by 31.3 percent between involved countries, indicative of the protectionist nature of GIs. The findings suggest that PTAs with those provisions may protect the agricultural products listed as GIs and suppresses the usage of similar product brands from other unauthorized sources. It is important to note that our analysis does not differentiate between the quality of goods, so the potential gains of trade in highly differentiated products and the associated welfare gains are not captured (Moschini, Menapace and Pick 2008; Huysmans and Swinnen 2019; Duvaleix et al. 2021; Curzi and Huysmans 2022). TBT provision 4 increases agricultural trade by 86.1 percent, enhancing mutual recognition of standards. The harmonization of standards increases trade flows between the PTA partners as it mitigates administrative costs associated with non-tariff measures imposed at borders (Santeramo and Lamonaca 2022; Disdier, Fontagné and Cadot 2014). Lastly, we find a positive impact of the Movement of Capitals provision 38 on agricultural trade. This provision aims to promote capital flows by ensuring the non-discriminatory application of the law. Our findings suggest that this PTA provision helps to lower barriers to cross-border investment, which can increase economic activity

[4] A detailed description of the binding PTA provisions is provided in Table 3.

and boost trade between countries (Schmitz and Helmberger 1970; Josling et al. 2010).[5]

5. Determinants of Binding PTA Provisions

5.1 Important Determinants of Most Binding Provisions

To identify the determinants of binding PTA provisions selected by Lasso regression, we estimate a Random Forest model for each provision on hundreds of potential determinants. Table 4 shows that Random Forests achieve excellent fit of the data, attaining an out-of-bag misclassification rate of 14% or less for all binding provisions. Having conducted many random permutations per determinant to construct the null distributions, we compute the p-value of the permutation importance (our variable importance measure of choice) for each potential determinant and each PTA provision. Figure 2 lists the economic, political, and geographic determinants that are relevant (p-value < 0.05) for three or more of the nine binding provisions.[6]

We find that contagion is a determinant with universal importance. This result extends the findings of a growing literature on the importance of contagion in PTA formation to the question of PTA design (Baldwin 1993; Egger and Larch 2008; Baldwin and Jaimovich 2012; Baccini and Dür 2012; Chen and Joshi 2010). Not only does the existence of a trade agreement between each of the nations in the country pair and other important trade partners matter for whether they choose to sign a PTA of their own, but the *content* of those agreements with third parties also matters for what the country pair chooses to include in their new trade deal. The average and the difference in the contagion indices are highly relevant, implying that the contrast in competitive pressures faced by the two trading partners matters for whether they can settle on including PTA provisions that foster agricultural trade.

In addition, we find a set of geographic indicators to be important as well. Two measures of the distance between the two countries are highly deterministic, extending insights on the relevance

[5] The impact of this provision on agricultural trade is indirect and primarily mediated through the activities of a complex global value chain. To exclusively examine the impact of this provision on agricultural trade, a second-stage or instrumental variable approach would be required. However, we do not investigate it further as it is outside the scope of this study.

[6] See Appendix Table A.2 for a detailed breakdown of this ranking.

of transportation costs (Baier and Bergstrand 2004; Bergstrand, Egger and Larch 2016). The combination of the country pair's continents, rather than the simple indicator variable of whether they share a continent, is also important for over half of the binding provisions, suggesting that different PTA design incentives are at play across different continents. The quality of governance is another relevant area, as reflected by the measures "voice and accountability", "regulatory quality", and "political stability". What matters is the *average* of the two measures—whether both countries have transparent and efficient governments—not the difference between the two, which would have pointed at agreements across regime types. This result extends the findings on the importance of political factors for PTA formation (Mansfield and Milner 2012; Baccini and Urpelainen 2014).

Surprisingly, another important determinant for including PTA provisions that foster agricultural trade is the energy use per capita, averaged within the country pair. While this measure correlates strongly with various economic development metrics, it is notable that it has the highest predictive power for including agriculture-relevant provisions in PTAs. It hints at the importance of development level and the intensity of mechanized production and trade in energy sources. Another factor highly related to the development level is the average human capital index of the two countries, highlighting the importance of both partners sharing the same development level (since the difference in human capital indices is irrelevant). This pattern shows that while the differential development level is important for PTA formation (Manger 2009; Büthe and Milner 2008; Baccini and Urpelainen 2014), it is irrelevant for including PTA provisions pertinent to agricultural trade. Another important economic determinant is the average of the consumer price level between the country pair, pointing to the importance of the real exchange rate of the country pair *relative to everyone else* (rather than their bilateral real exchange rate). The last of the selected economic determinants is the average hours worked per person, hinting at the relevance of the intensity of labor supply.

5.2 Idiosyncratic Determinants of Specific PTA Provisions

Some country-pair characteristics are of interest not only because they are important determinants of every PTA provision but also because they are of idiosyncratic importance to just one of them. While Export Taxes provision 18, Movement of Capitals provision 37, State Enterprises provision

43, and TBT provision 4 have no unique determinants, the rest do.

The Competition Policy provision 23 has only a few important determinants. Governmental quality plays an even larger role in including this PTA provision. Averages of “Government effectiveness” and “Rule of law” measures are uniquely significant for it. Gamso and Grosse (2021) showed PTAs need to be deeper if they want to attract foreign investment when domestic property rights protection is lacking: a similar mechanism may be at play here, incentivizing negotiators to protect competition through international regulation if exporters are concerned that the existing regulation in the foreign market leaves them at a disadvantage. The “Frontier technology readiness index” is also relevant to the Competition Policy provision 23. These suggest that effective governance and the ability of trading partners to adopt new technologies matter greatly for their propensity to insist on transparency in competition policy.

Measures of trade in manufacturing and natural resources are important for Export Taxes provision 15: net energy imports, manufacturing imports, exports, and the average and the difference in the two countries’ natural resources rents as a share of GDP. Natural resource-rich commodity exporters are likely to face different calculus in relying on export taxes as a source of revenue. In parallel with Competition Policy provision 23, several political competitiveness metrics are highly relevant: “Checks and balances index” and “Fractionalization of legislature”. A potential explanation is an extension of a result by Mansfield and Milner (2012) that countries with a larger number of veto players sign fewer PTAs: the same mechanism could hinder negotiators’ ability to stipulate restrictions on the government’s ability to levy export taxes in the agreements that do get signed. Both the average and difference in inflation matter, too, hinting at the relevance of domestic monetary policy concerns for the two nations’ willingness to abandon export taxes.

The determinants of IPR provisions 58 and 88 are curious because of their lack of determinants, as only contagion and distance measures are significant for these PTA provisions. The RF model cannot identify any pattern in which countries decide to implement these intellectual property rights provisions relevant to agricultural trade beyond spillovers of these provisions from their trading partners.

The volume of bilateral trade in manufacturing is uniquely important for Movement of Capitals provision 38, hinting at the relevance of this sector for the nations' desire to normalize capital account regulation.

5.3 Effect of Important Determinants on Selected Provisions

While RF models excel in identifying important PTA determinants (in our case, country-pair characteristics) of an outcome (the presence of selected provisions in PTAs), they are ill-equipped to summarize the overall effect of each predictor on the outcome (Mullainathan and Spiess 2017). Because RFs capture all kinds of non-linearities and interactions in the data, they lack the concept of a single coefficient describing the sign and magnitude of a predictor's linear effect on the outcome. To aid interpretability, we conduct an auxiliary exercise to obtain an overview of the signs of linear treatment effects for selected determinants. First, we run a Poisson regression of the number (out of nine) of binding provisions included in a PTA on the determinants found by the Random Forests to be important for including said provisions. We include determinants that are significantly important (permutation importance p-value below 5%) for over half of binding provisions.[7] The results from this exercise are presented in the first column ("Count") of Table 5. The second and third columns present the results from a hurdle model that can fit count data better by splitting the problem into two.[8] First, a logistic regression model is estimated to predict whether *any* of the binding provisions are included in the PTA (column "Any"). Then a Poisson regression is estimated on the count of binding provisions, truncating the data only to include country-pairs that have at least one binding provision (column "Count (≥ 1)").

Despite the complexity of PTA design and its irreducibility to a single index, some clear patterns emerge. Expectedly, the average contagion index is associated with a strong positive propensity for including most provisions. If country A's exports compete for the import market of country B with country C's exports, the inclusion of a certain provision in an agreement between countries B and C pressures country A to follow suit or risk losing market share. However, the difference in

---

[7] This cutoff is stricter than the one used in Figure 2, to avoid losing an excessive amount of observations to missing values, which is something that the Random Forest can deal with but ordinary least squares regression cannot.

[8] For discussions of hurdle models and closely related zero-inflated models, see Lambert (1992); Mullahy (1986); Feng (2021).

contagion index values between the two trading partners has a strong negative effect on whether any binding provisions are included, implying that both partners have to face similar competitive pressures or the agreement will fall through. Like the average level of contagion, the coefficient on the two nations' energy use per capita is also strongly positive, capturing the greater likelihood that two economically powerful nations will include agriculture-relevant provisions in their agreement.

The exercises in Table 5 offer an imperfect interpretation of the RF results. First, logistic and Poisson regressions only include observations with no missing values in the important determinants, drastically reducing the sample size (a limitation that RFs overcome). Secondly, it is limited to the determinants found to be important for most provisions and not the hundreds of determinants included in the RFs. Finally, the three linear regressions, by necessity, impose linearity on the estimated relationships. However, a given feature of a country-pair may make having a particular provision in the PTA more likely in some circumstances but less likely in others. RFs can easily capture this heterogeneity, whether the econometrician anticipated it. Therefore, while our RF results identify the factors that are the most important determinants of PTA provisions relevant to agricultural trade, regressions in Table 5 come up with the best *linear* approximation of how much these determinants matter.

6. Conclusion

Modern ML techniques are increasingly relevant in applied economics research because they can help researchers and policymakers analyze vast amounts of data and understand complex relationships (Varian 2014; Mullainathan and Spiess 2017). When used in a manner consistent with economic theory, ML models can provide researchers with valuable information, attracting the attention of policymakers and providing them with novel insights. In our application, we explore critical issues in agricultural trade within the broader framework of economic integration policies. By applying ML methods to analyze extensive PTA and global trade datasets, we identify the PTA provisions that foster agricultural trade. We expand the three-way gravity model relying on plug-in Lasso regularized regression to evaluate the influence of binding PTA provisions on agricultural trade (Yotov et al. 2016; Breinlich et al. 2022). This analysis informs an investigation of the socio-economic and political determinants associated with binding PTA provisions. Subsequently, we employ RF models to identify the key economic, political, social, and geographic factors that influence the inclusion of those PTA provisions (Breiman 2001; Ziegler and König 2014). This analysis allows us to pinpoint the most important determinants of PTA provisions that promote agricultural trade.

In the first stage, a plug-in Lasso regression is used to identify PTA provisions that promote agricultural trade. Those PTA provisions are related to competition policy, export taxes, IPR, movement of capital, state enterprises, and TBT. The post-Lasso results indicate that eight of the nine identified PTA provisions have a statistically significant impact on agricultural trade at conventional levels of statistical significance. Intuitively, these findings are consistent with the related literature as they highlight the importance of specific policy frameworks that can affect agricultural trade (Scoppola, Raimondi and Olper 2018; Duvaleix et al. 2021; He 2022; Raimondi et al. 2023). The first-stage results are then extended into RF models to identify the economic, political, social, and geographic factors related to the presence of these provisions in bilateral and regional trade agreements. Of the nine provisions identified in the first stage, all are significantly associated with contagion effects, consistent with the extant literature on PTAs, but extended to specific PTA provisions in this paper (Baldwin 1993; Baldwin and Jaimovich 2012). In addition to contagion, gravity and governance variables are strongly associated with PTA provisions meaningful to agricultural trade at conventional levels of statistical significance. Energy use and human capital index are highly relevant as well and can be framed in the context of structural transformation. These findings can be interpreted similarly to the standard gravity variables, whereby countries that share similar characteristics are more-or-less likely to engage in PTAs with similar provisions (Baier and Bergstrand 2004; Egger and Larch 2008; Bergstrand, Egger and Larch 2016). To extend these results, we regress the most important determinants identified by the RF model on the presence and number of agricultural-relevant PTA provisions. Resulting estimates indicate positive associations for contagion (mean) and energy use, which are highly statistically significant, while an inverse association is observed for contagion (difference). Simply put, trade provisions related to agriculture are more likely to be shared between countries that are similar and/or aligned, while misalignment in this regard is associated with the divergence of these provisions.

The findings from the two-part approach using Lasso and RF models offer novel insights into

PTA provisions that drive agricultural trade and the factors that influence whether or not such provisions are included in trade deals. The two models complement each other, each providing part of the complex picture. The Lasso model combs through hundreds of provisions and singles out the ones that are most impactful for agricultural trade. The Random Forest model, in turn, wades through hundreds of potential determinants of including these provisions and identifies the ones most impactful for PTA design. Our two-pronged approach is thus a first step in summarizing the complex factors related to PTA provisions that promote agricultural trade. Moreover, since much of the world's population relies upon agricultural production and trade, determining the effects of broader trade policies on agricultural trade could be particularly interesting to policymakers. Identifying capabilities to support or mitigate drivers of certain provisions, such as contagion, is a critical factor in dealing with many of today's global challenges, such as world hunger and global climate change.

References

Alston, Julian M, Colin A Carter, Daniel A Sumner, and Richard Gray. 1997. "Third-country effects and second-best grain trade policies: Export subsidies and bilateral liberalization." *American Journal of Agricultural Economics*, 79(4): 1300–1310.

Altmann, André, Laura Toloşi, Oliver Sander, and Thomas Lengauer. 2010. "Permutation Importance: A Corrected Feature Importance Measure." *Bioinformatics*, 26(10): 1340–1347.

Anderson, James E. 2011. "The gravity model." *Annual Review of Economics*, 3: 133–160.

Anderson, James E, and Yoto V Yotov. 2010. "The changing incidence of geography." *American Economic Review*, 100(5): 2157–86.

Anderson, Kym, Will Martin, and Ernesto Valenzuela. 2006. "The relative importance of global agricultural subsidies and market access." *World Trade Review*, 5(3): 357—-376.

Baccini, Leonardo. 2019. "The Economics and Politics of Preferential Trade Agreements." *Annual Review of Political Science*, 22(1): 75–92.

Baccini, Leonardo, and Andreas Dür. 2012. "The New Regionalism and Policy Interdependence." *British Journal of Political Science*, 42(1): 57–79.

Baccini, Leonardo, and Johannes Urpelainen. 2014. "International Institutions and Domestic Politics: Can Preferential Trading Agreements Help Leaders Promote Economic Reform?" *Journal of Politics*, 76(1): 195–214.

Baccini, Leonardo, Andreas Dür, and Manfred Elsig. 2018. "Intra-Industry Trade, Global Value Chains, and Preferential Tariff Liberalization." *International Studies Quarterly*, 62(2): 329–340.

Baier, Scott L., and Jeffrey H. Bergstrand. 2004. "Economic Determinants of Free Trade Agreements." *Journal of International Economics*, 64(1): 29–63.

Baier, Scott L, and Jeffrey H Bergstrand. 2007. "Do free trade agreements actually increase members' international trade?" *Journal of International Economics*, 71(1): 72–95.

Baldwin, Richard. 1993. "A Domino Theory of Regionalism." *NBER Working Paper 4465.*

Baldwin, Richard, and Dany Jaimovich. 2012. "Are Free Trade Agreements Contagious?" *Journal of International Economics*, 88(1): 1–16.

Belloni, Alexandre, Daniel Chen, Victor Chernozhukov, and Christian Hansen. 2012. "Sparse models and methods for optimal instruments with an application to eminent domain." *Econometrica*, 80(6): 2369–2429.

Belloni, Alexandre, Victor Chernozhukov, Christian Hansen, and Damian Kozbur. 2016. "Inference in high-dimensional panel models with an application to gun control." *Journal of Business & Economic Statistics*, 34(4): 590–605.

Bergstrand, Jeffrey H., Peter Egger, and Mario Larch. 2016. "Economic Determinants of the Timing of Preferential Trade Agreement Formations and Enlargements." *Economic Inquiry*, 54(1): 315–341.

Biau, Gérard, and Erwan Scornet. 2016. "A Random Forest Guided Tour." *TEST*, 25(2): 197–227.

Bischl, Bernd, Martin Binder, Michel Lang, Tobias Pielok, Jakob Richter, Stefan Coors, Janek Thomas, Theresa Ullmann, Marc Becker, Anne-Laure Boulesteix, Difan Deng, and Marius Lindauer. 2021. "Hyperparameter Optimization: Foundations, Algorithms, Best Practices and Open Challenges." `http://arxiv.org/abs/2107.05847`.

Breiman, Leo. 2001. "Random Forests." *Machine Learning*, 45(1): 5–32.

Breinlich, Holger, Valentina Corradi, Nadia Rocha, Michele Ruta, JMC Santos Silva, and Thomas Zylkin. 2022. "Machine learning in international trade research - evaluating the impact of trade agreements." *CEPR Discussion Paper No. DP17325.*

Buccirossi, Paolo, Stéphan Marette, and Alessandra Schiavina. 2002. "Competition policy and the agribusiness sector in the European Union." *European Review of Agricultural Economics*, 29(3): 373–397.

Büthe, Tim, and Helen V. Milner. 2008. "The Politics of Foreign Direct Investment into Developing Countries: Increasing FDI through International Trade Agreements?" *American Journal of Political Science*, 52(4): 741–762.

Cameron, A. Colin, and Douglas L. Miller. 2015. "A practitioner's guide to cluster-robust inference." *Journal of Human Resources*, 50(2): 317–372.

Carter, Colin A, and Caroline Gunning-Trant. 2010. "US trade remedy law and agriculture: Trade diversion and investigation effects." *Canadian Journal of Economics*, 43(1): 97–126.

Carter, Colin A, and Sandro Steinbach. 2018. "Trade diversion and the initiation effect: A case study of US trade remedies in agriculture." *NBER Working Paper 24745.*

Chase, Kerry A. 2008. "Protecting Free Trade: The Political Economy of Rules of Origin." *International Organization*, 62(3): 507–530.

Chen, Maggie Xiaoyang, and Sumit Joshi. 2010. "Third-Country Effects on the Formation of Free Trade Agreements." *Journal of International Economics*, 82(2): 238–248.

Conte, Maddalena, Pierre Cotterlaz, and Thierry Mayer. 2022. "The CEPII Gravity Database." *CEPII Working Paper 2022-05*.

Correia, Sergio, Paulo Guimarães, and Tom Zylkin. 2020. "Fast Poisson estimation with high-dimensional fixed effects." *The Stata Journal*, 20(1): 95–115.

Cruz, Cesi, Philip Keefer, and Carlos Scartascini. 2021. "The Database of Political Institutions 2020." `https://publications.iadb.org/en/node/29752`.

Curzi, Daniele, and Martijn Huysmans. 2022. "The impact of protecting EU geographical indications in trade agreements." *American Journal of Agricultural Economics*, 104(1): 364–384.

Disdier, Anne-Célia, Lionel Fontagné, and Mondher Mimouni. 2008. "The impact of regulations on agricultural trade: Evidence from the SPS and TBT agreements." *American Journal of Agricultural Economics*, 90(2): 336–350.

Disdier, Anne-Célia, Lionel Fontagné, and Olivier Cadot. 2014. "North-South standards Harmonization and International Trade." *The World Bank Economic Review*, 29(2): 327–352.

Duvaleix, Sabine, Charlotte Emlinger, Carl Gaigné, and Karine Latouche. 2021. "Geographical indications and trade: Firm-level evidence from the French cheese industry." *Food Policy*, 102: 102118.

Dür, Andreas. 2007. "EU Trade Policy as Protection for Exporters: The Agreements with Mexico and Chile." *Journal of Common Market Studies*, 45(4): 833–855.

Dür, Andreas, Leonardo Baccini, and Manfred Elsig. 2014. "The Design of International Trade Agreements: Introducing a New Dataset." *Review of International Organization*, 9(3): 353–375.

Egger, Peter, and Mario Larch. 2008. "Interdependent Preferential Trade Agreement Memberships: An Empirical Analysis." *Journal of International Economics*, 76(2): 384–399.

Facchini, Giovanni, Peri Silva, and Gerald Willmann. 2021. "The Political Economy of Preferential Trade Agreements: An Empirical Investigation." *The Economic Journal*, 131(640): 3207–3240.

Fally, Thibault. 2015. "Structural gravity and fixed effects." *Journal of International Economics*, 97(1): 76–85.

Falvey, Rod, and Neil Foster-McGregor. 2018. "North-South foreign direct investment and bilateral investment treaties." *The World Economy*, 41(1): 2–28.

Feenstra, Robert C., Robert Inklaar, and Marcel P. Timmer. 2015. "The Next Generation of the Penn World Table." *American Economic Review*, 105(10): 3150–3182.

Feng, Cindy Xin. 2021. "A comparison of zero-inflated and hurdle models for modeling zero-inflated count data." *Journal of Statistical Distributions and Applications*, 8(1): 8.

Fulponi, Linda, Matthew Shearer, and Juliana Almeida. 2011. "Regional trade agreements - Treatment of agriculture." *OECD Food, Agriculture and Fisheries Papers 44*.

Gamso, Jonas, and Robert Grosse. 2021. "Trade Agreement Depth, Foreign Direct Investment, and the Moderating Role of Property Rights." *Journal of International Business Policy*, 4(2): 308–325.

Gaulier, Guillaume, and Soledad Zignago. 2010. "BACI: International Trade Database at the Product-Level (the 1994-2007 Version)." `http://www.ssrn.com/abstract=1994500`.

Grant, Jason H, and Dayton M Lambert. 2008. "Do regional trade agreements increase members' agricultural trade?" *American Journal of Agricultural Economics*, 90(3): 765–782.

Grossman, Gene M., and Elhanan Helpman. 1995. "The Politics of Free-Trade Agreements." *American Economic Review*, 85(4): 667–690.

He, Xi. 2022. "Political and economic determinants of export restrictions in the agricultural and food sector." *Agricultural Economics*, 53(3): 439–453.

Hofmann, Claudia, Alberto Osnago, and Michele Ruta. 2017. *Horizontal Depth: A New Database on the Content of Preferential Trade Agreements.* World Bank, Washington, DC.

Hofmann, Claudia, Alberto Osnago, and Michele Ruta. 2019. "The Content of Preferential Trade Agreements." *World Trade Review*, 18(3): 365–398.

Huysmans, Martijn, and Johan Swinnen. 2019. "No terroir in the cold? A note on the geography of geographical indications." *Journal of Agricultural Economics*, 70(2): 550–559.

International Monetary Fund. 2023. "CDIS." *https://data.imf.org/*.

Ishwaran, Hemant, Udaya B. Kogalur, Eugene H. Blackstone, and Michael S. Lauer. 2008. "Random Survival Forests." *The Annals of Applied Statistics*, 2(3).

Josling, Tim. 2006. "The war on terroir: Geographical indications as a transatlantic trade conflict." *Journal of Agricultural Economics*, 57(3): 337–363.

Josling, Tim, Kym Anderson, Andrew Schmitz, and Stefan Tangermann. 2010. "Understanding international trade in agricultural products: one hundred years of contributions by agricultural economists." *American Journal of Agricultural Economics*, 92(2): 424–446.

Kaufmann, Daniel, Aart Kraay, and Massimo Mastruzzi. 2010. *The Worldwide Governance Indicators.* World Bank, Washington, DC.

Kim, Dongin, and Sandro Steinbach. 2023. "Preferential trading in agriculture: New insights from a structural gravity analysis and machine learning." *SSRN Working Paper No. 4412389*, 1–45.

Kucik, Jeffrey. 2012. "The Domestic Politics of Institutional Design: Producer Preferences over Trade Agreement Rules." *Economics & Politics*, 24(2): 95–118.

Lambert, Diane. 1992. "Zero-Inflated Poisson Regression, with an Application to Defects in Manufacturing." *Technometrics*, 34(1): 1.

Lechner, Lisa. 2016. "The Domestic Battle over the Design of Non-Trade Issues in Preferential Trade Agreements." *Review of International Political Economy*, 23(5): 840–871.

Lee, Woori, Nadia Rocha, and Michele Ruta. 2021. "Trade facilitation provisions in preferential trade agreements." *World Bank Working Paper 9674*.

Limão, N. 2016. "Preferential Trade Agreements." In *Handbook of Commercial Policy*. 279–367. Elsevier.

Ling, Charles X, Jin Huang, and Harry Zhang. 2003. "AUC: a Statistically Consistent and more Discriminating Measure than Accuracy." Vol. 3, 519–524.

Manger, Mark S. 2009. *Investing in Protection: The Politics of Preferential Trade Agreements between North and South.* . 1 ed., Cambridge University Press.

Mansfield, Edward D., and Helen V. Milner. 2012. *Votes, Vetoes, and the Political Economy of International Trade Agreements.* Princeton University Press.

Martin, Will. 2018. "A research agenda for international agricultural trade." *Applied Economic Perspectives and Policy*, 40(1): 155–173.

Martin, Will, and Kym Anderson. 2012. "Export restrictions and price insulation during commodity price booms." *American Journal of Agricultural Economics*, 94(2): 422–427.

Mattoo, Aaditya, Nadia Rocha, and Michele Ruta. 2020. "The Evolution of Deep Trade Agreements." In *Handbook of Deep Trade Agreements*. The World Bank.

Mayer, Thierry, and Soledad Zignago. 2011. "Notes on CEPII's Distances Measures: The GeoDist Database." `http://www.ssrn.com/abstract=1994531`.

McCalla, Alex F. 1992. "GATT, preferential/regional trading blocs and agricultural trade." *Review of International Economics*, 1(1): 73–89.

Melitz, Jacques, and Farid Toubal. 2014. "Native Language, Spoken Language, Translation and Trade." *Journal of International Economics*, 93(2): 351–363.

Moschini, GianCarlo, Luisa Menapace, and Daniel Pick. 2008. "Geographical indications and the competitive provision of quality in agricultural markets." *American Journal of Agricultural Economics*, 90(3): 794–812.

Mullahy, John. 1986. "Specification and testing of some modified count data models." *Journal of Econometrics*, 33(3): 341–365.

Mullainathan, Sendhil, and Jann Spiess. 2017. "Machine Learning: An Applied Econometric Approach." *Journal of Economic Perspectives*, 31(2): 87–106.

Murina, Marina, and Alessandro Nicita. 2017. "Trading with conditions: The effect of sanitary and phytosanitary measures on the agricultural exports from low-income countries." *The World Economy*, 40(1): 168–181.

Peterson, Everett, Jason Grant, Donna Roberts, and Vuko Karov. 2013. "Evaluating the trade restrictiveness of phytosanitary measures on US fresh fruit and vegetable imports." *American Journal of Agricultural Economics*, 95(4): 842–858.

Raess, Damian, Andreas Dür, and Dora Sari. 2018. "Protecting Labor Rights in Preferential Trade Agreements: The Role of Trade Unions, Left Governments, and Skilled Labor." *Review of International Organization*, 13(2): 143–162.

Raimondi, Valentina, Andreea Piriu, Johan Swinnen, and Alessandro Olper. 2023. "Impact of global value chains on tariffs and non-tariff measures in agriculture and food." *Food Policy*, 118: 102469.

Rodrik, Dani. 2018. "What do trade agreements really do?" *Journal of Economic Perspectives*, 32(2): 73–90.

Ruppel, Fred J, Fred O Boadu, and E Wesley F Peterson. 1991. "Federalism, opportunism, and multilateral trade negotiations in agriculture." *American Journal of Agricultural Economics*, 73(4): 1009–1019.

Santeramo, Fabio Gaetano, and Emilia 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(3): 595–617.

Santeramo, Fabio Gaetano, and Emilia Lamonaca. 2022. "Standards and regulatory cooperation in regional trade agreements: What the effects on trade?" *Applied Economic Perspectives and Policy*, 44(4): 1682–1701.

Scheffer, Judi. 2002. "Dealing with Missing Data."

Schmitz, Andrew, and Peter Helmberger. 1970. "Factor mobility and international trade: the case of complementarity." *American Economic Review*, 60(4): 761–767.

Schonlau, Matthias, and Rosie Yuyan Zou. 2020. "The Random Forest Algorithm for Statistical Learning." *The Stata Journal: Promoting communications on statistics and Stata*, 20(1): 3–29.

Scoppola, Margherita, Valentina Raimondi, and Alessandro Olper. 2018. "The impact of EU trade preferences on the extensive and intensive margins of agricultural and food products." *Agricultural Economics*, 49(2): 251–263.

Sheldon, Ian M, Daniel CK Chow, and William McGuire. 2018. "Trade liberalization and constraints on moves to protectionism: Multilateralism vs. regionalism." *American Journal of Agricultural Economics*, 100(5): 1375–1390.

Silva, JMC Santos, and Silvana Tenreyro. 2006. "The log of gravity." *The Review of Economics and statistics*, 88(4): 641–658.

Strobl, Carolin, Anne-Laure Boulesteix, Achim Zeileis, and Torsten Hothorn. 2007. "Bias in Random Forest Variable Importance Measures: Illustrations, Sources and a Solution." *BMC Bioinformatics*, 8(1): 25.

Sun, Lin, and Michael R Reed. 2010. "Impacts of free trade agreements on agricultural trade creation and trade diversion." *American Journal of Agricultural Economics*, 92(5): 1351–1363.

Tang, Fei, and Hemant Ishwaran. 2017. "Random Forest Missing Data Algorithms." *Statistical Analysis and Data Mining: The ASA Data Science Journal*, 10(6): 363–377.

The World Bank. 2023*a*. "WITS." *https://wits.worldbank.org/*.

The World Bank. 2023*b*. "World Development Indicators." *https://data.worldbank.org/*.

Thompson-Lipponen, Clara, and Jared Greenville. 2019. "The Evolution of the Treatment of Agriculture in Preferential Trade Agreements." *OECD Working Paper 126*.

UN Comtrade. 2023. "United Nations Comtrade Database." *https://comtradeplus.un.org/*.

United Nations Conference on Trade and Development. 2023. "UNCTADstat." *https://unctadstat.unctad.org/*.

Varian, Hal R. 2014. "Big Data: New Tricks for Econometrics." *Journal of Economic Perspectives*, 28(2): 3–28.

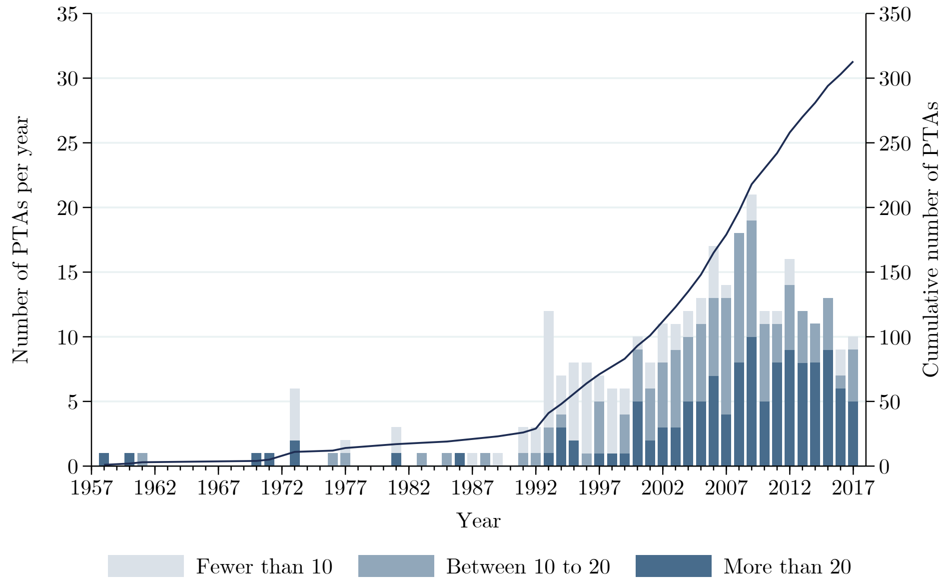
Weidner, Martin, and Thomas Zylkin. 2021. "Bias and consistency in three-way gravity models." *Journal of International Economics*, 132: 103513.

WTO. 2023. "Regional Trade Agreements Database." http://rtais.wto.org/UI/PublicMaintainRTAHome.aspx.

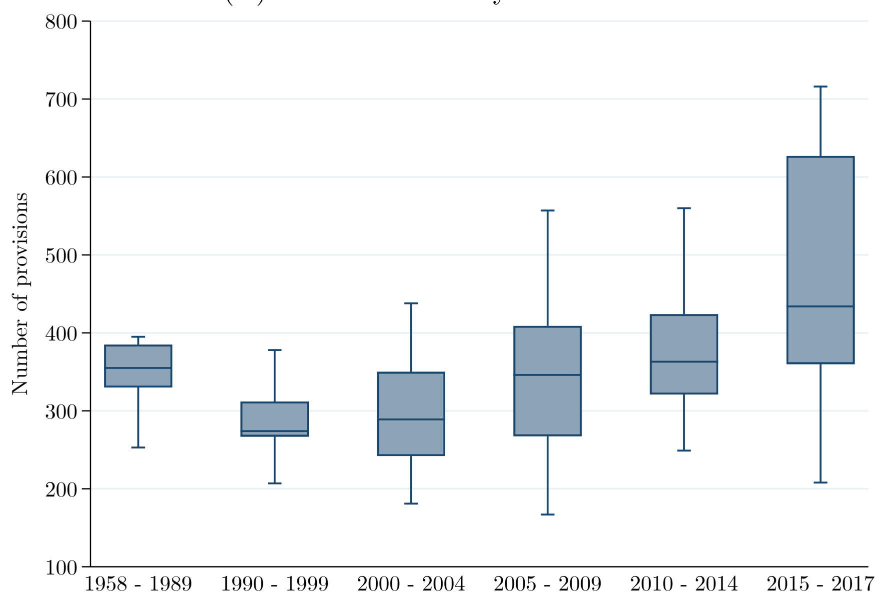
Yotov, Yoto V, Roberta Piermartini, Mario Larch, et al. 2016. "An advanced guide to trade policy analysis." *WTO iLibrary*.

Ziegler, Andreas, and Inke R. König. 2014. "Mining Data with Random Forests: Current Options for Real-World Applications: Mining Data with Random Forests." *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 4(1): 55–63.

Figures and Tables



Panel (A): Number of Policy Areas Covered in PTAs



Panel (B): Evolution of PTA Provisions

Figure 1: The Evolution of PTAs.

Note. The figure shows the evolution of the number and design of PTA provisions. Panel (A) depicts PTA enforcement from 1958 to 2017. The various shades indicate the number of policy areas covered in PTAs. We include a total of 52 policy areas according to Hofmann, Osnago and Ruta (2017). Panel (B) shows the range and average of the number of provisions of PTAs enforced during the same period. The number of PTA provisions comes from Mattoo, Rocha and Ruta (2020).

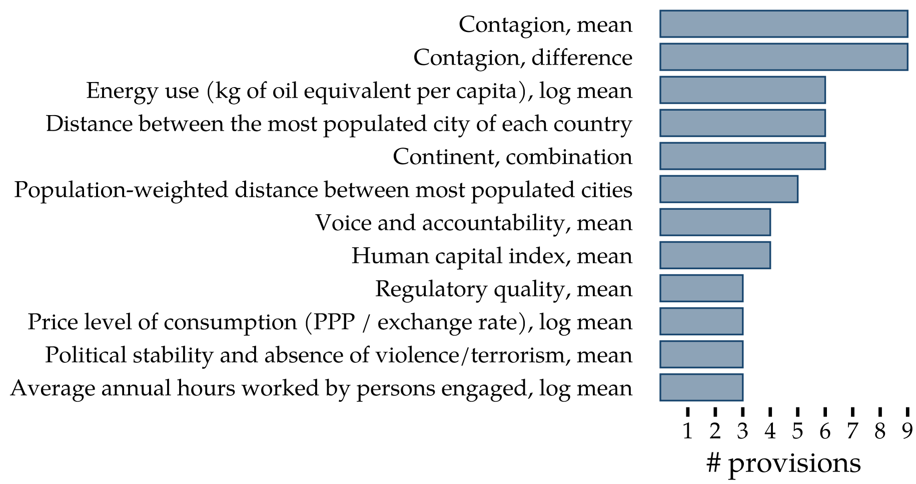


Figure 2: Important Determinants of Binding PTA Provisions.

Note. The figure lists the major determinants of binding PTA provisions. "# provisions" denotes the number of binding provisions for which each determinant is significantly important (the p-value of the permutation variable importance measure is below 0.05). Only the determinants relevant to three or more provisions are displayed.

Table 1: Distribution of Provisions by Policy Area.

Policy Area	Number of Provisions	Essential Provisions	Mean	Standard Deviation
Anti-dumping	39	6	0.2	0.8
Competition Policy	35	14	4.1	3.2
Countervailing Duties	14	5	0.2	0.8
Environmental Laws	49	27	2.1	3.3
Export Taxes	45	23	4.6	3.8
IPR	120	67	5.7	10.2
Investment	57	15	3.3	4.6
Labor Market Regulations	18	12	1.5	3.2
Migration	30	3	0.4	0.7
Movement of Capital	94	8	1.5	2.3
Public Procurement	100	5	0.9	1.3
Rules of Origin	38	19	15.7	12.1
SPS	59	24	2.0	2.6
Services	64	21	3.2	2.7
State-Owned Enterprises	53	13	2.3	2.6
Subsidies	36	13	3.0	2.0
TBT	34	19	0.9	1.9
Trade Facilitation	52	11	3.7	2.5
Total	937	305	-	-

Note. The table lists the 18 policy areas covered in the PTA dataset. The second column presents the count of PTA provisions by policy area according to Mattoo, Rocha and Ruta (2020). In the third column, we include the essential provisions utilized for the plug-in Lasso analysis. The fourth and fifth columns display the mean and standard deviation for the number of provisions observed across all agreements.

Table 2: Impact of PTA Provisions on Agricultural Trade.

	(1) PPML	(2) Lasso	(3) Post-Lasso	(4) PPML
PTA	0.307***			-0.005
	(0.071)			(0.064)
Competition Policy - 23		0.222	0.236***	0.236***
			(0.080)	(0.080)
Export Taxes - 15		0.003	0.047	0.047
			(0.083)	(0.083)
Export Taxes - 18		0.109	0.312***	0.314***
			(0.071)	(0.072)
IPR - 58		-0.012	-0.376***	-0.376***
			(0.142)	(0.142)
IPR - 88		-0.060	-0.327***	-0.327***
			(0.095)	(0.095)
Movement of Capitals - 37		0.099	0.250*	0.249*
			(0.140)	(0.140)
Movement of Capitals - 38		0.021	0.200***	0.201***
			(0.067)	(0.067)
State Enterprises - 43		0.630	0.881***	0.883***
			(0.100)	(0.102)
TBT - 4		0.025	0.621***	0.621***
			(0.178)	(0.178)
Observations	368,227		368,227	368,227
Pseudo R^2	0.939		0.941	0.941

Note. The table presents the plug-in Lasso regression results in Column (2) and the post-estimation results in Column (3). We also show the results of the PTA dummy specification in Column (1) and the comprehensive model that includes the PTA dummy and the selected provisions in Column (4). Asterisks denote statistical significance at * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Heteroskedasticity-robust standard errors clustered at the exporter-importer level are reported in parentheses.

Table 3: Binding PTA Provisions and Descriptions.

Variable	Abbreviated	Description
Competition Policy - 23	CP 23	Does the agreement contain provisions that promote transparency?
Export Taxes - 15	ET 15	Requires phase out of existing export taxes, without reference to exceptions within the provision.
Export Taxes - 18	ET 18	Prohibits an increase in the rate of any existing export tax.
IPR - 58	IPR 58	Geographical Indication: Designates that any parties meeting a particular specification may use a GI without registering independently
IPR - 88	IPR 88	Industrial Design: Provides minimum term of protection.
Movement of Capitals - 37	MoC 37	Non-discriminatory Application of Law: Does the transfer provision explicitly exclude 'good faith and non-discriminatory application of its laws' governing capital account regulations?
Movement of Capitals - 38	MoC 38	Non-discriminatory Application of Law: Does the agreement contain Country annexes with specific transfer reservations by individual parties?
STE - 43	STE 43	Does the agreement include any other specific discipline for certain sectors or objectives?
TBT - 4	TBT 4	Integration Approach – Standards: Is mutual recognition in force?

Note. The table lists the binding PTA provisions and their descriptions identified by the plug-in Lasso regression.

Table 4: Random Forest Misclassification Error.

PTA Provision	OOB Misclassification
Competition Policy - 23	0.14
Export Taxes - 15	0.11
Export Taxes - 18	0.14
IPR - 58	0.02
IPR - 88	0.07
Movement of Capitals - 37	0.04
Movement of Capitals - 38	0.04
State Enterprises - 43	0.07
TBT - 4	0.03

Note. The table reports the out-of-bag (OOB) misclassification error of provision-specific RFs: the fraction of country-pair-year observations with PTAs for which the estimated RF predicts the presence of the particular PTA provision incorrectly, computed "out-of-bag" (for each observation, only predictions from trees that do *not* include the observation in their bootstrap sample are considered).

Table 5: Effect of Most Important Determinants on the Presence and Number of PTA Provisions Relevant to Agricultural Trade.

		Hurdle Model	
	Count	Any	Count (≥ 1)
Contagion, mean	0.057***	36.669***	0.589***
	(0.009)	(10.116)	(0.205)
Contagion, difference	-0.009	-23.057***	-0.123
	(0.014)	(6.384)	(0.168)
Energy use (kg of oil equivalent per capita), log mean	0.195***	0.525***	0.138***
	(0.019)	(0.080)	(0.027)
Distance between the most populated city of each country	0.774**	-3.518**	0.000***
	(0.338)	(1.534)	(0.000)
Population-weighted distance between most-populated cities	-0.941***	3.555**	0.000***
	(0.344)	(1.550)	(0.000)
Observations	3,450	3,450	2,358
Pseudo R^2	0.910	0.583	0.754

Note. The table presents results of regressions of the number or presence of binding provision in PTAs on the most important determinants identified by RF models. Column "Count" presents the results of a Poisson regression on the number of agriculture-relevant provisions in a PTA on the determinants identified by the RFs to be important (p-value < 0.05) for over half of the selected provisions. Column "Any" presents the results of a logistic regression on whether any agriculture-relevant provisions are present. Column "Count (≥ 1)" presents results of a Poisson regression on the count of agriculture-relevant provisions in a PTA, conditional on there being some. The latter two columns together constitute a hurdle model. All three regressions include the "Continent, combination" determinant: coefficients of its combinations are omitted for conciseness. All columns report standardized coefficients, making magnitudes comparable within columns (but not across since different models are used). Asterisks denote statistical significance at * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Heteroskedasticity-robust standard errors are reported in parentheses.

Appendix Figures and Tables

Table A.1: List of Potential Determinants Included in the RF Models.

Variable	Aggregators
CEPII BACI + UN Comtrade	
Pair's bilateral intra-industry trade index	
Pair's bilateral trade	
Pair's bilateral trade in agriculture	
Pair's bilateral trade in manufacturing	
Pair's bilateral trade in services	
Share of bilateral trade in agriculture	
Share of bilateral trade in manufacturing	
Share of bilateral trade in services	
Share of trade in agriculture	mean, difference
Share of trade in manufacturing	mean, difference
Share of trade in services	mean, difference
Total trade	log mean, log diff.
Value of exports, agriculture	log mean, log diff.
Value of exports, manufacturing	log mean, log diff.
Value of imports, agriculture	log mean, log diff.
Value of imports, manufacturing	log mean, log diff.
Value of imports, services	log mean, log diff.
CEPII Geodist	
Continent	same, combination
Landlocked	same, combination
CEPII Gravity	
Distance between the most populated city of each country	
EU member	same
GATT member	same
Historical origin of the legal system	same
Pair ever in a colonial dependency relationship	
Pair ever in a colonial sibling relationship	
Pair is contiguous	
Pair shares common language spoken by 9%+ of population	
Pair shares common legal system origins	
Pair shares common official/primary language	
Pair's trade share imbalance	
Pair's trade share in their trade with everyone	

Population-weighted distance between most populated cities	
Religious proximity index	
Trade imbalance, relative	
WTO member	same
CEPII Language	
Pair's linguistic proximity (LP1)	
Pair's linguistic proximity (LP2)	
Composite	
Contagion	mean, difference
Database of Political Institutions	
Checks and balances index	mean, difference
Executive branch elected indirectly	same, combination
Executive branch is nationalist	same, combination
Executive branch is religious	same, combination
Executive branch is rural	same, combination
Exeuctive branch is regionalist	same, combination
Fractionalization of legislature	mean, difference
Fractionalization of opposition	mean, difference
Fractionalization of the executive	mean, difference
Ideological position of the executive branch	same, combination
Incumbent leader is serving final term	same, combination
Incumbent leader still in office	same, combination
Largest party in the executive is right-leaning	same, combination
Lax checks and balances index	mean, difference
Legislature has multiple parties	same, combination
Legislature is bicameral	same, combination
Military has a role in government	same, combination
Number of years the incumbent leader's been in office	mean, difference
Political system	same, combination
IMF CDIS	
FDI imbalance	mean, difference
FDI inflow	log mean, log diff.
FDI outflow	log mean, log diff.
Pair's FDI share imbalance	
Pair's FDI share in their FDI with everyone	
Pair's relative FDI imbalance	
Total FDI	log mean, log diff.
Penn World Table	
Average annual hours worked by persons engaged	log mean, log diff.

Capital services levels, PPP	log mean, log diff.
Capital stock depreciation rate	mean, difference
Capital stock, PPP	log mean, log diff.
Government consumption share in GDP, PPP	mean, difference
Gross capital formation share in GDP, PPP	mean, difference
Household consumption share in GDP, PPP	mean, difference
Human capital index	mean, difference
Number of persons engaged	log mean, log diff.
Population	log mean, log diff.
Price level of consumption (PPP / exchange rate)	log mean, log diff.
Real GDP, PPP	log mean, log diff.
Real consumption, PPP	log mean, log diff.
Real domestic absorption, PPP	log mean, log diff.
Real internal rate of return	mean, difference
Share of labor compensation in GDP	mean, difference
TFP level, PPP	log mean, log diff.
UNCTAD	
Container port throughput	log mean, log diff.
Frontier technology readiness index	mean, difference
Pair's liner connectivity index	
WITS	
Avg weighted tariff pair levies on each other	
Difference in weighted tariff pair levies on each other	
World Development Indicators	
Agricultural land (% of land area)	mean, difference
Arable land (% of land area)	mean, difference
Average time to clear exports through customs	mean, difference
Bribery incidence	mean, difference
Business extent of disclosure index	mean, difference
CPIA building human resources rating	mean, difference
CPIA business regulatory environment rating	mean, difference
CPIA business regulatory environment rating	mean, difference
CPIA debt policy rating	mean, difference
CPIA economic management cluster average	mean, difference
CPIA efficiency of revenue mobilization rating	mean, difference
CPIA equity of public resource use rating	mean, difference
CPIA financial sector rating	mean, difference
CPIA fiscal policy rating	mean, difference
CPIA gender equality rating	mean, difference

CPIA macroeconomic management rating	mean, difference
CPIA policies for social inclusion/equity cluster average	mean, difference
CPIA policy and institutions for environmental sustainability	mean, difference
CPIA property rights and rule-based governance rating	mean, difference
CPIA public sector management and institutions cluster average	mean, difference
CPIA quality of budgetary and financial management rating	mean, difference
CPIA quality of public administration rating	mean, difference
CPIA social protection rating	mean, difference
CPIA structural policies cluster average	mean, difference
CPIA trade rating	mean, difference
CPIA transparency, accountability, and corruption in the public sector rating	mean, difference
Central government debt, total (% of GDP)	mean, difference
Cost of business start-up procedures (% of GNI per capita)	mean, difference
Current account balance (% of GDP)	mean, difference
Depth of credit information index	mean, difference
Ease of doing business score	mean, difference
Educational attainment: % of population 25+ completed upper secondary	mean, difference
Educational attainment: % of population 25+ with Bachelor's	mean, difference
Energy imports, net (% of energy use)	mean, difference
Energy use (kg of oil equivalent per capita)	log mean, log diff.
Firms formally registered when operations started (% of firms)	mean, difference
Firms that spend on R&D (% of firms)	mean, difference
Firms using banks to finance working capital (% of firms)	mean, difference
Foreign direct investment, net inflows (% of GDP)	mean, difference
High-technology exports (% of manufactured exports)	mean, difference
Human capital index	mean, difference
Industrial design applications, resident, by count	log mean, log diff.
Inflation, consumer prices (annual %)	mean, difference
International migrant stock (% of population)	mean, difference
Labor force	log mean, log diff.
Labor force participation rate	mean, difference
Land area	log mean, log diff.
Market capitalization of listed domestic companies (% of GDP)	mean, difference
Net lending/borrowing (% of GDP)	mean, difference
Net migration	mean, difference
Patent applications, residents	log mean, log diff.
Personal remittances, received (% of GDP)	mean, difference
Research and development expenditure (% of GDP)	mean, difference
Rural land area	log mean, log diff.
Statistical capacity score	mean, difference

Statistical performance indicators	mean, difference
Stocks traded, total value (% of GDP)	mean, difference
Strength of legal rights index	mean, difference
Total greenhouse gas emissions (kt of CO2 equivalent)	log mean, log diff.
Total natural resources rents (% of GDP)	mean, difference
Trademark applications, resident, by count	log mean, log diff.
Unemployment, total (% of total labor force)	mean, difference
Urbal land area	log mean, log diff.
Urban population (% of total population)	mean, difference
World Governance Indicators	
Control of Corruption	mean, difference
Government effectiveness	mean, difference
Political stability and absence of violence/terrorism	mean, difference
Regulatory quality	mean, difference
Rule of law	mean, difference
Voice and accountability	mean, difference

Note. The table lists the potentials determinants included in the random forest models. Variables are grouped by source of the measure of the underlying data used to construct the measure. For variables measured at the country-year level, the "aggregators" column reports how they were aggregated to the country-pair-year level: mean (or log mean) and difference (or log difference) for numerical variables, same (an indicator of whether the two countries' values are the same) and combination (a combination of the two countries' values) for binary or categorical variables.

Table A.2: P-Values of the Major Determinants of Binding PTA Provisions.

Determinant	CP 23	ET 15	ET 18	IPR 58	IPR 88	MoC 37	MoC 38	SE 43	TBT 4
Contagion, mean	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
Contagion, difference	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
Energy use (kg of oil equivalent per capita), log mean	**0.00**	**0.00**	**0.00**	1.00	1.00	**0.00**	0.20	**0.00**	**0.00**
Distance between the most populated city of each country	**0.00**	**0.00**	**0.00**	0.82	**0.00**	**0.04**	0.64	1.00	**0.03**
Continent, combination	**0.00**	**0.00**	**0.00**	**0.00**	0.20	**0.00**	0.19	1.00	**0.00**
Population-weighted distance between most populated cities	**0.00**	**0.00**	**0.00**	0.94	**0.00**	**0.07**	0.74	1.00	**0.00**
Voice and accountability, mean	**0.00**	**0.00**	**0.00**	1.00	1.00	0.17	0.10	**0.00**	0.23
Human capital index, mean	**0.00**	**0.00**	**0.00**	1.00	1.00	0.87	0.28	**0.00**	1.00
Regulatory quality, mean	**0.00**	**0.00**	**0.00**	1.00	1.00	0.98	0.40	1.00	0.93
Price level of consumption (PPP / exchange rate), log mean	0.80	**0.00**	**0.00**	0.99	0.67	1.00	0.24	**0.00**	1.00
Political stability and absence of violence/terrorism, mean	**0.01**	**0.01**	**0.00**	1.00	1.00	1.00	0.23	1.00	1.00
Average annual hours worked by persons engaged, log mean	**0.02**	**0.00**	**0.00**	1.00	1.00	1.00	0.66	1.00	1.00

Note. The table shows p-values of variable importance for the major determinants of binding PTA provisions fostering agricultural trade. The order of determinants follows Figure 2, and only the determinants relevant for three or more provisions are displayed. Individual provision columns use abbreviated versions of provision names in Table 3 and display the p-values of the permutation variable importance measures for selected determinants. The p-value color coding indicates **p-value $<$ 0.01**, **p-value $<$ 0.05**, and p-value $<$ 0.1.