Determinants of PTA Design: Insights from Machine Learning*

Stepan Gordeev† Sandro Steinbach‡

This version: March 24, 2024

Abstract

Preferential trade agreements (PTAs) have emerged as the dominant form of international trade governance. Provisions included in PTAs are increasingly numerous, broad in their purview, deep in their scope, and varied between agreements. We study the economic, political, and geographic determinants of PTA design differences. For each of the hundreds of classified PTA provisions, we consider 287 country-pair characteristics as potential determinants, covering many individual mechanisms the literature has studied. We employ random forests, a supervised machine learning technique, to handle this high dimensionality and complexity. We use a robust variable importance measure to identify the most critical determinants of the inclusion of each PTA provision. Contagion due to competition for export markets, geographic proximity, and governance quality emerge as essential determinants of PTA design. These results motivate future exploration of individual mechanisms our exercise points to.

Keywords: Preferential trade agreements, machine learning, provisions, trade integration

JEL Codes: F13; F14; F15

* We thank Deexith Reddy for excellent research assistance. We are also grateful to Armen Khederlarian, Dongin Kim, Julia Seither, and the participants of the ICEE at PSU-Altoona for their comments and suggestions. 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.

† Agricultural and Resource Economics, University of Connecticut, Email: stepan.gordeev@uconn.edu, Corresponding Author.

‡ Agribusiness and Applied Economics, North Dakota State University, Email: sandro.steinbach@ndsu.edu.

1 Introduction

Preferential Trade Agreements (PTAs) increasingly dominate international trade governance, while multilateral negotiations through the World Trade Organization have given some way to bilateral or regional PTA negotiations, whose number has grown from less than 50 in the early 1990s to over 360 in 2023 (WTO 2023).[1] This expansion has occurred despite tariffs being at their lowest-ever levels between most countries. Indeed, modern PTAs are no longer dealing mainly with tariff reductions. Instead, they are increasingly broad in the number of non-tariff policy areas their provisions cover, such as intellectual property rights protection, environmental laws, or public procurement rules. They are also increasingly deep in how far they go in harmonizing the rules of international trade. Those agreements vary widely in their provisions, reflecting different quantitative levels and qualitative features of trade liberalization and regulatory integration that signatories commit to. Given the growing role of PTAs, making sense of their increasing complexity and diversity is crucial for understanding modern trade institutions. However, isolating the determinants of PTA differences is challenging because PTA design is complex, and its determinants could have unexpected non-linear and interacting effects. This paper uses random forests, a supervised machine learning technique, to handle this complexity and identify the economic, geographic, cultural, and political factors that predict differences in PTA design. This exercise motivates and informs future research that would delve deeper into individual mechanisms related to potential determinants identified by random forests.

Studying the determinants of PTA design requires its systematization. Two existing datasets provide an expansive classification of provisions included in preferential trade agreements: Design of Trade Agreements, abbreviated as DESTA below (Dür, Baccini, and Elsig 2014), and Deep Trade Agreements, abbreviated as DTA (Mattoo, Rocha, and Ruta 2020). These two teams devel-

[1] Although regional integration is on the rise, the WTO still plays an important role for international trade governance. Recent WTO achievements include the implementation of the Information Technology Agreement II (2015) and the Trade Facilitation Agreement (2017), ongoing negotiations on the Electronic Commerce Initiative (2020), and operationalization of the Multi-Party Interim Appeal Arbitration Arrangement (2020). Additionally, the conclusion of the Services Domestic Regulation (2021) and the Investment Facilitation for Development (2023) initiatives underscore the WTO's adaptability and continued relevance in addressing evolving global trade challenges.

oped classifications of PTA provisions and manually encoded whether each classified provision was included in a given PTA. We use these classified provisions as measures of PTA design differences. For each classified PTA provision, we construct a statistical model that predicts whether the provision is included in a given PTA based on the characteristics of the pair of countries that signed the agreement. Whether a pair of countries choose to add a particular provision to their PTA can be dictated by a broad range of factors: their domestic economic and political situations, existing trade relationships, shared cultural or institutional features, and many others. We assemble 287 factors observed at country- and country-pair-level, collected from dozens of data sources. The existing literature informs the list of factors we consider on the determinants of PTA formation and the nascent literature on PTA design differences. Once the statistical model is constructed, we identify the factors that contribute the most to the model's predictive capacity and are thus the prime candidates for the important determinants of PTA design.

We employ random forests both for the estimation of each statistical model and for identifying factors that are most predictive of each provision. The random forest algorithm, developed by Breiman (2001), is an effective machine learning tool for prediction.[2] A random forest is constructed by "growing" many individual decision trees and aggregating their predictions. This algorithm is gaining popularity in economics due to its excellent predictive performance in a wide variety of problems, many of which the classic Ordinary Least Squares regression struggles with (Varian 2014; Mullainathan and Spiess 2017). Random forests have several advantages that make them well-suited for identifying the important determinants of PTA provisions among hundreds of potential factors. First, the algorithm deals well with high-dimensional data in which the number of predictors (potential determinants) is large without underdeterminacy. Second, random forests naturally adapt to non-linearities and interactions between predictors in the data without requiring the econometrician to impose a flexible parametric structure ex-ante, which would be infeasible with our high-dimensional data. Third, extensions of random forests allow for missing values without invalidating the variable importance measures: a crucial advantage given the many datasets with disparate coverage that we source the potential determinant variables from

[2] See Biau and Scornet (2016), Schonlau and Zou (2020), or Ziegler and König (2014) for excellent reviews.

(Tang and Ishwaran 2017). Finally, random forests have well-developed procedures for variable selection, facilitating our task of identifying the important determinants of PTA provisions among a multitude of potential factors: we employ the permutation importance method developed by Altmann et al. (2010).

In the initial exercise, we estimate a random forest to predict whether a PTA exists between a given pair of countries in a given period, which is a question the literature has extensively studied using conventional econometric techniques (Baier and Bergstrand 2004; Bergstrand, Egger, and Larch 2016). The random forest identifies geographic proximity, average regulatory quality, trade volume, and domestic political regimes as the most predictive of whether countries form a PTA. In the main exercise, we then estimate random forests to predict the inclusion of each classified PTA provision into an agreement between a particular country pair and aggregate the results. Firstly, we identify the country-pair characteristics that are the best predictors of the largest number of provisions, making them influence the overall PTA design the most. Several measures of geographic proximity are highly predictive of differences in PTA design, suggesting that neighbors are interested in different dimensions of trade integration compared to remote partners. Interdependence of PTA content between trading partners or nations competing for the same export markets emerges as another strong predictor of PTA design. This finding extends the existing results showing the importance of contagion for PTA formation (Baldwin 1993; Baldwin and Jaimovich 2012; Chen and Joshi 2010). Secondly, we repeat this analysis at a more fine-grained level, summarizing the critical determinants of common policy areas like intellectual property protection or anti-dumping regulation. Metrics of government quality and political competitiveness are highly relevant for several of these areas. Throughout this analysis, we link our results to individual mechanisms studied by the literature.

While we employ random forests to identify the factors most predictive of PTA formation and differences in PTA design, leading us to often refer to them as "determinants", it is important to stress that our empirical method does not identify causal effects of these variables on PTA outcomes. A variable that is crucial for the prediction of PTA formation or inclusion of certain provisions may contribute to prediction because it truly causes the PTA outcome, because it itself is affected by the PTA outcome (or its anticipation), or because it is correlated with another omitted predictor. Some of the determinants identified by random forests and discussed above link to individual mechanisms of PTA formation and design studied in the literature, but they cannot prove the importance of these mechanisms. Therefore, we view these machine-learning-derived results as primarily suggesting the most fruitful avenues of further research that can hone in on identifying the causal effect of each individual suggested mechanism one at a time.

A growing literature explores the economic and political determinants of PTA formation, identifying factors that make it more likely for a pair of countries to have signed an agreement. Conceptualizing PTAs as a dichotomous variable, whether an agreement between the two countries exists or not, loses the complexity of their design but makes the analysis feasible. Baier and Bergstrand (2004) show that simple geographic (like distance) and economic (like GDP level and similarity) factors related to the potential welfare benefits of trade liberalization can explain most existing PTAs. A significant strand of the literature highlights the importance of domestic interest groups, primarily exporters, supporting a PTA: Baldwin (1993) suggests that a newly signed PTA between two foreign countries hurts the competitiveness of domestic exporters, who push their government to sign a PTA of their own. Such interdependence, or contagion, in PTA formation is relevant empirically (Egger and Larch 2008; Baldwin and Jaimovich 2012; Baccini and Dür 2012; Chen and Joshi 2010). Multi-national corporations may be no less critical in lobbying for PTAs: MNCs in developed nations seek to open new markets for investment, while developing countries, in turn, seek to attract this investment (Baccini, Dür, and Elsig 2018; Manger 2009; Büthe and Milner 2008; Gamso and Grosse 2021). Domestic politics also seem to play a significant role: countries with autocracies or a large number of veto players sign fewer PTAs (Mansfield and Milner 2012), while leaders in developing countries may use PTAs as a way to commit to reform in a competitive political environment (Baccini and Urpelainen 2014). While most studies in this literature hypothesize a particular mechanism of PTA formation and then test it empirically, some combine a range of potential determinants informed by the literature in their empirical model to achieve the best performance in predicting PTA formation or its timing (Baier and Bergstrand 2004; Bergstrand, Egger, and Larch 2016). Our paper makes several contributions to this second strand of the literature. First, it considers a broader selection of potential economic, political, geographic, and cultural determinants of PTA formation, leveraging the ability of random forests to consider many predictors and their interactions. Second, it explicitly conducts variable selection to identify which potential factors are the most critical determinants of PTA formation, leveraging variable importance measures developed for random forests. Our most novel contribution, however, is not to the question of PTA formation but to PTA design.

Recently, the literature has shifted from exploring whether a PTA exists between a pair of countries to understanding why PTAs differ in their design. Existing studies primarily follow two approaches to dealing with PTA design complexity. Some reduce this complexity to a single index of PTA breadth or depth (like the number of provisions included) and analyze the determinants of variation in this measure across agreements (Gamso and Grosse 2021; Mattoo, Rocha, and Ruta 2020; Orefice and Rocha 2014; Hofmann, Osnago, and Ruta 2019). Others deal with the complexity by focusing on a single policy area, like rules of origin, escape clauses, or labor protection, and exploring a particular mechanism determining whether a provision covering that area is included in the PTA (Raess, Dür, and Sari 2018; Kucik 2012; Lechner 2016). We contribute to this literature by analyzing PTA design while keeping its complexity largely intact: we identify the critical determinants of hundreds of individual PTA provisions and use them to suggest the most relevant mechanisms affecting the entirety of PTA design. Our broad exercise is designed not to delve deep into a particular mechanism but to identify several areas of important determinants, informing future research into individual mechanisms. Understanding the differences in the contents of PTAs can help identify the breadth of objectives that negotiating parties seek and diagnose the obstacles to deeper international integration.

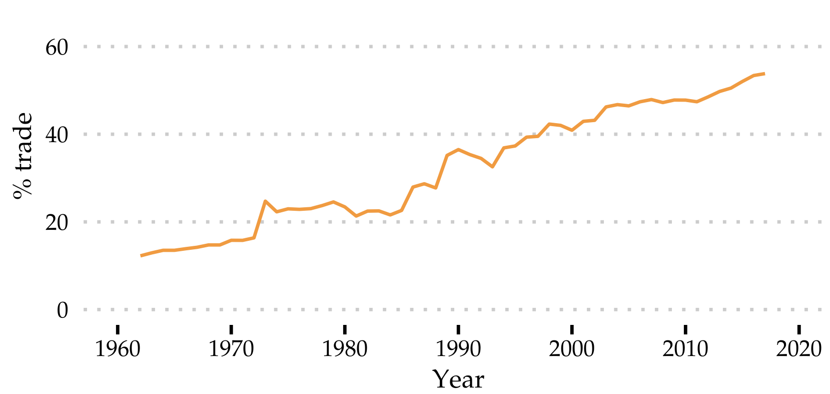
Our contributions to the two strands of the literature discussed above are made possible by employing advanced machine-learning methods. PTA design is high-dimensional, the number of potential determinants is significant, and these determinants tend to have non-linear and interacting effects (Baccini 2019). Classical econometric techniques tend to struggle with these features, while machine learning tends to excel in comparison (Varian 2014; Mullainathan and Spiess 2017), making it potentially effective in studying PTAs. This approach is leveraged by Breinlich et al. (2021) and Kim and Steinbach (2023), who apply machine learning techniques (lasso and several extensions) to identify PTA provisions that are most important for increasing trade flows. Our paper instead uses another machine learning technique (random forests coupled with several extensions) to understand the causes, not consequences, of PTA design differences. Machine learning methods applied to high-dimensional and non-linear economic problems can identify promising directions for future theoretical and econometric research to focus on.

2 Background

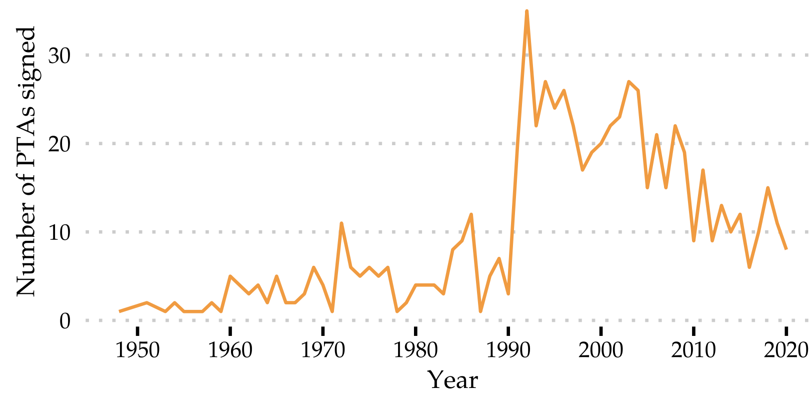
Multilateral trade negotiations through the WTO have largely stalled. The Doha Round of negotiations launched in 2001, but the member countries have not yet agreed. Instead, treaties with restrictive membership have proliferated to fill the void. A Preferential Trade Agreement (PTA) is an international treaty with restrictive membership that seeks to improve the mutual market access of its members (Limão 2016). As Figure 1a shows, the fraction of global trade between country pairs sharing a trade agreement grew steadily from 12% in the early 1960s to over 50% in 2023.[3] Figure 1b highlights the drastic upward jump in the number of new PTAs signed each year (as counted in DESTA), from below 10 in most years between 1950 and 1990 to 20-30 in most years in the 1990s. The formation of new agreements has slowed since then but is still elevated.

Not only are preferential trade agreements becoming more widespread, but they are also changing. Figure 1c shows the drastic change in the average number of provisions (as classified in DESTA) included in the agreement. Until the 1990s, most PTAs had no more than 10-20 provisions. Since the 1990s, however, that number has skyrocketed to over 100. Trade agreements no longer deal primarily with negotiating trade cost reductions. Instead, they are increasingly concerned with harmonizing a far broader set of international trade rules and domestic regulation relevant to trade: intellectual property rights protection, government procurement rules, environmental regulation, and many other policy areas (Mattoo, Rocha, and Ruta 2020). Due to their increasing complexity, agreements are also becoming more varied in their design. As Figure 1d shows, the variance in the set of provisions included in PTAs skyrocketed in the 1990s. Figures 1b and 1d also indicate that the complexity of new agreements has hit a snag in the last couple of years. It is too

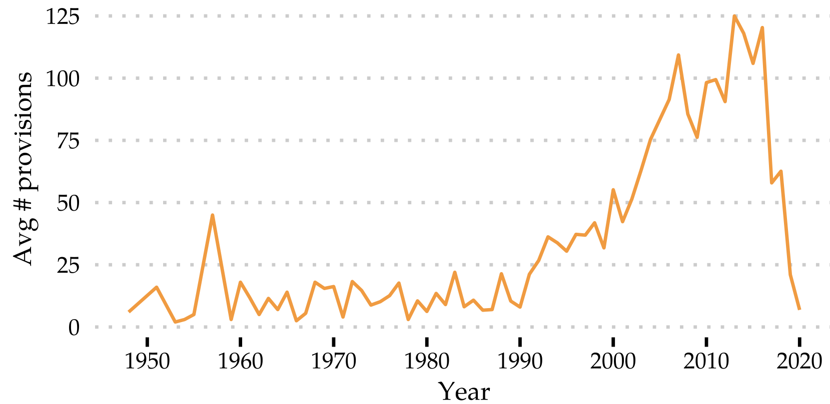
[3] Appendix Figure A.1 displays the fraction of country-pairs sharing an agreement, unweighted by their trade: it was growing steadily between the 1950s and 1970s but has skyrocketed starting in the 1990s from less than 5% in 1990 to 18% in 2017.



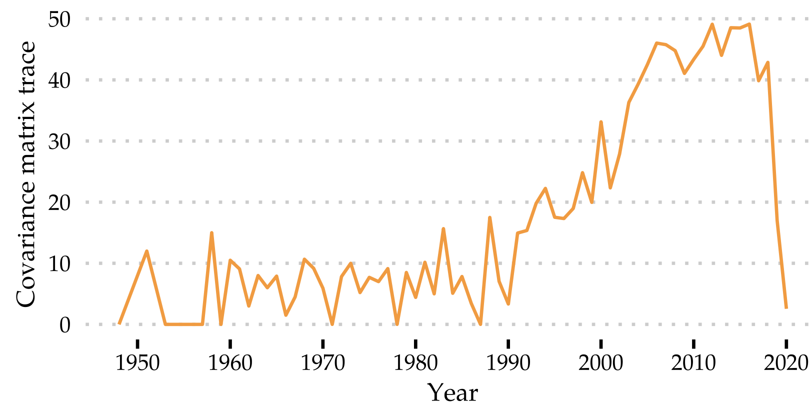
(a) Share of Global Trade Happening between PTA Members



(b) Number of PTAs Signed per Year



(c) Average Number of Provisions in a PTA over Time



(d) Variance of Provisions Included in PTAs over Time

Figure 1: Evolution of PTAs

Note. Panel (a) displays the share of all bilateral international trade that occurred between pairs of countries sharing a PTA (as classified in EIA), by year. Panels (b)-(d) are based on PTAs and provisions as classified in DESTA. Panel (b) displays the number of new PTAs signed per year. Panel (c) displays the average number of provisions in newly signed PTAs, by year. Panel (d) displays the change in the variance of provisions included in newly signed PTAs. Variance is measured as the trace of the covariance matrix of vectors indicating the inclusion of provisions in each PTA, which is equivalent to the sum of variances of the binary provision inclusion variable across provisions.

early to say if that is a temporary dip or another regime change. Therefore, our paper focuses on the trend of growing complexity that lasted from the 1990s to at least the 2010s.[4]

The estimated impact effects of PTAs on trade far exceed what can be expected from the moderate trade cost reductions they include, suggesting that the non-tariff provisions account for much of the trade-boosting effect of PTAs (Limão 2016). Furthermore, the growing breadth of PTA provisions reflects the scope of their impact. PTAs have been found not only to boost trade but also foreign direct investment (Baltagi, Egger, and Pfaffermayr 2008), survival of democratic regimes (Liu and Ornelas 2014), and even human rights protection (Hafner-Burton 2013)—as long as relevant provisions are part of the agreement. Understanding the determinants of provision inclusion can help understand the breadth of outcomes that signatories are seeking—and why they may seek some outcomes and not others. Understanding the determinants of this increased complexity and diversity of PTA design also matters in understanding its effects. Estimating the impact of PTA formation or design differences on trade or other outcomes suffers from the endogeneity of countries' decisions to form those PTAs. A strand of the literature tackles this problem by instrumenting PTA formation with important determinants of PTA formation that are plausibly exogenous to the outcome variable being studied. A similar approach has been used for studying the effect of PTA depth (Osnago, Rocha, and Ruta 2017; Mattoo, Mulabdic, and Ruta 2022). Identifying a wide range of essential determinants not only of PTA formation but also design can thus aid future work in this line by providing potential instruments to choose from.

3 Empirical Methods

3.1 Random Forests

Classification Trees and Random Forests—The random forest is a supervised machine learning algorithm introduced by Breiman (2001). Analogously to a regression, the algorithm constructs a statistical model that predicts the value of an outcome variable based on the values of provided

[4] The patterns of PTA evolution discussed in this section using the DESTA dataset are robust to using the DTA dataset: see Appendix Figures A.2, A.3, and A.4, the DTA analogs of Figures 1b, 1c, and 1d respectively.

predictor variables. In our application, the prediction task is binary classification: the forest must predict whether a given country-pair has a particular PTA provision. A random forest is an ensemble model comprised of many smaller models, namely classification trees. Each tree is trained on a random bootstrap sample of the data, where the sample is drawn with replacement, and its size equals the size of the original data. To come up with a single prediction of the outcome variable for a given observation, the random forest collects the predictions of all individual trees and picks the majority prediction. Each classification tree in a random forest consists of nodes that repeatedly partition the data into branches. Each node splits the data into two groups based on the value of a particular predictor variable. The predictor variable and value to split on are picked to optimize some goodness-of-fit measure, i.e., to achieve the highest contrast between the two branches in the expected value of the outcome variable. Not all predictor variables are considered in the search, however: at each node, only a random subset M_{try} of predictor variables is picked for the search. Then, all possible cutoff values (for continuous variables) or all partitions of groups into two sets (for categorical variables) are searched over to find the best split.

Figure 2 provides a stylized illustration. At the first node, the algorithm chooses to split on the value of continuous variable X_3 at cutoff a, sending observations with $X_3 \leq a$ to the left branch and those with $X_3 > a$ to the right one. Once the data is split into two branches, the process repeats at the two newly created nodes. The left branch is split again on the categorical variable X_1, and the right branch on the binary variable X_7. With each consecutive split, the number of observations going down each branch to the following node drops. Once the number of observations at a node falls to some cutoff value *nodesize*, the branching stops: the majority value of the binary outcome variable of observations at this terminal node provides the prediction of the branch.[5] To obtain the tree's prediction of the outcome value of a new observation, the observation is "dropped" down the tree, following the branches according to its predictor values. Then, the observation arrives at one of the terminal nodes, generating the prediction. There are two sources of randomness in a random forest: the random bootstrap sample used by each tree and the random selection of predictors to consider for the role of the splitting variable at each node. These two features

[5] If the outcome variable is continuous, the tree is called a regression tree, and the average of values in the final node provides the prediction.

mitigate the propensity of individual trees to overfit and allow random forests to achieve high out-of-sample predictive performance compared to other machine learning methods (Caruana and Niculescu-Mizil 2006).

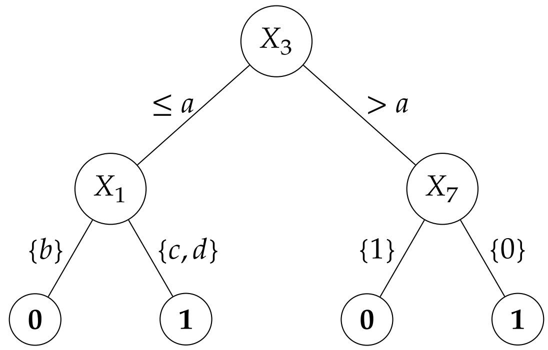


Figure 2: Illustration of a Classification Tree

Note. The figure provides an example of a simple classification tree with two splits levels. Based on the observable characteristics $\{X_i\}_i$, the tree predicts whether the outcome variable is 0 or 1 for each observation. Only three characteristics are used as predictors: X_1, X_3, and X_7.

Random Forests are Well-Suited for Studying the Determinants of PTA Formation and Design—Random forests have several advantages that make them a better choice for identifying the essential determinants of PTA design than ordinary least squares (OLS) regression or alternative machine learning methods (Ziegler and König 2014; Schonlau and Zou 2020). Firstly, random forests naturally adapt to non-linearities and interactions between predictors in the data without requiring the econometrician to impose the parametric structure ex-ante. This feature is critical to allow us to consider hundreds of potential determinants without hypothesizing any particular relationships beforehand. Allowing for a broad array of possible interactions and non-linearities between 287 potential determinants in OLS would balloon the number of predictors and make estimation impossible. Random forests adapt to interactions and non-linearities they find on the fly through the flexible structure of repeated node splits. In Section 5.2 below, we show that random forests significantly outperform logistic regression in out-of-sample prediction of PTA provisions.

Secondly, one extension of random forests, described below, allows for the presence of missing values without relying on imputation that may have spurious effects on variable importance measures. This feature is essential for our exercise, which combines hundreds of potential determinants from many data sources: few country-pair-year observations have no missing values in any of these. OLS would require restricting the sample to this small subset of observations or restricting the set of predictors only to those with universal coverage. The method of handling missing values in random forests allows us to avoid these extremes and maximize the information utilized in estimation.

Finally, random forests have well-developed procedures for variable selection, facilitating our task of identifying the critical determinants of PTA provisions among many potential factors discussed in Section 3.2. An extensive literature on variable selection measures for random forests distinguishes them from many alternative machine learning techniques (e.g. Strobl et al. 2007; Gregorutti, Michel, and Saint-Pierre 2017; Altmann et al. 2010; Nembrini, König, and Wright 2018; Janitza, Celik, and Boulesteix 2018).

At the same time, random forests have limitations compared to conventional econometric methods. Their theoretical properties are poorly understood: their black box nature complicates mathematical analysis (Biau and Scornet 2016). There are limited results on the consistency of random forests with particular data-generating processes and certain simplifications in the algorithm itself (Scornet, Biau, and Vert 2015; Scornet 2016). There are also negative results identifying examples in which random forests are not consistent (Tang, Garreau, and von Luxburg 2018). Furthermore, while random forests excel at prediction, the way they arrive at this prediction is not transparent: we return to this issue in Section 5.

Dealing with Missing Observations—Collecting hundreds of variables to serve as potential determinants from many different sources inevitably runs into the problem of disparate data coverage. Various data sources cover different country-year combinations, and the gaps in their data availability do not overlap perfectly. With each added variable, the share of country-pair-year observations with missing values in at least some potential determinants inexorably grows. To deal with this data challenge, we employ on-the-fly-imputation proposed by Ishwaran et al. (2008) and Tang and Ishwaran (2017) that allows random forests to work with missing data without relying on imputation (despite its name, this technique involves no imputation when used purely for handling missing values). When constructing each tree node, the modified algorithm ignores missing values when finding the best splitting point and calculating the splitting statistic. Once the node is constructed, each missing value is (temporarily) replaced with a random draw from the distribution of non-missing values of this variable to determine which branch to send it to. The replacements are reset immediately so as not to influence further splits.

This approach avoids the drawbacks of the two standard solutions to the problem of missing values (Scheffer 2002). The first standard solution is to throw out the missing data, which means using rows (country-pair-year observations) with no missing values in any of the columns (potential determinants) or only using columns with no missing values in any of the rows. Both would throw out most of the determinants we observe, leaving only a small subset of the data suffering from severe selection. This method is unsuitable for our application because considering many potential determinants is one of the critical objectives. In contrast, on-the-fly-imputation allows us to maximize the information used since it does not discard observations with some missing values. The second conventional solution to the problem of missing data is to impute the missing values. This approach does not discard any information, but it complicates the interpretation of variable importance measures since imputed values of a given variable now contain information on relationships with other variables absent from the raw data. This method is unsuitable because it precludes reliable identification of the critical determinants of PTA formation and design. On-the-fly-imputation, in contrast, does not perturb the variable importance measures since only the non-missing values are used to create splits.

Tuning—The random forest algorithm has three key parameters: N, the number of trees in the forest; M_{try}, the number of predictors randomly considered as splitting candidates at each node; and *nodesize*, the number of observations in each cell of the tree below which the cell is not split further. Higher N is always better but more computationally costly. We set $N = 500$ as raising it to 1,000 yields only a negligible improvement in predictive performance. We tune M_{try} and *nodesize* using k-fold tuning on a 2-dimensional parameter grid.[6] At each grid point, the data is split into k sub-samples, or folds. We use $k = 3$. Consequently, three random forests are estimated: each uses two of the folds as training data and one as testing data. The computed prediction error is thus out-of-sample for each of the forests. The prediction error is averaged across the three random

[6] See Bischl et al. (2021) for an overview of this and other tuning methods.

forests. The parameter grid point with the lowest prediction error is picked. We conduct the tuning procedure for each provision separately.[7]

The final choice is the splitting statistic the algorithm seeks to optimize when searching for the best split at each tree node. We use the area under the receiver operating characteristic curve (AUC-ROC), which has the interpretation of the probability that a random PTA that does have the provision in question and a random PTA that does not have the provision are both classified correctly: (Ling, Huang, and Zhang 2003). The more conventional default is Gini impurity, which can be interpreted as the probability that a random PTA is classified wrongly (has the provision in question but was sent down the "0" branch, or vice versa) at the split (Biau and Scornet 2016). However, because the data we work with is highly imbalanced (for most provisions, the share of PTAs that don't have the provision far exceeds the share that do), we find that AUC-ROC produces considerably better predictive performance (measured with misclassification error) of the whole forest for almost all provisions. By equally weighting the true positives and negatives, AUC-ROC works better with imbalanced data and produces better predictive performance of the whole forest—even though we measure this performance with the unweighted misclassification rate.

3.2 Variable Importance Measures

Constructing the random forest models and assessing their performance in predicting PTA formation and design is a necessary first step of our analysis but not its ultimate goal. What is more central is identifying the country-level and country-pair-level characteristics that are the most predictive of PTA formation and design. Random forests are largely appealing for our purposes because of the well-developed literature on variable importance measures (VIMs). These measures rank the predictor variables by their contribution to the predictive performance of the forest. We select Permutation Importance developed by Altmann et al. (2010) as the variable importance measure: it involves computing the Mean Decrease Accuracy of each variable and then comparing this value to a null distribution of Mean Decrease Accuracies obtained through repeated permutation.

[7] Although we use tuned parameters in our analysis, the procedure provided only a small improvement in predictive performance relative to the conventional defaults of $M_{try} = \sqrt{M}$, where M is the total number of predictor variables, and $nodesize = 1$.

We describe both components below.

Mean Decrease Accuracy—Mean Decrease Accuracy (MDA) is a commonly used variable importance measure proposed in the original Breiman (2001) paper.[8] For a given predictor X, it randomly reshuffles the vector of values of X associated with each observation. For each tree, it then computes the prediction error obtained by running this fake data (with all variables but X retaining their true order) down the tree. It obtains the difference between this fake error (using the permuted X) and the original prediction error (without X being permuted). The average difference across all trees is the Mean Decrease Accuracy of variable X. We use the out-of-bag misclassification rate as the conventional measure of prediction error used by the algorithm. For each observation, only the trees that did not happen to have the observation in their bootstrap training sample (i.e., the observation is "out-of-bag" for them) participate in generating the random forest's prediction, exploiting the fact that each bootstrap sample leaves out $\frac{1}{e} \approx 36.8\%$ of observations on average. This approach is less prone to overfitting than using the unadjusted misclassification rate. MDA is commonly used for random forests and other applications, but it has two limitations that are of particular concern to us. Firstly, it is mechanically biased in favor of predictors that offer many potential splitting points (Strobl et al. 2007).[9] Continuous variables or categorical variables with many categories receive a higher value regardless of how informative they are of the outcome variable. Secondly, MDA is biased against predictors that belong to clusters of correlated predictors (Gregorutti, Michel, and Saint-Pierre 2017). Finally, MDA allows one to rank variables by importance but provides no natural cutoffs for splitting variables into "important" and "unimportant" ones.

Permutation Importance—Altmann et al. (2010) developed Permutation Importance, a method of computing p-values for any VIM while simultaneously removing its biases. The core idea is an application of randomization-based inference: the method permutes the outcome variable vector (randomly reshuffling the mapping between observations and outcome variable values) many times, re-estimating the random forest and the VIM of each variable on every permutation. For a

[8] Mean Decrease Accuracy is sometimes called Permutation Importance. We elect the former name to avoid confusion with the method developed by Altmann et al. (2010), also called Permutation Importance.

[9] Mean Decrease Impurity, another commonly used VIM, is even more susceptible to this bias.

given predictor X, the random permutation breaks any association X had with the outcome variable. The distribution of X's MDA values across these permuted random forests thus composes the distribution of null importances. The position of X's MDA value from the original non-permuted random forest in this null distribution allows us to compute its p-value directly: the p-value captures the probability of observing X's MDA importance if X had no association with the outcome. This p-value constitutes the corrected variable importance measure: it does not suffer from the bias of conventional measures identified by Strobl et al. (2007), does not suffer from the bias against correlated predictors identified by Gregorutti, Michel, and Saint-Pierre (2017), and provides an absolute metric of variable importance. One drawback of this method is its computational cost, as each random forest of interest requires constructing many random forests (we conduct 100 permutations for each provision). See Janitza, Celik, and Boulesteix (2018) and Nembrini, König, and Wright (2018) for faster heuristic methods that also remove biases and provide p-values while yielding results similar to Altmann et al. (2010). Still, we use the technique by Altmann et al. (2010) as it is more theoretically grounded than these heuristic methods. The Variable Importance Measure we use is the p-value of the Permutation Importance, which uses Mean Decrease Accuracy within each permuted iteration.

4 Data

4.1 Preferential Trade Agreements and Their Provisions

We rely on three complementary datasets cataloging and classifying PTAs. We exploit the broad coverage of the EIA dataset in our analysis of the determinants of PTA formation. We then explore the features and determinants of PTA design using the classified provisions of DESTA and DTA, relying on the former for our primary analysis due to its larger sample size.

EIA—The most complete database of trade agreements between country pairs is the NSF-Kellogg Institute Data Base on Economic Integration Agreements (EIA) (NSF-Kellogg Institute 2021). It indexes agreements between every country pair from 1950 to 2017. The level of integration categorizes each trade deal. We use this dataset to measure whether a PTA exists between a given country pair in a given year. EIA covers the largest number of agreements but provides no information on the provisions of each agreement, precluding analysis of variation in PTA design. The following two datasets fill this gap.

DESTA—Design of International Trade Agreements (Dür, Baccini, and Elsig 2014), or DESTA, provides a manual classification of provisions included in 697 PTAs. The team behind the dataset developed a classification of 313 provisions, enumerating whether each provision is included in each PTA. The dataset is regularly updated: we use the 2022 vintage.

DTA—Deep Trade Agreements (Mattoo, Rocha, and Ruta 2020), or DTA, provides a more fine-grained look at the contents of agreements, classifying PTA clauses into 937 provisions. The cost of this level of detail is a smaller sample size of classified PTAs, which is 274. This smaller number of agreements is why we rely on DTA for robustness checks rather than the primary analysis.

4.2 Potential Determinants

We assemble an extensive array of observable characteristics of countries and country pairs. The random forest algorithm will search among these factors for essential determinants of PTA formation and differences in PTA design. The inclusion of many potential determinants is motivated by existing studies on the importance of those variables for PTA formation. Our empirical strategy allows us to verify their importance for PTA formation in the presence of many other potential factors and to study their ability to explain individual provisions in signed PTAs. Other potential determinants are motivated by the growing number of provisions discussing a particular topic. For example, the rise of intellectual property rights protection clauses leads us to include several measures of countries' innovation activity, human capital, and property rights protection, as well as the differentials in these measures between the two PTA signatories. Below, we summarize the main groups of potential determinants, the motivation for including them, and the data sources that provide them.

Economy—Early studies like Baier and Bergstrand (2004) showed that simple economic variables like GDP level and similarity can explain much of PTA formation. To explain PTA design, however, we need to include more detailed measures. PTA provisions are often sector-specific, so our analysis includes sectoral agriculture, manufacturing, and services shares. More and more provisions protect labor rights, intellectual property, and the environment. So, we include measures of labor share in compensation, inequality, innovation, human capital, emissions, energy use and sources, and natural resource endowments. Most of these aggregate and sectoral measures are available from *Penn World Tables* (Feenstra, Inklaar, and Timmer 2015) and *World Development Indicators* (The World Bank 2023b).

Proximity—The proximity of two countries to each other is one of the most significant predictors of PTA formation (Baier and Bergstrand 2004; Bergstrand, Egger, and Larch 2016), so we include measures of physical proximity and shipping connectivity. Many provisions attempt to harmonize institutions between two countries, motivating us to include measures of linguistic and institutional similarity and shared colonial past: these may influence how much regulatory harmonization is needed. These measures of geographic, institutional, and cultural proximity are available from *CEPII Gravity* (Conte, Cotterlaz, and Mayer 2022), *CEPII Language* (Melitz and Toubal 2014), *GeoDist* (Mayer and Zignago 2011), *UNCTADstat* (United Nations Conference on Trade and Development 2023).

Trade—We include measures of existing trade flows, tariff levels, and trade openness, as the extent of existing trade and remaining tariff trade barriers may influence whether a PTA focuses on tariffs or non-tariff regulation. We include unilateral and bilateral trade imbalance, as some studies find a prominent role of signs and magnitudes of trade imbalances in explaining PTA formation and features (Grossman and Helpman 1995; Kucik 2012; Facchini, Silva, and Willmann 2021). A growing literature shifts attention to the role of multinational corporations and global value chains in leading the push for greater trade integration (Chase 2008; Baccini, Dür, and Elsig 2018; Manger 2009; Büthe and Milner 2008; Gamso and Grosse 2021; Manger 2015; Raimondi et al. 2023), leading us to include measures of foreign direct investment flows and intra-industry trade. These measures are constructed from data available from *UN Comtrade* (United Nations 2023), *WITS* (The World Bank 2023a), *BACI* (Gaulier and Zignago 2010), *CEPII Gravity* (Conte, Cotterlaz, and Mayer 2022), *IMF CDIS* (International Monetary Fund 2023).

Interdependence—A large literature has documented contagion and interdependence of agreements, likely due to exporter lobbying, as an important determinant of PTA formation (Baldwin 1993; Baldwin and Jaimovich 2012; Baccini and Dür 2012; Egger and Larch 2008; Chen and Joshi 2010). If country j is an important export destination for countries i and k, and country k signs a PTA with country j, then country i has the incentive to emulate its "competitor" and also sign an agreement with country j, or else risk losing competitiveness in this export market. To capture this channel, we include a measure of contagion following this literature, primarily Baldwin and Jaimovich (2012). For country i considering an agreement with country j in time t, the contagion index of PTA property p is the number of agreements with this property that j has with other nations k, weighted by the importance of imports from k for j and by the importance of j as an export market for i:

$$\text{Contagion}_{p,ij,t} = \left(\frac{\text{bilateral exports}_{ij,t}}{\text{total exports}_{i,t}}\right) \sum_{k \neq i,j,t} \left(\frac{\text{bilateral exports}_{kj,t}}{\text{total imports}_{j,t}}\right) \mathbb{1}_{p,jk,t} \tag{1}$$

where $\mathbb{1}_{p,jk,t}$ is an indicator function encoding the presence of PTA property p between countries j and k in year t. To represent the mechanism of PTA contagion studied in the formation literature, we construct the Contagion index using $p =$ existence of a PTA. Our data, however, also allows us to extend this analysis to the question of interdependence of provisions. If countries j and k include a particular provision in their agreement, protecting its market share may require i not only to sign an agreement of its own with j but also to include similar provisions. Therefore, when estimating a random forest on the determinants of PTA formation or a particular provision, we include the contagion index for the property in question. When using the contagion index as a predictor of the presence of provision p in the PTA between i and j signed in year t, we lag the index by one year to avoid its contamination by countries k that are joining the same treaty (which is possible for PTAs with > 2 members).

Interdependence may be driven not only by competitive pressures but also by simple imitation. Modern trade agreements are highly complex, and a nation entering a new agreement may lift part of the structure from its existing agreements with third countries. We adopt a measure of such "template" effects used in the literature (Osnago, Rocha, and Ruta 2017; Mattoo, Mulabdic, and Ruta 2022) to our setting. Highly related to the contagion index, the template index for country i considering an agreement with country j in time t with property p is the number of agreements with this property that i has with other nations k, weighted by the importance of k as an export market for i:

$$\text{Template}_{p,ij,t} = \sum_{i,k\neq j,t} \left(\frac{\text{bilateral exports}_{ik,t}}{\text{total exports}_{i,t}} \right) \mathbb{1}_{p,ik,t} \tag{2}$$

Politics—Many studies emphasize the *political* aspect in the political economy of PTA formation and design. Domestic political pressures (or lack thereof) and leanings appear to play a large role in the formation of PTAs (Mansfield and Milner 2012; Baccini and Urpelainen 2014; Raess, Dür, and Sari 2018). So, we test the role they may also play in PTA design by including a broad range of measures of domestic political regime and political competitiveness. An increasing number of PTA provisions focus on harmonizing domestic regulations relating to property rights, government procurement, etc. (Gamso and Grosse 2021; Lechner 2016). So, we include measures of the quality of governance in these fields. These variables are compiled from the *Database of Political Institutions* (Cruz, Keefer, and Scartascini 2021), *Worldwide Governance Indicators* (Kaufmann, Kraay, and Mastruzzi 2010), and *World Development Indicators* (The World Bank 2023b).

Country-Pair Measures—Some of the measures discussed above (like bilateral trade or FDI flows) are inherently measured at the country-pair level. Most potential determinants, however, are measured at the country level and need to be aggregated to the country-pair level before they can be used as explanatory variables in country-pair-level PTA formation and design analysis. Generally, we construct two aggregated variables from each country-level measure: an average of the two countries' values (either in levels or in logs, depending on context) and a difference (in levels or logs). This simple construction lets us capture a broad scope of mechanisms previously discussed in the literature. A clear example is the diversity of results on the effects of GDP levels and differences. Baier and Bergstrand (2004) find that country-pairs whose GDPs are bigger (high average) or more similar (low difference) are more likely to form PTAs. At the same time, Baccini and Urpelainen (2014) show that domestic political pressures may lead developing countries' governments to sign PTAs with partners much richer than themselves (high difference). Orefice and Rocha (2014) find that PTAs are deeper when members trade more intermediate inputs, especially if the PTA is across development levels (high difference). Our specification allows us to flexibly capture such effects and interactions—not only for GDP but for all variables we collect—and to explore their role in influencing PTA formation and the specifics of PTA design.

Appendix Table A.1 shows the complete list of 287 assembled potential determinants. We have

17,678 country-pair-year observations for the primary analysis of each random forest. Each observation captures a particular PTA signed by a country-pair in a year.[10] Of the 313 classified DESTA provisions, a sizable portion has either no variation across agreements (either all or none of the agreements have the provision in question) or minimal variation. In such cases, random forest output is hardly interpretable as the model has too little variation to train on. Therefore, we discard provisions present in less than 10% or more than 90% of country-pair-year observations, leaving 119 provisions for the analysis.

5 Results

5.1 Determinants of PTA Formation

Before tackling the complexity of PTA design, we apply the random forest algorithm to identify the critical determinants of PTA formation. In this case, the algorithm's objective is to predict whether a given country-pair shares a PTA in a given year. Because most potential determinants we collected vary only slowly over time, we reduce the time dimension to five-year intervals, averaging the value of each determinant within the interval and treating a country-pair as having a PTA in that interval signed at any point up to the end of the interval.

Modifications to the Random Forest Algorithm—Due to its country-pair-period structure, the data used for the formation analysis consists of 480,738 observations, making classical random forests (which we use for finding the determinants of PTA design in Section 5.2 below) infeasibly computationally costly. We introduce two simplifications to the random forest algorithm that render the problem feasible. First, the procedure of finding the optimal splitting point at each tree node is simplified. Instead of considering all possible splitting points, only ten random splitting points for each

[10] Following the literature, we treat multilateral PTAs as sets of bilateral agreements between pairs of their members. This approach fits all PTA information into a rectangular table, facilitating empirical analysis and mapping results to earlier studies. At the same time, essential lessons will likely be learned from treating multilateral PTAs as singular entities, preserving the high-dimensionality of their members' characteristics. Tackling this problem with machine learning techniques will be a fruitful avenue for future research. As another simplification, each agreement between a country pair is treated as an independent observation even if the pair previously had another agreement that got superseded.

variable are considered. Ishwaran (2015) finds that this simplification tends to attain predictive performance that is no worse (and potentially better) than the default algorithm. Second, we simplify the procedure of computing the variable importance: only a random 10% subsample of the data is used to come up with MDA values on each iteration of Altmann et al. (2010)'s method. We find that the results are not sensitive to the chosen subsample size. Note that the provision-level random forests discussed in Section 5.2 are computationally simpler and do not require these two simplifications.

Furthermore, the PTA formation exercise is extremely imbalanced: of the 480,738 country-pair-period observations in the final data, only 16,683 have an active PTA. In such highly imbalanced data, regular random forests focus on achieving high predictive performance for the 96.5% majority class (country pairs without a PTA) at the expense of the 3.5% minority class (country pairs with a PTA). To overcome this performance discrepancy, we employ a quantile classifier for random forests, developed by O'Brien and Ishwaran (2019), which effectively boosts the predictions of the minority class: the modified algorithm minimizes not the overall unweighted misclassification error, but the sum of the within-class misclassification errors.

Performance—Table 1 presents the performance of the quantile classifier random forest in predicting the presence of PTAs, measured with out-of-bag misclassification error. The misclassification error is the share of observations for which the forest came up with an incorrect prediction: the given country-pair had an agreement in a given period, but the forest predicted that it did not, or vice versa. The error is computed out-of-bag (OOB): to come up with a prediction for each observation in the forest, only the trees that did not have this observation in their random bootstrap training sample are used. This approach is a compromise solution between using unadjusted misclassification (which we avoid as it would be overly optimistic due to overfitting) or splitting the data into training and testing samples and using the latter only for performance evaluation (which we avoid as it would stretch our sample size too thinly). The random forest achieves a 25.5% overall error with a reasonable balance between the two classes: 26% of the majority class and 2% of the minority class are misclassified.[11]

Table 1: Out-of-Bag Misclassification in Imbalanced Formation Random Forest

OOB Misclassification			
Overall	0 (Absent)	1 (Present)	Share of 1s
0.255	0.264	0.021	0.035

Note. The table presents statistics on the predictive performance of the random forest predicting PTA formation using the quantile classifier. The misclassification error is the share of country-pair-period observations for which the random forest incorrectly predicted the presence/absence of a PTA. It is computed out-of-bag: for each observation, only the trees that did not have this observation in their training sample are used to develop a prediction. Columns "0 (Absent)" and "1 (Present)" show misclassification only for country-pair-period observations that had or did not have an agreement within the period, respectively.

Significant Determinants—Table 2 lists the country-pair characteristics whose p-value variable importance is below 0.01, ranked from most to least significant. These characteristics are the most important for the random forest's ability to predict PTA formation: we call these the significant determinants of PTA formation. Several distinct groups of determinants emerge within the ranking. The determinants within each group are highly related. Identifying important determinants even when some form clusters of correlated variables is another advantage of the permutation importance method we adopt (Altmann et al. 2010). Geographic characteristics like the distance between the two nations, an indicator of whether they share a continent, and the exact combination of their continents are among the variables with the highest ability to predict PTA formation.[12] This esti-

[11] Appendix Table A.2 shows the performance of the "regular" (i.e. not quantile classifier) random forest. It achieves a better misclassification rate of only 3% but at the cost of significant imbalance: fully 89% of the minority class is misclassified. The balance of performance can be summarized by the G-mean measure, which is often used as a measure of predictive performance for imbalanced data. It is defined as the geometric mean of sensitivity and specificity:

$$\text{G-mean} = \sqrt{\text{sensitivity} \times \text{specificity}} = \sqrt{(1 - \text{ OOB misclassification of 1s}) \times (1 - \text{ OOB misclassification of 0s})}$$

The G-mean is 0.33 for the regular formation random forest and 0.85 for the quantile classifier ("imbalanced") formation random forest, capturing the far better balance of performance of the latter.

[12] The combination of a country pair's continents is a categorical variable with a category for each possible pair of continents, e.g., "Europe, Africa"

mate supports the findings of earlier studies that identify the importance of geographic proximity for PTA formation (Baier and Bergstrand 2004; Bergstrand, Egger, and Larch 2016). The random forest likewise demonstrates the importance of contagion of PTAs between trading partners: the average and difference of the contagion index between the two nations are the second and third most significant determinants (Baldwin 1993; Baldwin and Jaimovich 2012; Baccini and Dür 2012; Egger and Larch 2008; Chen and Joshi 2010).

Several metrics of the domestic political situation and regime are highly predictive of PTA formation. Features of the legislative and executive branches and the overall political system appear among the most significant determinants. These findings echo the theoretical and empirical results of Mansfield and Milner (2012) showing that the domestic political regime and the number of veto players significantly affect the propensity of a nation to enter into PTAs. Another critical determinant identified by the random forest is whether the leaders of those countries are in their final term, supporting the result of Baccini and Urpelainen (2014, 2015) on the tenure of the current leader being an essential driver of whether the leader seeks PTA membership as a way to commit to reforms. Several determinants in Table 2 speak to the average regulatory quality of the two nations: the averages of "Voice and accountability", "Regulatory quality", and "Ease of doing business" are all important for predicting PTA formation. Regulatory convergence is a significant goal of modern PTAs (Polanco Lazo and Sauvé 2018). However, the random forest suggests that what matters the most for PTA formation is the *average level* of regulatory quality in the two nations rather than the *differential* in regulatory qualities. Predictably, several measures of trade volume appear as well. The average overall volume of trade done by two nations, the volume of their bilateral trade, and the share of bilateral trade in their overall trade are all highly predictive of PTA formation. Moreover, the bilateral intra-industry trade index is significant, supporting the existing findings that global value chains are important contributors to PTA formation, driven primarily by multi-national corporations lobbying for increased economic integration (Baccini, Dür, and Elsig 2018; Manger 2009).

Linear Effects of Significant Determinants—In discussing the results of Table 2, we sought to connect the significant determinants identified by the random forest to the mechanisms that the literature has studied theoretically or using conventional econometric techniques. Note, however, that we

Table 2: Significant Determinants of PTA Formation

Determinant	P-value VIM
Population-weighted distance between most populated cities	<0.01
Contagion, mean	<0.01
Contagion, difference	<0.01
Continent, combination	<0.01
Continent, same	<0.01
Price level of consumption (PPP / exchange rate), log mean	<0.01
Pair's bilateral trade in manufacturing	<0.01
Human capital index (PWT), mean	<0.01
Pair's bilateral trade in agriculture	<0.01
Executive branch is rural, same	<0.01
Pair's bilateral trade	<0.01
Value of exports, agriculture, log mean	<0.01
Legislature is bicameral, combination	<0.01
Average annual hours worked by persons engaged, log mean	<0.01
Pair's trade share in their trade with everyone	<0.01
Executive branch elected indirectly, combination	<0.01
Voice and accountability, mean	<0.01
Pair's bilateral intra-industry trade index	<0.01
Pair's bilateral trade in agriculture, log mean	<0.01
Regulatory quality, mean	<0.01
Pair's bilateral trade in manufacturing, log mean	<0.01
Exeuctive branch is regionalist, same	<0.01
Value of imports, manufacturing, log mean	<0.01
Value of imports, agriculture, log mean	<0.01
Value of imports, services, log mean	<0.01
Fractionalization of legislature, mean	<0.01
Political system, combination	<0.01
Pair's bilateral trade in services, log mean	<0.01
Value of exports, manufacturing, log mean	<0.01
Incumbent leader is serving final term, same	<0.01
Capital stock, PPP, log mean	<0.01
Urban population (% of total population), mean	<0.01
Ease of doing business score, mean	<0.01

Note. The table lists the country-pair characteristics that are significant determinants (permutation importance p-value < 0.01) of PTA formation, ranked from most to least significant. All insignificant variables are omitted.

Table 3: Effect of Most Significant Determinants on the Likelihood of PTA Formation

	PTA exists
Population-weighted distance between most populated cities	−1.11***
	(0.04)
Contagion, mean	3.06***
	(0.20)
Contagion, difference	−2.83***
	(0.19)
Price level of consumption (PPP / exchange rate), log mean	0.90***
	(0.03)
Pair's bilateral trade in manufacturing	−0.34***
	(0.05)
Human capital index (PWT), mean	0.57***
	(0.03)
Pair's bilateral trade in agriculture	−0.12***
	(0.04)
Executive branch is rural, same	0.22***
	(0.05)
N	48,560
Pseudo R^2	0.39

Note. The table presents results of a logistic regression of the presence/absence of a PTA in country-pair-period on the most significant determinants identified by the random forest. The specification includes the "Continent, combination" (and thus implicitly "Continent, same") determinant: the coefficients of combinations are omitted for conciseness. Pseudo R^2 reports Nagelkerke R^2. All coefficients are standardized. Robust standard errors are in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

did not discuss the signs and magnitudes of the effects that these determinants have on the predicted likelihood of PTA formation—neither did we compare the directions of these effects to the existing findings. The reason is that the random forest is effective at predicting and identifying the variables that matter most for its predictive performance, but it cannot provide a single easily interpretable coefficient to summarize the effect each variable has on the prediction. Conventional linear regression yields such coefficients by imposing linearity on the data, which the random forest does not do. To aid interpretation, we conduct an auxiliary linear exercise. Table 3 presents the results of logistic regression of PTA formation on the ten most significant determinants, as selected by the random forest in Table 2. The number of determinants and observations we can include in this linear regression is necessarily limited due to missing values: the more variables are included, the more observations are lost to missing data.

Even though the logistic regression does not allow for complex non-linearities and interactions that the random forest does, it still yields some valuable insights into the role of the most critical determinants of PTA formation. Nations farther apart are not natural trading partners and are thus predicted to be less likely to sign an agreement, echoing the result of Baier and Bergstrand (2004). Manufacturing and agriculture trade is negatively associated with the likelihood of PTA formation. A greater average contagion index, which captures the competitive pressures to liberalize, is likewise positively related to this likelihood, supporting the findings of the contagion literature (Baldwin and Jaimovich 2012). At the same time, a large differential in contagion indices between the two nations is associated with PTA formation being *less* likely. One interpretation of this result is that both nations must be facing competition in each other's export market from countries that have already signed a PTA. If only country i faces pressures to defend its share in j's market but j does not face the same in reverse, j may see no incentive to enter into an agreement with i. Similarity in the executive branch is also associated with a greater likelihood of successfully reaching an agreement.

Most Likely New PTAs—The formation random forest can be used not only for identifying the most important determinants of PTA formation but also for predicting the most likely PTAs to be formed. We use the formation random forest to predict the presence or absence of a PTA between country pairs that do not have a recorded trade agreement in the EIA dataset in the last five-year period in the data (2015 to 2020). Table 4 displays the ten country pairs with the highest probability of forming a PTA, according to the random forest.

Table 4: Most Likely New PTAs

	Country Pair	PTA Probability
1	Dominican Republic, Panama	0.47
2	Colombia, Costa Rica	0.46
3	Bosnia & Herzegovina, Slovenia	0.45
4	Colombia, Dominican Republic	0.45
5	Norway, Russia	0.45
6	Albania, Greece	0.45
7	Ecuador, Panama	0.44
8	Australia, Germany	0.43
9	Albania, Spain	0.42
10	Austria, Bosnia & Herzegovina	0.42

Note. The table lists the ten country pairs that do not currently have a PTA (as recorded in the EIA dataset) but are most likely to have one as predicted by the PTA formation random forest.

Most country pairs in this list are actually making progress toward having a PTA, supporting the validity of the model's predictions. Colombia and Costa Rica (second-highest probability of PTA formation) have, in fact, already entered into a free trade agreement that went into force in 2016: it simply has not been recorded in EIA yet. Two other entries on the list (third and tenth) involve Bosnia and Herzegovina as well as two European Union members: Bosnia and Herzegovina applied for EU membership in 2016 and has been a candidate for accession since 2022. Similarly, the sixth and ninth entries involve Albania and two EU members: Albania is likewise a candidate for EU accession and is in active negotiations as of the writing of this paper. Colombia and the Dominican Republic are another entry on the list. Colombia has an agreement on trade cooperation with the Caribbean Community (CAPRICOM): while the Dominican Republic is not a member of CAPRICOM, it applied for associate membership in 2024. Ecuador and Panama also appear on the list: they are currently negotiating a partial scope agreement, which started in 2021. Finally, Australia and Germany appear on the list: Australia and the EU have been negotiating a free trade agreement since 2018. Of the ten most likely country pairs to form a PTA as predicted by the random forest, only two (Dominican Republic & Panama, Norway & Russia) are not currently making progress toward one.

5.2 Determinants of PTA Design

While random forests yield helpful insights into the determinants of PTA formation, it is even more fruitful to apply them to the determinants of PTA design, which is more complex and diverse. In this section, we present the result of provision-level random forests that seek to predict the inclusion of each classified DESTA provision into PTAs—and to identify the country-pair characteristics most essential for this prediction. We restrict these provision-level random forests to the sample of country pairs that *have* formed an agreement, comparing those that included a particular provision in their PTA to those that didn't.

Performance—First, we evaluate the goodness of fit of the random forest constructed for each provision classified in DESTA. The "overall" column of Table 5 displays the distribution of the misclassification error across provision-level random forests. The median provision's OOB misclassification is only 16.0%, and even the 75th percentile error is less than one quarter: random forests correctly predict the presence or absence of a provision in a given agreement in most cases. At the same time, the overall error masks greater heterogeneity between the two classes in the data: zeroes (provision is absent from the agreement) and ones (provision is present in the agreement). The absence of a provision (the majority class for almost all provisions) is predicted correctly in almost all cases: the median error is just 0.6%. But the presence of a provision (the minority class for most: its median share is 21.2%) is predicted far less accurately: for the median provision, 70.2% of "ones" are misclassified. On the one hand, these figures mean that random forests extract quite a bit of information on the determinants of PTA design from the data, excelling at overall prediction and correctly predicting the presence of provisions in a fair share of cases despite the strictness of the out-of-bag measure. On the other hand, there is plenty of variation in PTA design that the random forests cannot rationalize even when considering almost three hundred observable characteristics of the signatories of each agreement.[13]

Table 5: Out-of-Bag Misclassification in Provision-Level Random Forests

	OOB Misclassification			
	Overall	0 (Absent)	1 (Present)	Share of 1s
25th %-ile	0.120	0.000	0.390	0.148
median	**0.160**	**0.006**	**0.702**	**0.212**
75th %-ile	0.226	0.034	0.948	0.359

Note. The table presents statistics on the predictive performance of provision-level random forests. The misclassification error is the share of country-pair-year observations for which the random forest predicted the presence/absence of a provision incorrectly. It is computed out-of-bag: for each observation, only the trees that did not have this observation in their training sample are used to develop a prediction. Columns "0 (Absent)" and "1 (Present)" show misclassification only for country-pair-year observations that had or did not have the provision in their agreement, respectively.

The quantile classifier (O'Brien and Ishwaran 2019), which we employed for the formation analysis, does improve the balance of performance in the case of provision-level analysis as well, as can be seen in Appendix Table A.3. It does so, however, at the cost of a significant drop in the overall predictive performance and, more importantly, in the forest's ability to discriminate between es-

[13] The overall performance and distribution across classes is similar for the DTA classification of provisions, see Appendix Table A.4.

sential determinants of PTA design (Appendix Figure A.5). Therefore, we rely on the conventional random forests for the headline provision-level analysis below.

Performance Comparison across Models—To better contextualize the performance of the random forest, we compare it to two alternative models: a single classification tree and a logistic regression. Because out-of-bag misclassification can only be defined for a random forest, we estimate all three models on randomly selected $\frac{2}{3}$ of the data and evaluate their performance on the remaining $\frac{1}{3}$.[14] All models are run for each individual DESTA provision. The single classification tree is identical to the component trees estimated within the random forest, except that it uses the entire training data rather than a bootstrap sub-sample. Because the logistic regression cannot deal with missing values, it cannot be run on the same sample as the two tree-based methods, and so we estimate it only on the predictors that have fewer than 20% missing values, leaving us with 56 predictors and 61% complete observations. To make a fair comparison between the methods, we conducted two exercises. In the first, "Full Sample", the tree and the forest are estimated on complete data, and the logistic regression's misclassification measure treats all observations that are not in its 61% sample as misclassified. In the second, "Non-NA Sub-Sample," all three models are estimated using 56 predictors and 61% complete data: the rest is not used in calculating misclassification. Results are presented in Table 6. For both exercises, the random forest outperforms individual trees and logistic regression.

Table 6: Performance Comparison: Random Forest vs Alternative Models

	Full Sample			Non-NA Sub-Sample		
	Random Forest	Tree	Logit	Random Forest	Tree	Logit
25th %-ile	0.123	0.212	0.432	0.055	0.107	0.079
median	**0.164**	**0.277**	**0.456**	**0.079**	**0.145**	**0.114**
75th %-ile	0.227	0.376	0.477	0.115	0.211	0.157

Note. The table presents statistics on the predictive performance of provision-level random forests, individual trees, and logistic regressions. All are estimated on random 2/3 of the data and evaluated on the remaining 1/3. "Full Sample" counts observations that the logistic regression cannot make a prediction for (because of missing values) as misclassified. "Non-NA Sub-Sample" estimates all models only on 56 predictors and observations with no missing values.

[14] For this reason, the random forests' performance metrics in this exercise do not exactly match the metrics of our benchmark forests in Table 5.

Important Determinants of Overall Design—Several country-pair characteristics stand out as essential determinants of overall PTA design. Figure 3 displays the ten factors that are significant determinants of the largest share of provisions as classified in DESTA.[15] Random forests estimated using the quantile classifier or using the DTA classification of provisions display a similar ranking of determinants: see Appendix Figures A.5 and A.7 respectively. We call a variable a significant determinant of a provision if its permutation importance p-value is below 1%. The variable encoding the combination of the trading partners' continents is the most universally important determinant of PTA provisions, significantly predictive of the inclusion of over 90% of provisions classified in DESTA: its top rank is consistent across alternative specifications (Appendix Figures A.5 and A.7). A measure of the distance between the two trading partners is also among the top determinants for the two alternative specifications. Geographic factors have been shown to be important for PTA formation in prior literature (Baier and Bergstrand 2004; Bergstrand, Egger, and Larch 2016) and confirmed by random forests in Section 5.1 are essential for predicting the content of said PTAs as well. The interpretation of these findings, however, is different. Not only do neighbors appear more prone to signing agreements (which is captured by the literature's results on PTA formation), but even conditional on agreeing to sign one, neighbors may face different needs when liberalizing their trade compared to two geographically remote trading partners.

Interdependence of provisions across agreements has some of the most universal predictive power. The presence of a provision in the two trading partners' existing agreements with third nations (captured by the template index) matters for whether the two partners choose to include this provision in an agreement of their own, as does the difference in the partners' exposure to the provision. The contrast in the competitive pressures the two partners face for each other's market from third countries (captured by the difference in the contagion index) is also crucial for predicting whether the provision makes it into their PTA. This result extends the findings of an extensive literature and our results in Section 5.1 on the importance of interdependence for PTA formation (Baldwin 1993; Baldwin and Jaimovich 2012; Baccini and Dür 2012; Egger and Larch 2008; Chen and Joshi 2010), showing that it is highly relevant for understanding differences in

[15] Appendix Figure A.6 plots the distribution of potential determinants over the share of provisions each is a significant determinant of.

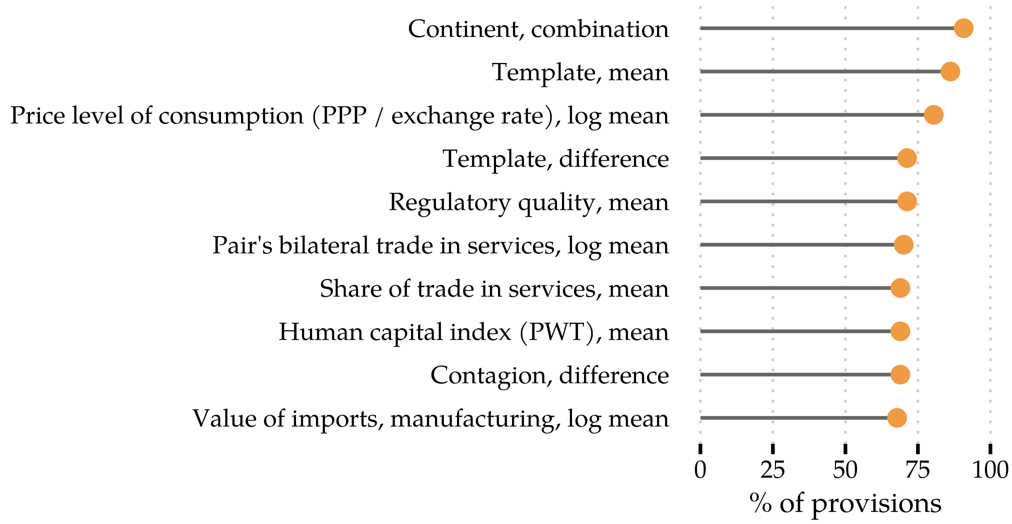


Figure 3: Top Ten Determinants of Provisions

Note. The figure displays the country-pair characteristics ranked by the share of provisions each is a significant determinant of. Only the top ten determinants are displayed. For a given variable, "% of provisions" measures the share of provisions for which the p-value of its variable importance measure is below 1%.

PTA design as well. The average consumption price level in the two countries is another crucial determinant. It speaks to the real exchange rate of the two nations: not between each other, which would be captured by the difference in price levels (which is not significantly influential), but their average real exchange rate with the rest of the world. It extends prior findings that real exchange rate movements (or fear thereof) may limit nations' desire to commit to integration (Fernández-Arias, Panizza, and Stein 2004). The human capital index (as constructed by Penn World Tables) is also highly relevant. The importance of the development level for PTA design is not surprising: what is interesting is that random forests found this development metric the most informative. At the same time, random forests do not find the *difference* in development levels (by any metric) to be particularly relevant for PTA design, in contrast to the existing literature finding it to be essential for PTA formation (Manger 2009; Büthe and Milner 2008; Baccini and Urpelainen 2014).

While average relative Foreign Direct Investment (FDI) inflows do not make the top 10 using the DESTA classification, they are an important determinant of over half of DTA provisions (Appendix Figure A.7). Their significance is particularly stark as FDI flows were not an essential determinant of PTA formation in Section 5.1. This result supports the literature's focus on multinational corporations as one of the primary drivers of the increasing depth of recent agreements, shifting their focus from simple trade liberalization through tariff reduction to FDI liberalization through regulation harmonization (Manger 2009; Baccini, Dür, and Elsig 2018; Gamso and Grosse 2021; Kim et al. 2019). Another highly predictive determinant relates to the average quality of governance of the two nations, as captured by the World Governance Indicators' "Regulatory quality" metric. Predictably, the regulatory apparatus of the trading partners matters for their ability and will to impose international regulation on each other (Polanco Lazo and Sauvé 2018). This observation extends the findings of Gamso and Grosse (2021) and Lechner (2016) on the importance of domestic regulation for PTA provisions sought in negotiations. At the same time, internal political factors—like political competitiveness, leaning of the current government, or the tenure of the current leader—have been previously shown to be essential for PTA formation (Mansfield and Milner 2012; Baccini and Urpelainen 2014) and confirmed by random forests in Section 5.1, but the random forests do not find them to be particularly informative for PTA design.

Another way to compare country characteristics in their ability to predict the content of PTAs is to estimate the random forest with different subsets of the list of potential determinants and compare the attained predictive performance. Table 7 presents the results of this exercise, comparing the out-of-bag misclassification for the median provision across several alternative variable sets. Estimating a random forest while excluding the top ten determinants listed in Figure 3 raises the misclassification rate from 16% (first column) to 19% (second column): the top determinants are important but not the only potent predictors of differences in PTA design. Next, we split all potential determinants into six topic areas: economy, geography, history & culture, interdependence, politics, and trade. Estimating a set of random forests using only the potential determinants from one area at a time, we find that interdependence (contagion and template indices) and geography produce the highest predictive performance. These two areas outperformed the benchmark set of random forests that included all variables. Politics is the next most predictive area. Trade and economic variables are further down the ranking. History & culture offer the least explanatory power of all topic areas. This ranking broadly mirrors the insights from the permutation variable importance measure summarized in Figure 3.

Important Determinants of Provisions by Policy Area—PTA provisions cover many trade issues. What is essential for one area may be irrelevant for another. DESTA groups provisions by policy area

Table 7: Performance Comparison: Alternative Sets of Potential Determinants

All	Excluding Top 10	Economy	Geography	History, Culture	Interdependence	Politics	Trade
0.160	0.186	0.191	0.148	0.210	0.137	0.179	0.187

Note. The table presents the OOB misclassification for the median provision across several alternative variable sets. "All" is identical to the benchmark specification in Table 5. "Excluding Top 10" omits the ten most important determinants displayed in Figure 3. The remaining sets split all variables into six topics and include only predictors related to one topic at a time.

they fit in. We repeat the exercise of identifying the most important determinants within these areas. Table 8 lists up to three most important determinants for each provision area, discarding determinants that appear in the overall ranking in Figure 3. Thus, we obtain a list of *uniquely* important determinants by policy area that are highly predictive for understanding PTA design within this area but few others.

Table 8: Uniquely Important Determinants of Each Provision Area

Determinant	Share of provisions
Capital Movement And Exchange Rates	
Cost of business start-up procedures (% of GNI per capita), mean	4/6
Government effectiveness, mean	4/6
Human capital index (WDI), mean	4/6
Competition	
Foreign direct investment, net inflows (% of GDP), mean	4/5
Human capital index (WDI), mean	4/5
Voice and accountability, mean	4/5
Dispute Settlement	
Value of imports, agriculture, log mean	13/16
Urban population (% of total population), mean	12/16
Real consumption, PPP, log mean	12/16
Intellectual Property Rights	
Contagion, mean	4/5
Population-weighted distance between most populated cities	4/5
Inflation, consumer prices (annual %), mean	3/5
Investments	
Frontier technology readiness index, mean	11/12
Contagion, mean	10/12
Population-weighted distance between most populated cities	10/12
Public Procurement	
Ease of doing business score, mean	2/2
Rule of law, mean	2/2

Voice and accountability, mean	2/2
Regulatory Co-Operation And Transparency	
Business extent of disclosure index, mean	2/2
CPIA structural policies cluster average, difference	2/2
Control of corruption, mean	2/2
Services	
Human capital index (WDI), mean	14/15
Statistical performance indicators, mean	14/15
Value of imports, services, log mean	11/15
Technical Barriers To Trade	
Voice and accountability, mean	4/5
Population-weighted distance between most populated cities	4/5
Value of imports, manufacturing, log diff.	4/5
Temporary Entry Of Business Persons	
Business extent of disclosure index, mean	5/5
Control of corruption, mean	5/5
Energy imports, net (% of energy use), mean	5/5
Trade Defense Instruments	
Pair's bilateral trade in manufacturing, log mean	8/8
Value of imports, services, log mean	7/8
Real GDP, PPP, log mean	7/8

Note. The table presents determinants ranked by the number of provisions each is significant for, by provision area. The top 10 determinants by the overall number of provisions they are important for are excluded, leaving variables that are uniquely important for each area. Provision areas are presented as categorized in DESTA, dropping areas with one or no provisions in our sample. For a given variable, "Share of provisions" counts the provisions in each area for which the p-value of its variable importance measure is below 1%. Within each area, variables are ranked by this share and only the top 3 are displayed, or fewer if the area lacks unique significant determinants.

Measures of domestic administrative quality are even more relevant for some areas. Several metrics from the World Governance Indicators ("Government effectiveness", "Voice and Accountability", etc.), the Database of Political Institutions ("Checks and balances") or the World Bank's Doing Business survey ("Ease of doing business", "Cost of business start-up procedures", "Business extent of disclosure index") emerge as some of the top predictors of provisions on *capital movement*, *competition*, *public procurement*, *regulatory co-operation and transparency*, *technical barriers to trade*, and *temporary entry of business persons*. Additional measures of trade matter for some areas: services trade is a strong predictor of provisions in the "services" and "trade defense instruments" areas, while agricultural trade is important for the "dispute settlement" area. Provisions in the "investments" area are predicted by the frontier technology readiness of both partners.

5.3 Important Determinants of Provisions That Are Most Impactful for Trade

The exercises above consider all provisions classified in DESTA (conditional on each provision having sufficient variation). Not all provisions matter equally for trade outcomes, and those that do may be of particular interest to researchers and policymakers. Breinlich et al. (2021) use a version of the lasso regularized regression to identify a list of DTA provisions with the largest effects on trade outcomes. In this section, we explore the important determinants of the provisions they identified. We use the seven provisions the plug-in PPML-lasso regression produces as in Table 5 of Breinlich et al. (2021).[16] Table 9 lists the seventeen country-pair characteristics that are important determinants of the inclusion of all seven provisions that are impactful for trade outcomes. Overall, factors important for overall PTA design (Section 5.2)—measures of geographic proximity, regulatory quality, and contagion—emerge as leading determinants of including trade-relevant provisions. Average energy use, energy imports, and natural resource rents are also highly relevant for these selected provisions, analogously to the *services* and *competition* policy areas.

To help interpret *how* the identified significant determinants of all trade-relevant provisions affect their inclusion, we conduct an auxiliary linear regression exercise akin to that done in Section 5.1 for the question of PTA formation. Column "Count" of Table 9 presents a Poisson regression of the number of trade-relevant provisions included in a PTA on their essential determinants identified by Random Forests. For a more fine-grained look, we apply a hurdle model that tends to fit count data better by splitting the problem into two: first, it uses a logistic regression to predict whether the count is above zero or not; second, it uses a Poisson regression to predict the value of the count provided it is non-zero (Lambert 1992; Mullahy 1986; Feng 2021). In our application, the first stage, in column "Any" of Table 9, captures the determinants of whether *any* trade-relevant provisions were included. The second stage, in column "Count (≥ 1)", captures the determinants of *how many* trade-relevant provisions were included in PTAs that have at least some.

[16] The method they use identifies eight provisions, but two are perfectly collinear, and thus we drop one. Furthermore, while lasso identifies these seven provisions as sufficient to explain the trade effects of PTAs, post-lasso estimation shows that only three have a significantly non-zero effect on trade. Finally, they also developed the iceberg lasso method. It identifies a greater number of impactful provisions, but these are highly collinear with the eight identified by the plug-in PPML-Lasso.

The exercise yields fewer interpretable insights than its analog in Section 5.1. "Regulatory quality" is negatively associated with the propensity of two nations to have *any* trade-relevant provisions in their agreement but positively associated with the *number* of these provisions. Greater distance reduces the expected number of trade-relevant provisions as long as there are some. Most puzzlingly, the estimated effects of the aggregates of the contagion index have unexpected signs for the "any" specification. Greater contagion pressures shared by both nations (captured by "Contagion, mean") appear to harm a PTA's chances of having any trade-relevant provisions, and the differential in contagion pressures between the two nations (captured by "Contagion, difference") appears to boost them. Both are counter to theoretical mechanisms developed in the literature on contagion cited above and our results on PTA formation in Section 5.1. These puzzling results underscore that random forests can pick up complex non-linear and interacting relationships between potential determinants and PTA design differences that are not easily summarized with a linear approximation. Furthermore, the random forests could use all of the available data thanks to the on-the-fly-imputation method, while the logistic regression loses most observations due to the disparate sets of missing values in the seventeen selected variables.

6 Conclusion

Modern preferential trade agreements are increasingly complex and diverse. Existing research has studied many factors that determine PTA formation and design. At the same time, those studies consider only a limited number of determinants at once and reduce PTA design to the presence of a specific provision or a single index of PTA depth. Machine learning techniques provide one avenue to overcome these limitations and compare many potential determinants of PTA design while preserving the high dimensionality of design differences. We apply random forests to select the country-pair characteristics with the highest predictive power, including various provisions in PTAs, thus identifying the critical determinants of PTA design. Several categories of determinants emerge as essential for most provisions. These determinants are the interdependence of PTA features across trading partners, measures of geographic proximity, regulatory quality, and FDI flows. These findings point to several of the mechanisms of PTA formation or design that the literature has studied (Baldwin and Jaimovich 2012; Gamso and Grosse 2021; Mansfield and Milner

Table 9: Effects of Most Important Determinants on the Presence and Number of PTA Provisions Most Impactful for Trade

		Hurdle Model	
	Count	Any	Count (≥ 1)
Arable land (% of land area), mean	0.01	0.53	0.00
	(0.01)	(0.32)	(0.00)
Average annual hours worked by persons engaged, log mean	0.01	0.18	0.01*
	(0.01)	(0.71)	(0.01)
Contagion, difference	−0.01	26.86**	0.00
	(0.01)	(10.89)	(0.00)
Contagion, mean	0.01*	−38.79**	0.00
	(0.00)	(15.89)	(0.00)
Energy imports, net (% of energy use), mean	0.13**	−0.76	0.00
	(0.05)	(0.57)	(0.03)
Energy use (kg of oil equivalent per capita), log diff.	−0.03**	−0.23	−0.02***
	(0.01)	(0.45)	(0.01)
Energy use (kg of oil equivalent per capita), log mean	0.00	0.58	0.02
	(0.03)	(0.69)	(0.02)
Human capital index (PWT), mean	0.05*	−1.16	−0.01
	(0.03)	(1.04)	(0.01)
Political stability and absence of violence/terrorism, mean	0.04**	−1.76***	0.00
	(0.01)	(0.63)	(0.01)
Population-weighted distance between most populated cities	0.07	0.50	−0.01
	(0.06)	(0.92)	(0.03)
Regulatory quality, difference	0.00	−0.43	0.00
	(0.01)	(0.36)	(0.01)
Regulatory quality, mean	−0.02	2.30**	0.04**
	(0.03)	(1.05)	(0.02)
Template, mean	0.00	2.49***	0.00
	(0.01)	(0.94)	(0.01)
Total natural resources rents (% of GDP), difference	−0.37***	2.73***	−0.06
	(0.07)	(0.98)	(0.06)
Total natural resources rents (% of GDP), mean	0.41***	−3.38**	0.06
	(0.10)	(1.61)	(0.08)
Voice and accountability, mean	0.08	−2.43**	−0.02
	(0.05)	(0.99)	(0.03)
N	1,161	1,161	1,056
Pseudo R^2	0.93	0.71	0.97

Note. The table presents results of regressions of the number and presence of most impactful provisions identified by Breinlich et al. (2021) on the most important determinants identified in this section. Column "Count": Poisson regression on the number of selected provisions in a PTA. Column "Any": logistic regression on whether any selected provisions are present in a PTA. Column "Count (≥ 1)": Poisson regression on the count of selected provisions, conditional on it being non-zero. All three regressions include the "Continent, combination" determinant: coefficients of its combinations are omitted for conciseness. Pseudo R^2 reports Nagelkerke R^2. All columns report standardized coefficients, making magnitudes comparable within columns (but not across since different models are used). Robust standard errors are in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

2012). At the same time, many other determinants linked to previously studied mechanisms are not identified by random forests to be particularly informative of PTA design differences, particularly those related to internal political situation and economic differences across partners (e.g. Baccini and Urpelainen 2014; Bergstrand, Egger, and Larch 2016). This difference suggests that such factors could be relevant for PTA formation but less critical for the contents that negotiators settle on.

There are several directions for future research that can build on our methodology and findings. The first direction is studying the determinants of PTA formation and design while retaining the high dimensionality of PTA membership, which could yield important insights into multilateral agreements. Our approach represented multilateral PTAs as sets of bilateral agreements to facilitate empirical analysis with random forests at the cost of ignoring multilateral considerations like the size and composition of a PTA's membership. One alternative approach is to conduct OLS or RF analysis at the multilateral agreement level, using various moments of each characteristic across members as predictors. Another is to use machine learning algorithms that support variable unordered feature sets (reflecting variable PTA membership sizes), like CDANs (Gardner, Elhami, and Selmic 2019). The second research direction should investigate how commitments within PTAs surpass those established by the WTO. This is particularly pertinent in light of recent developments and initiatives within the WTO, such as the ongoing negotiations on the Joint Statement Initiative on Electronic Commerce (ongoing), the operational Multi-Party Interim Appeal Arbitration Arrangement (2020), the concluded negotiations of the Services Domestic Regulation initiative (2021), and the Investment Facilitation for Development initiative (2023). Understanding these dynamics will offer insightful perspectives on the evolving landscape of international trade agreements. Lastly, the third direction for future research is to delve deeper into the individual mechanisms identified by random forests as important for PTA formation and design. Random forests excel at picking up all kinds of non-linearities and interactions between variables in the data but do so at the cost of the interpretability of individual mechanisms. Although they highlight the country-pair characteristics that are most relevant to PTA design, they leave it for conventional econometric methods to tease out *how* or *why* each characteristic is related to PTA design. Thus, our results motivate further research focusing on individual mechanisms that relate to essential determinants suggested by the random forests.

References

Altmann, André, Laura Toloşi, Oliver Sander, and Thomas Lengauer. 2010. "Permutation Importance: A Corrected Feature Importance Measure." *Bioinformatics* 26, no. 10 (15, 2010): 1340–1347.

Baccini, Leonardo. 2019. "The Economics and Politics of Preferential Trade Agreements." *Annual Review of Political Science* 22, no. 1 (11, 2019): 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, Andreas Dür, and Manfred Elsig. 2018. "Intra-Industry Trade, Global Value Chains, and Preferential Tariff Liberalization." *International Studies Quarterly* 62, no. 2 (1, 2018): 329–340.

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

———. 2015. *Cutting the Gordian Knot of Economic Reform: When and How International Institutions Help.* New York, NY: Oxford University Press.

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

Baldwin, Richard. 1993. *A Domino Theory of Regionalism.* w4465. Cambridge, MA: National Bureau of Economic Research, September.

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

Baltagi, Badi H., Peter Egger, and Michael Pfaffermayr. 2008. "Estimating Regional Trade Agreement Effects on FDI in an Interdependent World." *Journal of Econometrics* 145 (1-2): 194–208.

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, et al. 2021. "Hyperparameter Optimization: Foundations, Algorithms, Best Practices and Open Challenges." Preprint, November 24, 2021. Accessed March 25, 2023. arXiv: `2107.05847 [cs, stat]`. 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, J.M.C. Santos Silva, and Tom Zylkin. 2021. *Machine Learning in International Trade Research: Evaluating the Impact of Trade Agreements.* Policy Research Working Paper 9629. The World Bank.

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.

Caruana, Rich, and Alexandru Niculescu-Mizil. 2006. "An Empirical Comparison of Supervised Learning Algorithms." In *Proceedings of the 23rd International Conference on Machine Learning - ICML '06,* 161–168. The 23rd International Conference. Pittsburgh, Pennsylvania: ACM Press.

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 N°2022-05. July.

Cruz, Cesi, Philip Keefer, and Carlos Scartascini. 2021. *The Database of Political Institutions 2020 (DPI2020).* Inter-American Development Bank, January.

Dür, Andreas, Leonardo Baccini, and Manfred Elsig. 2014. "The Design of International Trade Agreements: Introducing a New Dataset." *The Review of International Organizations* 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, no. 640 (8, 2021): 3207–3240.

Feenstra, Robert C., Robert Inklaar, and Marcel P. Timmer. 2015. "The Next Generation of the Penn World Table." *American Economic Review* 105, no. 10 (1, 2015): 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.

Fernández-Arias, Eduardo, Ugo Panizza, and Ernesto Stein. 2004. "Trade Agreements, Exchange Rate Disagreements." In *Monetary Unions and Hard Pegs,* 1st ed., edited by Volbert Alexander, George M. Von Furstenberg, and Jacques Mélitz, 135–150. Oxford University PressOxford, March 25, 2004.

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.

Gardner, Andrew, Neshat Elhami, and Rastko R. Selmic. 2019. "Classifying Unordered Feature Sets with Convolutional Deep Averaging Networks." In *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC),* 3447–3453. 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC). Bari, Italy: IEEE, October.

Gaulier, Guillaume, and Soledad Zignago. 2010. *BACI: International Trade Database at the Product-Level (the 1994-2007 Version).* CEPII Working Paper N°2010-23.

Gregorutti, Baptiste, Bertrand Michel, and Philippe Saint-Pierre. 2017. "Correlation and Variable Importance in Random Forests." *Statistics and Computing* 27 (3): 659–678.

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

Hafner-Burton, Emilie M. 2013. *Forced to Be Good: Why Trade Agreements Boost Human Rights.* Ithaca, NY: Cornell University Press, January 1, 2013.

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

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

Ishwaran, Hemant. 2015. "The Effect of Splitting on Random Forests." *Machine Learning* 99 (1): 75–118.

Ishwaran, Hemant, Udaya B. Kogalur, Eugene H. Blackstone, and Michael S. Lauer. 2008. "Random Survival Forests." *The Annals of Applied Statistics* 2, no. 3 (1, 2008).

Janitza, Silke, Ender Celik, and Anne-Laure Boulesteix. 2018. "A Computationally Fast Variable Importance Test for Random Forests for High-Dimensional Data." *Advances in Data Analysis and Classification* 12 (4): 885–915.

Kaufmann, Daniel, Aart Kraay, and Massimo Mastruzzi. 2010. *The Worldwide Governance Indicators.* World Bank Policy Research Working Paper No. 5430. The World Bank.

Kim, Dongin, and Sandro Steinbach. 2023. *Preferential Trading in Agriculture: New Insights from a Structural Gravity Analysis and Machine Learning.* Working paper.

Kim, In Song, Helen V Milner, Thomas Bernauer, Iain Osgood, Gabriele Spilker, and Dustin Tingley. 2019. "Firms and Global Value Chains: Identifying Firms' Multidimensional Trade Preferences." *International Studies Quarterly* 63, no. 1 (1, 2019): 153–167.

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, no. 5 (2, 2016): 840–871.

Limão, N. 2016. "Preferential Trade Agreements." In *Handbook of Commercial Policy,* 1:279–367. Elsevier.

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

Liu, Xuepeng, and Emanuel Ornelas. 2014. "Free Trade Agreements and the Consolidation of Democracy." *American Economic Journal: Macroeconomics* 6, no. 2 (1, 2014): 29–70.

Manger, Mark. 2009. *Investing in Protection: The Politics of Preferential Trade Agreements between North and South.* 1st ed. Cambridge University Press, September 3, 2009.

———. 2015. "PTA Design, Tariffs and Intra-Industry Trade." In *Trade Cooperation,* 1st ed., edited by Andreas Dür and Manfred Elsig, 195–217. Cambridge University Press, January 8, 2015.

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

Mattoo, Aaditya, Alen Mulabdic, and Michele Ruta. 2022. "Trade Creation and Trade Diversion in Deep Agreements." *Canadian Journal of Economics/Revue canadienne d'économique* 55 (3): 1598–1637.

Mattoo, Aaditya, Nadia Rocha, and Michele Ruta. 2020. "The Evolution of Deep Trade Agreements." In *Handbook of Deep Trade Agreements.* The World Bank, September 23, 2020.

Mayer, Thierry, and Soledad Zignago. 2011. *Notes on CEPII's Distances Measures: The GeoDist Database.* CEPII Working Paper N°2011-25.

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

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, no. 2 (1, 2017): 87–106.

Nembrini, Stefano, Inke R König, and Marvin N Wright. 2018. "The Revival of the Gini Importance?" Edited by Alfonso Valencia. *Bioinformatics* 34, no. 21 (1, 2018): 3711–3718.

NSF-Kellogg Institute. 2021. "Database on Economic Integration Agreements." https://kellogg.nd.edu/nsf-kellogg-institute-data-base-economic-integration-agreements.

O'Brien, Robert, and Hemant Ishwaran. 2019. "A Random Forests Quantile Classifier for Class Imbalanced Data." *Pattern Recognition* 90:232–249.

Orefice, Gianluca, and Nadia Rocha. 2014. "Deep Integration and Production Networks: An Empirical Analysis." *The World Economy* 37 (1): 106–136.

Osnago, Alberto, Nadia Rocha, and Michele Ruta. 2017. "Do Deep Trade Agreements Boost Vertical FDI?" *The World Bank Economic Review* 30 (Supplement 1): S119–S125.

Polanco Lazo, Rodrigo, and Pierre Sauvé. 2018. "The Treatment of Regulatory Convergence in Preferential Trade Agreements." *World Trade Review* 17 (4): 575–607.

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." *The Review of International Organizations* 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.

Scheffer, Judi. 2002. "Dealing with Missing Data." *Research Letters in the Information and Mathematical Sciences,* no. 3, 153–160.

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.

Scornet, Erwan. 2016. "On the Asymptotics of Random Forests." *Journal of Multivariate Analysis* 146:72–83.

Scornet, Erwan, Gérard Biau, and Jean-Philippe Vert. 2015. "Consistency of Random Forests." *The Annals of Statistics* 43, no. 4 (1, 2015).

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.

Tang, Cheng, Damien Garreau, and Ulrike von Luxburg. 2018. "When Do Random Forests Fail?" *Advances in Neural Information Processing Systems* 31.

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. 2023a. "WITS." https://wits.worldbank.org/.

———. 2023b. "World Development Indicators." https://data.worldbank.org/.

United Nations. 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, no. 2 (1, 2014): 3–28.

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

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.

Appendix Figures and Tables

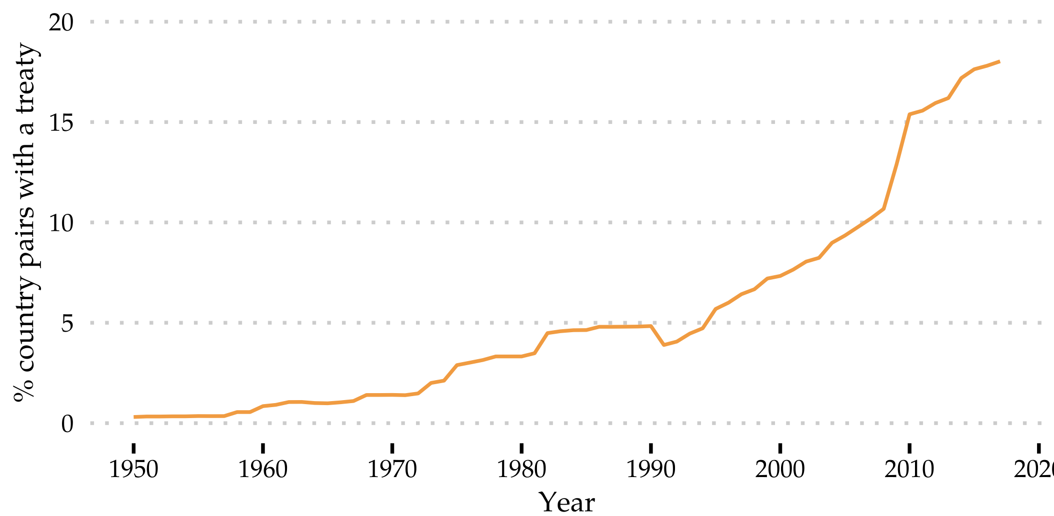


Figure A.1: Number of Country Pairs with an Active Agreement

Note. The figure displays the percentage share of all country pairs that were members of the same PTA (as classified in EIA) in a given year.

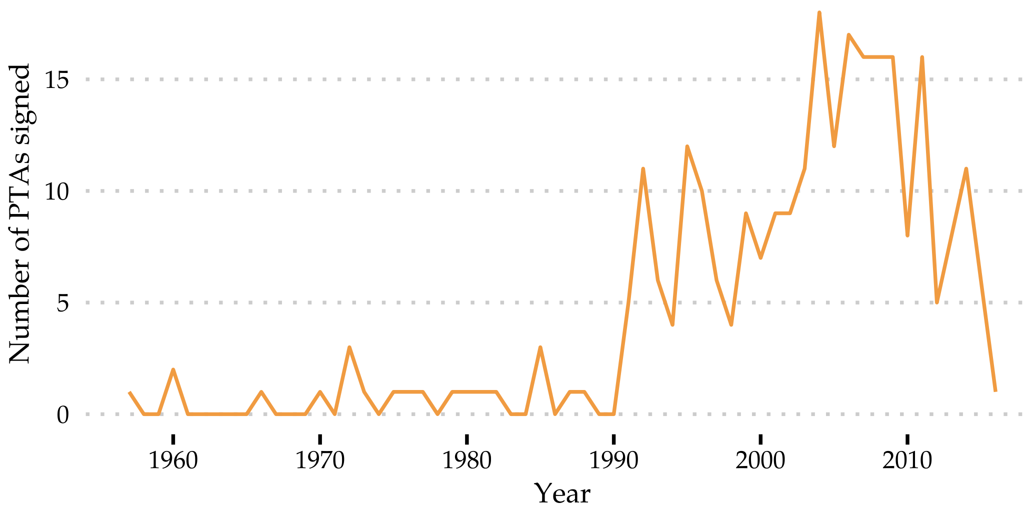


Figure A.2: Number of PTAs Signed per Year (Data: DTA)

Note. The figure is the analog of Figure 1b, using PTAs as classified in DTA rather than DESTA. The figure displays the number of new PTAs signed per year.

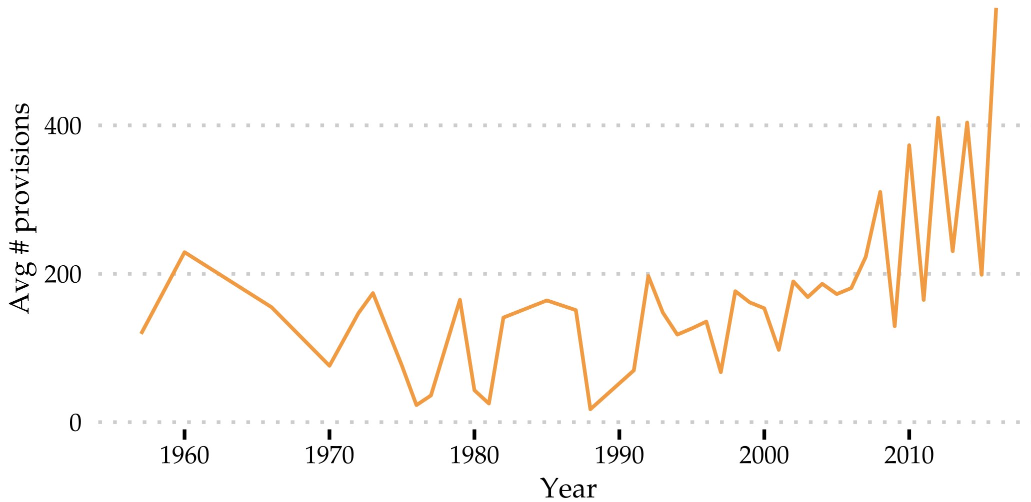


Figure A.3: Average Number of Provisions in a PTA over Time (Data: DTA)

Note. The figure is the analog of Figure 1c, using PTAs and provisions as classified in DTA rather than DESTA. The figure displays the average number of provisions in newly signed PTAs by year.

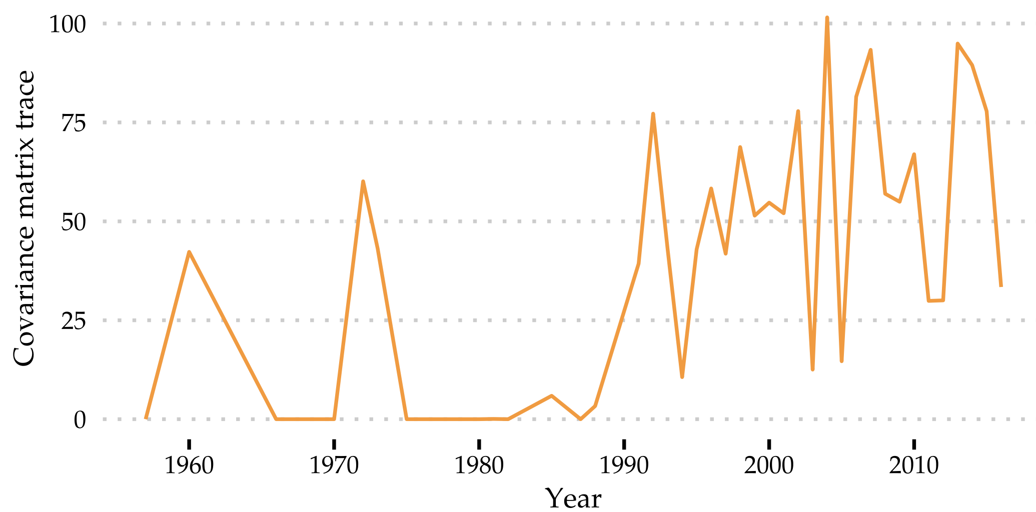


Figure A.4: Variance of Provisions Included in PTAs over Time (Data: DTA)

Note. The figure is the analog of Figure 1d, using PTAs and provisions as classified in DTA rather than DESTA. Variance is measured as the trace of the covariance matrix of vectors indicating the inclusion of provisions in each PTA, which is equivalent to the sum of variances of the binary provision inclusion variable across provisions.

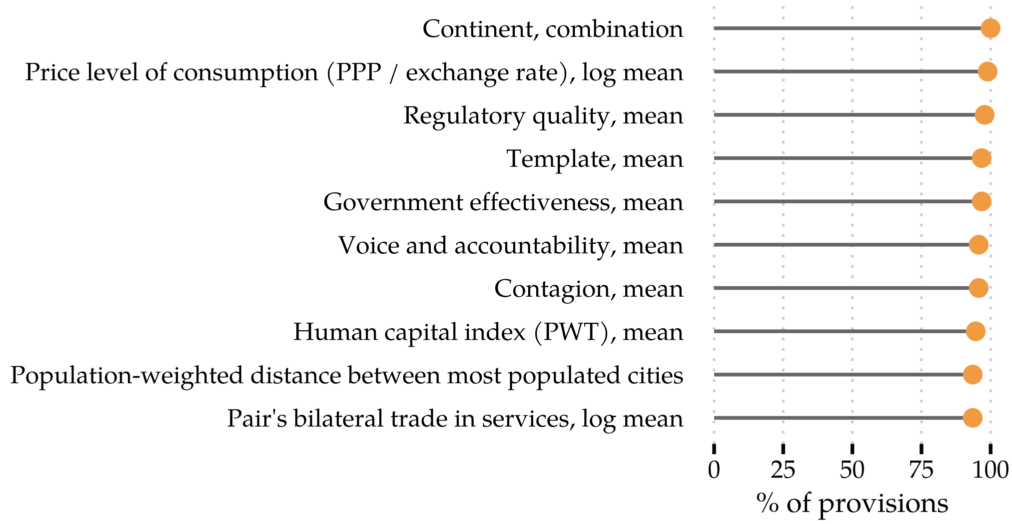


Figure A.5: Top Determinants of Provisions, Imbalanced Random Forests

Note. The figure displays the country-pair characteristics ranked by the share of provisions each is a significant determinant of, using the imbalanced classifier random forests. For a given variable, "% of provisions" measures the share of provisions for which the p-value of its variable importance measure is below 1%.

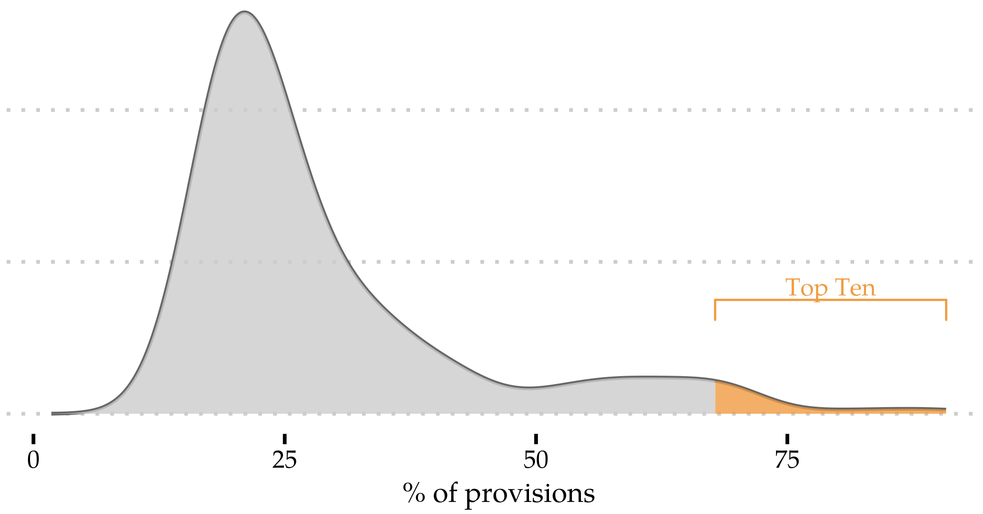


Figure A.6: Distribution of Determinants

Note. The figure displays the distribution of country-pair characteristics over the share of provisions each is a significant determinant of, using the DESTA classification. For a given variable, "% of provisions" measures the share of provisions for which the p-value of its variable importance measure is below 1%. The top ten determinants in Figure 3 are highlighted.

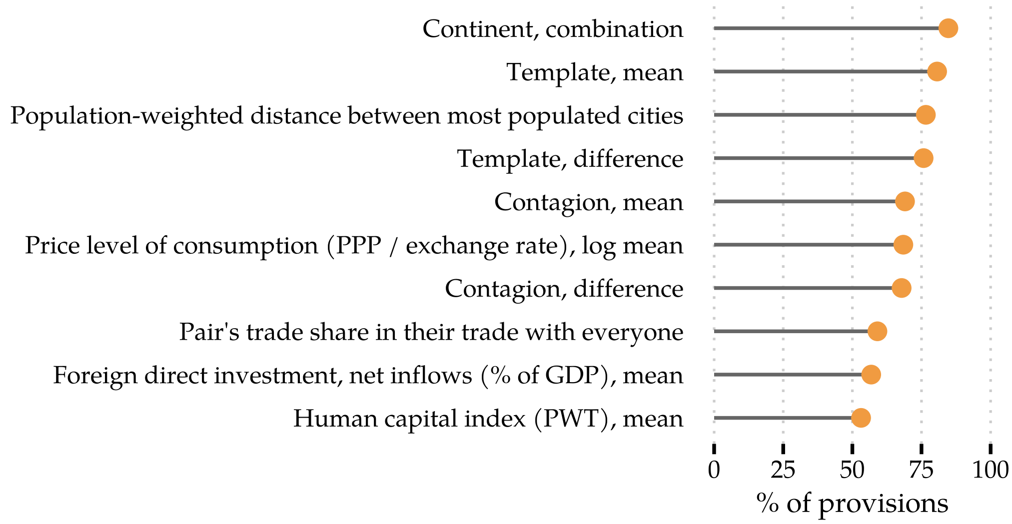


Figure A.7: Top Ten Determinants of Provisions, DTA

Note. The figure displays the country-pair characteristics ranked by the share of provisions each is a significant determinant of, using the DTA classification. For a given variable, "% of provisions" measures the share of provisions for which the p-value of its variable importance measure is below 1%.

Table A.1: List of Potential Determinants Used in Random Forests, Grouped by Data Source

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	
EU member	same
GATT member	same
WTO 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	
CEPII Language	
Pair's linguistic proximity (LP1)	
Pair's linguistic proximity (LP2)	
Composite	
Contagion	mean, difference
Template	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	
TFP level, PPP	log mean, log diff.
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 (PWT)	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
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	
CPIA building human resources 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
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
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 (WDI)	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. Variables are grouped by the source the measure was taken from or whose data it's constructed from. The "aggregators" column lists ways in which country-year variables were aggregated between the two countries to form country-pair-year variables. Options are: mean (in levels or in logs) and difference (level or log) for numeric variables, same (a binary indicator for whether the two countries' values match) and combination (a category for each possible combination of values) for categorical variables. Bilateral variables measured at the country-pair-level originally do not require such aggregation, and so "aggregators" blank for them.

Table A.2: Out-of-Bag Misclassification in Regular Formation Random Forest

OOB Misclassification			
Overall	0 (Absent)	1 (Present)	Share of 1s
0.031	0.000	0.892	0.035

Note. The table presents statistics on the predictive performance of the random forest predicting PTA formation. The misclassification error is the share of country-pair-period observations for which the random forest predicted the presence/absence of a PTA incorrectly. It is computed out-of-bag: for each observation, only the trees that did not have this observation in their training sample are used to develop a prediction. Columns "0 (Absent)" and "1 (Present)" show misclassification only for country-pair-period observations that had or did not have an agreement within the period, respectively.

Table A.3: Out-of-Bag Misclassification in Provision-Level Imbalanced Random Forests

	OOB Misclassification			
	Overall	0 (Absent)	1 (Present)	Share of 1s
25th %-ile	0.197	0.212	0.055	0.161
median	**0.252**	**0.282**	**0.115**	**0.229**
75th %-ile	0.311	0.342	0.178	0.402

Note. The table presents statistics on the predictive performance of provision-level random forests using the quantile classifier. The misclassification error is the share of country-pair-year observations for which the random forest predicted the presence/absence of a provision incorrectly. It is computed out-of-bag: for each observation, only the trees that did not have this observation in their training sample are used to develop a prediction. Columns "0 (Absent)" and "1 (Present)" show misclassification only for country-pair-year observations that had or did not have the provision in their agreement, respectively.

Table A.4: Out-of-Bag Misclassification in Provision-Level Random Forests, DTA

	OOB Misclassification			
	Overall	0 (Absent)	1 (Present)	Share of 1s
25th %-ile	0.107	0.000	0.267	0.152
median	**0.143**	**0.001**	**0.705**	**0.212**
75th %-ile	0.190	0.038	0.906	0.379

Note. The table presents statistics on the predictive performance of provision-level random forests using DTA classification. The misclassification error is the share of country-pair-year observations for which the random forest predicted the presence/absence of a provision incorrectly. It is computed out-of-bag: for each observation, only the trees that did not have this observation in their training sample are used to develop a prediction. Columns "0 (Absent)" and "1 (Present)" show misclassification only for country-pair-year observations that had or did not have the provision in their agreement, respectively.