

Exam of Microeconometrics

# **The effectiveness of environmental provisions in regional trade agreements**

Replication of paper by Abman, R., Lundberg, C., and  
Ruta, M. (2024)

Giovanni Parrillo - 642728  
Giovanni Stivella - 642210

21 January 2026

## Introduction

This paper aims to replicate the findings of Abman, R., Lundberg, C., and Ruta, M. (2024) on the effectiveness of environmental provisions in Regional Trade Agreements (RTAs).

In the original paper, the authors considered different Regional Trade Agreements between Countries or group of Countries and looked at whether the inclusion of environmental provisions aimed at reducing the environmental impact of trade agreements resulted effective.

In particular, they looked at the impact of Regional Trade Agreements on deforestation in Countries involved in the agreements.

The rationale behind the study is that the implementation of RTAs lead to increased trade flows between the Countries involved, with particular effect on the trade of agricultural goods: higher demand for agricultural goods incentives deforestation in order to convert to agriculture land previously devoted to forest.

On the other side, the inclusion of environmental provisions should be aimed at reducing the environmental impact of RTAs, and in particular they should mitigate the effect on deforestation.

When analysing these trends, however, problems of endogeneity arise. In fact, we expect environmental provisions to be implemented by Countries which are more concerned with deforestation or biodiversity loss, possibly because they are more affected by these phenomena.

In order to deal with the problem of endogeneity, the authors use propensity score matching: based on other covariates, they estimate the probability that a Regional Trade Agreement includes an environmental provision; then, they match RTAs with environmental provision with RTAs without environmental provisions but close propensity score.

It is important to note that RTAs are considered as units: that is the default for most of the paper, where variables are analysed at RTA level.

As the number of covariates is quite high, a LASSO estimation is used while computing the propensity score: this methodology allows researchers to perform feature selection.

In the following sections, we will present and briefly comment the results obtained by performing the same analysis as the one conducted by Abman, R., Lundberg, C., and Ruta, M. (2024) on another dataset.

## Propensity Score Estimation

In order to estimate the propensity score, we performed a LASSO regression using as dependent variable a dummy equal to 1 if the RTA includes an environmental provision, and 0 otherwise.

LASSO regression is a method of regression in which a different minimisation problem is used: instead of minimising the sum of squares of residuals, the quantity to be minimised also includes the sum of absolute values of coefficients multiplied by a penalisation term  $\lambda$ :

$$\min_{\beta} \left( \sum_{i=1}^n (y_i - X_i \beta)^2 + \lambda \sum_{j=1}^p |\beta_j| \right)$$

We can think of  $\lambda$  as a penalisation term because it penalises the inclusion of a high number of coefficients: when  $\lambda = 0$ , the LASSO regression coincides with the OLS regression, while with higher values of  $\lambda$ , the inclusion of more coefficients is penalised and therefore the mechanism of feature selection takes place.

In order to choose the best value of penalisation term  $\lambda$ , we adopted the same strategy used in the paper: using the value that minimizes the mean  $k$ -fold cross-validation error.

In fact, the use of penalisation term  $\lambda$  has two effects on the Mean Squared Error in the estimation of the parameters: on one side, it reduces variance by constraining the estimated coefficients; on the other side, it introduces a bias in the estimation. While we can't observe the variance and the bias of the estimator, we can estimate them through cross-validation dividing iteratively the sample in testing and training set. However, we will not explore in detail the methodology as it did not involve any particular intervention by us: using *glmnet* library, we were able to obtain automatically  $\lambda$  that minimises estimated MSE, and use it to get estimated coefficients.

To be more precise, instead of using LASSO regression directly, we used a LASSO logistic regression, as the dependent variable is a dummy variable. Its formulation is similar to the one of LASSO regression:

$$\max_{\beta} \left( \sum_{i=1}^n l(\mathbf{1}[y_i], \mathbf{X}_i \beta | \beta) - \lambda \sum_{j=1}^p |\beta_j| \right)$$

where  $l(\cdot)$  is the log-likelihood function of the logistic regression.

As candidate variables, we used the same variables used in the original paper: the maximum and average biodiversity indices across agreement signatories, the number of tropical signatories and an indicator for whether any signatories are in the tropics, the total forest cover of signatories in the year 2000, the total landarea of signatories, the total percent of land area among signatories covered by forest in 2000, and finally a set of regional categories that include an indicator for any signatories belonging to the region as well as the numbers of members

in the region (e.g., North America, Central and South America); furthermore, we included individual country indicators taking values of unity, if the country is a signatory to the RTA and zero otherwise, which capture country-specific factors, as well as any other country-level factors related to the political economy of trade negotiations. Finally, we included indicators for the presence of different combinations of developed and developing counterparties

After performing the LASSO logistic regression, we obtained the propensity scores for each RTA as the predicted values of the dependent variable using the estimated coefficients. These propensity scores allowed us to move to the actual estimation of the model.

## Main model estimation

The main models estimated in the paper are difference-in-differences models with fixed effects, written as follows:

$$y_{gt} = \beta_1 \mathbf{1}[\text{Post\_RTA}_{gt}] + \beta_2 \mathbf{1}[\text{Post\_RTA}_{gt}] \times \mathbf{1}[\text{Enviro\_RTA}_g] + \alpha_g + \gamma_t + \varepsilon_{gt}$$

where  $y_{gt}$  is the outcome variable of interest,  $\alpha_g$  is the RTA fixed-effect and  $\gamma_t$  is the time fixed-effect.

The variable  $\mathbf{1}[\text{Post\_RTA}_{gt}]$  is a dummy variable equal to 1 if the year  $t$  is after the implementation of the RTA  $g$ , and 0 otherwise. The variable  $\mathbf{1}[\text{Enviro\_RTA}_g]$  is a dummy variable equal to 1 if the RTA  $g$  includes an environmental provision, and 0 otherwise.

Consequently, the coefficient  $\beta_1$  captures the effect of RTAs without environmental provisions on the outcome variable, while the coefficient  $\beta_2$  captures the additional effect of RTAs with environmental provisions. Therefore, the total effect of RTAs with environmental provisions is given by the sum of  $\beta_1$  and  $\beta_2$ .

In order to give causal interpretation to the estimated coefficients, we need that some assumptions hold. First of all, we need the Stable Unit Treatment Value Assumption (SUTVA) to hold: this assumption requires that the potential outcomes for any unit do not vary with the treatments assigned to other units, and that for each unit, there are no different forms or versions of each treatment level, which lead to different potential outcomes.

The authors do not discuss in detail the possibility that different levels of environmental provisions may lead to different potential outcomes, therefore we will assume that all environmental provisions in our dataset have similar effects on the outcome variable.

As for the absence of spillover effects, the authors discuss more in detail this point. The main variable of interest is deforestation, which tends to be mainly driven by agricultural land expansion; RTAs, therefore, usually increase deforestation because signatory Countries experience higher demand for agricultural goods which is addressed by expanding agricultural land. In this context, spillover effects may arise when RTAs displace demand for agricultural good from non-signatory to signatory Countries. However, deforestation is almost irreversible: therefore, while the effect of RTAs in signatory Countries is to increase deforestation, non-signatory Countries will not experience a decrease in deforestation because the land that has already been deforested cannot be reforested.

The second assumption that needs to be analysed is the parallel trends assumption.

## Standard errors corrected with sparse cross-cluster correlations

We develop an extension to the standard approach to clustered covariance matrices. We allow for sparse cross-cluster correlation in the covariance matrix induced by overlapping cluster membership. Leveraging the structural information about the presence and degree of cluster overlap, we develop a covariance matrix that allows for within-country correlations in our RTA-level triple-difference model. Consider the standard matrix representation of equation (2)

$$Y = X\beta + \varepsilon$$

then the covariance matrix of the vector of coefficient estimates  $\hat{\beta}$  is given by

$$\text{Var}[\hat{\beta}] = (X'X)^{-1}\hat{V}(X'X)^{-1}$$

where  $\hat{V}$  is the cluster-robust covariance matrix with sparse cross-cluster correlations, which we define as

$$\hat{V} = \sum_{g=1}^n \sum_{h=1}^n w_{g,h} X_g' \hat{\varepsilon}_g \hat{\varepsilon}_h' X_h,$$

where  $g$  and  $h$  index agreements,  $n$  is the number of agreements in our matched sample, and  $w_{g,h}$  is a weighting function that allows for sparse cross-cluster correlation. Let  $n_g$  and  $n_h$  denote the number of countries that are party to trade agreements  $g$  and  $h$ , respectively, and  $G$  and  $H$  denote the set of parties to each respective agreement. Then, our weight function is given by

$$w_{g,h} = \frac{1}{n_g} \sum_{k \in G} \mathbf{1}[k \in H],$$

where  $\mathbf{1}[k \in G] = 1$  if country  $k$  is a party to agreement  $g$ . If there is no membership overlap between agreements  $g$  and  $h$ ,  $w_{g,h} = 0$ , and we restrict the cross-cluster correlation between cluster  $g$  and  $h$  to be zero. If there is overlap, we allow for cross-cluster correlation but weight it by the degree of overlap, that is, the share of total members to agreement  $g$  that are also members to agreement  $h$ . Note that our cluster-robust covariance matrix with sparse cross-cluster correlations is a generalization of the standard cluster-robust covariance matrix and nests the standard approach—if there is no overlapping agreement membership,  $w_{g,h} = 1$  if  $g = h$  and  $w_{g,h} = 0$  for  $g \neq h$ .

The structural source of cross-cluster correlation in our experimental setting—cluster membership overlap—suggests that these cross-cluster correlations will be positive. Consequently, our approach should yield larger standard errors and, hence, represents a more conservative approach to inference relative to standard cluster-robust methods.

**Model estimation for deforestation**

**Parallel trends**

**Results**

## **Model estimation for other variables**

## **Inclusion of enforcement mechanisms**