
Taylor & Francis
Taylor & Francis Group

Statistics and Public Policy

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/uspp20

# Explaining Central Government’s Tax Revenue Categories through the Bradley-Terry Regression Trunk Model

**Alessio Baldassarre, Antonio D’Ambrosio & Claudio Conversano**

**To cite this article:** Alessio Baldassarre, Antonio D’Ambrosio & Claudio Conversano (2024) Explaining Central Government’s Tax Revenue Categories through the Bradley-Terry Regression Trunk Model, Statistics and Public Policy, 11:1, 2356507, DOI: 10.1080/2330443X.2024.2356507

**To link to this article:** https://doi.org/10.1080/2330443X.2024.2356507

STATISTICS AND PUBLIC POLICY
2024, VOL. 11, NO. 1, 1–15
https://doi.org/10.1080/2330443X.2024.2356507

Taylor & Francis
Taylor & Francis Group

OPEN ACCESS Check for updates

# Explaining Central Government's Tax Revenue Categories through the Bradley-Terry Regression Trunk Model

Alessio Baldassarre[a], Antonio D'Ambrosio[b], and Claudio Conversano[c]

[a]Department of Finance, Italian Ministry of Economy and Finance, Roma, Italy; [b]Department of Economics and Statistics, University of Naples Federico II, Naples, Italy; [c]Department of Business and Economics, University of Cagliari, Cagliari, Italy

**ABSTRACT**

The Bradley-Terry Regression Trunk (BTRT) model combines the log-linear Bradley-Terry model, including subject-specific covariates, with a particular tree-based model, the so-called regression trunk. It aims to consider simultaneously the main effects and the interaction effects of covariates on data expressed as paired comparisons. We apply this model to financial data expressed as rankings and then transformed into paired comparisons. Tax revenues differentiated by category represent the statistical units of the analysis (i.e., taxes on income, social security contributions, taxes on property, and taxes on goods and services). We combine data from OECD, World Bank, and IMF databases for the year 2018 to investigate the effect size of socio-economic covariates and their interaction on the composition of tax revenues for a set of 100 countries worldwide. We also present a comparison with a more established method proposed in tax determinants literature and with two alternative models used for matched pairs. Finally, we discuss the implications of reported results for stakeholders and policymakers.

**ARTICLE HISTORY**
Received March 2022
Accepted May 2024

**KEYWORDS**
Cross-country analysis; Ordinal data; Paired comparisons; Public finance; Rankings; Regression tree; STIMA; Tax determinants

## 1. Introduction

Public finance studies the government's role in the economic system (Gruber 2004), emphasizing the importance of maintaining a balanced budget by managing the gap between public revenues and expenditures. This balance became crucial after the 2007 financial crisis in the United States, which later escalated into a sovereign debt crisis. The objective of public authorities is to achieve a balance between revenues and expenditures to ensure positive outcomes and prevent negative ones (Jain 1989). Examining revenues and expenditures separately shows their effect on a country's Gross Domestic Product (GDP): government spending directly affects GDP, whereas taxation has a contrasting effect. Nonetheless, the complex interaction between tax revenues and public spending with economic factors is not fully understood. Some studies have investigated the relationships between taxation and government expenditure. Manage and Marlow (1986) demonstrated that taxation leads to expenditure at the state level, with this causality becoming bidirectional in the short term. Anderson, Wallace, and Warner (1986) found that government spending leads to increased taxation.

The marked interest in understanding the determinants of tax revenues led to the pivotal question of the sources of variance in tax revenues among countries. Kaldor (1963) observed that tax revenues are lower in underdeveloped countries than in developed ones, a disparity attributed to the capability of populations to pay taxes from their income surplus beyond basic subsistence needs (Boukbech, Bousselhami, and Ezzahid 2019). Emerging economies, therefore, face limitations in converting national income into tax revenues to fulfill community needs without inciting severe social unrest. It follows logically that a country's ability to generate tax resources is positively correlated with its level of development, suggesting that more developed nations have a higher capacity to levy taxes (Brun and Diakite 2016).

Within the discussed context, we present a modeling approach to analyze the relationship between tax revenues and government expenditure, focusing on tax revenue components and the diverse characteristics of countries. We employ the Bradley-Terry Regression Trunk (BTRT) model (Baldassarre et al. 2022) to examine the impact of socio-economic covariates and their interactions on tax revenue composition across 100 countries globally. The most significant covariates are identified using the Bradley-Terry-Luce Lasso method and are used in the BTRT model to group countries based on the BTRT's output. BTRT facilitates the incorporation into the model of country-specific fiscal system heterogeneity and covariate interactions. Notably, the Environmental Performance Index (EPI) is employed as an indicator of a country's development level and economic growth. Our findings reveal that EPI significantly influences tax revenue composition by interacting with government gross debt, health expenditure, and employment rate. The BTRT algorithm selects EPI as the primary splitting covariate. This interaction notably enhances the model's fit, highlighting the critical role of EPI in delineating tax revenue structures.

**CONTACT** Claudio Conversano conversa@unica.it Department of Business and Economics, University of Cagliari, Viale S. Ignazio 17, 09123 Cagliari, Italy.

The article is organized as follows. Section 2 presents previous literature. Section 3 presents the original dataset. Section 4 reports the key features of both the BTRT model and algorithm. Section 5 focuses on variable selection. Empirical results are shown in Section 6. Next, the BTRT model is compared with a more established model in tax determinants literature (Section 6.1), as well as with both the basic log-linear Bradley-Terry model without subject-specific covariates and the Bradley-Terry tree model (Section 6.2). Finally, Section 7 ends the article with final remarks and advices to policymakers.

## 2. Background and Motivations

The relationship between socioeconomic explanatory variables and tax revenues (response variable) has been investigated by analyzing cross-sectional or panel data. The goal has been to find the determinants of tax revenues. Main findings suggest the most important explanatory variables are the per capita GDP (Gupta 2007; Pessino and Fenochietto 2010), the peculiarities of the industries reproduced in the sectoral composition of GDP (Chelliah 1971; Chelliah, Baas, and Kelly 1975; Tait, Gratz, and Eichengreen 1979; Piancastelli 2001; Karagoz 2013), external factors like the levels of Foreign Direct Investment (FDI), trade (Cassou 1997; Gupta 2007; Bird, Martinez-Vazquez, and Torgler 2008), and public debt (Teera and Hudson 2004), as well as policy makers' choices like exchange rate, inflation rules (a.k.a. Keynes-Oliveira-Tanzi effect) and financial-fiscal policies (Tanzi 1989; Brun and Diakite 2016). Other studies analyze the role of government efficiency and institutional factors like political stability and political and civil rights (Bird, Martinez-Vazquez, and Torgler 2008; Martin and Uribe Teran 2010). On the social side, the impact of the educational level (as a share of public expenditure on education), illiteracy rate, and population growth on tax revenues have been studied (Piancastelli 2001; Wallace and Bahl 2005; Pessino and Fenochietto 2010). Accountability and civil and political rights are also considered determinants of tax revenues (Bird, Martinez-Vazquez, and Torgler 2008; Martin and Uribe Teran 2010). Lastly, corruption, entry regulations, and the rule of law play an important role in explaining differences in tax revenues (Bird, Martinez-Vazquez, and Torgler 2014).

Different methodological approaches are used for similar purposes. Feltenstein and Cyan (2012) apply the dynamic general equilibrium model, whilst other studies use econometric techniques. Lotz and Morss (1967) proposed the first cross-sectional study on international tax ratios: the tax effort concept is introduced and it is demonstrated that per capita income and trade share are determinants of the tax share. Studies analyzing panel data use static fixed and random effect models based on the generalized method of moments for dynamic panels. Apart from these, Pessino and Fenochietto (2010) introduces a panel version of a stochastic tax frontier model.

As for the sample size, Teera and Hudson (2004) and Pessino and Fenochietto (2010) consider large samples of countries selected for the geographical location or the income level. Other studies are based on a restricted sample of countries. Based on the idea that the research for tax determinants may be inadequate if a heterogeneous group of countries is observed, as the composition of tax revenues can be different in low, middle, and high-income countries, Castro and Camarillo (2014) analyze the 34 countries that are part of the Organization for Economic Co-operation and Development (OECD) only and, for them, consider lagged values of the tax revenues over the 10 years 2001–2011.

Our empirical analysis is aimed at finding the tax revenue determinants taking into account the heterogeneity of countries, with a focus on the relationship between tax revenues and government expenditure. The design of the empirical study is peculiar and differs from those previously presented. Following the OECD classification, we study the determinants of tax revenues by decomposing them into four categories: taxes on income, social security contributions—SSC, taxes on property, and taxes on goods. This decomposition-based approach represents one of the major advantages of this analysis: the disaggregation of tax revenues allows the identification of different effects of the same covariate on different tax revenue categories. A single socio-economic characteristic could affect in opposite direction the size of two or more different tax revenue categories. In addition, the tax categories are paired compared on their size, which means that the tax revenue categories are first ordered according to their size and next compared to each other. To quantify for each country the effects of socioeconomic covariates on the size of each specific tax revenues category, the variable tax revenues is re-coded into a set of categorical variables (i.e., tax revenues are transformed into rankings and next into paired comparisons) to apply the Bradley-Terry model for matched pairs (Bradley and Terry 1952) using the log-linear formulation with subject-specific covariates introduced in Dittrich, Hatzinger, and Katzenbeisser (1998).

The use of paired comparisons is reasonable when comparing countries with different fiscal systems. The OECD tax revenues classification defines a common ground where data from different countries are compared. In line with this, paired comparisons work on sizes, and not on numerical values of tax revenues, thus, facilitating the comparability between countries. Most of the studies on tax revenue determinants use panel data since they assume that tax revenues change over time. Since the order of tax revenues by categories is more stable, our study utilizes cross-sectional data and estimates tax determinants through paired comparisons in a cross-country dataset for a single year. We integrate data from OECD, International Monetary Fund (IMF), and the World Bank, and focus on 2018. The heterogeneity among countries is included in the BTRT model also concerning the covariates: their values are expressed as a percentage of GDP. Government expenditure is decomposed following the Classification Of the Functions Of Government (COFOG) to capture the effect of each type of government expenditure on each category of tax revenues.

To model possible interaction effects in addition to the main effects we resort to BTRT. It allows us to define interaction effects when no a-priori hypotheses are available, and produces a small tree, called a trunk, representing a fair compromise between a straightforward interpretation of the interaction effects and an easy-to-read partition of countries based on their socioeconomic characteristics and the order of their tax revenues. Thus, our study focuses on finding interaction effects on tax determinants in a data-driven manner.

Few works in public finance studies focus on paired comparison data. Dittrich, Katzenbeisser, and Reisinger (2000) analyze rank-ordered preference through Bradley-Terry type models with categorical subject-specific covariates. The (complete) rankings are transformed into paired comparisons on which a tree-based log-linear model is estimated. Data partitioning considers model deviance and a forward selection and backward elimination procedure. This approach is well suited for hypothesis-based modeling. However, when no a priori hypotheses are known it requires the arbitrary introduction of higher-order interactions. Instead, Strobl, Wickelmaier, and Zeileis (2011) propose a tree-based classifier where paired comparisons are treated as response variables in a Bradley-Terry tree (BTtree) model. BTtree defines interactions when no a priori hypothesis is known based on a model-based recursive partitioning approach (Zeileis, Hothorn, and Hornik 2008): Splits are selected assessing the instability of the parameters of the basic Bradley-Terry model. The tree returns in each terminal node the preference scales deriving from the object-related parameters. Nevertheless, BTtree lacks of information about how the subject-specific covariates affect the judges' preferences. This semi-parametric model does not return beta coefficients neither for the main effects nor for the interaction effects.

BTRT is preferable to BTtree and Dittrich's model as it accounts for the data-driven definition of interaction effects between subject-specific covariates. It results in a completely parametric model: each covariate that enters the main effect part of the model and each interaction effect has an associated regression parameter. Thus, the interpretation of results is highly facilitated as it derives both from the inspection of the tree and the interpretation of the estimated regression parameters.

## 3. Data

We consider 100 countries and their associated tax revenues by category for the year 2018. To ensure consistency across countries and granularity of tax revenue categories (OECD 2018a), data are taken from the Global Revenue Statistics Database (OECD 2018b), which uses the OECD classification of taxes. Taxes are classified by their base in: income and profits (heading 1000), compulsory Social Security Contribution (SSC, head. 2000), payroll and workforce (head. 3000), property (head. 4000), goods and services (head. 5000), other taxes (head. 6000). All categories are expressed in terms of the level of taxation through the tax-to-GDP ratio, computed by the ratio of the nominal tax revenue of a country and its nominal GDP for the year. This ratio is used in cross-country research studies where the aim is to compare tax levels across countries with different degrees of development.

The numerical variables of tax revenue categories are transformed into rankings assigning values from 1 to 6 to each tax category: the category generating the highest (lowest) tax revenue is assigned a value of 1 (6). Rankings are first transformed into paired comparisons between tax categories and next into consensus ranking, which represents the best compromise between a set of rankings and corresponds to that ranking in the permutation space that maximizes the sum of correlations between itself and all the other rankings. Operationally, consensus rankings maximize the extended correlation coefficient $\tau_x$ (Emond and Mason 2002). The R package `ConsRank` (D'Ambrosio et al. 2019) is used to compute $\tau_x$, and the result obtained for the tax categories is: (goods and services > income and profits > compulsory SSCs > property > workforce = other taxes). The categories taxes on the workforce and other taxes are both ranked in the last position. In the original data, in most cases, they present values equal to 0. For this reason, we do not consider them. Thus, our analysis covers 100 countries and 4 tax revenue categories.

Initial data on tax revenues are integrated with other socio-economic covariates, in prevalence expressed in terms of GDP, that explain variations in tax revenues and are obtained by combining data from IMF, OECD, and World Bank databases. These covariates are (see Table 1 in the Appendix, supplementary materials): current account balance; employment and unemployment rate; government gross debt; government net lending/borrowing; gross fixed capital formation; change in the gross domestic product (in %); gross national savings; the volume of exports and imports of goods and services; account ownership at a financial institution; subsidies and other transfers; interest payments; compensation of employees; value-added of agriculture, services, industry, and manufacturing; population density; final consumption expenditure; banking nonperforming loans; claims on central government; households consumption; government consumption; trade volume; environmental performance index; government expenditure for military, education, health, and other (as a residual category). The last covariate is obtained by subtracting military expenses, education expenses, and healthcare expenses from the total government expenditure and is intended as a proxy of the other public expenditure items included in the COFOG classification.

The variable location (OECD database) is used to assign the $H = 100$ countries to four geo-economic areas: OECD countries (36), Africa (29), Asia (15), and South America (21). This variable is used to impute missing values for the other covariates. Missing values are replaced with the median value of the covariate computed in the specific area (see Table 1 in the Appendix, supplementary materials). Observed countries for each area are displayed in the Appendix, supplementary materials (Figures A1–A4) together with the tax revenues composition for each country.

## 4. Road to the Bradley-Terry Regression Trunk (BTRT)

Bradley and Terry (1952) proposed a method to derive a latent preference scale from paired comparison data when no natural measuring scale is available. This model has been successfully used in Dittrich et al. (2006), Choisel and Wickelmaier (2007), and Rodriguez Montequin et al. (2020).

Hereinafter, the terms rankings and orderings are used interchangeably. Rankings are numerical vectors that assign values to objects based on their size. Usually, they represent preferences expressed by $H$ individuals (or judges). In our study, they are intended as orders of magnitude of tax revenues differentiated by tax category for $H$ countries. Instead, orderings are nominal lists of objects that express the same concept as rankings.

The paired comparison method splits the ordering process into a series of evaluations carried out on two objects at a time. To address the problem of determining the scale values of a set of objects on a preference continuum that is not directly observable, each pair is compared and a decision is made based on which of the two objects is preferred. When the number of objects is $n_o$, the number of paired comparisons is equal to $\binom{n_o}{2}$.

The basic Bradley-Terry (BT) model computes the probability $\pi_{(ij)i}$ that an object $i$ is ranked above another object $j$ (Agresti 1990) as

$$\pi_{(ij)i} = \frac{\pi_i}{\pi_i + \pi_j}, \tag{1}$$

where $\pi_i$ and $\pi_j$ are nonnegative worth parameters describing the location of objects on the overall ranking scale. When $\sum_{i=1}^{n_o} \pi_i = 1$ is imposed, $\pi_i$ and $\pi_j$ can be regarded as probabilities: for example, $\pi_i$ is the probability that the object $i$ is ranked above all the other objects by considering all the rankings expressed by or related to $H$ individuals.

The BT model is expressed as a quasi-symmetry logistic model with parameters $\lambda_i^O$

$$\text{logit}(\pi_{(ij)i}) = \log\left(\frac{\pi_{(ij)i}}{\pi_{(ij)j}}\right) = \lambda_i^O - \lambda_j^O, \tag{2}$$

where $\lambda_i^O$ and $\lambda_j^O$ are object-parameters related to $\pi_i$'s in (1) by $\lambda_i^O = \frac{1}{2}\ln(\pi_i)$. The superscript $O$ refers to object-specific parameters. The estimated probability in (2) is computed as $\hat{\pi}_{(ij)i} = \frac{\exp(\hat{\lambda}_i^O - \hat{\lambda}_j^O)}{1+\exp(\hat{\lambda}_i^O - \hat{\lambda}_j^O)}$, and $\hat{\pi}_{(ij)i} = \frac{1}{2}$ if $\lambda_i^O = \lambda_j^O$ (the two objects are tied). Equation (2) is equivalent to (1), and the identifiability of these two models requires a restriction on the parameters related to the last object $n_o$, such as $\lambda_{n_o}^O = 0$ and $\sum_i^{n_o} \pi_i = 1$.

Sinclair (1982) extended (1) introducing a log-linear version (LLBT): in comparing object $i$ with object $j$, the Poisson-distributed random variables $y_{(ij)i}$ and $y_{(ij)j}$ are the number of times in which the objects $i$ and $j$ are ranked above the other objects, respectively. In total, $2\binom{n_o}{2}$ expected counts are estimated in a design matrix that is used to synthesize the information about all the considered rankings. Sometimes, it is reasonable to assume that the rankings associated with judges or individuals depend on their features. For example, the preferences expressed by judges usually depend on their characteristics. This issue characterizes the present study as well: the tax revenues composition of a country $h$ and their order is associated with the country's socio-economic structure. In the BT model literature, these features are called subject-specific covariates to differentiate them from the object-specific covariates and subject-object-specific covariates (Schauberger 2015).

The LLBT model accounts for multiple subject-specific covariates. The BT model can include categorical or continuous covariates. Considering numerical covariates only, as in our application, the LLBT for the $h$th judge and objects $i$ and $j$ is

$$\log m\left(y_{(ij)i,h}\right) = \mu_{ij,h} + y_{(ij)i,h}(\lambda_{i,h}^O - \lambda_{j,h}^O). \tag{3}$$

LLBT with subject-specific covariates requires a contingency table for each judge, or individual, and each covariate value; $\lambda_{i,h}^O$ can be re-parameterized through a linear formulation

$$\lambda_{i,h}^O = \lambda_i^O + \sum_{p=1}^{P} \beta_{ip} x_{p,h}, \tag{4}$$

where $x_{p,h}$ is the continuous covariate $x_p$ ($p = 1, \ldots, P$) observed for individual $h$, and $\beta_{ip}$ is the effect of $x_p$ on the location of the object $i$ in a ranking $h$. The intercept $\lambda_i^O$ is the location of the object $i$ in the overall consensus ranking, which includes the rankings associated with all the $H$ individuals.

The deviance of the model is computed as the deviance of a fitted Poisson regression

$$D = 2\sum_{h=1}^{H} y_{ij,h} \times \log\left(\frac{y_{ij,h}}{\hat{y}_{ij,h}}\right), \tag{5}$$

where $y_{ij,h}$ is the value of the comparison $ij$ for a judge $h$, and $\hat{y}_{ij,h}$ is the predicted value based on the estimated model parameters.

### 4.1. Combining STIMA, Regression Trunk, and BTRT

The LLBT model with subject-specific covariates is expressed for ordinal data through a regression model for paired comparisons. The aim is to estimate in a data-driven way the main effects part of the model and, if present, the interaction effects part by implementing the Simultaneous Threshold Interaction Modeling Algorithm (STIMA) (Dusseldorp, Conversano, and Os 2010; Conversano and Dusseldorp 2017) within the BTRT model. STIMA uses the regression trunk methodology (Dusseldorp and Meulman 2004). Interaction effects are considered when two or more covariates do not combine additively (de González and Cox 2007), or when they have a combined effect over and above their additive combination (Cohen et al. 2022, p. 257). STIMA simultaneously estimates both main and interaction effects by growing a small tree called a regression trunk, which is next pruned to avoid overfitting. The covariates can interact with each other so that the interaction effects are considered as a particular kind of nonadditivity. The STIMA's recursive partitioning algorithm is based on the integration of Generalized Linear Models (GLM) (McCullagh and Nelder 1989) and Classification And Regression Trees (CART) (Breiman et al. 1984).

The BTRT model combines the LLBT model with subject-specific covariates (4) with STIMA, when the response derives from the outcome of paired comparisons. The estimated object-parameters $\hat{\lambda}_{i,h}$ for the object $i$ and the individual $h$ in the node $t$ are[1]

$$\hat{\lambda}_{i,h} = \hat{\lambda}_i + \sum_{p=1}^{P} \hat{\beta}_{i,p} x_{p,h} + \sum_{t=1}^{T-1} \hat{\beta}_{i,P+t} I\{(x_{1,h}, \ldots, x_{P,h}) \in t\} \tag{6}$$

The main effect part is $\sum_{p=1}^{P} \hat{\beta}_{i,p} x_{p,h}$, whilst $\sum_{t=1}^{T-1} \hat{\beta}_{i,P+t} I\{(x_{1,h}, \ldots, x_{P,h}) \in t\}$ are the interaction effects estimated for individuals in each terminal node $T$. The interaction terms are $T-1$ since one terminal node is treated as the reference group. The intercept $\hat{\lambda}_i$ quantifies the overall location of the object $i$ for all the individuals of the trunk. To estimate the intercept in each terminal node it is necessary to add $\hat{\beta}_{i,P+t}$ to $\hat{\lambda}_i$. The overall number of parameters includes the number of intercepts $\lambda_i$, plus all the main effects $\beta_{i,P}$ and the interaction effects $\beta_{i,P+t}$. Since one object and one terminal node are considered as reference categories, the number of model parameters is $(n_o-1)+[P\times(n_o-1)]+[(T-1)\times(n_o-1)]$.

[1] The superscript $O$ used in (4) is intentionally left out to lighten notation.

**Algorithm 1:** Bradley-Terry Regression Trunk

```
input : X (matrix of covariates), Y paired comparisons;
initialize: estimate the main effects model in the root node of the trunk: λ̂_{i,h} = λ̂_i + Σ_{p=1}^{P} β̂_{i,p} x_{p,h};
while splitting rule causes a decrease in model deviance or sample size within each child is greater or equal to 5 do
    Split search for each current node t_c;
    for p = 1 → P do
        find the dichotomous variable z_{ijp,t} that minimizes the local log-likelihood deviance of the model;
    end
    detect z*_{ijp,t} as the quantity that minimizes the global log-likelihood deviance of the model;
    create child nodes t_{c+1} and t_{c+2};
    z*_{ijp,t} updates the indicator function I{(x_{1,h}, ..., x_{P,h})} ∈ t and the model including the last threshold interaction effect is
     re-estimated:
```

$$\hat{\lambda}_{i,h} = \hat{\lambda}_i + \sum_{i=1}^{P} \hat{\beta}_{i,p} x_{p,h} + \sum_{t=1}^{T-1} \hat{\beta}_{i,P+t} I\{(x_{1,h}, \ldots, x_{P,h}) \in t\}$$

```
end
output: Bradley-Terry Regression Trunk
```

Table 1. Design matrix with two judges/individuals, three objects, and one continuous subject-specific covariate.

| *Response* | $\mu$ | $\lambda_A^O$ | $\lambda_B^O$ | $\lambda_C^O$ | $x_1$ |
|---|---|---|---|---|---|
| $y_{AB} = 1$ | 1 | −1 | 1 | 0 | 23 |
| $y_{AB} = 0$ | 1 | 1 | −1 | 0 | 23 |
| $y_{AC} = 1$ | 2 | −1 | 0 | 1 | 23 |
| $y_{AC} = 0$ | 2 | 1 | 0 | −1 | 23 |
| $y_{BC} = 1$ | 3 | 0 | 1 | −1 | 23 |
| $y_{BC} = 0$ | 3 | 0 | −1 | 1 | 23 |
| $y_{AB} = 0$ | 1 | −1 | 1 | 0 | 24 |
| $y_{AB} = 1$ | 1 | 1 | −1 | 0 | 24 |
| $y_{AC} = 0$ | 2 | −1 | 0 | 1 | 24 |
| $y_{AC} = 1$ | 2 | 1 | 0 | −1 | 24 |
| $y_{BC} = 0$ | 3 | 0 | 1 | −1 | 24 |
| $y_{BC} = 1$ | 3 | 0 | −1 | 1 | 24 |

*NOTE*: The first column indicates if the object $i$ is preferred ($y_{ij} = 1$) or not ($y_{ij} = 0$) in a certain preference for each pair of objects $ij$. The second column serves as an index for the $n \times (n-1)/2$ comparisons. Preferences are expressed in the next three columns, and finally, the covariate $x_1$ is shown in the last column. In this example, the two individuals present opposite rankings, BCA and ACB, respectively.

## 4.2. The BTRT Algorithm

The BTRT algorithm, described in Algorithm 1, follows the STIMA algorithm with some adjustments due to pairwise comparisons.

First, a design matrix composed of $n = n_o \times (n_o - 1) \times H$ rows is defined. Indeed, the total number of rows is $2 \times (n_o \times (n_o - 1)/2) \times H$, that is, the product between the number of comparing objects, that is 2 when ties are not allowed, the number of paired comparisons ($n_o \times (n_o - 1)/2$), and the number of individuals $H$. If ties are considered, the model can be extended by incorporating undecidedness parameters. Table 1 is an example of a design matrix with two subjects, three objects, and one continuous subject-specific covariate.

A generalized linear model with a log link and Poisson distribution is fitted to the root node to compute the main effects $\beta_p$. The splitting procedure is applied to the root node, and the resulting child nodes, by searching for a threshold interaction between the $P$ covariates assuming they have a combined effect on the objects' order besides their individual (main) effect. The best candidate split maximizes the decrease in deviance of a Poisson regression model when moving from a parent node to the two possible child nodes. Once found, the best split defines a dichotomous variable $z^*_{ijp,t}$ that updates the indicator function $I(\cdot)$ in (6). Next, all regression coefficients are re-estimated. The number of dichotomous variables $z^*_{ijp,t}$ is equal to the number of splits leading to node $t$.

Two main approaches to growing the BTRT trunk are presented in Baldassarre et al. (2022): One Split Only (OSO) and Multiple Splitting (MS). In OSO, all the covariates are used as splitting covariates only once. In MS, each covariate can be used more than once to split a node. Generally, MS performs better than OSO in reducing deviance, although OSO is more appropriate when the goal is obtaining an easier-to-read partition. For our analysis, results suggest MS represents a fair compromise between goodness-of-fit and interpretation. The selected trunk shown in Section 6 presents few splits and fits the data satisfactorily.

The BTRT stopping criterion is based on the prior definition of a minimum number of individuals (minimum bucket size) in a node. In the default BTRT implementation, it is set to 5 but it can be changed based on the desired depth of the trunk.

Once the entire trunk is grown, pruning based on $V$-fold cross-validation and the $c \cdot SE$ rule is applied. The cross-validation deviance $D^{cv}$ for a Poisson distribution and its associated standard error $SE^{cv}$ are computed in each step of the splitting procedure.[2] $D^{cv}$ initially follows a decreasing trend and then begins to grow after a certain number of splits. The $c \cdot SE$ rule works as follows: Let $t^* \in [1, T]$ be the size of the regression trunk with the lowest $D^{cv}$, say $D^{cv}_{t^*}$. The best size of the BTRT trunk $t^{**}$ corresponds to the minimum value of $t$ such that $D^{cv}_{t^{**}} \leq D^{cv}_{t^*} + c \cdot SE^{cv}_{t^*}$, where $c \in [0, 1]$. The optimal choice of the pruning parameter $c$ has been investigated in Baldassarre et al. (2022): the proportion of times BTRT captures a true interaction

[2] See eqs. (14) and (15) in Baldassarre et al. (2022).

effect is higher than 0.8 when $0.5 < c \leq 1$, and it is advised to increase $c$ as the number of objects, rankings, and the effect size of the covariates on the orderings increase. In general, the higher the value of $c$, the more weight is given to the standard error so that the selected trunk has a reduced size. In this case, since BTRT includes main effects and interaction effects within the same model, it would be advisable to choose conservative values of $c$ (i.e., $0.5 < c \leq 1$).

## 5. Bradley-Terry-Luce Lasso for Covariates Selection

We consider a wide set of covariates as initial candidate variables for BTRT. However, one of the main drawbacks of applying the BT model with subject-specific covariates is the number of parameters to estimate. This drawback is accentuated when the BTRT model is used: in addition to the intercepts for each tax category, also main effects and interaction effects between covariates are included. This issue is solved by following a penalized likelihood approach instead of ordinary ML estimation. In this way, it is possible to first consider a high number of candidate covariates, and next, reduce the number of involved parameters by selecting the relevant ones only.

Lasso regression (Tibshirani 1996) is commonly used for variable selection in presence of numerical covariates. In case of paired comparisons the penalty terms used in standard Lasso have to adapt the specific structure of the model (Schauberger and Tutz 2017), and the method for variable selection that adapts to paired comparison is the Bradley-Terry-Luce Lasso (BTLL) (Schauberger 2015). We use BTLL to select the most influential country-specific covariates to be included in BTRT. In BTLL, the BT model is reinterpreted according to Luce's axiom (Luce 1959) of independence from irrelevant alternatives. It states that other objects do not influence the decision between two objects. BTLL is based on the maximization of the penalized log-likelihood:

$$l_p(\xi) = l(\xi) - \Lambda J(\xi), \tag{7}$$

where $l(\xi)$ is the log-likelihood of the vector of model parameters ($\xi$), $J(\xi)$ is a penalty term and $\Lambda$ is a tuning parameter that quantifies the penalty degree. If $\Lambda = 0$, $l_p(\xi)$ reduces to the standard maximum likelihood estimate. In BTLL, the penalty for the set $x_p$ of $P$ subject-specific covariates that share the same effect size on the response is

$$\phi_P(\beta_{i,1}, \ldots, \beta_{i,P}) = \sum_{p=1}^{P} \sum_{i<j} |\beta_{i,p} - \beta_{j,p}| \tag{8}$$

If $\Lambda \to \infty$, all the effect sizes of $x_p$ are merged into a single one, and $x_p$ is removed from the model as all its effect sizes tend to zero. Instead, when $\Lambda = 0$ BTLL reduces to BT with subject-specific covariates. In general, for a finite value of $\Lambda$ some of the covariates are removed, whilst the remaining are still identified in the model. The penalty terms $\phi_P$ can be weighted based on the adaptive lasso principle (Zou 2006), and next combined as

$$J(\xi) = \sum_{l=1}^{L} \phi_l P_l, \tag{9}$$

where $\psi_l$ represents penalty-specific weights. Thus, all penalty terms $P_l$ are combined into a unique penalty $J(\xi)$ controlled by the tuning parameter $\Lambda$. The comparability of different penalty terms requires two conditions: (a) all subject-specific covariates have to be scaled to allow the comparison of their effect sizes; (b) the weights $\psi_l$ have to be assigned according to the number of penalties and free parameters they include (Bondell and Reich 2009; Oelker, Pößnecker, and Tutz 2015). The penalty is interpreted as a Lasso-type fusion penalty rather than a simple Lasso (Schauberger and Tutz 2017). Similar approaches have been used in GLM (Bondell and Reich 2009; Gertheiss and Tutz 2010), and in paired comparison models (Masarotto and Varin 2012; Tutz and Schauberger 2015). Since the information matrix used in penalized likelihood approaches does not lead to standard errors or confidence intervals, we apply bootstrap to the cross-validated model to compute confidence intervals that account for the additional variance originating from the model selection procedure.

To select covariates for BTRT we use the `BTLLasso` *R* package (Schauberger and Tutz 2019), which penalizes the Fisher scoring with a $L_1$ penalty. The optimal tuning parameters $\Lambda$ is selected through 10-fold cross-validation. For the main effect part of the BTRT model (root node), BTLL returns the cross-validated $\Lambda$ as the optimal value for the size of penalties associated with the $\beta_{i,p}$s; 13 out of 30 subject-specific covariates with an effect size equal to zero at least on two tax revenue categories are removed from the model. Descriptive statistics for the remaining covariates are reported in Table 2.

The final dataset includes $H = 100$ countries, $n_o = 4$ tax revenue categories, and 17 covariates. Interestingly, the covariates related to public expenditure remain in the model as the size of their main effect on the composition of tax revenues can not be overlooked.

The major drawback of our approach is the computational cost of the bootstrap procedure. In addition, BTLL selects the best covariates by only considering their main effects on each tax revenues category. Nevertheless, it still constitutes the more appropriate covariate selection procedure when dealing with the Bradley-Terry model.

## 6. Results

Figure 1 and Table 3 describe the BTRT pruned regression trunk. The CART's node numbering criterion is used: each terminal node $R_t$ is the region of the predictor space obtained from the recursive partitioning procedure. For each $R_t$, the number of countries $H$, the consensus ranking $C$ and the correlation coefficient $\tau_x$ between rankings are reported. $C$ is computed maximizing $\tau_x$ inside the nodes. In Figure 1, it represents a summary measure for each $R_t$ that results from the minimization of distances between preferences without considering the characteristics of the countries (the subject-specific covariates).

BTRT confirms EPI ($x_{13}$) is a key covariate as it is selected to split the first node: 76 countries are assigned to node 2 (EPI $\leq$ 0.7) and 24 to node 3 (EPI $>$ 0.7). The second splitting covariate is the government health spending ($x_4$) leading to nodes 4 and 5 that are further split according to EPI and Employment rate ($x_1$), respectively, to obtain terminal nodes $R_2$–$R_5$. The employment rate is a useful indicator of current labour market conditions as it is unaffected by voluntary changes in labour

**Table 2.** Descriptive statistics for the subject-specific covariates selected by the BTLL model.

| Covariates | $x_p$ | # missing | Mean | SD | Median | Min | Max | Range | Skew | Kurtosis | SE |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Employment rate | $x_1$ | 2 | 0.59 | 0.09 | 0.59 | 0.39 | 0.85 | 0.45 | 0.05 | 0.27 | 0.01 |
| Government gross debt | $x_2$ | 1 | 0.59 | 0.35 | 0.50 | 0.09 | 2.38 | 2.29 | 2.12 | 6.96 | 0.03 |
| Savings | $x_3$ | 5 | 0.21 | 0.08 | 0.21 | 0.02 | 0.46 | 0.44 | 0.53 | 0.48 | 0.01 |
| Exports | $x_4$ | 13 | 0.04 | 0.06 | 0.05 | −0.36 | 0.17 | 0.54 | −3.59 | 22.69 | 0.01 |
| Imports | $x_5$ | 11 | 0.04 | 0.06 | 0.05 | −0.23 | 0.18 | 0.41 | −1.66 | 6.12 | 0.01 |
| Interests | $x_6$ | 21 | 0.08 | 0.05 | 0.07 | 0.00 | 0.29 | 0.29 | 1.34 | 2.98 | 0.01 |
| Agric added-value | $x_7$ | 2 | 0.09 | 0.09 | 0.05 | 0.00 | 0.45 | 0.45 | 1.69 | 2.79 | 0.01 |
| Services added-value | $x_8$ | 21 | 0.57 | 0.10 | 0.57 | 0.29 | 0.79 | 0.50 | −0.40 | −0.20 | 0.01 |
| Final consumption | $x_9$ | 8 | 0.78 | 0.11 | 0.79 | 0.43 | 1.12 | 0.69 | −0.49 | 1.88 | 0.01 |
| Bank NPloans | $x_{10}$ | 16 | 0.04 | 0.06 | 0.03 | 0.00 | 0.42 | 0.42 | 4.77 | 25.75 | 0.01 |
| Claims centr gov | $x_{11}$ | 4 | 0.12 | 0.18 | 0.08 | −0.14 | 1.42 | 1.56 | 3.85 | 23.62 | 0.02 |
| Gov consumption | $x_{12}$ | 8 | 0.16 | 0.05 | 0.16 | 0.04 | 0.39 | 0.35 | 0.63 | 2.42 | 0.01 |
| EPI | $x_{13}$ | 0 | 0.59 | 0.16 | 0.59 | 0.00 | 0.87 | 0.87 | −0.93 | 2.21 | 0.02 |
| Military spending | $x_{14}$ | 2 | 0.01 | 0.01 | 0.01 | 0.00 | 0.05 | 0.05 | 1.44 | 3.72 | 0.00 |
| Education spending | $x_{15}$ | 33 | 0.04 | 0.01 | 0.04 | 0.01 | 0.08 | 0.07 | 0.72 | 0.25 | 0.00 |
| Health spending | $x_{16}$ | 0 | 0.07 | 0.03 | 0.07 | 0.02 | 0.17 | 0.15 | 0.74 | 0.96 | 0.00 |
| Other spending | $x_{17}$ | 9 | 0.20 | 0.10 | 0.20 | 0.03 | 0.76 | 0.73 | 1.77 | 7.89 | 0.01 |

*NOTE*: # missing = n. of missing values; SD = standard deviation; Min = minimum; Max = maximum; Skew = skewness; SE = standard error.

**Table 3.** Pruned BTRT regression trunk.

| | Node n. | *Splitting covariate* | *Split Point* | *Model Deviance* |
|---|---|---|---|---|
| | | main effects (no splits) | | 313 |
| bestsplit1 | 1 | $x_{15}$ (Environmental Performance Index) | 0.70 | 271 |
| bestsplit3 | 2 | $x_4$ (Health spending) | 0.06 | 213 |
| bestsplit5 | 4 | $x_8$ (Environmental Performance Index) | 0.56 | 171 |
| bestsplit4 | 5 | $x_8$ (Employment rate) | 0.51 | 190 |
| bestsplit2 | 3 | $x_2$ (GDP % change) | 0.40 | 238 |

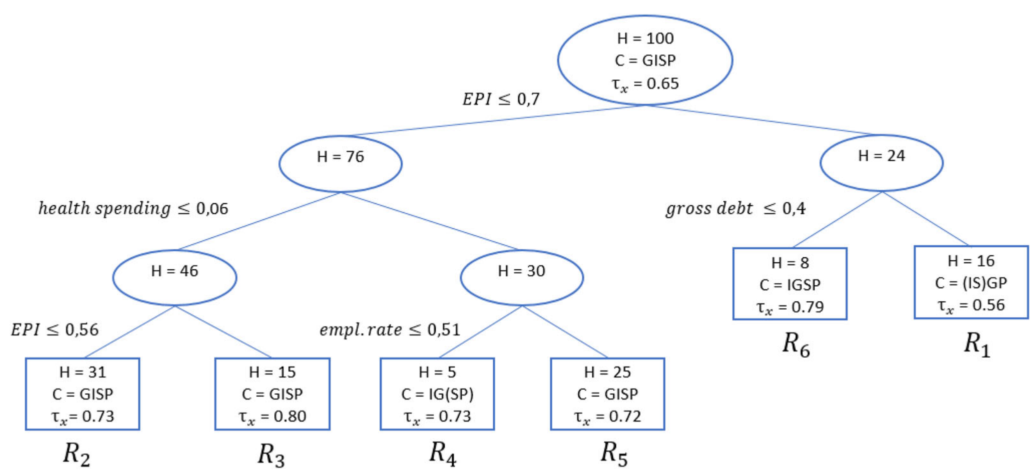


**Figure 1.** Pruned regression trunk.
*NOTE*: In each terminal node $R_1, \ldots, R_6$ and in the root note, the number of countries $H$, the consensus ordering $C$, and the correlation coefficient $\tau_x$ are shown. Taxes are named as follows: I - Taxes on income, S - Social security contributions, P - Taxes on property, and G - Taxes on goods. When two tax categories are in brackets, it means they are in a tie.

force participation. The unemployment rate is influenced by the labour force size, decreasing when workers stop job-seeking and potentially increasing during labour market recovery as more individuals join the labour force. The third splitting covariate is Gross debt ($x_2$) which splits node 3 into terminal nodes $R_1$ and $R_6$. Thus, BTRT finds a first-order interaction between EPI and Gross debt ($R_1$ and $R_6$), another first-order interaction between EPI and Health spending ($R_2$ and $R_3$), and a third-order interaction between EPI, Health spending and Employment rate ($R_4$ and $R_5$). Overall, in five terminal nodes $\tau_x \geq 0.7$ indicates a broad consensus. $R_2$, $R_3$, and $R_5$ show a similar consensus ranking: for them, taxes on goods have a substantial effect on the composition of tax revenues. In $R_1$, $R_4$ and $R_6$ taxes on income are the most influential item of tax revenues. Ties are observed in $R_1$ and $R_4$.

Table 4 summarizes the BTRT pruning procedure. Results of *model*6 (*mod*6) are reported even if the pruning indicates *model*5 (*mod*5) as the final model. Adding another split to the *mod*5 trunk causes the deviance $D$ and the cross-validation deviance $D^{cv}$ to decrease. The $D^{cv}$ values are mean values of deviances computed on each row of the design matrix, thus, they are much smaller than the $D$ values, computed as a sum (5). The $c \cdot SE$ rule with $c = 0.5$ leads to prune *mod*6 since $0.3565 + 0.5 \times 0.0014 = 0.3572 > 0.3567$. Thus, the final

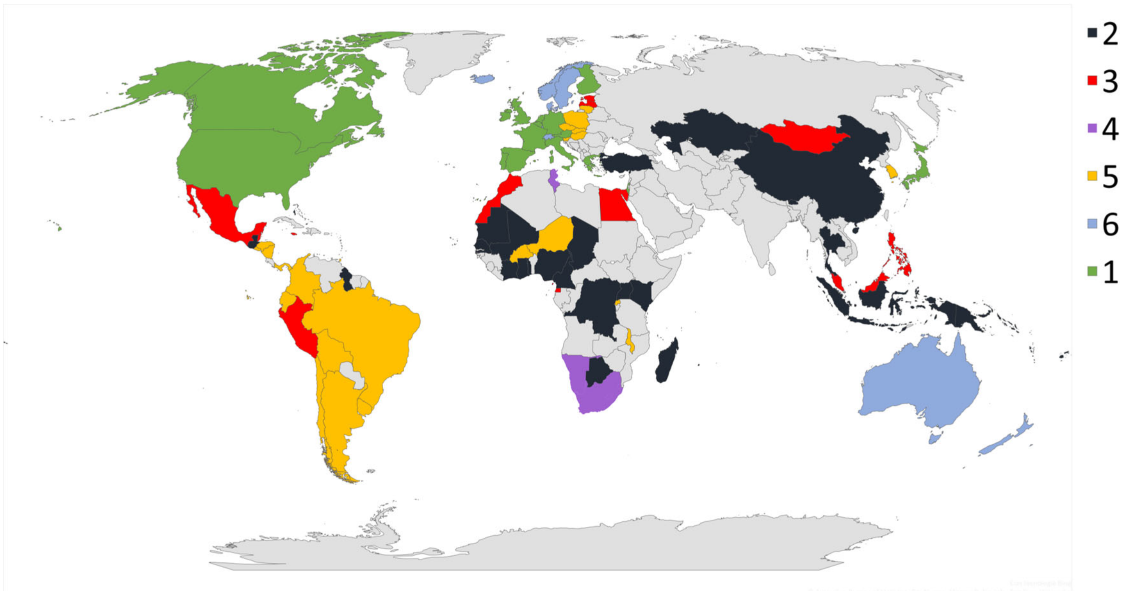


**Figure 2.** Countries' distribution map in each region (terminal node) $R_1, \ldots, R_6$, where $R_1$ is the reference region in the BTRT model.

**Table 4.** Pruned BTRT regression trunk: 10-fold cross-validation results.

| | $D$ | $D^{CV}$ | $SE^{CV}$ |
|---|---|---|---|
| mod0 | 313.0671 | 0.3775 | 0.0010 |
| mod1 | 271.9198 | 0.3682 | 0.0010 |
| mod2 | 238.8955 | 0.3659 | 0.0011 |
| mod3 | 213.5469 | 0.3617 | 0.0011 |
| mod4 | 190.8353 | 0.3594 | 0.0013 |
| mod5 | 171.2580 | 0.3567 | 0.0013 |
| mod6 | 136.5742 | 0.3565 | 0.0014 |

*NOTE*: $D$ = model deviance for a Poisson regression; $D^{CV}$ = casewise cross-validation deviance (eq. (14) Baldassarre et al. 2022); $SE^{CV}$ = standard error of $D^{CV}$ (eq. (15) Baldassarre et al. 2022)

trunk corresponds to *mod*5 with five splits and $T = 6$ terminal nodes.

Figure 2 shows the world map with the final breakdown of the countries in the regions $R_1, \ldots, R_6$. Most of the OECD countries with higher EPI and gross public debt are in $R_1$ (reference region). $R_2$ includes 31 countries with lower EPI and health expenditure. Besides Turkey, they are all located in Asia and Africa. $R_3$ includes 15 countries located on different continents (e.g., Egypt, Morocco, Mongolia, and Peru) that have poor EPI but spend more than 6% of GDP on health. $R_4$ consists of five countries, all located in South Africa except Tunisia, characterized by low EPI and employment rates, but with health spending larger than 6% of GDP. Countries in $R_5$, in prevalence located in South America and Central Africa, share the same characteristics of those in $R_4$ but have a higher labour force participation rate. Finally, $R_6$ includes the most virtuous countries in terms of EPI and public budget management.

In general, Figure 2 provides a proxy for groups of countries that follow a common fiscal policy. For example, $R_1$ includes all OECD member countries with fairly similar fiscal policies (e.g., high levels of taxation on income, goods and services, and social security contribution; lower taxation on property). As discussed in Caselli and Reynaud (2020), Buera, Monge-Naranjo, and Primiceri (2011), and Giuliano, Mishra, and Spilimbergo (2013), the intuition is that tax reforms in neighbouring countries affect the adoption of reforms through equal tax burdens and imitation effects.

Besides the inspection of terminal nodes, BTRT provides the estimated regression coefficients for each tax revenue. These are reported in Table 5 for taxes on income ($o_1$), compulsory SSC ($o_2$), and taxes on property ($o_3$). For the reference category taxes on goods ($o_4$) they are set to zero. Some interesting considerations emerge:

• as for $o_1$ (taxes on income), military spending ($x_{14}$) and exports ($x_4$) have a strong impact on its size. The more is the level of $x_{14}$ or $x_4$, the lower the probability that taxes on income are higher than taxes on goods ($o_4$, reference category). On the contrary, imports ($x_5$) and health expenditures ($x_{16}$) have a positive effect on $o_1$. Then, for taxes on income, all the coefficients of the interaction effects except for $R_6$ (OECD members) have a negative sign.

• As for SSC ($o_2$), savings ($x_3$), EPI ($x_{13}$), and other spending ($x_{17}$) have a positive and strong impact on its size. For instance, the higher $x_{13}$, the lower the log odds that SSC is higher than taxes on goods. EPI has a positive and high impact on $o_2$, despite the negative effect observed for $o_1$. In most developed countries SSCs are likely higher than taxes on goods. In addition, government consumption ($x_{12}$) has a strong and negative effect on SSC. Finally, the interaction term for $R_4$ ($R_6$) has a positive (negative) effect on $o_2$;

• Taxes on property ($o_3$) has a strong tendency to be the last ranked object as it shows the lowest value for the estimated intercept. As for the main effects, export levels ($x_4$) and government consumption ($x_{12}$) have a strong and negative effect on the size of taxes on the property. On the contrary, military spending ($x_{14}$) and other spending ($x_{17}$) have a positive impact on $o_3$. All the interaction effects have a positive impact on the comparison between taxes on property and taxes on goods and services.

Table 6 reports for each terminal node the worth parameters $\pi_i$ (6) and the mean value of the incidence of each tax category for all the countries inside the node. The worth parameters are

**Table 5.** BTRT estimated coefficients.

| | $\hat{\lambda}_{o_1,h}$ | $\hat{\lambda}_{o_2,h}$ | $\hat{\lambda}_{o_3,h}$ | Interaction effect |
|---|---|---|---|---|
| $\hat{\lambda}_i$ | 13.15* (5.33) | −33.63** (11.83) | −50.09 (115.79) | |
| $\hat{\beta}_{i,x1}$ | −4.16 (2.77) | 1.24 (4.62) | 8.73 (4.54) | |
| $\hat{\beta}_{i,x2}$ | 0.71 (0.93) | −5.28** (1.94) | −2.99 (1.98) | |
| $\hat{\beta}_{i,x3}$ | 5.45 (3.68) | 20.95** (7.56) | 15.78 (8.15) | |
| $\hat{\beta}_{i,x4}$ | −16.68** (5.58) | −5.02 (9.13) | −32.94* (12.96) | |
| $\hat{\beta}_{i,x5}$ | 14.89* (5.43) | −11.30 (7.38) | 5.96 (8.51) | |
| $\hat{\beta}_{i,x6}$ | −15.22* (7.38) | −2.61 (10.94) | 4.06 (12.10) | |
| $\hat{\beta}_{i,x7}$ | −10.12* (4.48) | 2.40 (10.70) | −15.69 (12.72) | |
| $\hat{\beta}_{i,x8}$ | −8.05* (3.78) | 16.25* (6.72) | 11.85 (7.51) | |
| $\hat{\beta}_{i,x9}$ | 2.82 (3.39) | 5.23 (5.50) | 7.94 (6.69) | |
| $\hat{\beta}_{i,x10}$ | −4.54 (4.09) | 16.92* (7.60) | −0.19 (10.47) | |
| $\hat{\beta}_{i,x11}$ | 0.33 (1.60) | 7.91** (3.05) | 9.32* (4.04) | |
| $\hat{\beta}_{i,x12}$ | −11.06 (7.49) | −28.28* (12.56) | −51.68** (16.17) | |
| $\hat{\beta}_{i,x13}$ | −6.54* (2.65) | 22.30*** (6.52) | 24.17** (7.61) | |
| $\hat{\beta}_{i,x14}$ | −25.63 (25.38) | 6.32 (33.37) | 137.84** (53.16) | |
| $\hat{\beta}_{i,x15}$ | −3.98 (21.42) | −65.68 (38.44) | −45.98 (43.92) | |
| $\hat{\beta}_{i,x16}$ | 10.26 (13.88) | 20.80 (17.44) | 5.51 (25.97) | |
| $\hat{\beta}_{i,x17}$ | −4.15 (2.63) | 22.81** (7.01) | 33.23*** (8.64) | |
| $\hat{\beta}_{i,R_2}$ | −3.21* (1.32) | 0.83 (1.88) | 14.30 (115.09) | $R_2 = I(\text{EPI} \leq 0.56, \text{Health sp.} \leq 0.06)$ |
| $\hat{\beta}_{i,R_3}$ | −2.25* (1.06) | 0.50 (1.42) | 11.58 (115.09) | $R_3 = I(0.56 < \text{EPI} \leq 0.7, \text{Health sp.} \leq 0.06)$ |
| $\hat{\beta}_{i,R_4}$ | −0.77 (1.32) | 4.87* (2.22) | 19.19 (115.10) | $R_4 = I(\text{EPI} \leq .7, \text{Health sp.} > .06, \text{Empl. rate} \leq .51)$ |
| $\hat{\beta}_{i,R_5}$ | −3.24*** (0.96) | 3.31* (1.29) | 6.00 (153.15) | $R_5 = I(\text{EPI} \leq .7, \text{Health sp.} > .06, \text{Empl. rate} > .51)$ |
| $\hat{\beta}_{i,R_6}$ | 7.93 (201.22) | −4.26* (1.73) | 8.35 (115.10) | $R_6 = I(\text{EPI} > 0.7, \text{Gross debt} \leq 0.4)$ |

*NOTE*: Estimated coefficients associated to the objects taxes on income $o_1$, social security contributions $o_2$, taxes on property $o_3$, and taxes on goods and services $o_4$ (reference category, that is, coefficients equal to zero). Standard errors are in parenthesis and the stars "$*$" associated with some estimated coefficients indicate they are different from zero with a $p$-value $\leq 0.001$ ("$***$"), $\leq 0.01$ ("$**$"), and $\leq 0.05$ ("$*$"), respectively.

**Table 6.** Mean worth parameters $\pi_i$ for each tax revenue category in each terminal node of the pruned regression trunk.

| | $\pi_{income}$ | $\pi_{social}$ | $\pi_{property}$ | $\pi_{goods}$ |
|---|---|---|---|---|
| $R_2$ | 0.04713 | 0.00000 | 0.00000 | 0.95287 |
| $R_3$ | 0.12403 | 0.00011 | 0.00001 | 0.87585 |
| $R_4$ | 0.96217 | 0.00000 | 0.00000 | 0.03783 |
| $R_5$ | 0.02498 | 0.00276 | 0.00000 | 0.97226 |
| $R_6$ | 1.00000 | 0.00000 | 0.00000 | 0.00000 |
| $R_1$ | 0.40127 | 0.39267 | 0.00000 | 0.20605 |

*Notes*: Each $\pi_i$ corresponds to the average value of the worth parameters associated with each country in the specific terminal node.

computed based on the relationship between $\pi_i$ and $\lambda_i$. They are more informative than the consensus ranking $C$. For instance, the regions $R_2$ and $R_3$ in Figure 1 show the same consensus ranking. These orderings are in line with the order of the $\pi_i$s in Table 6, but with a relevant difference: in $R_2$, SSC and taxes on property are very close to each other, while in $R_3$ the difference among them is larger. The $\pi_i$s provide additional information about the composition of the estimated ordering given the characteristics of the countries falling in a terminal node. Note that OECD countries (most of them in green in Figure 2 and belonging to $R_1$) are characterized by the highest EPI and government gross debt. Compared to countries in other terminal nodes, they show a more balanced composition of tax revenues, with direct income taxes and social contributions dominating. Taxes on goods and services play a secondary role, whilst property taxes make up a residual share of government revenues. This interpretation is useful for researchers and policy-makers: a country with high gross debt and high environmental performances is likely to have tax revenue compositions similar to that of OECD countries.

### 6.1. Comparing BTRT Results with Tax Determinants Literature

Results presented in Section 6 are compared with those obtainable from the model specified in Gupta (2007), whose study is characterized by the same objectives and structure of data and represents a milestone in the tax determinants literature, and is based on the study of Lotz and Morss (1967), one of the first attempts to estimate tax determinants.

Lotz and Morss (1967) estimate the tax ratio for 52 developing countries using Ordinary Least Squares (OLS) and including the per capita income and openness to foreign trade as covariates. Their dataset is cross-sectional since tax ratios are computed as the average ratio for the years 1963–1965. The estimated tax ratios are compared, to create groups of countries with similar tax performance. Similarly, Gupta (2007) examines the main determinants of central government tax revenue using panel data for 105 developing countries over 25 years. The dependent variable is the total central government tax revenue (as a percentage of GDP). As in our analysis, Gupta focuses on the impact of debt, agricultural value-added, and trade on the determination of tax revenues. In addition, he uses the disaggregation of tax revenue categories as covariates, whereas in BTRT these are transformed them into paired comparisons. Moreover, in line with most of the studies on tax determinants, Gupta investigates the effect of socio-economic covariates on total tax revenue through a panel regression with fixed and random effects that include some institutional factors as additional covariates. Instead, BTRT is orientated toward the investigation of the relationship between tax revenue category and government expenditures by function (COFOG). To achieve comparability of results, the model proposed in Gupta (2007) is

**Table 7.** OLS estimated coefficients with interaction effects.

| | Estimate | Std. Error | $t$ value | Pr(>\|t\|) |
|---|---|---|---|---|
| Intercepts | −2.10 | 0.21 | −9.85 | 0.00 |
| Employment rate | 0.01 | 0.10 | 0.10 | 0.92 |
| Government gross debt | 0.03 | 0.04 | 2.06 | 0.04 |
| Savings | −0.11 | 0.16 | −0.68 | 0.50 |
| Exports | 0.11 | 0.21 | 0.52 | 0.60 |
| Imports | −0.03 | 0.20 | −0.42 | 0.68 |
| Interests | −0.29 | 0.21 | −1.39 | 0.17 |
| Agriculture value added | −4.58 | 0.81 | −5.62 | 0.00 |
| Services | −0.22 | 0.14 | −1.54 | 0.13 |
| Final consumption | −0.20 | 0.14 | −1.40 | 0.17 |
| Bank NP loans | −0.27 | 0.20 | −1.40 | 0.17 |
| Claims centr gov | −0.16 | 0.07 | −2.19 | 0.03 |
| Gov consumption | 0.23 | 0.57 | 0.41 | 0.68 |
| EPI | 0.35 | 0.08 | 4.32 | 0.00 |
| Military spending | 6.08 | 5.14 | 1.18 | 0.24 |
| Education spending | 2.84 | 0.83 | 3.42 | 0.00 |
| Health spending | 2.42 | 0.45 | 5.32 | 0.00 |
| Other spending | 2.80 | 0.12 | 24.17 | 0.00 |
| Agriculture VA:Gov consumption | 30.15 | 5.84 | 5.16 | 0.00 |
| Agriculture VA:Military sp | 190.34 | 36.34 | 5.24 | 0.00 |
| Gov cons:Military sp | −8.01 | 27.54 | −0.29 | 0.77 |
| Agriculture VA:Gov cons:Military sp | −1313.86 | 260.58 | −5.04 | 0.00 |

applied as follows: (a) The dependent variable is the logarithm of the total central government tax revenue instead of tax revenue categories. The logarithmic term accounts for the heterogeneity of total tax revenues across countries. As confirmed by BTRT, low-income countries typically have tax revenues between 10% and 20% of GDP, while for high-income countries it rises to 40% (see also Besley and Persson (2014)); (b) OLS is estimated on the same set of countries used for BTRT; (c) All the data refer to the year 2018; (d) The OLS model includes the same set of covariates used for BTRT (Table 2).

Beside the main effects, an exhaustive search of all the possible cross-product interaction effects up to the third order is performed. The selected OLS model is the one minimizing the Bayesian Information Criterion (BIC). It is compared with BTRT (Table 5) to highlight the additional information supplied by BTRT when interpreting the interaction effects.

Table 7 shows the OLS estimated coefficients. To understand how disaggregation of tax revenues by category improves interpretation of the interaction effects, values reported in Table 7 are compared with Table 5. For both OLS and BTRT the null hypothesis that the correct model includes only the main effects part is rejected in favor of the model with interaction effects (for OLS the test statistic is $\chi^2_4 = 36.09$; $p < .000001$, whilst for BTRT we get $\chi^2_{15} = 141.81$; $p < .000001$). Important differences leading to different findings arise when interpreting the estimated coefficients. Gupta (2007) reports that central government gross debt has a positive impact on total tax revenue. This result is confirmed in the estimated OLS (Table 7, $\hat{\beta}_3 = 0.09$; $p = 0.04$) but $\hat{\beta}_3$ is borderline significant. This positive effect is due to the behaviour of countries with the burden of future loan repayments, which induces policymakers to raise tax rates. BTRT reports the same result in terms of the direction of the causal effect for all tax categories. For $x_3$, the increase in gross debt has a positive and significant effect on social security contributions ($o_2$).

Consistent with Lotz and Morss (1967) and Gupta (2007) BTRT outlines the positive relationship between openness and revenue performance. Having an efficient tax system is critical, especially for countries noncompetitive in exports and natural resources (Karagoz 2013). The OLS estimated coefficients for exports and imports are not significant. On the contrary, BTRT confirms the findings of Karagoz (2013): there is a negative (positive) relationship between exports (imports) and some tax categories: exports ($x_4$) influence negatively taxes on income and property, whilst imports ($x_5$) influence positively taxes on income only.

As for the effect of agriculture value-added on tax revenue, Gupta (2007) reports a strong negative impact (in line with Leuthold 1991; Stotsky and WoldeMariam 1997; Eltony 2001). Likewise, Agbeyegbe, Stotsky, and WoldeMariam (2006) find a negative association between tax revenue and agricultural share, but not in Sub-Saharan countries where agricultural products are extensively exported and in turn increase tax revenues. However, also for these countries, the sign of the agricultural sector share is negative when income tax revenue is used as the response variable. OLS results in Table 7 show this negative association ($\hat{\beta}_8 = -4.58$; $p = 0.00$). The unique cross-product interaction of the third order found by OLS involves agriculture value added together with government consumption and military spending. It indicates a relevant decrease in tax revenues ($\hat{\beta}_{21} = -1334.55$; $p = 0.00$), at the net of the positive effect on tax revenues of agriculture combined with either government consumption or military spending (see estimated coefficient of the second-order interactions). The same finding is not reported by BTRT: agriculture ($x_7$) hurts the taxes on income only. On the supply side, BTRT shows that for a country with a large agricultural sector, it is difficult or politically infeasible to raise taxes, especially if the agricultural sector is mostly subsistence or if it is organized on a small scale basis (Agbeyegbe, Stotsky, and WoldeMariam 2006). On the demand side, if the agricultural sector is leading, there is no need for high expenditures for the country. The need for public goods and services, which in most cases leads to raised tax rates, tends to be relatively urban-based (Gupta 2007).

The EPI index, used as a proxy for the country and human development, has a positive effect in OLS ($\hat{\beta}_{13} = 0.35$; $p = 0.00$). In Gupta (2007), the GDP per capita is used as a proxy for development and has a positive effect on tax revenue. This finding is consistent with Chelliah (1971) and demonstrates that a high level of development corresponds to a higher capacity to collect and pay taxes. BTRT shows similar results, with one unique difference: the proxy for the degree of development ($x_{13}$) has a positive impact on every tax revenue category except for income taxes. This result demonstrates that taxes from broad-based sources such as VAT and income taxes are lower in case of greater access to other forms of revenue (Besley and Persson 2014). BTRT provides additional findings from threshold interactions induced by EPI (from $\hat{\beta}_{i,R_2}$ to $\hat{\beta}_{i,R_6}$ in Table 5). In $R_2$ and $R_3$, low values of EPI interact with health spending to reduce taxes on income. Conversely, in $R_4$ and $R_5$ the average value of EPI interacts with health spending and employment rate to increase taxes on social security contributions (and reduce taxes on income when the employment rate is high). In $R_6$, countries with high levels of EPI and reduced gross debt experience a reduction in taxes on social security contributions.

As for government expenditure by categories, OLS reports a positive association for education spending, health spending,

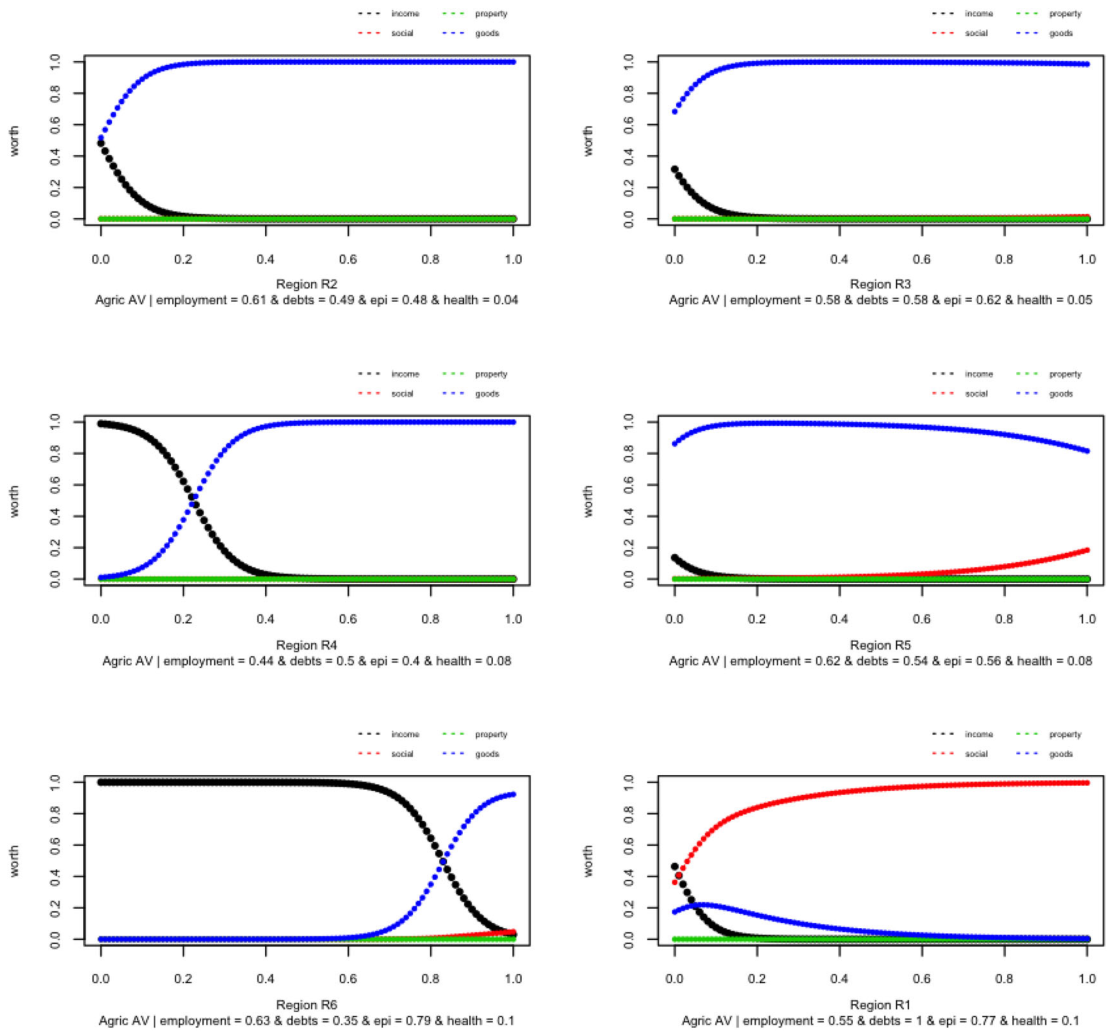


**Figure 3.** Mean worth parameters $\pi_i$ conditioned to the share of agricultural sector to GDP.
*NOTE*: The worth parameters are computed for $R_1, \ldots, R_6$ by fixing all covariates (except for the share of the agricultural sector) to the average level of the group. The order of graphs within the figure follows the trunk in Figure 1: $R_2$ top left, $R_3$ top right, $R_4$ center left, $R_5$ center right, $R_6$ bottom left, $R_1$ bottom right. The x-axis shows simulated values of the covariate agricultural share within the standardized range [0, 1]. Finally, for each plot, the mean values of the covariates resulting from the splitting procedure of the regression trunk are reported. For the sake of clarity, the mean values of all the other country-specific covariates are not reported.

and other spending. In the main effects part their $\hat{\beta}$ have the largest positive impact on the total tax revenue. For example, an increase in health expenditure is compensated by an increase in total tax revenue of about 2.42%. BTRT does not always estimate the same positive effect: in the main effects part, a consistent increase in taxes on property is caused by an increase in military spending ($x_{14}$). Interestingly, the interaction effects induced by $\hat{\beta}_{i,R_4}$ to $\hat{\beta}_{i,R_5}$ indicate that higher health spending induces an increase in taxes on social security contributions (and a decrease in taxes on income) in countries with reduced (increased) employment rate and EPI below 0.70. The positive associations involving health spending indicate that high public health spending leads to lower household consumption of health services, which in turn leads to higher disposable income and, consequently, to a wider tax base for income tax (Jiang 2021).

It is possible to explore the BTRT results from a different perspective by focusing on the threshold interaction effects that lead to the terminal nodes of the trunk. Additional information derived from BTRT is obtained from the worth parameters estimated in each terminal node, which are conditioned on a specific country-specific covariate. The worth parameters allow us to investigate how the composition of tax revenues varies as the magnitude of a covariate changes in a group of countries with similar socioeconomic characteristics. To demonstrate that this additional information is useful for policymakers, we fix all the covariates at the mean value observed in each terminal node and compute the worth parameters for each tax category as the size of agricultural value-added to GDP varies. This corresponds to simulating agricultural sector size values within the range [0, 1].

Figure 3 shows the results for each terminal node. It highlights some common features of all the groups identified by the terminal nodes. In each node, as the value added by the agricultural sector increases relative to GDP, income tax revenue (black line) decreases. This decrease is consistent with the BTRT estimated coefficients for the share of the agricultural sector. Moreover, in all terminal nodes except for $R_1$ (third row, bottom right), when the contribution of the agricultural sector to the economy is high, countries prefer taxes on goods and services (blue line), whilst resources from property taxes and social contributions are lower (red and green lines, respectively). The only case when this situation does not occur is for countries belonging to $R_1$ (countries belonging to OECD). Here, when the agricultural sector is important, other things being equal, governments prefer to raise resources through social contributions, whilst goods and services and property taxes have approximately the same degree of preference. Finally, countries in $R_6$ (third line, bottom left), which are those more virtuous in terms of debt management and environmental performance, raise their resources mainly through income taxes until the share of the agricultural sector is lower than about 80% of its maximum value. If this share exceeds 80% most of the fiscal resources of central governments come from taxes on goods and services.

The analysis of the composition of tax revenues for different groups of countries based on the size of their agricultural sector can be repeated for any covariate. Once the final trunk and the groups of countries with similar tax revenues have been identified, BTRT further investigates how the composition of government budget revenues varies as long as a country-specific covariate varies within the same group. This investigation can be done by comparing the groups of countries in the terminal nodes of the tree against each other.

To conclude, BTRT is preferable to OLS because of its tree-based structure: an OLS coefficient is not case-specific, as it refers to a conditional expectation. Thus, it is more interesting to estimate specific coefficients for different subsets of observations. In this respect, BTRT identifies automatically a homogeneous subset of cases corresponding to groups of countries with which the same set of coefficients and similar orders of tax revenues are associated. This homogeneous subset identification helps in explaining the structure of, and the differences between, the fiscal and economic systems of the subsets of countries.

## 6.2. Comparing BTRT with LLBT and BTtree

BTRT is compared with LLBT (Sinclair 1982) without country-specific covariates and BTtree (Strobl, Wickelmaier, and Zeileis 2011). LLBT is similar to the BT model of Dittrich, Katzenbeisser, and Reisinger (2000).

Both models are implemented in the `prefmod` R package (Hatzinger and Dittrich 2012), which does not allow for the inclusion of numerical subject-specific covariates. Thus, these models assume the covariates do not affect the order of magnitude of the tax revenue categories. Their main aim is to find a consensus ranking based on a probabilistic approach. Moreover, LLBT is not based on recursive partitioning, thus, different results based on the characteristics of individuals are not available and only the average value of $\lambda$ for each individual is computed. The LLBT's estimated worth parameters (Table 8) are: $\pi_1 = 0.33$, $\pi_2 = 0.11$, $\pi_3 = 0.04$, and $\pi_4 = 0.50$. The consensus ranking in the root node of the BTRT trunk reproduces the same ordering of the items: Table 8 (fourth object) in the first position; taxes on income (first object) in the second position; taxes on property (third object) in the third position and the SSC (second object) in the last position.

The BTtree model, implemented in the `psychotree` R package (Strobl, Wickelmaier, and Zeileis 2011), is instead capable of integrating recursive partitioning and the BT model: splits are found semi-parametrically searching for instability of the basic BT model object parameters.

**Table 8.** Log-linear Bradley-Terry model without subject-specific covariates: results.

| | Estimate | Std. Error | z value | Pr(>\|z\|) |
|---|---|---|---|---|
| (Intercept) | 3.61 | 0.05 | 67.24 | 0.00 |
| o1 | −0.20 | 0.07 | −2.72 | 0.00 |
| o2 | −0.73 | 0.08 | −9.24 | 0.00 |
| o3 | −1.26 | 0.08 | −15.37 | 0.00 |
| o4 | 0.00 | 0.00 | 0.00 | 0.00 |

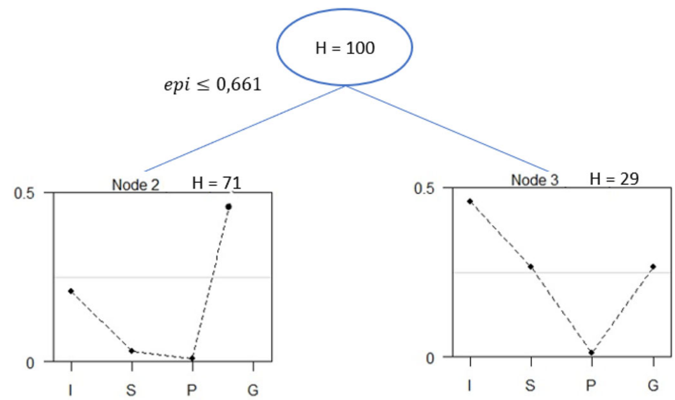


**Figure 4.** Output of BTtree.
*NOTE*: *H* is the number of countries, *I* = taxes on income, *S* = social security contributions, *P* = taxes on property, and *G* = taxes on goods and services. The plot shows the estimated worth parameters in both terminal nodes.

The tree obtained from BTtree (Figure 4) consists of two terminal nodes, with EPI being the splitting covariate. Results align with those derived from BTRT. BTtree does not identify interactions among covariates. After fitting a BT model, it verifies whether the item order established in the root node remains consistent across all observations. Covariates that introduce instability into the item order, and thus into the $\hat{\lambda}$, are prioritized. Employing EPI as the initial split covariate allows for the generation of two distinct orders: For nations where $EPI \leq 0.661$, income tax revenues are inferior to those from goods. In contrast, for $EPI > 0.661$, the situation inversely applies. Additionally, in the left child node, income taxes occupy the second position, while taxes on goods take precedence. Conversely, in the right child node, this order is reversed.

Interaction effects in BTRT allow to further differentiation of the subgroups based on the ordering of their income by taxation. Thus, the final output of BTtree contains less information than those contained in BTRT. BTtree provides the preference scales in each group of the partition that derive from the order of object-related parameters, but it does not specify how the subject-specific covariates affect the judges' preferences. Therefore, this semi-parametric model returns beta coefficients neither for the main effects nor for the interaction effects. In addition, as pointed out in Strobl, Wickelmaier, and Zeileis (2011), the testing procedure for the split search is challenging (Wiedermann, Frick, and Merkle 2021). BTtree is based on the M-fluctuation test which is a score-based procedure (Zeileis and Hornik 2007), whereas BTRT is based on the easy-to-compute decrease in deviance.

## 7. Discussion

This study innovatively examined global tax revenue categories, diverging from traditional analyses, to assess how countries' characteristics influence their tax revenue composition. A specialized dataset was created by merging information from the OECD, IMF, and World Bank for 2018. Tax revenues for 100 countries were classified into four categories and ranked. The Bradley-Terry Regression Trunk (BTRT) model, a novel probabilistic method for analyzing preference rankings, was applied. BTRT allows for an assessment of tax categories across a broad range of countries through pairwise comparisons. The BTRT model segments individuals by considering their tax revenue rankings and the causal link between these rankings and individual characteristics. It operates on the premise that the hierarchy of tax revenues within a country is influenced by its socio-economic traits. Thus, for each country, determinants of tax revenue were modeled as covariates. Our findings added to the discourse on the interplay between tax revenues and government spending by examining four key government expenditure categories: military, education, health, and a residual group encompassing all other types of government spending.

In our application, BTRT partitioned 100 countries into six groups based on several socio-economic covariates. It identified important interactions between these covariates, including both first-order interactions and complex interactions. Moreover, BTRT computed coefficients that helped in ordering tax revenues, shedding light on the impact of significant covariates and their interactions. Unlike traditional Bradley-Terry models which focus on preference data, BTRT offers a more nuanced understanding by probabilistically assessing the causal relationships between covariates and the outcomes in paired comparisons, facilitating the grouping of data into homogenous clusters from the model's terminal nodes. This approach provides a more actionable analysis compared to standard Bradley-Terry models. Compared to Strobl's semi-parametric BTtree, BTRT provides precise estimates of the impact of subject-specific covariates in paired comparisons. Our empirical results revealed several advantages of BTRT: it accommodates a diverse array of countries and fiscal systems thanks to its paired comparison structure; it enables the identification of varied effects of covariates on different tax revenue categories; its fully parametric nature facilitates the detection of threshold interactions, enhancing model fit; and its tree structure offers a clear visualization of countries grouped by similar fiscal revenue characteristics.

In conclusion, this article presents a public finance application that provides valuable insights for policymakers, exploring the impact of economic shocks on tax revenue structures and offering guidance for countries looking to modify their tax compositions. The study suggests that changes in a country's socio-economic characteristics can influence tax revenue amounts without necessitating new tax policy regulations, which are often slow to implement. The article also discusses the limitations of discretionary fiscal policy due to recognition and implementation delays, advocating for the effectiveness of automatic stabilizers that respond immediately to economic changes. Using the BTRT model, the study identifies comparable fiscal policies across diverse fiscal systems by analyzing socioeconomic traits.

## Supplementary Materials

The Appendix includes the descriptive statistics for the subject specific covariates (Table 1) and the tax revenues composition by country in OECD (Figure A1), Africa (Figure A2), Asia (Figure A3) and South America (Figure A4).

## Acknowledgments

Authors thank Nicola Nazzaro (University of Naples Federico II) for help assistance in data collection. Research work of C. Conversano is supported by the European Union - NextGenerationEU, in the framework of the GRINS - Growing Resilient, INclusive and Sustainable project (GRINS PE00000018 – CUP F53C22000760007). Research work of A. D'Amborsio is supported by the European Union - NextGenerationEU, in the framework of the ECSC - Centro Nazionale HPC, Big Data e Quantum Computing (ECSC E63C22000980007 – CUP CN_00000013).

## Disclosure Statement

The authors declare they have no conflict of interest and the findings, interpretations, and conclusions expressed herein do not necessarily reflect the view of the institutions of affiliation.

## ORCID

Antonio D'Ambrosio http://orcid.org/0000-0002-1905-037X

## References

Agbeyegbe, T. D., Stotsky, J., and WoldeMariam, A. (2006), "Trade Liberalization, Exchange Rate Changes, and Tax Revenue in Sub-Saharan Africa," *Journal of Asian Economics*, 17, 261–284. [10]

Agresti, A. (1990), *Categorical Data Analysis*, Hoboken, NJ: A Wiley-Interscience Publication, Wiley. [4]

Anderson, W., Wallace, M. S., and Warner, J. T. (1986), "Government Spending and Taxation: What Causes What?" *Southern Economic Journal*, 52, 630–639. [1]

Baldassarre, A., Dusseldorp, E., D'Ambrosio, A., Rooij, M. d., and Conversano, C. (2022), "The Bradley–Terry Regression Trunk Approach for Modeling Preference Data with Small Trees," *Psychometrika*, 88, 1443–1465. [1,5,8]

Besley, T., and Persson, T. (2014), "Why Do Developing Countries Tax So Little?" *Journal of Economic Perspectives*, 28, 99–120. [10]

Bird, R. M., Martinez-Vazquez, J., and Torgler, B. (2008), "Tax Effort in Developing Countries and High Income Countries: The Impact of Corruption, Voice and Accountability," *Economic Analysis and Policy*, 38, 55–71. [2]

Bird, R. M., Martinez-Vazquez, J., and Torgler, B. (2014), "Societal Institutions and Tax Effort in Developing Countries," *Annals of Economics and Finance*, 15, 185–230. [2]

Bondell, H. D., and Reich, B. J. (2009), "Simultaneous Factor Selection and Collapsing Levels in Anova," *Biometrics*, 65, 169–177. [6]

Boukbech, R., Bousselhami, A., and Ezzahid, E. (2019), "Determinants of Tax Revenues: Evidence From a Sample of Lower Middle Income Countries," *Applied Economics and Finance*, 6, 11–20. [1]

Bradley, R. A., and Terry, M. E. (1952), "Rank Analysis of Incomplete Block Designs: I. The Method of Paired Comparisons," *Biometrika*, 39, 324–345. [2,3]

Breiman, L., Friedman, J. H., Olshen, R. A., and Stone, C. J. (1984), *Classification and Regression Trees*, Belmont, CA: Wadsworth. [4]

Brun, J., and Diakite, M. (2016), "Tax Potential and Tax Effort: An Empirical Estimation for Non-Resource Tax Revenue and vat's Revenue," Working Papers 201610, CERDI. [1,2]

Buera, F. J., Monge-Naranjo, A., and Primiceri, G. E. (2011), "Learning the Wealth of Nations," *Econometrica*, 79, 1–45. [8]

Caselli, F., and Reynaud, J. (2020), "Do Fiscal Rules Cause Better Fiscal Balances? A New Instrumental Variable Strategy," *European Journal of Political Economy*, 63, 101873. [8]

Cassou, S. P. (1997), "The Link between Tax Rates and Foreign Direct Investment," *Applied Economics*, 29, 1295–1301. [2]

Castro, G., and Camarillo, D. (2014), "Determinants of Tax Revenue in OECD Countries over the Period 2001-2011," *Contaduria y Administracion*, 59, 35–59. [2]

Chelliah, R. H., Baas, H. J., and Kelly, M. R. (1975), "Tax Ratios and Tax Effort in Developing Countries, 1969-71," *IMF Staff Papers*, 1975(001), A006. [2]

Chelliah, R. J. (1971), "Trends in Taxation in Developing Countries," *IMF Staff Papers*, 1971(002), A002. [2,10]

Choisel, S., and Wickelmaier, F. (2007), "Evaluation of Multichannel Reproduced Sound: Scaling Auditory Attributes Underlying Listener Preference," *The Journal of the Acoustical Society of America*, 121, 388–400. [3]

Cohen, J., Cohen, P., West, S. G., and Aiken, L. S. (2022), *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*, New York: Routledge. [4]

Conversano, C., and Dusseldorp, E. (2017), "Modeling Threshold Interaction Effects through the Logistic Classification Trunk," *Journal of Classification*, 34, 399–426. [4]

D'Ambrosio, A., Iorio, C., Staiano, M., and Siciliano, R. (2019), "Median Constrained Bucket Order Rank Aggregation," *Computational Statistics*, 34, 787–802. [3]

de González, A. B., and Cox, D. R. (2007), "Interpretation of Interaction: A Review," *The Annals of Applied Statistics*, 1, 371–385. [4]

Dittrich, R., Francis, B., Hatzinger, R., and Katzenbeisser, W. (2006), "Modelling Dependency in Multivariate Paired Comparisons: A Log-Linear Approach," *Mathematical Social Sciences*, 52, 197–209. [3]

Dittrich, R., Hatzinger, R., and Katzenbeisser, W. (1998), "Modelling the Effect of Subject-Specific Covariates in Paired Comparison Studies with an Application to University Rankings," *Journal of the Royal Statistical Society*, Series C, 47, 511–525. [2]

Dittrich, R., Katzenbeisser, W., and Reisinger, H. (2000), "The Analysis of Rank Ordered Preference Data based on Bradley-Terry Type Models," *Or Spektrum*, 22, 117–134. [3,12]

Dusseldorp, E., Conversano, C., and Os, B. J. V. (2010), "Combining An Additive and Tree-based Regression Model Simultaneously: Stima," *Journal of Computational and Graphical Statistics*, 19, 514–530. [4]

Dusseldorp, E., and Meulman, J. J. (2004), "The Regression Trunk Approach to Discover Treatment Covariate Interaction," *Psychometrika*, 69, 355–374. [4]

Eltony, N. (2001), *The Determinants of Tax Effort in Arab Countries*, Kuwait: Arab Planning Institute. [10]

Emond, E. J., and Mason, D. W. (2002), "A New Rank Correlation Coefficient with Application to the Consensus Ranking Problem," *Journal of Multi-Criteria Decision Analysis*, 11, 17–28. [3]

Feltenstein, A., and Cyan, M. (2012), "A Computational General Equilibrium Approach to Sectoral Analysis for Tax Potential: An Application to Pakistan," nternational Center for Public Policy Working Paper Series, at AYSPS, GSU paper1226, International Center for Public Policy, Andrew Young School of Policy Studies, Georgia State University. [2]

Gertheiss, J., and Tutz, G. (2010), "Sparse Modeling of Categorial Explanatory Variables," *The Annals of Applied Statistics*, 4, 2150–2180. [6]

Giuliano, P., Mishra, P., and Spilimbergo, A. (2013), "Democracy and Reforms: Evidence from a New Dataset," *American Economic Journal: Macroeconomics*, 5, 179–204. [8]

Gruber, J. (2004), "*Public Finance and Public Policy*, New York: Worth Publishers, [2005, 2004] ©2005. [1]

Gupta, A. S. (2007), "Determinants of Tax Revenue Efforts in Developing Countries," Working Papers 07/184, International Monetary Fund. [2,9,10]

Hatzinger, R., and Dittrich, R. (2012), "prefmod: An r Package for Modeling Preferences based on Paired Comparisons, Rankings, or Ratings," *JSS Journal of Statistical Software*, 48, 1–31. [12]

Jain, P. (1989), *Economics of Public Finance*, Dehli, India : Atlantic Publishers. [1]

Jiang, T. Y. (2021), "Drelationship between Government Expenditures on Health and Residents' Consumption: New Evidence from China based on the Bootstrap Rolling-Window Causality Test," *Frontiers in Public Health*, 9, 710147. [11]

Kaldor, N. (1963), "Taxation for Economic Development," *The Journal of Modern African Studies*, 1, 7–23. [1]

Karagoz, K. (2013), "Determinants of Tax Revenue: Does Sctorial Composition Matter?" *Journal of Finance, Accounting and Management*, 4, 50–63. [2,10]

Leuthold, J. H. (1991), "Tax Shares in Developing Economies a Panel Study," *Journal of Development Economics*, 35, 173–185. [10]

Lotz, J. R., and Morss, E. R. (1967), "Measuring Tax Effort in Developing Countries," *IMF Staff Papers*, 1967(003), A004. [2,9,10]

Luce, R. D. (1959), *Individual Choice Behavior: A Theoretical Analysis*, New York: Wiley. [6]

Manage, N., and Marlow, M. L. (1986), "The Causal Relation between Federal Expenditures and Receipts," *Southern Economic Journal*, 52, 617–629. [1]

Martin, F., and Uribe Teran, C. (2010), "Economic and Institutional Determinants of Fiscal Effort in Latin America," *Investigacion Economica*, 69, 85–113. [2]

Masarotto, G., and Varin, C. (2012), "The Ranking Lasso and its Application to Sport Tournaments," *The Annals of Applied Statistics*, 6, 1949–1970. [6]

McCullagh, P., and Nelder, J. (1989), *Generalized Linear Models* (2nd ed.), Chapman and Hall/CRC Monographs on Statistics and Applied Probability Series, New York: Chapman & Hall. [4]

OECD. (2018a), "Constructing the Global Revenue Statistics Database," Technical Paper, Organization for Economic Co-operation and Development. [3]

——— (2018b), "Global Revenue Statistics Database," Technical Report, Organization for Economic Co-operation and Development. [3]

Oelker, M.-R., Pößnecker, W., and Tutz, G. (2015), "Selection and Fusion of Categorical Predictors with l0-type Penalties," *Statistical Modelling*, 15, 389–410. [6]

Pessino, C., and Fenochietto, R. (2010), "Determining Countries' Tax Effort," *Hacienda Publica Espanola*, 195, 65–87. [2]

Piancastelli, M. (2001), "Measuring the Tax Effort of Developed and Developing Countries: Cross Country Panel Data Analysis," *IPEA Working Paper* (818). [2]

Rodriguez Montequin, V., Villanueva Balsera, J. M., Diaz Piloneta, M., and Alvarez Perez, C. (2020), "A Bradley-Terry Model-based Approach to Prioritize the Balance Scorecard Driving Factors: The Case Study of a Financial Software Factory," *Mathematics*, 8, 276. [3]

Schauberger, G., and Tutz, G. (2017), "Subject-Specific Modelling of Paired Comparison Data: A Lasso-Type Penalty Approach," *Statistical Modelling*, 17, 223–243. [6]

Schauberger, G., and Tutz, G. (2019), "BTLLasso: A Common Framework and Software Package for the Inclusion and Selection of Covariates in Bradley-Terry Models," *Journal of Statistical Software*, 88, 1–29. [6]

Schauberger, G. J. (2015), *Regularization Methods for Item Response and Paired Comparison Models*, Göttingen: Cuvillier Verlag. [4,6]

Sinclair, C. D. (1982), "Glim for Preference." [4,12]

Stotsky, M. J. G., and WoldeMariam, M. A. (1997), *Tax Effort in Sub-Saharan Africa*, Washington, DC: International Monetary Fund. [10]

Strobl, C., Wickelmaier, F., and Zeileis, A. (2011), "Accounting for Individual Differences in Bradley-Terry Models by Means of Recursive Partitioning," *Journal of Educational and Behavioral Statistics*, 36, 135–153. [3,12]

Tait, A. A., Gratz, W. L., and Eichengreen, B. J. (1979), "International Comparisons of Taxation for Selected Developing Countries, 1972-1976," *IMF Staff Papers*, 1979(001), A005. [2]

Tanzi, V. (1989), "The Impact of Macroeconomic Policies on the Level of Taxation and the Fiscal Balance in Developing Countries," *IMF Staff Papers*, 1989(003), A005. [2]

Teera, J. M., and Hudson, J. (2004), "Tax Performance: A Comparative Study," *Journal of International Development*, 16, 785–802. [2]

Tibshirani, R. (1996), "Regression Shrinkage and Selection via the Lasso," *Journal of the Royal Statistical Society*, Series B, 58, 267–288. [6]

Tutz, G., and Schauberger, G. (2015), "A Penalty Approach to Differential Item Functioning in Rasch Models," *Psychometrika*, 80, 21–43. [6]

Wallace, S., and Bahl, R. W. (2005), "Public Financing in Developing and Transition Countries," *Public Budgeting & Finance*, 25, 83–98. [2]

Wiedermann, W., Frick, U., and Merkle, E. C. (2021), "Detecting Heterogeneity of Intervention Effects in Comparative Judgments," *Prevention Science*, 32, 846–854. [12]

Zeileis, A., and Hornik, K. (2007), "Generalized m-fluctuation Tests for Parameter Instability," *Statistica Neerlandica*, 61, 488–508. [12]

Zeileis, A., Hothorn, T., and Hornik, K. (2008), "Model-based Recursive Partitioning," *Journal of Computational and Graphical Statistics*, 17, 492–514. [3]

Zou, H. (2006), "The Adaptive Lasso and its Oracle Properties," *Journal of the American Statistical Association*, 101, 1418–1429. [6]