

Università degli Studi del Molise
Dipartimento di Economia



ECONOMICS AND STATISTICS DISCUSSION PAPER
No. 098/24

Is Local Taxation Predictable? A Machine Learning Approach

Nicola Caravaggio
Giuliano Resce
Idola Francesca Spanò

The Economics and Statistics Discussion Papers are preliminary materials circulated to stimulate discussion and critical comment. The views expressed in the papers are solely the responsibility of the authors.

(This page intentionally left blank.)

Is Local Taxation Predictable? A Machine Learning Approach

Nicola Caravaggio^{a,†}, Giuliano Resce^{a,††},
Idola Francesca Spanò^{a,†††}

^a*Department of Economics, University of Molise, Campobasso*

[†]nicola.caravaggio@unimol.it, ^{††}giuliano.resce@unimol.it,
^{†††}ispano@unimol.it

September 16, 2024

Abstract

This paper investigates determinants of local tax policy, with a particular focus on personal income tax rates in Italian municipalities. By employing seven Machine Learning (ML) algorithms, we assess and predict tax rate decisions, identifying Random Forest as the most accurate model. Results underscore the critical influence of demographic dynamics, fiscal health, socioeconomic conditions, and institutional quality on tax policy formulation. The findings not only showcase the power of ML in enhancing predictive precision in public finance but also provide actionable insights for policymakers and stakeholders, enabling more informed decision-making and the mitigation of fiscal uncertainties.

Keywords: Local taxation, Machine learning, Municipalities.

JEL Classification: C53, H24, H71.

Acknowledgements

The authors wishes to express their sincere gratitude to Drs. Agapito Emanuele Santangelo and Pietro Cruciata for their insightful comments. Furthermore, the authors are grateful for the precious feedback received during the 36th annual conference of the Italian Association of Public Economics (SIEP).

Funding Statement

We acknowledge funding from Next Generation EU, in the context of the National Recovery and Resilience Plan, Investment PE8 – Project Age-It: “Ageing Well in an Ageing Society” [DM 1557 11.10.2022]. The views and opinions expressed are only those of the authors and do not necessarily reflect those of the European Union or the European Commission. Neither the European Union nor the European Commission can be held responsible for them. This resource was co-financed by the Next Generation EU.

1 Introduction

Local taxation is a topic of increasing interest, mainly since many governments have implemented fiscal decentralization policies, devolving increasing decision-making powers on expenditures and revenues to sub-national levels over recent decades (Garman et al., 2001; Hooghe et al., 2010; Boetti et al., 2012; Hansjörg and Junghun, 2016). In the decentralization process, local taxation decision-making directly affects the autonomy and accountability of local governments, as local authorities’ independence is linked to their ability to finance expenditures from local taxes. Reliance on local taxes enhances the accountability of local policy makers to their constituents, promoting prudent budgeting and fiscal responsibility. Local governments often face budget gaps between revenue and expenses, requiring support from central governments for equalization or control with either state-polycentric or regional-federalist structures. This reliance can reduce fiscal responsibility, as local officials might spend freely without needing to justify tax increases to residents (Devarajan et al., 2009).

The exploration of factors influencing local taxation decision-making has been relatively overlooked in existing literature, except for some relevant political economy studies highlighting the role of political incentives in the process (*e.g.*, Bordignon et al., 2017; Bracco et al., 2019; Ferraresi, 2021b). This paper investigates the determinants of local taxation by implementing a Machine Learning (ML) approach. Although traditional statistical and econometric tools have historically been implemented for exploring these policy issues, recent economic literature suggests a shift toward leveraging advancements in ML for addressing policy problems. This perspective contends that the potential of ML surpasses that of conventional econometric techniques for addressing policy-related prediction problems (Kleinberg et al., 2015). In the context of public finance, ML techniques are gaining traction in evaluating the efficacy of public programs and spending (Andini et al., 2018; Kasy, 2018) and forecasting the potential bankruptcy of local governments (Antulov-Fantulin et al., 2021). In the specific field of taxation, ML techniques have been applied for determining optimal tax rates (Kasy, 2018), predicting tax crimes, detecting

tax fraud, evasion, and avoidance (De Roux et al., 2018; Ippolito and Lozano, 2020; Zumaya et al., 2021; Masrom et al., 2022), tax planning, and resolving tax disputes (Alarie and Xue Griffin, 2022). Additionally, ML is implemented to optimize tax administration policies (Battiston et al., 2024), evaluate the effectiveness of tax systems and reforms (Andini et al., 2018; Abrell et al., 2022), and to predict tax treaty formations (Erokhin and Zagler, 2024).

Accurate predictions of local taxation offer several market benefits. They enable citizens, businesses, and other stakeholders to forecast expenses and make informed decisions. These predictions can guide investment locations and help companies manage operating costs effectively. Additionally, accurate predictions can reduce financial uncertainty, allowing stakeholders to better mitigate risks associated with tax changes and allocate resources efficiently for growth and development initiatives. More generally, predicting local taxation decision-making with ML can reduce information asymmetries by offering clearer insights into tax policy formulation and implementation, allowing for a better understanding of the reasoning behind tax decisions, and enhancing transparency.

As a case study, we consider the personal income tax (PIT) of the Italian municipalities, a surcharge on personal income introduced in 1999. Municipalities have the discretion to set this tax between 0 and 0.8 percent. In 2019, revenues from the municipal PIT amount to 4,670 millions of euros which accounts for nearly 22% of total personal income tax revenues raised in Italy (21,498 millions of euros). By considering the primary source of local revenues, the property tax, for the same year revenues amount to 16.301 euros, instead (Ministero dell’Economia e delle Finanze, 2019).

We predict municipalities that have the maximum personal income tax rate. The analysis relies on administrative, demographic, socioeconomic, institutional, and fiscal municipality characteristics. Starting from the set of features at time $t - 1$, we predict our target variable in time t by implementing seven different ML algorithms: Elastic Net, Gradient Boosting, k-Nearest Neighbors, Naïve Bayes, Neural Network, Random Forest, and Support Vector Machine. Results show that local taxation can be accurately predicted out of the sample,

and we identified Random Forest as the best algorithm for this task. Among the features with the highest predictive importance, we find demographics, financial, socio-economic, and institutional quality indicators.

The remainder of the paper is structured as follows: Section 2 analyzes the concepts of decentralization and local taxation in the literature; Section 3 presents the institutional setting and the data; Section 4 describes the methodology adopted while Section 5 evaluates the predictive task and identifies the core determinants of local taxation; Section 6 concludes the paper highlighting policy implications.

2 Decentralization and local taxation in the literature

The optimal level of fiscal autonomy that should be granted to local authorities has garnered significant interest in both the public debate and the academic world. Traditional claims in favor of fiscal decentralization emphasized a closer alignment between citizens' preferences and public policies as well as enhanced accountability of locally elected policy makers (Weingast, 2009), while those in disagreement leverage the associated drawbacks, such as worse quality of public services, increased disparities and instabilities (Prud'Homme, 1995).

In the economics literature, the impact of fiscal decentralization has been extensively analyzed (see Martínez-Vázquez et al., 2017 for a comprehensive review). One strand of literature has been focused on the effects on the provision of public services, providing empirical evidence of a positive impact of decentralized systems on academic performance (Sebastián and Schargrotsky, 2001; Faguet, 2004; Clark, 2009; Falch and Fischer, 2012), health (Jiménez-Rubio, 2011; Porcelli, 2014), and infrastructure (Bardhan and Mookherjee, 2006). Mixed results emerge from other fields of studies concerning income inequality and poverty (Sacchi and Salotti, 2014; Hung and Thanh, 2022; Pietrovito et al., 2023; Tselios, 2023), development and economic growth (Bardhan, 2002; Martinez-Vazquez and McNab, 2006; Baskaran and Feld, 2013;

Gemmell et al., 2013; Filippetti and Sacchi, 2016). More recently, Bucci et al. (2024) show that fiscal decentralization has a positive impact on municipality efficiency in the Italian context.

Local governments' autonomy heavily relies on having adequate resources to perform their functions, making the taxation aspects highly significant (Resce, 2022). In this regard, many researchers have investigated the impact of variations in the local tax system on a variety of outcomes, directing their attention toward both individual and firm taxes. Individual-level taxes, such as personal income and property taxes, have been a focal point, with a strong emphasis on the link between tax rates and mobility—both national and international (see Kleven et al., 2020 for a recent review). Studies targeting specific segments of the population, such as individuals aged between 21 and 64 years (Liebig et al., 2007), high-income foreign workers (Schmidheiny and Slotwinski, 2018), football players (Kleven et al., 2013), the elderly (Shan, 2010), and rich taxpayers (Young et al., 2016; Martinez, 2017) have provided evidence of a significant effect of taxation on individuals' decisions to relocate. The Italian context mirrors these results. Rubolino (2020) finds that variations in local income tax significantly affect the location of the tax base and the probability of changing residence for a broad segment of workers.

Shifting the attention to how corporate taxation shapes a variety of firms' decisions, Becker et al. (2012) highlight how higher municipal business tax rates deter the attraction of new multinational corporations. Flipping the direction of the analyzed impact, Langenmayr and Simmler (2021) show that governments set higher corporate tax rates in response to declining capital mobility. Poli (2019) finds that Italian private companies strategically shifted profits between fiscal years to capitalize on a temporary decrease in the standard tax rate. This adaptability highlights how firms exploit variations in local tax systems to minimize their tax burden.

With special reference to fiscal decision-making within the Italian framework, some researchers have focused their attention on the strategic interactions observed in tax collection among local governments. The introduction of a local tax on personal income in 1999 marked a crucial step in enhancing the

fiscal autonomy of Italian municipalities and regions. This tax empowers local authorities to levy the tax and set the local tax rate within a state-mandated range, thereby enabling them to tailor revenue collection to their specific needs. However, because the central government has retained full powers over personal income tax (PIT), it has been described as a “less visible and transparent” tax instrument than others (*e.g.*, the municipal property tax), in the sense that citizens might have problems in disentangling the part of total PIT due to municipal decisions (Bordignon and Minelli, 2001).

In this regard, Bordignon et al. (2017) examined the role of political incentives on resource allocation and governance dynamics at the municipal level. The study demonstrated that mayors in their first term in office are inclined to adopt a greater surcharge on PIT—the less transparent tool—and reduce the property tax rate—the more transparent source of tax revenue—more significantly in comparison to mayors in their final term to enhance their chances of re-election. Similar results are obtained by Ferraresi (2021b) who enriches the analysis by exploring spatial interactions among neighbouring municipalities. He found that municipalities tend to increase the PIT surcharge in the lead up to an election. Moreover, his study provides evidence of positive and sizable spatial interactions between municipalities for the PIT surcharge, whereas decisions on property tax (here identified as more salient tax for citizens) appear to be not spatially correlated.¹

The rise of flat income tax system in several countries has given rise to significant debate among researchers and policymakers in recent decades about its impact on economic growth, income distribution, and labor market performance (Keen et al., 2008; Jacobs et al., 2010; Ji, 2013; Remeta et al., 2015). By means of a simulation of three different scenarios in Italy, Socci et al. (2022) provides evidence of a negative effect on real GDP, coupled with an

¹The PIT surcharge is considered a less salient tax instrument for local policy makers (Bracco et al., 2019) and probably one of the least known by citizens. In fact, according to a survey conducted by Censis (2011) about the most detested taxes by Italians, the national PIT was ranked third while its municipal surcharge was not even mentioned in the ranking. This outcome was also confirmed by another survey conducted by an Italian national newspaper over the period 1999–2006 (Bordignon et al., 2017) and, more recently, by a survey conducted by the Associazione Contribuenti Italiani (2022).

increase in inequality, with a strong advantage for rich households. Analyzing the evolution of the Italian PIT since its introduction in 1974, [Baldini \(2021\)](#) finds that, after 40 years and contrary to expectations, it has become more redistributive and slightly more progressive.

Rather than drawing causal inferences, the present paper adopts a machine learning (ML) approach to analyze the predictability of municipal decisions regarding personal income tax. In this regard, we frame the local taxation decision-making as a prediction policy problem ([Kleinberg et al., 2015](#)). Predicting local taxation decision-making helps in reducing information asymmetries by providing clearer insights into how tax policies are formulated and implemented. A ML approach in this task has the potential to enable citizens, businesses and other stakeholders to better understand the reasoning behind tax decisions. The Italian context provides an excellent opportunity for investigating this issue, given the high discretion municipalities enjoy in setting personal income surcharge tax rates. This discretion is accompanied by considerable heterogeneity in tax revenues and public expenditures ([Bracco et al., 2015](#); [Lagravinese et al., 2018](#); [Bucci et al., 2024](#)).

3 Institutional setting and Data

Italian Municipalities derive their revenue from a combination of self-generated sources and transfers from higher levels of government, primarily the central administrations ([Bracco et al., 2015, 2019](#)). These self-generated sources, that is own taxes and fees, can be further categorized into three major types: (i) own municipal tax (IMU), a property tax levied on all real estate assets, excluding agricultural land and primary residences; (ii) a surcharge on the national personal income tax (IRPEF); (iii) waste tax (TARI), a waste disposal tax that varies based on the property type and the number of occupants. Tax revenues are mostly used to deliver public services ([Bucci et al., 2024](#)). In addition to these sources there is the fee for the occupation of public spaces and areas (*Canone patrimoniale di concessione, autorizzazione o esposizione pubblicitaria*) and the possibility for municipalities to establish specific taxes (*di scopo*) aimed

at the construction of public works. Eventually, capital municipalities, unions of municipalities, and the municipalities included in the regional lists of tourist locations or cities of art can establish a tourist tax.²

3.1 Municipal personal income tax

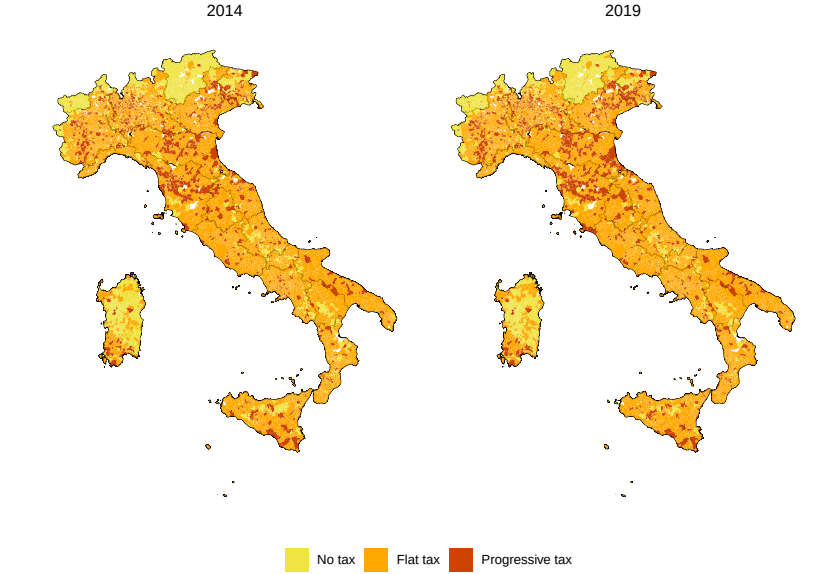
Italian municipalities have gained an additional revenue source since 1999. Legislative Decree 360/1998 grants them the discretionary power to establish a surcharge on the national PIT. This local surcharge, known as *Addizionale IRPEF*, has a maximum rate of 0.8 percent, except for cases expressly provided for by the law, such as the case of Rome, which can levy a rate of up to 0.9 percent. Further flexibility was introduced in 2007, allowing municipalities to implement an exemption threshold based on income levels. This means that the surcharge does not apply to residents whose income falls below or meets the established limit. Finally, since 2011, municipalities have had the option of a single flat rate or a progressive tax scheme for the surcharge. If they choose the latter system, the tax brackets must mirror the national IRPEF structure, with rates increasing progressively alongside income levels.

Figure 1 compares the taxation system adopted by municipalities in 2014 (left panel) and 2019 (right panel). In 2014, the majority of the municipalities (70.34%) adopted a flat tax. Progressive taxation is adopted by almost 12.49% and a larger proportion of municipalities (17.17%) opted out of any surcharge. By 2019, the proportion of municipalities adopting a flat rate slightly increases (70.76%). The greater variation is observed between municipalities adopting a progressive tax system and those not levying any surcharge; the former increases to 14.24%, while the latter decreases to 15%.

Figure 2 shows the share of municipalities that levies the highest PIT rate. Over the period under consideration, this proportion saw a sharp rise from 24.35% in 2014 to 28.8% in 2015, followed by a period of relative stability.

²In the period to which our analysis refers (2018–2019), local governments could rely on another tax instrument, a tribute for indivisible services (TASI) imposed on the use of indivisible public services, such as street lighting, waste collection, and public transportation. This tax has been abolished in 2020, compensated for by a slight increase in the IMU tax rate.

Figure 1: *Municipality PIT schemes*



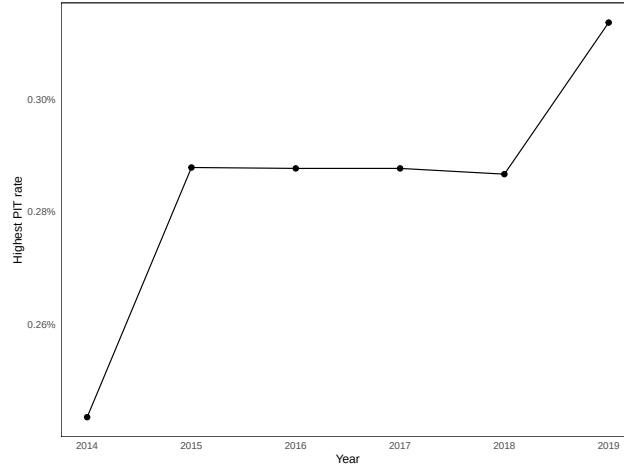
Notes: The figure displays the evolution in the municipality PIT schemes. It shows whether a municipality has set a PIT surcharge and, if so, whether it is a flat rate or a progressive tax scheme. Source: Authors' elaborations on data from the Italian Ministry of Economy and Finance and Istat.

After 2018, the share of municipalities levying the highest PIT rate significantly increases to 31.4%t.

The reason for our interest in investigating the role of a variety of characteristics as potential key determinants of municipal taxation stems from differences in tax burdens across municipalities. To elaborate on this aspect further, Figure 3 presents the proportion of municipalities levying the highest PIT for two groups: municipalities with a proportion of population aged 65 and over (namely, an older population) greater than its median value (23.64) and those with a share of older population below the median. Although in both groups the proportion of municipalities adopting the highest taxation increases over time, those with a younger population present a higher tax burden than those with a larger elderly population in all periods.

At the same time, the data also reveals differences in taxation among municipalities with different levels of accessibility to essential services: high schools, hospitals, and railway stations (Barca, 2019). In particular, in Figure

Figure 2: *Share of municipalities with the highest PIT rate*



Notes: The figure displays the evolution in the proportion of municipalities that levy the highest PIT rate over the period 2014-2019.
Source: Authors' elaborations on data from the Italian Ministry of Economy and Finance and Istat.

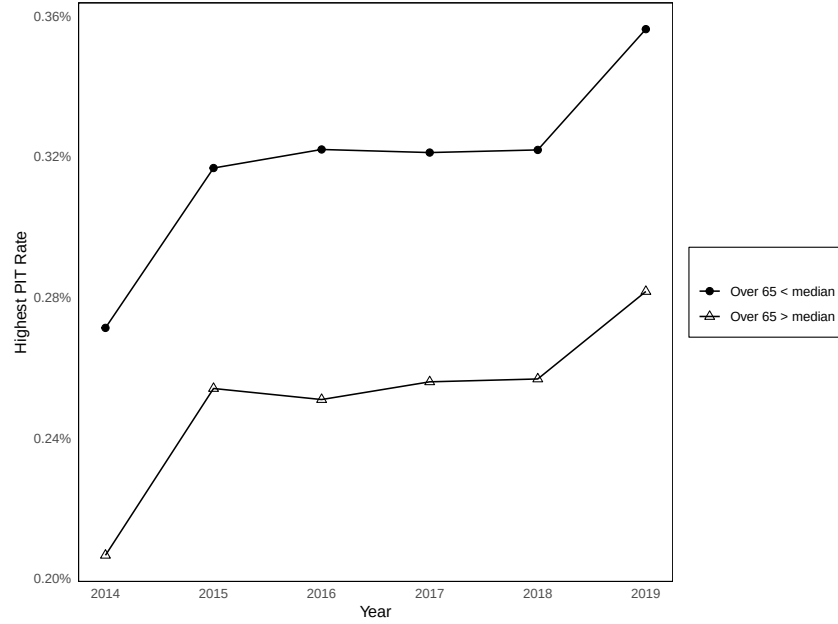
4 we plot the trend in the proportion of municipalities levying the highest PIT among municipalities located in inner areas—defined according to the distance from centers offering essential services, divided into intermediate (small distance), peripheral (medium distance), and extremely peripheral (high distance)—and those other municipalities located in the non-inner area.³ Municipalities in inner areas always present a lower proportion of those adopting the highest PIT rate compared to other municipalities. In addition, the proportion of municipalities with higher tax rates significantly decreases as the distance from centers increases.

3.2 Data

The analysis relies on a rich dataset of Italian municipalities drawn from a combination of different public sources of data provided by the Italian Ministry of Economy and Finance (MEF, 2024), the Italian Ministry of the Interior (MINT, 2024a,b), the Italian Presidency of the Council of Ministers (PCM, 2024)

³We adopted the classification of the National Strategy for Internal Areas (*Strategia Nazionale Aree Interne* – SNAI) for period 2021–2027. See Monturano et al. (2023) for a detailed description of inner areas classification.

Figure 3: *Share of population over 65 and highest PIT rate*

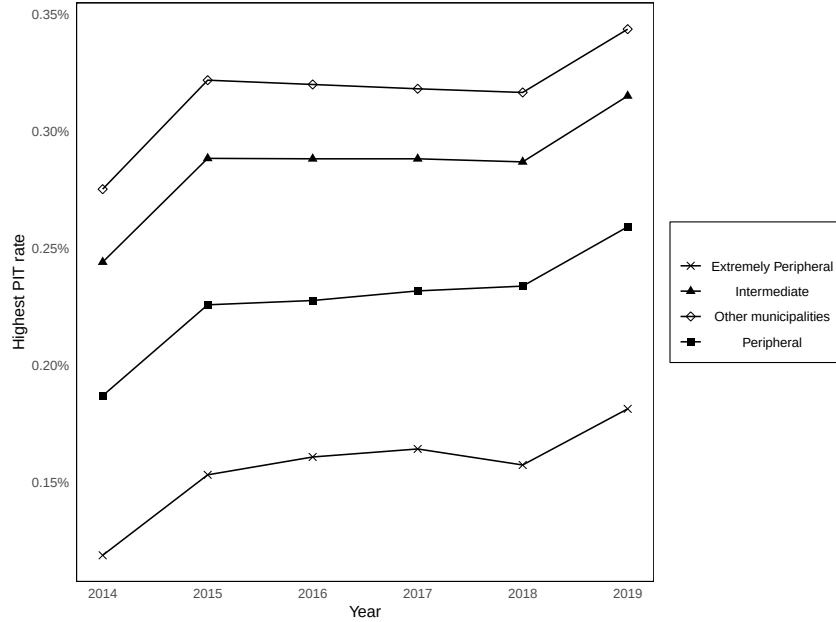


Notes: The figure displays the evolution in the proportion of municipalities adopting the highest PIT rate over the period 2014-2019 according to the proportion of population aged 65 and over. Source: Authors' elaborations on data from the Italian Ministry of Economy and Finance and Istat.

the Italian National Institute of Statistics ([Istat, 2024a,b,c](#)), and [Fondazione Ca' Foscari \(2024\)](#). It provides a full range of municipal information that we organized into four main categories:

1. *Demographics and Geography (DG)*, it includes information on population size, land consumption and population density, whether a municipal is classified as inner area or a center, proportion of population aged 65 and over, mortality, and natality rate. Data for this category is mainly sourced from "*Demo - Demografia in cifre*" ([Istat, 2024b](#)) and "*A Misura di Comune*" ([Istat, 2024a](#)), a specific tool tailored to municipal data;
2. *Socioeconomics (SE)*, this category focuses on labor market outcomes and human capital information, such as employment and unemployment rates, the density of active local units, municipal expenditures on social interventions and services for the elderly, and controls on the education level of the population;

Figure 4: *Distance from essential services and highest PIT tax rate*



Notes: The figure displays the proportion of municipalities adopting the highest PIT rate over the period 2014-2019 for inner areas and centers.
Source: Authors' elaborations on data from the Italian Ministry of Economy and Finance and Istat.

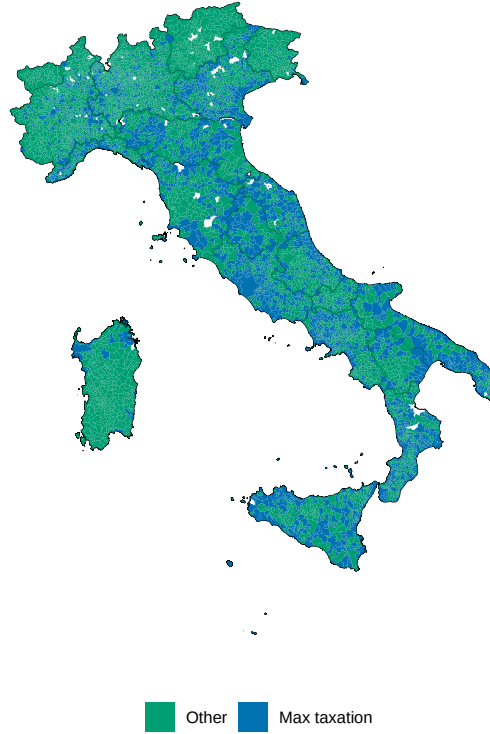
3. *Institutional (IS)*, this section covers a wide range of information on local administrator characteristics including gender, age, and level of education of the mayor retrieved primarily from [MINT \(2024a\)](#). Institutional quality is assessed both at the municipal level, through an assemblage of information on the municipal council's members and employees and, at the provincial level, through the indices developed by [Nifo and Vecchione \(2024\)](#);
4. *Fiscal and Financial (FF)*, this category includes data on the municipal taxation system, taxable income, stock of taxpayers and local tax rate. Note that we focus on personal income taxation only. It also includes a range of parameters to identify the municipal's structural deficit.

The full range of variables used in the analysis is listed in Table [A1](#) in the [Appendix](#) with relative descriptive statistics.⁴

⁴Due to missing values in some municipal characteristics, we restrict our analysis to data

The dependent variable Tax_i is a binary variable that takes the value of 1 if a municipality adopts the highest PIT, and the value of 0 otherwise. In our sample, roughly 31.36% of municipalities levy the highest PIT rate. Figure 5 shows the spatial distribution of these municipalities.

Figure 5: *Municipals with the highest tax rate (2019)*



Notes: The figure displays the spatial distribution of municipalities levying the highest tax rate.
Source: Authors' elaborations on data from the Italian Ministry of Economy and Finance and Istat.

4 Methodology

The prediction task is formulated as follows: for each municipality i , based on the set of features $\{DG, SE, IS, FF\}_i$ (which refer to 2018), it finds the function $f(\cdot)$ (machine learning model) that predicts the probability of having

for the years 2018 and 2019. This results in a sample of 7,825 municipalities (98.87% of all Italian municipalities active between 2018 and 2019).

the highest flat rate (in 2019) Tax_i (*i.e.*, equal or greater than 0.8%):

$$\{DG, SE, IS, FF\}_{i,t-1} \xrightarrow{f(\cdot)} Tax_{i,t} \quad (1)$$

Following the predominant approach in the literature using ML models, we randomly divide the database as 70% for training (train) and 30% for the out-of-sample testing set (test).⁵ The hyper-parameter optimization is only applied to the training set using a repeated (5 times) five-fold cross-validation.⁶ To carry out our analysis, we use seven different ML predictive algorithms. In addition, we also considered a more "classical" logistic regression (LR) as baseline model for completeness. The ML algorithms implemented are presented below:

Elastic Net (EN), a regression statistical method that performs features selection and regularization with a mix of L1 (LASSO-type) and L2 (Ridge-type) penalization to reduce over-fitting and increase prediction accuracy and interpretability (Tibshirani, 1996; Zou and Hastie, 2005);

Gradient boosting (GBM), an ensemble method which works in an iterative way whereby, at each stage, a new learner tries to correct the pseudo-residual of its predecessors (Friedman, 2001);

k-Nearest Neighbors (KNN), a non-parametric algorithm which classifies a data point based on the majority class of its k-nearest neighbors in the feature space (Dasarathy, 1991);

Naïve Bayes (NB), a probabilistic algorithm based on Bayes' theorem, assuming independence between features (Lewis, 1998);

Neural Network (NN), a model that uses a set of connected input/output units in which each connection has a weight associated and learns by adjusting the weights to predict the correct class label of the given inputs (Venables and Ripley, 2002; Ripley, 2014);

⁵Even though the split was causal, a sample coherence was established to have a proper representativeness of all municipalities between the train and test samples.

⁶To account for the relatively high imbalance between the two analyzed classes (0 and 1), we weighted individuals appropriately to rebalance our classes in the training set.

Random Forest (RF), a family of randomized tree-based classifier decision trees that uses different random subsets of the features at each split in the tree (Breiman, 2001);

Support Vector Machine (SVM), an algorithm which aims to find a hyperplane in a high-dimensional feature space that best fits the data points while minimizing the error between the predicted and actual target values (Drucker et al., 1996).

The performance of our classification prediction—the probability of having the highest tax rate—is assessed by analyzing the Receiver Operating Characteristics (ROC) curve (Fawcett, 2006). In our binary classification problem, the ROC curve serves as a visual representation of a classifier’s performance across various discrimination thresholds by illustrating the trade-off between the true positive rate (sensitivity), on the *y-axis*, and the false positive rate (1 - specificity), on the *x-axis*. In ML, particularly in classification tasks, the ROC curve is a valuable tool for assessing the diagnostic capability of a model. As the classification threshold is adjusted, the ROC curve showcases how well the model distinguishes between positive and negative instances. In scenarios where the classification task is entirely unpredictable, the ROC curve aligns with the diagonal line, resulting in an Area Under the Curve (AUC) of 0.5. A model that performs perfectly would yield an AUC of 1, indicating flawless discrimination between classes. In essence, a higher AUC identifies a more predictive model.⁷

5 Results

This section presents the results of the model predicting local taxation. The focus will be on two main aspects: the predictability of our dependent variable (Section 5.1) and the features’ importance of independent variables used for predictions (Section 5.2).

⁷The analysis has been conducted in R using the package `caret`.

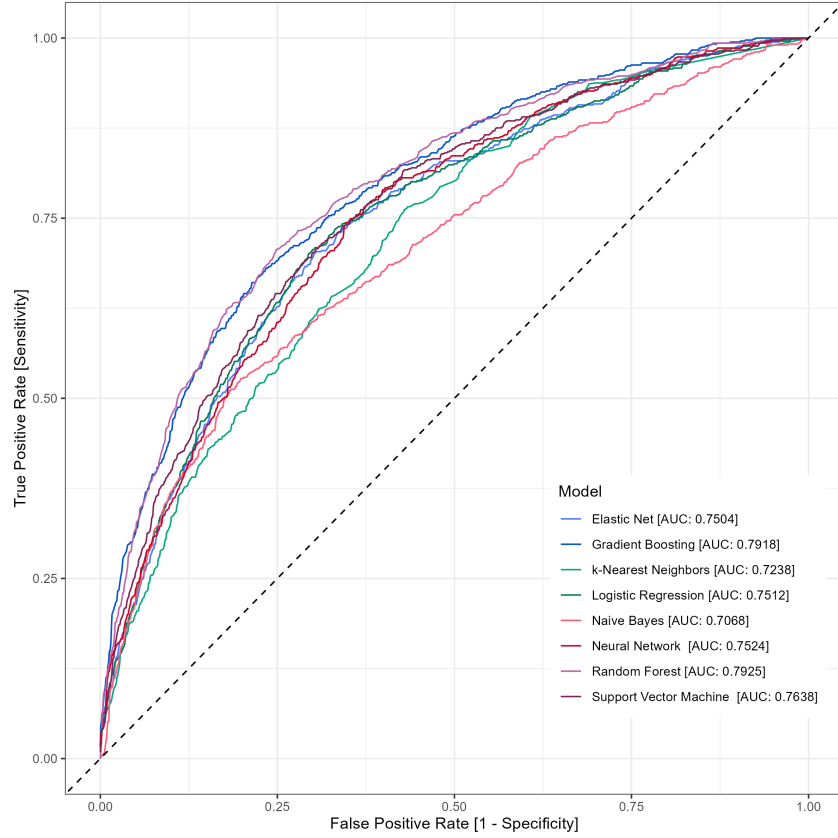
5.1 The prediction task

The AUC in Figure 6 shows that the RF model outperforms all the others.⁸ This is also confirmed by looking at other measures of performance shown in Table 1. In particular, the accuracy of RF is statistically higher than the no-information rate. This is in line with previous empirical applications, confirming that tree-based models are among the more competitive methods for structured binary tasks (Antulov-Fantulin et al., 2021; Resce and Vaquero-Piñeiro, 2022). The second-best performer model is GBM, still within the framework of tree-based algorithms. Furthermore, RF and GBM also have the highest value of the F1 score, a performance index suitable for classification tasks with unbalanced data, as in our case. These results show that local tax rates can be accurately predicted using local-level lagged features.

Thereafter, once the proficient predictability of the models has been verified, we performed the DeLong test to determine whether the AUCs of two models are statistically significantly different (DeLong et al., 1988). We compared the ROC curve of our best model (RF) with all other algorithms and results are shown in Table 2. We do not reject the null hypothesis at 1% only for GBM. This means that, compared to our baseline logistic regression analysis, the ML approach results are better in terms of predictability.

⁸Here are reported the best-tuned hyperparameters for each implemented algorithm. EN: `alpha` (balancing between Lasso-L1 and Ridge-L2 regularization) = 0.55, `lambda` (regularization's strength) = 0.002491188. GB: `interaction.depth` (maximum depth of the trees) = 3, `n.minobsinnode` (minimum number of observations in a node) = 10, `n.trees` (number of trees to fit) = 150, `shrinkage` (learning rate) = 0.1. KNN: `kmax` (maximum number of neighbors) = 9, `distance` (distance between data points) = 2 (Euclidean), `kernel` = optimal (Epanechnikov kernel). NB: `adjust` (smoothing adjustment to density estimation, to avoid zero probabilities) = 1, `kernel` = TRUE (use of kernel density estimation), `laplace` = 0 (No Laplace smoothing applied). NN: `decay` (control for weight decay) = 0.0001, `size` (number of units, or neurons, in the hidden layer) = 5. RF: `ntree` (number of trees to fit) = 500, `mtry` (number of features randomly selected in each split) = 36. SVM: `sigma` (controls decision boundary curvature) = 0.009981415, `C` (penalty for misclassification) = 1.

Figure 6: *Models' ROC curves*



Notes: models trained on 70% of observations and tested on the remaining 30%. The ROC curves are calculated on the test set (30% of the data, the same for each algorithm).

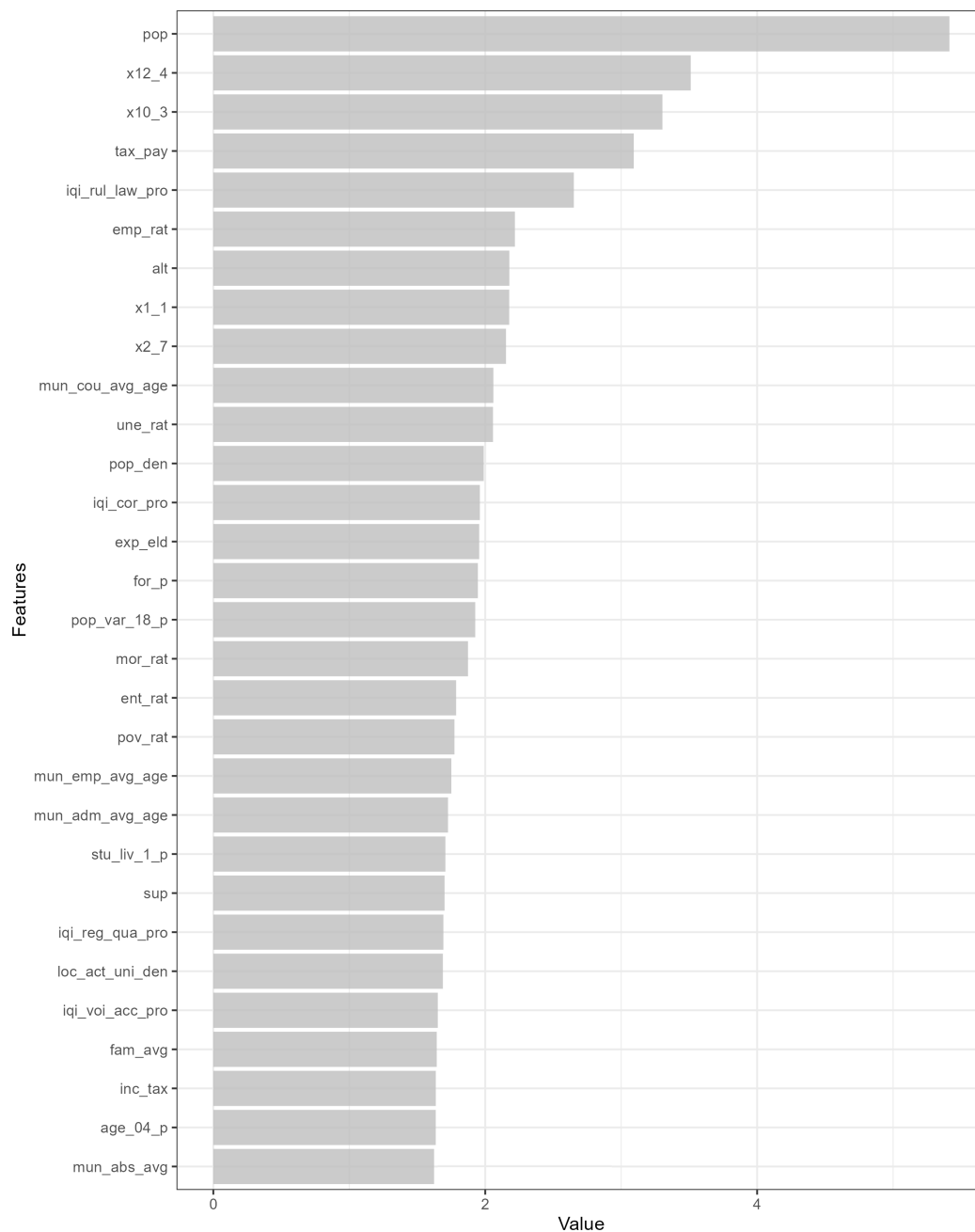
5.2 The feature importance

Once we identified the best ML model and verified that its predictability performance outperformed the basic logistic regression, we scored our features based on their predictability importance (meaning how much the model's performance decreases when a specific feature is permuted during the prediction process) by relying only on the best performing algorithm (RF). The graphical representation in Figure 7 helps to better understand the core determinants in predicting taxation decisions, where all features' importance has been standardized.⁹

Demographics factors take the main stage in the importance ranking, with

⁹The feature importance is calculated over the entire data frame (train and test) by selecting the best hyperparameters obtained from the trained model.

Figure 7: *Features importance with RF*



Notes: The feature importance is calculated over the entire data frame (train and test) by selecting the best hyper parameters obtained from the trained model of the best selected ML algorithm. The figures reports only the first 30 features sorted by their predictive importance.

Table 1: *Models' performances*

Metrics	Logistic Regression	Gradient Boosting	Elastic Net	k-Nearest Neighbors	Naive Bayes	Neural Network	Random Forest	Support Vector Machine
AUC	0.75117	0.79179	0.75045	0.72384	0.70684	0.75244	0.79254	0.76382
Accuracy	0.71672	0.75160	0.71289	0.70651	0.60868	0.72820	0.76818	0.74734
Kappa	0.12965	0.28035	0.10936	0.29237	0.22410	0.20616	0.40254	0.31370
Accuracy Lower 95%	0.69803	0.73361	0.69413	0.68764	0.58861	0.70973	0.75058	0.72926
Accuracy Upper 95%	0.73486	0.76895	0.73111	0.72486	0.62847	0.74610	0.78511	0.76480
No Information Rate	0.69290	0.69290	0.69290	0.69290	0.69290	0.69290	0.69290	0.69290
Accuracy p-value	0.00626	0.00000	0.01836	0.07908	1.00000	0.00010	0.00000	0.00000
Mcnemar p-value	0.00000	0.00000	0.00000	0.01334	0.00000	0.00000	0.00000	0.00000
Sensitivity	0.98281	0.97115	0.98588	0.80847	0.57029	0.95948	0.90301	0.92511
Specificity	0.11634	0.25623	0.09695	0.47645	0.69529	0.20637	0.46399	0.34626
Pos Pred Value	0.71505	0.74658	0.71125	0.77699	0.80853	0.73174	0.79171	0.76150
Neg Pred Value	0.75000	0.79741	0.75269	0.52439	0.41764	0.69302	0.67951	0.67204
Precision	0.71505	0.74658	0.71125	0.77699	0.80853	0.73174	0.79171	0.76150
Recall	0.98281	0.97115	0.98588	0.80847	0.57029	0.95948	0.90301	0.92511
F1	0.82782	0.84418	0.82634	0.79242	0.66883	0.83028	0.84371	0.83537
Prevalence	0.69290	0.69290	0.69290	0.69290	0.69290	0.69290	0.69290	0.69290
Detection Rate	0.68099	0.67291	0.68311	0.56019	0.39515	0.66482	0.62569	0.64100
Detection Prevalence	0.95236	0.90132	0.96044	0.72097	0.48873	0.90855	0.79030	0.84177
Balanced Accuracy	0.54958	0.61369	0.54142	0.64246	0.63279	0.58293	0.68350	0.63568

Notes: results are estimated on the confusion matrix, which shows a cross-tabulation of the observed and predicted classes, generating the predicted classes based on the typical 50% cutoff for the probabilities (Kuhn et al., 2020).

total resident population (**pop**) being the most important feature. The variables related to the financial stability of municipalities follow in terms of importance, in the order sustainability deficit borne by the year (**x12_4**) and financial debts sustainability (**x10_3**), respectively. Within fiscal and financial data we can find also budget rigidity (**x1_1**), the proportion of own-source revenue collected relative to forecasts (**x2_7**) as well as the number of taxpayers (**tax_pay**). Within the group of factors associated with demography and geography we have the municipality altitude (**mun_alt**) and size (**sup**), the population density (**pop_den**), the share foreign residents (**for_p**), the mortality rate (**mor_rat**), and the variation in the population (**pop_var_18_p**). Indirectly linked to the age structure of the population we noticed an interesting important predictive features: the share of local welfare expenditure for elderly (**exp_eld**). Other important features pertain to the socio-economics category, whereby employment rate (**emp_rat**) among the first top ten features followed by unemployment (**une_rat**) and entrepreneurship (**ent_rat**) rates as well as by incidence of poverty (**pov_rat**). Of particular interest are the institutional quality indices, both at the provincial and municipal level, identified as important features

Table 2: *Models' DeLong test, Random Forest against all other algorithms*

Model	stat z	p-value
Logistic Regression	5.0912	0.0000
Gradient Boosting	0.1396	0.8890
Elastic Net	5.3248	0.0000
k-Nearest Neighbors	6.4475	0.0000
Naive Bayes	9.1749	0.0000
Neural Network	4.8757	0.0000
Support Vector Machine	4.1408	0.0000

in the predictive task. These include rule of law, regulatory quality, and corruption control at provincial level, and the average age of municipal council members (`mun_cou_avg_age`) and employees (`mun_emp_avg_age`) as well as the average number of absences (`mun_abs_avg`) at municipal level.

Once we identified the most important features in predicting local taxation, we were interested in unveiling the relationship between those features and our target variable. In fact, results shown in Figure 7 only provide information about the predictive importance, not about the sign. Consequently, in order to disentangle those relationships, we relied on partial dependency plots (PDPs) without the need for a predefined mathematical model for the functional relationship. PDPs are obtained by evaluating the model predictions while varying the selected feature over its range holding all other features constant at their average (Hastie et al., 2009). The PDPs for a selected group of the most important features are shown in Figure 8.

The first feature in terms of importance, population (`pop`), show a positive S-shape relationship with the probability of having the maximum PIT. Furthermore, 5,000 is approximately the number of citizens which represent the threshold.¹⁰ Although not reported in Figure 8, the variable of tax payers (`tax_pay`) shows a similar pattern (rightly with a lower threshold) while the positive association of population density (`pop_den`) with our target variable has a concave shape. These three features stress how the size of the municipal-

¹⁰In our sample municipalities with more than 5,000 inhabitants are just over 30%.

ity is a crucial element in pushing local administration to maximize the tax burden through the PIT surcharge. These results are somewhat in line with the research of [Bracco et al. \(2019\)](#), who focused on larger Italian municipalities where they stressed how local politicians rely intensively on less salient tax instruments, such as the PIT, in more competitive jurisdictions. Furthermore, it also fits within the wider range of literature which investigates the relationship between fiscal decentralization and local government size (*e.g.*, [Liberati and Sacchi, 2013](#)). Still within the group of demographic factors, it is interesting to observe the positive relationship between the share of foreign residents (`for_p`) and the municipality PIT investigated. This could be associated with a possible lower awareness for foreign inhabitants of the Italian fiscal system which may lead local policymaker for further rely on the PIT surcharge.

An interesting variable, indirectly, associated with the group of demographic features, is the share of welfare expenditure at municipal level for elderly (`exp_eld`). An higher value for this variable would be either associated to a larger share of elderly population in the municipality or a greater importance given to them by the local policy maker. The opposite in the case of low shares. The association of this feature with our target variable depicts a U-shape pattern which means that the maximum probability of taxation is associated with municipalities with either a relatively particularly young or elderly population. Because the specific income tax under consideration is not directly observable by taxpayers, local policymakers may take advantage of the lowered awareness and knowledge of these two groups of citizens by deliberately bringing the PIT tax rate to its maximum. For example, [Reschovsky \(1989\)](#) shows how local taxation in Massachusetts is higher for elderly people.

The features of employment (`emp_rat`) and unemployment rate (`une_rat`) show a similar but mirrored trend. The former has a negative association with our target variable, stressing how municipalities where the job market is particularly flourishing and dynamic tend not to apply maximum income taxation. The opposite occurs for municipalities characterized by higher levels of unemployment. Since we are investigating personal income taxation, the higher the presence of an unemployed person, the lower the tax base from

which resources can be drawn. Therefore, to guarantee the same resources, the local policymaker is compelled to 'squeeze' the remaining tax base to the maximum. The opposite happens for municipalities where a large part of the population is employed which implies a greater tax base on which to rely. Despite a different framework, it is worth stressing how higher (labor) taxation increases unemployment as demonstrated by [Daveri and Tabellini \(2000\)](#), hence this positive association between taxes and unemployment rate.

By looking at some institutional indices (at the provincial level) we find a negative association of both rule of law (`iqi_rul_law_pro`) and voice accountability (`iqi_voi_acc_pro`) with taxation. In the first case the variable accounts for crime against persons or property, magistrate productivity, trial times, tax evasion, and shadow economy (to be interpreted in reverse order). These elements may lead to cases where policymakers are more self-oriented, hence prone to push local income taxation. Furthermore, the presence of tax evasion leads to a lower tax base which, in turn, similar to the unemployment rate, may lead to the necessity to further rely on the remaining taxable incomes with higher rates. Moreover, there is a negative relationship between tax progressivity and tax evasion as shown by [Rubolino \(2023\)](#). In the second feature, instead, the variable is constructed by considering citizens' degree of participation in public elections, civic and social associations, the number of social cooperatives, the INVALSI¹¹ test, and the cultural liveliness measured in terms of books published. This variable could be interpreted as the political participation of citizens as well as their civil knowledge. Therefore, these areas may be characterized by greater citizen control over the actions of local politicians which in turn reduces the likelihood that they will disproportionately increase local taxation. This result follows the literature which stresses the higher voice requested by citizens when governments try to tax them ([Weigel, 2020](#)). Furthermore, the awareness and participation of citizens, which embody

¹¹The INVALSI national tests (National Institute for the Evaluation of the Educational and Training System – *Istituto Nazionale per la valutazione del sistema educativo di istruzione e di formazione*) are written tests carried out every year for all Italian students for specific classes. INVALSI aims to evaluate the student's skill levels in Italian, Mathematics, and English.

their influence on local policy makers, may lead them to even reduce maximum levels of the PIT, a tool used by mayors in exchange for a reduction in local property tax—much more visible to citizens—(Bordignon et al., 2017), even in order to be re-elected (Bracco et al., 2019; Ferraresi, 2021a).

Still focusing on the institutional setting, we report the PDP for the average number of absences (sick leaves) from work for public employees¹² at municipal level (`mun_abs_avg`) which could represent a proxy for local institutional quality. The relationship with our target variable is positive with a concave shape stressing how low-effective institutions (more absenteeism from work) are associated with higher taxation. This is the result of a lesser degree of interest on the part of the local administration in the benefit of its citizens and, consequently, a lower probability of having a reduced PIT rate. Still analyzing local administrations, the average age of municipal council members (`mun_cou_avg_age`) also shows a positive association—albeit with a slightly U-like shape—with the probability of having a maximum PIT rate. An older local ruling local class may be more inclined to rely on less visible instruments to increase the municipal budget such as PIT rather than seek other opportunities such as national and international calls. These two features, especially the first one, could be proxies of less efficient local governments as stressed in the theoretical framework of Bordignon et al. (2017) where less competent mayors tended to raise taxation. Interestingly, the political party affiliation does not emerge within this framework as a core determinant in the use of PIT, and this is in line with the results of Bracco et al. (2019).

Finally, the last two variables for which the PDPs are shown are associated with municipal budgets. The consolidated law on the organization of local authorities (*Testo unico delle leggi sull'ordinamento degli enti locali – TUEL*) of 2000 defined a parametric procedure to identify municipalities under financial distress and automatically trigger corrective measures. The measurement identifies eight parameters with specific threshold levels. If at least four of those indicators are not complied with, an infringement procedure is initiated.¹³ We

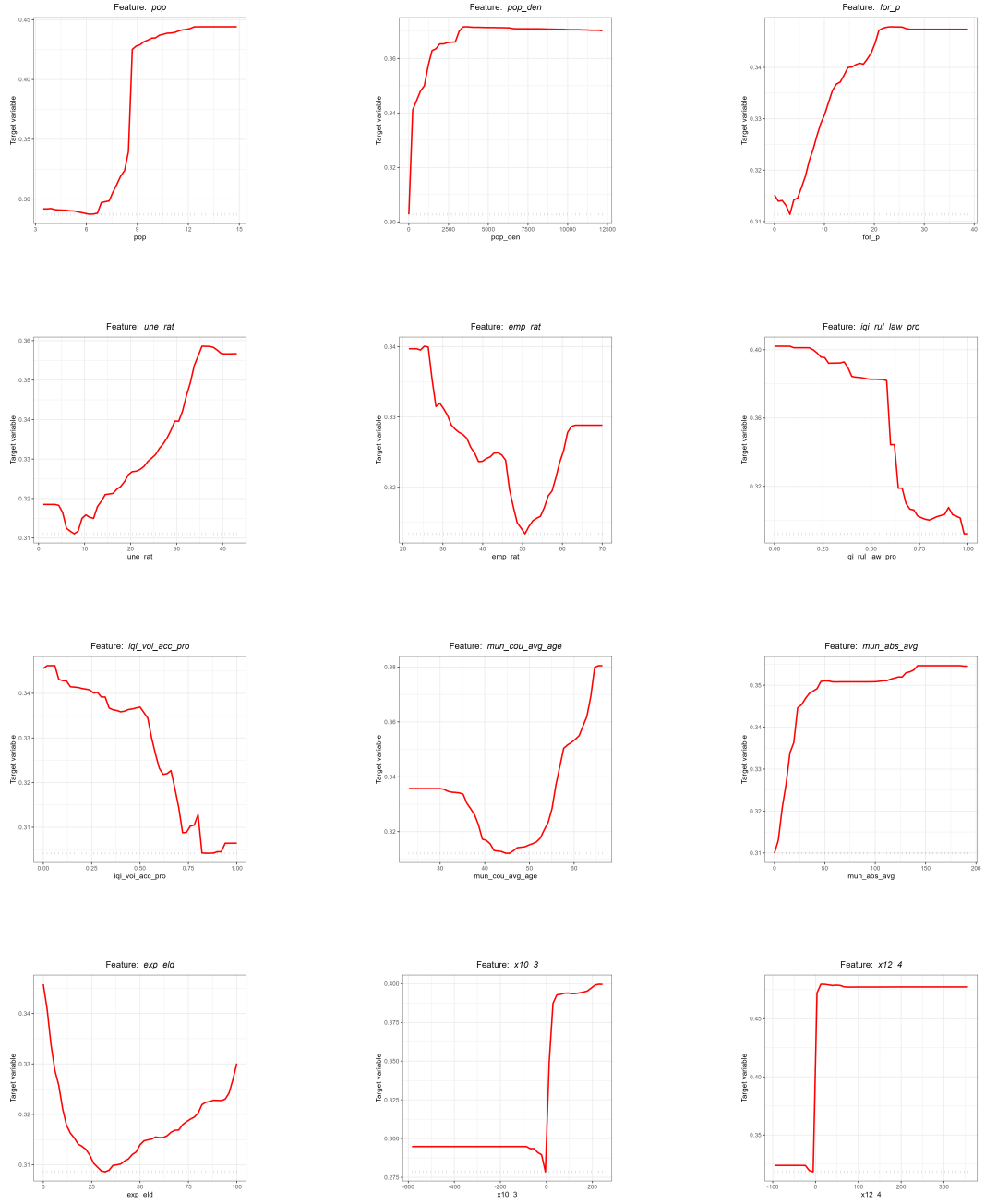
¹²The data considers workers employed with a contract which falls within a specific group of national collective work contracts ([CCNL Regioni e autonomie locali](#)).

¹³If a municipality ends up declaring bankruptcy it follows that: (i) impossibility of taking

showed the relationship of the two with the highest importance in determining our target variable. The first one is the municipality deficit sustainability borne by the financial year (**x12_4**), the second most important features among all. It shows an abrupt increase just before 0, hence close to the threshold level, which is set equal to 1,2%. A similar pattern emerges also for the other financial index reported, the sustainability of financial debts (**x10_3**) shows a steep vertical growth right after 0, by considering that the threshold is set equal to 16% for this indicator. Before achieving these alarming thresholds, municipalities tend to show a higher probability of setting the highest level of income taxation. Therefore, the closer they are to situations of financial distress (as defined by the TUEL), municipalities tend to have more burdensome taxes on citizens to raise cash and thus avoid recovery or insolvency plans. This result resonates with the Israeli case investigated by [Ben-Bassat et al. \(2016\)](#) where soft-budget constraint led to an increase of local property tax collection. Nonetheless, for the Italian case, the presence of soft-budget constraints leads to general inefficiencies ([Bordignon, 2000](#)). Therefore, the positive association emerging here could represent a legacy of this peculiar characteristic within a more stringent budget surveillance for local administration.

out new mortgages, (ii) limitations on current spending, (iii) freeze on new hiring, (iv) and the setting of municipal tax rates at their ceiling.

Figure 8: *Partial dependence plots for selected features*



6 Conclusions

This paper investigates the predictability of local taxation by employing machine learning (ML) techniques. As a case study, we consider the income tax rate of Italian municipalities (maximum surcharge on personal income tax). Results show that ML algorithms, particularly Random Forest, offer superior predictive accuracy over traditional econometric methods. This finding highlights the relevance of ML in public finance, providing a robust framework for predicting tax policy outcomes and informing stakeholders.

Feature importance reveals that demographic, financial, socioeconomic, and institutional quality indicators are key determinants of local tax decisions. These insights contribute to a deeper understanding of the complexities involved in fiscal decentralization and the autonomy of local governments. As such, the study emphasizes the importance of leveraging data-driven approaches to enhance transparency and accountability in tax policy formulation.

The implications of our findings are significant for policymakers. By identifying the factors that most influence local tax rates, governments can tailor interventions to promote fiscal responsibility and equity. Moreover, accurate predictions of tax policies can aid citizens and businesses in making informed decisions, thereby reducing financial uncertainty and fostering economic growth.

From an economic perspective, the emergence of variables such as population, foreign population proportion, and institutional quality indicators (voice and accountability and rule of law) as strong predictors of the likelihood of levying the highest tax rate and, more importantly, the direction of their impact, highlight that policymakers might tailor taxation based on demographics and civic engagement. This result sheds light on the crucial role of citizen participation in promoting transparency and shaping fairer taxation policies that better reflect the needs of the local community.

Future research could extend this approach to other forms of taxation or explore the integration of real-time data for dynamic policy analysis. As ML continues to evolve, its application in public finance promises to drive more efficient and effective governance.

References

- Abrell, J., Kosch, M., and Rausch, S. (2022). How effective is carbon pricing?—A machine learning approach to policy evaluation. *Journal of Environmental Economics and Management*, 112:102589.
- Alarie, B. and Xue Griffin, B. (2022). Using machine learning to crack the tax code. *Tax Notes Federal*, page 661.
- Andini, M., Ciani, E., de Blasio, G., D’Ignazio, A., and Salvestrini, V. (2018). Targeting with machine learning: An application to a tax rebate program in Italy. *Journal of Economic Behavior & Organization*, 156:86–102.
- Antulov-Fantulin, N., Lagravinese, R., and Resce, G. (2021). Predicting bankruptcy of local government: A machine learning approach. *Journal of Economic Behavior & Organization*, 183:681–699.
- Associazione Contribuenti Italiani (2022). La top ten delle imposte più odiate nel 2022. Press release of August 8, 2022.
- Baldini, M. (2021). Redistribution and progressivity of the Italian personal income tax, 40 years later. *Fiscal studies*, 42(2):345–366.
- Barca, F. (2019). Place-based policy and politics. *Renewal: A Journal of Social Democracy*, 27(1):84–95.
- Bardhan, P. (2002). Decentralization of Governance and Development. *Journal of Economic Perspectives*, 16:185–205.
- Bardhan, P. and Mookherjee, D. (2006). Decentralisation and accountability in infrastructure delivery in developing countries. *Economic Journal*, 116:101–127.
- Baskaran, T. and Feld, L. P. (2013). Fiscal Decentralization and Economic Growth in OECD Countries: Is There a Relationship? *Public Finance Review*, 41:421–445.

- Battiston, P., Gamba, S., and Santoro, A. (2024). Machine learning and the optimization of prediction-based policies. *Technological Forecasting and Social Change*, 199:123080.
- Becker, S. O., Egger, P. H., and Merlo, V. (2012). How low business tax rates attract MNE activity: Municipality-level evidence from Germany. *Journal of Public Economics*, 96:698–711.
- Ben-Bassat, A., Dahan, M., and Klor, E. F. (2016). Is centralization a solution to the soft budget constraint problem? *European Journal of Political Economy*, 45:57–75.
- Boetti, L., Piacenza, M., and Turati, G. (2012). Decentralization and local governments’ performance: How does fiscal autonomy affect spending efficiency? *FinanzArchiv/Public Finance Analysis*, pages 269–302.
- Bordignon, M. (2000). Problems of soft budget constraints in intergovernmental relationships: the case of Italy. IDB Working Paper, No. 398.
- Bordignon, M., Grembi, V., and Piazza, S. (2017). Who do you blame in local finance? An analysis of municipal financing in Italy. *European Journal of Political Economy*, 49:146–163.
- Bordignon, M. and Minelli, E. (2001). Rules transparency and political accountability. *Journal of Public Economics*, 80:73–98.
- Bracco, E., Lockwood, B., Porcelli, F., and Redoano, M. (2015). Intergovernmental grants as signals and the alignment effect: Theory and evidence. *Journal of Public Economics*, 123:78–91.
- Bracco, E., Porcelli, F., and Redoano, M. (2019). Political competition, tax salience and accountability. Theory and evidence from Italy. *European Journal of Political Economy*, 58:138–163.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45:5–32.

- Bucci, V., Ferrara, G., and Resce, G. (2024). Local government spending efficiency and fiscal decentralization: evidence from Italian municipalities. *Applied Economics*, 56(5):599–614.
- Censis (2011). *45° Rapporto sulla situazione sociale del Paese/2011*. Franco Angeli, Rome, IT. Centro Studi Investimenti Sociali.
- Clark, D. (2009). The performance and competitive effects of school autonomy. *Journal of Political Economy*, 117:745–782.
- Dasarathy, B. V. (1991). Nearest neighbor (NN) norms: NN pattern classification techniques. *IEEE Computer Society Tutorial*.
- Daveri, F. and Tabellini, G. (2000). Unemployment, growth and taxation in industrial countries. *Economic policy*, 15(30):48–104.
- De Roux, D., Perez, B., Moreno, A., Villamil, M. d. P., and Figueroa, C. (2018). Tax fraud detection for under-reporting declarations using an unsupervised machine learning approach. In *KDD '18: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 215–222, New York, NY-US. Association for Computing Machinery.
- DeLong, E. R., DeLong, D. M., and Clarke-Pearson, D. L. (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*, pages 837–845.
- Devarajan, S., Khemani, S., and Shah, S. (2009). The politics of partial decentralization. In Ahmad, E. and Brosio, G., editors, *Does decentralization enhance service delivery and poverty reduction?* Edward Elgar Publishing, Cheltenham, UK.
- Drucker, H., Burges, C. J., Kaufman, L., Smola, A., and Vapnik, V. (1996). Support Vector Regression Machines. *Advances in neural information processing systems*, 9.

- Erokhin, D. and Zagler, M. (2024). Who will sign a Double Tax Treaty next? A prediction based on economic determinants and machine learning algorithms. *Economic Modelling*, 139:106819.
- Faguet, J. P. (2004). Does decentralization increase government responsiveness to local needs? Evidence from Bolivia. *Journal of Public Economics*, 88:867–893.
- Falch, T. and Fischer, J. A. (2012). Public sector decentralization and school performance: International evidence. *Economics Letters*, 114:276–279.
- Fawcett, T. (2006). An introduction to ROC analysis. *Pattern recognition letters*, 27(8):861–874.
- Ferraresi, M. (2021a). Evidence from Italy: how local governments manipulate tax and spend policies to help win re-election. LSE European Politics and Policy (EUROPP) blog.
- Ferraresi, M. (2021b). Political Budget Cycle, Tax Collection, and Yardstick Competition. *The B.E. Journal of Economic Analysis & Policy*, 21:1149–1161.
- Filippetti, A. and Sacchi, A. (2016). Decentralization and economic growth reconsidered: The role of regional authority. *Environment and Planning C: Government and Policy*, 34(8):1793–1824.
- Fondazione Ca’ Foscari (2024). Banca dati sulle criticità finanziarie dei comuni. Fondazione Università Ca’ Foscari. Retrieved from: https://mizar.unive.it/gsi/studi/public/elen_info.php [Accessed May 1, 2024].
- Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, pages 1189–1232.
- Garman, C., Haggard, S., and Willis, E. (2001). Fiscal decentralization: A political theory with Latin American cases. *World politics*, 53(2):205–236.

- Gemmell, N., Kneller, R., and Sanz, I. (2013). Fiscal decentralization and economic growth: Spending versus revenue decentralization. *Economic Inquiry*, 51:1915–1931.
- Hansjörg, B. and Junghun, K. (2016). *Fiscal federalism 2016 making decentralisation work: Making decentralisation work*. OECD publishing, Paris, FR.
- Hastie, T., Tibshirani, R., Friedman, J. H., and Friedman, J. H. (2009). *The Elements of Statistical Learning: data mining, inference, and prediction*. Springer, New York, NY-US.
- Hooghe, L., Marks, G., and Schakel, A. (2010). *The rise of regional authority: A comparative study of 42 democracies*. Routledge, London, UK.
- Hung, N. T. and Thanh, S. D. (2022). Threshold effects of fiscal decentralization on income inequality: Evidence from Vietnam. *Cogent Business and Management*, 9.
- Ippolito, A. and Lozano, A. C. G. (2020). Tax Crime Prediction with Machine Learning: A Case Study in the Municipality of São Paulo. In *Proceedings of the 22nd International Conference on Enterprise Information Systems (ICEIS 2020)*, volume 1, pages 452–459, Setúbal, PT. Science and Technology Publications.
- Istat (2024a). A misura di Comune. Istituto Nazionale di Statistica. Retrieved from: <https://www.istat.it/it/archivio/220004> [Accessed May 1, 2024].
- Istat (2024b). Demo - Demografia in cifre. Istituto Nazionale di Statistica. Retrieved from: <https://demo.istat.it/> [Accessed May 1, 2024].
- Istat (2024c). Principali statistiche geografiche sui comuni. Istituto Nazionale di Statistica. Retrieved from: <https://www.istat.it/classificazione/principali-statistiche-geografiche-sui-comuni/> [Accessed June 4, 2024].

- Jacobs, B., de Mooij, R. A., and Folmer, K. (2010). Flat income taxation, redistribution and labour market performance. *Applied Economics*, 42(25):3209–3220.
- Ji, K. (2013). *Essays on Tax Policy, Institutions, and Output*. PhD thesis, Tilburg University, Tilburg, NL.
- Jiménez-Rubio, D. (2011). The impact of fiscal decentralization on infant mortality rates: Evidence from OECD countries. *Social Science and Medicine*, 73:1401–1407.
- Kasy, M. (2018). Optimal taxation and insurance using machine learning—sufficient statistics and beyond. *Journal of Public Economics*, 167:205–219.
- Keen, M., Kim, Y., and Varsano, R. (2008). The “flat tax (es)”: principles and experience. *International Tax and Public Finance*, 15:712–751.
- Kleinberg, J., Ludwig, J., Mullainathan, S., and Obermeyer, Z. (2015). Prediction policy problems. *American Economic Review*, 105(5):491–495.
- Kleven, H., Landais, C., Muñoz, M., and Stantcheva, S. (2020). Taxation and migration: Evidence and policy implications. *Journal of Economic Perspectives*, 34:119–142.
- Kleven, H. J., Landais, C., and Saez, E. (2013). Taxation and international migration of superstars: Evidence from the european football market. *American Economic Review*, 103:1892–1924.
- Kuhn, M., Wing, J., Weston, S., Williams, A., Keefer, C., Engelhardt, A., Cooper, T., Mayer, Z., Kenkel, B., Benesty, M., Lescarbeau, R., Ziem, A., Scrucca, L., Tang, Y., Candan, C., and Hunt, T. (2020). Package ‘caret’. *The R Journal*, 223(7).
- Lagravinese, R., Liberati, P., and Sacchi, A. (2018). The growth and variability of regional taxes: an application to italy. *Regional Studies*, 52(3):416–429.

- Langenmayr, D. and Simmler, M. (2021). Firm mobility and jurisdictions' tax rate choices: Evidence from immobile firm entry. *Journal of Public Economics*, 204.
- Lewis, D. D. (1998). Naive (Bayes) at forty: The independence assumption in information retrieval. In Nédellec, C. and Rouveirol, C., editors, *European conference on Machine Learning*, pages 4–15, Berlin, DE. Springer.
- Liberati, P. and Sacchi, A. (2013). Tax decentralization and local government size. *Public Choice*, 157:183–205.
- Liebig, T., Puhani, P. A., and Sousa-Poza, A. (2007). Taxation and internal migration - Evidence from the Swiss census using community-level variation in income tax rates. *Journal of Regional Science*, 47:807–836.
- Martinez, I. Z. (2017). Beggar-Thy-Neighbour Tax Cuts: Mobility after a Local Income and Wealth Tax Reform in Switzerland. LISER Working Papers n. 2017-08.
- Martínez-Vázquez, J., Lago-Peñas, S., and Sacchi, A. (2017). The impact of fiscal decentralization: A survey. *Journal of Economic Surveys*, 31(4):1095–1129.
- Martinez-Vazquez, J. and McNab, R. M. (2006). Fiscal Decentralization, Macrostability, and Growth. *Hacienda Pública Española / Revista de Economía Pública*, 179:25–49.
- Masrom, S., Rahman, R. A., Mohamad, M., Abd Rahman, A. S., and Baharun, N. (2022). Machine learning of tax avoidance detection based on hybrid metaheuristics algorithms. *IAES International Journal of Artificial Intelligence*, 11(3):1153.
- MEF (2024). Addizionale comunale all'IRPEF. Ministero dell'Economia e delle Finanze. Retrieved from: https://www1.finanze.gov.it/finanze2/dipartimentopolitichefiscali/fiscalitalocale/nuova_addcomirpef/sceltaregione.htm [Accessed May 30, 2024].

- Ministero dell'Economia e delle Finanze (2019). Bollettino delle entrate tributarie 2019.
- MINT (2024a). Anagrafe degli amministratori locali e regionali. Ministero dell'Interno. Retrieved from: <https://dait.interno.gov.it/elezioni/anagrafe-amministratori> [Accessed May 30, 2024].
- MINT (2024b). Conto Annuale. Ministero dell'Interno. Retrieved from: <https://contoannuale.rgs.mef.gov.it/> [Accessed April 1, 2024].
- Monturano, G., Resce, G., and Ventura, M. (2023). The Impact of Italy's Strategy for Inner Areas on Depopulation and Industrial Growth: A Staggered Diff-in-Diff Analysis with Spatial Spillover Effects. SSRN's Research Paper Series n. 4575900.
- Nifo, A. and Vecchione, G. (2024). Institutional Quality Index. Retrieved from: <https://sites.google.com/site/institutionalqualityindex/home?authuser=0> [Accessed May 1, 2024].
- PCM (2024). Principali statistiche geografiche sui comuni. Presidenza del Consiglio dei Ministri. Retrieved from: <https://politichecoesione.governo.it/it/politica-di-coesione/strategie-tematiche-e-territoriali/strategie-territoriali/strategia-nazionale-aree-interne-snai/> [Accessed June 6, 2024].
- Petrovito, F., Pozzolo, A. F., Resce, G., and Scialà, A. (2023). Fiscal decentralization and income (re) distribution in OECD countries' regions. *Structural Change and Economic Dynamics*, 67:69–81.
- Poli, S. (2019). How Do Italian Private Companies Respond to a Reduction in the Corporate Standard Tax Rate? A Focus on the Ownership Gender Diversity Effect. *International Journal of Accounting and Financial Reporting*, 9:205.
- Porcelli, F. (2014). Electoral accountability and local government efficiency:

- Quasi-experimental evidence from the Italian health care sector reforms. *Economics of Governance*, 15:221–251.
- Prud’Homme, R. (1995). The dangers of decentralization. *The world bank research observer*, 10(2):201–220.
- Remeta, J., Perret, S., Jareš, M., and Brys, B. (2015). Moving beyond the flat tax-tax policy reform in the slovak republic. OECD Taxation Working Papers, No. 22.
- Resce, G. (2022). The impact of political and non-political officials on the financial management of local governments. *Journal of Policy Modeling*, 44(5):943–962.
- Resce, G. and Vaquero-Piñeiro, C. (2022). Predicting agri-food quality across space: A Machine Learning model for the acknowledgment of Geographical Indications. *Food Policy*, 112:102345.
- Reschovsky, A. (1989). State and local taxation of the elderly. *Journal of Aging & Social Policy*, 1(3-4):143–170.
- Ripley, B. D. (2014). *Pattern recognition and Neural Networks*. Cambridge University Press, Oxford, UK.
- Rubolino, E. (2020). Tax-induced transfer of residence: Evidence from tax decentralization in Italy. SSRN Electronic Journal n. 3710932.
- Rubolino, E. (2023). Does weak enforcement deter tax progressivity? *Journal of Public Economics*, 219.
- Sacchi, A. and Salotti, S. (2014). The effects of fiscal decentralization on household income inequality: some empirical evidence. *Spatial Economic Analysis*, 9:202–222.
- Schmidheiny, K. and Slotwinski, M. (2018). Tax-induced mobility: Evidence from a foreigners’ tax scheme in Switzerland.

- Sebastián, G. and Schargrodsky, E. (2001). Evaluating the Impact of School Decentralization on Education Quality. *Economia*, 2:275–314.
- Shan, H. (2010). Property taxes and elderly mobility. *Journal of Urban Economics*, 67:194–205.
- Socci, C., D’Andrea, S., Deriu, S., Pretaroli, R., and Severini, F. (2022). Does the personal income flat tax fit with economic growth and inequality in Italy? *Italian Economic Journal*, 8(3):523–548.
- Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288.
- Tselios, V. (2023). Does political decentralization affect income inequality? The role of governance quality. *Regional Studies*, 57:829–843.
- Venables, W. N. and Ripley, B. D. (2002). *Modern applied statistics with S*. Springer, New York, NY-US.
- Weigel, J. L. (2020). The participation dividend of taxation: How citizens in Congo engage more with the state when it tries to tax them. *The Quarterly Journal of Economics*, 135(4):1849–1903.
- Weingast, B. R. (2009). Second generation fiscal federalism: The implications of fiscal incentives. *Journal of Urban Economics*, 65:279–293.
- Young, C., Varner, C., Lurie, I. Z., and Prisinzano, R. (2016). Millionaire migration and taxation of the elite: Evidence from administrative data. *American Sociological Review*, 81:421–446.
- Zou, H. and Hastie, T. (2005). Regularization and variable selection via the Elastic Net. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 67(2):301–320.
- Zumaya, M., Guerrero, R., Islas, E., Pineda, O., Gershenson, C., Iñiguez, G., and Pineda, C. (2021). Identifying tax evasion in Mexico with tools from

network science and machine learning. *Corruption Networks: Concepts and Applications*, pages 89–113.

Appendix

Table A1: *Municipal Features: Descriptive Statistics*

Code	Description	Mean	Min	Max	Std. Dev.
target_variable	PIT rate equal or greater than 0.8%	0.3136	0	1	0.4639
<i>Demographics and geographical</i>					
pop (log)	Total resident population	7.8380	3.4657	14.8523	1.3641
pop_den	Population density	302.5312	0.7666	12186.4151	643.6625
pop_var_18_p	Percentage change in population from 2018	-1.6767	-23.4347	17.0418	2.5747
rip_Nord-West	Dummy variable (1 = for regions in the Nord-West)	0.3789	0	1	0.4851
rip_Nord-East	Dummy variable (1 = for regions in the Nord-East)	0.1742	0	1	0.3792
rip_South	Dummy variable (1 = for regions in the South)	0.227	0	1	0.4189
rip_Islands	Dummy variable (1 = for regions in the Islands)	0.0977	0	1	0.297
rip_Centre	Dummy variable (1 = for regions in the Centre)	0.1222	0	1	0.3276
alt	Altitude	354.3156	0	2035	295.9452
sup	Land surface	38.036	0.1206	1287.3895	50.6964
inn_are_d	Dummy variable (1 = inner area)	0.4844	0	1	0.4998
mun_uni_d	Dummy variable (1 = union of municipalities)	0.3764	0	1	0.4845
mun_mou_d	Dummy variable (1 = mountain municipalities)	0.4215	0	1	0.4938
mun_cap_pro_d	Dummy variable (1 = provincial capital municipality)	0.0138	0	1	0.1166
mun_lit_d	Dummy variable (1 = coastal municipality)	0.0817	0	1	0.2738
mun_isl_d	Dummy variable (1 = island municipality)	0.0042	0	1	0.0648
mun_cos_d	Dummy variable (1 = municipality in coastal zone)	0.1477	0	1	0.3549
fem_p	Female resident (%)	50.4035	35	59.0361	1.664
mal_p	Male resident (%)	49.5965	40.9639	65	1.664
for_p	Foreign resident (%)	6.6119	0	38.75	4.2092
nat_rat	Natality rate	6.6385	0	35.0877	2.7589
mor_rat	Mortality rate	12.5216	0	79.4702	5.7589
fer_rat	Fertility rate (live births/female aged 16-45)	3.3933	0	28.5714	1.4859
fam_avg	Average number of family members	2.2767	1.1165	3.1724	0.2451
pop_avg_y	Average age of residents	46.6239	34.97	67.58	3.3958
age_04_p	Resident population between 0 and 4 years (%)	3.5479	0	8.75	0.9974
age_65_ov_p	Resident population aged 65 and over (%)	24.7009	8.2298	66.6667	5.4212
agi_ind	Aging index	39.8358	11.3248	208	11.8017
lan_con_15_p	Land consumption (%)	9.9008	0.2478	91.0312	10.1153
urb_deg	Degree of urbanisation	2.605	1	3	0.5512
fra_ind_1819	Municipal fragility index	5.5103	1	10	2.8739
<i>Socioeconomics</i>					
emp_rat	Employment rate	44.8117	21.4559	70.0454	6.4167
une_rat	Unemployment rate	12.4446	0.9	43.0585	6.3647
ent_rat	Entrepreneurship rate	61.2835	2.2384	276.9456	21.4016
loc_act_uni_den	Active local units density	21.5779	0.0362	1114.3708	46.7531
pov_rat	Incidence of poverty (%)	31.871	11.1111	77.9817	10.2075
exp_eld	Expenditures on social interventions and services for the elderly	21.8286	0	100	17.0179
ban_bra	Bank branches (per inhabitants)	0.4342	0	7.6	0.4831
stu_liv_1_p	Population with at least a middle school certificate (%)	28.5753	0	49.6	4.1654
stu_liv_2_p	Population with at least a high school certificate (%)	2.9648	0	8.4	0.8426
stu_liv_3_p	Population with at least a degree certificate (%)	6.3163	0	22.6	2.3804

Continued

Code	Description	Mean	Min	Max	Std. Dev.
<i>Institutional</i>					
iqi_cor_pro	Corruption (IQI) - provincial level	0.8187	0	1	0.1794
iqi_gov_eff_pro	Government effectiveness (IQI) - provincial level	0.4147	0	1	0.1319
iqi_reg_qu_pro	Regulatory quality (IQI) - provincial level	0.388	0	1	0.1365
iqi_ru_law_pro	Rule of law (IQI) - provincial level	0.6466	0	1	0.1582
iqi_corr_pro	Corruption perception (IQI) - provincial level	0.6826	0	1	0.1796
was_sor	Proportion of waste sorting	58.908	0.0014	99.1197	21.4373
may_fem_d	Dummy variable (1 = female mayor)	0.1375	0	1	0.3443
may_stu	Mayor's years of education	11.1313	0	18	4.8396
may_deg_d	Dummy variable (1 = mayor with a degree)	0.4474	0	1	0.4972
may_age	Mayor's age	51.3039	0	90	12.9945
may_y	Mayor's years in office	3.9197	0	10	1.8485
pol_rep_fem_p	Female in the county council (% of total elected)	31.2239	0	100	13.5428
mun_cou_fem_p	Female in the municipal council (% of total council members)	30.4124	0	100	19.0214
mun_cou_avg_age	Average age of the council's members	46.5014	22	67.8333	4.7382
mun_adm_avg_age	Average age of municipal administrators	49.1104	23.3333	82.6667	6.9216
mun_dir_ov_emp_p	Proportion of executive over total municipal employees	0.1819	0	50	0.9627
mun_emp_avg_age	Average age of municipal employees	49.4866	0	63.8	11.8943
mun_deg_p	Proportion of employees with a bachelor's degree	18.2052	0	100	15.4038
mun_abs_avg	Average number of absences of public employees at municipal level due to sick leave	6.5103	0	192	7.9222
may_pol_CEN_LEFT	Dummy variable (1 = center-left political party)	0.0528	0	1	0.2236
may_pol_CEN_RIGHT	Dummy variable (1 = center-right political party)	0.03	0	1	0.1705
may_pol_Other	Dummy variable (1 = other political parties)	0.2113	0	1	0.4082
<i>Fiscal and Financial</i>					
tax_pay (log)	Taxpayers - municipal level	7.5031	3.2958	14.4792	1.3323
inc_tax	Taxable income - municipal level	17847.6111	6198.2353	50590.7287	3761.4777
x1_1	Budget rigidity	33.0779	0	399	13.5933
x2_7	Ratio of own revenue collection to definitive current budget forecasts	49.4254	0	7754	81.9014
x3_2	Advances closed for accounting purposes only	19.6965	-91	8638	184.1087
x10_3	Financial debts sustainability	8.635	-584	4041	29.2258
x12_4	Sustainability of the deficit actually charged to the financial year	1.3407	-132	868	11.9634
x13_1	Recognised and financed debts	9.3699	0	9510	201.604
x13_23	Recognised and under financing debts	3.7436	0	8500	119.3495