October 2025

Fondazione Eni Enrico Mattei

Working Paper

20.2025

Income Tax Treatment and Labour Supply in a multi-level hierarchical Difference-in-Differences model

Bruno Paolo Bosco, Carlo Federico Bosco, Paolo Maranzano

Income Tax Treatment and Labour Supply in a multi-level hierarchical Difference-in-Differences model

Bruno Paolo Bosco (Department of Economics, Management and Statistics, University of Milano-Bicocca), **Carlo Federico Bosco** (Independent Researcher), **Paolo Maranzano** (Department of Economics, Management and Statistics, University of Milano-Bicocca and Fondazione Eni Enrico Mattei)

Summary

Ignoring the possible hierarchical clustering of the data that frequently characterises the structure of labour markets implies that studies of the effects of income tax changes on labour supply use less than necessary information on the variability of the labour response. Estimation efficiency is reduced and relevant relationships affecting the agents' reaction to net wage changes remain undetected. Motivated by the desire to implement an estimation procedure that accommodates a nested hierarchical statistical structure of labour supply macro data into of a causal-effect framework, we propose a novel multilevel DiD model that can estimate labour responses to exogenous tax hikes taking the above hierarchical structure into consideration. Using Italy as a case study, we examine the labour response to exogenous income tax changes using a hierarchical DiD model modified to account for the existence of different sources of variation of the data (regional and provincial labour markets) as well as for various possible clustering of the data (territorial, age and gender). We compare results obtained from various nested and non-nested procedures and show that our multilevel variant of the DiD model generates gains in efficiency with respect to approaches that ignore the clustering nature of the labour data. The hierarchical multilevel DiD procedure permits to qualify labour response in terms of cluster membership and to shed light on aspects of the tax issues not highlighted by current literature.

Keywords: Income Taxation and Labour Supply; Tax Treatment Effect; Hierarchical Multilevel Panel Data DiD model; Linear mixed models for policy evaluation

JEL Classification: C21, C32; H24, H31, J38

Corresponding Author
Paolo Maranzano
Department of Economics, Management and Statistics, University of Milano-Bicocca
Via Bicocca degli Arcimboldi 8, Milan (Italy)
paolo.maranzano@unimib.it

Income Tax Treatment and Labour Supply in a multi-level hierarchical Difference-in-Differences model

Bruno Paolo Bosco[1] – Carlo Federico Bosco[2] – Paolo Maranzano[3]

ABSTRACT

Ignoring the possible hierarchical clustering of the data that frequently characterises the structure of labour markets implies that studies of the effects of income tax changes on labour supply use less than necessary information on the variability of the labour response. Estimation efficiency is reduced and relevant relationships affecting the agents' reaction to net wage changes remain undetected. Motivated by the desire to implement an estimation procedure that accommodates a nested hierarchical statistical structure of labour supply macro data into of a causal-effect framework, we propose a novel multilevel DiD model that can estimate labour responses to exogenous tax hikes taking the above hierarchical structure into consideration. Using Italy as a case study, we examine the labour response to exogenous income tax changes using a hierarchical DiD model modified to account for the existence of different sources of variation of the data (regional and provincial labour markets) as well as for various possible clustering of the data (territorial, age and gender). We compare results obtained from various nested and non-nested procedures and show that our multilevel variant of the DiD model generates gains in efficiency with respect to approaches that ignore the clustering nature of the labour data. The hierarchical multilevel DiD procedure permits to qualify labour response in terms of cluster membership and to shed light on aspects of the tax issues not highlighted by current literature.

A R T I C L E I N F O

Keywords: Income Taxation and Labour Supply; Tax Treatment Effect; Hierarchical Multilevel Panel Data DiD model; Linear mixed models for policy evaluation.

JEL codes: C21, C32; H24, H31, J38.

[1] Department of Economics, Management and Statistics @ University of Milano-Bicocca, bruno.bosco@unimib.it
[2] Independent Researcher, fede.bosco28@gmail.com
[3] Department of Economics, Management and Statistics @ University of Milano-Bicocca & Fondazione Eni Enrico Mattei, paolo.maranzano@unimib.it

1. Introduction

Labour supply response to income taxes is the subject of theoretical debates and empirical investigations. This paper offers a methodological contribution to the existing empirical literature by presenting a hierarchical multilevel Difference-in-Differences (DiD) analysis of labour markets' outcomes in response to income tax hikes. The need to consider the multilevel framework stems from two main considerations. On the one hand, local labour markets show heterogeneous socio-economic features resulting from the specific characteristics of their geography, sociology, and historical development that, combined with distinct tax and regulation rules lead to the generation of clusters of markets, with one cluster different from another. Consider for example the local labour markets existing in territories corresponding to a Province or a County belonging to a Region or to a State: local agents operating in that Province are subject to the regulation imposed by the Province and to the regulation dictated by the Region that Province belongs and are affected by uses, tradition, relationships specific of their territory. On the contrary, agents operating in another Province of the same Region are still subject to the common regional rules and regional taxation, but they must abide to different Province regulation, taxes, uses and practices. In other words, labour markets belonging to Provinces that belong to a particular Region have specific operational features resulting from practices, rules and taxes imposed by each provincial authority and, at the same time, they have common operational features engendered by the regional regulation and taxation that are valid for all the Provinces in the Region. In turn, the latter regulation changes across Regions, and overlaps the regulation implemented by the provincial authority existing in each regional territory. As a result, macro labour data recorded at the lower level (say, Provinces) present hierarchical clustering features: each cluster shares some cluster-specific features that are common to cluster members but are not shared by the member of some other cluster. Therefore, labour responses to taxation may depend upon the distribution of workers in Provinces of the same Region (implementing their own provincial rules) and upon the distributions of the Provinces in distinct Regions (implementing their own regional rules). The resulting nesting structure of the labour supply macro data sets resembles the nesting structure of many educational data sets. In the latter, the target response variable under investigation (generally the pupils' test performance) depends upon the pupils' idiosyncratic attitudes and effort as well as on the distribution of pupils in classes of the same school and on the distributions of the various schools in distinct scholar districts operating in their specific environment and implementing their own rules (Mejía-Rodríguez & Kyriakides, 2022; Sellström & Bremberg, 2006).

When studying the effects of exogenous labour policy measures, such as income tax hikes, the above-mentioned nesting structure of the data becomes pertinent because individual data can be "nested" in ways that enhance the probability of observing *similar responses to exogenous events by units belonging to the same cluster*. If this were the case, an intra-cluster correlation of error terms exists, and the assumption of independence may be violated and the standard errors of the estimations of labour response to policy changes (which is at the root of the estimation of labour supply elasticities) may be biased. Consequently, the policy interpretation of the estimation results based on methods that ignore the nesting structure can be misleading or deceptive.

The purpose of this paper is to propose a methodology useful to overcome the above shortcomings of more traditional empirical approaches to the study of taxation and labour supply and to implement a sounder estimation procedure of the labour response to changes in income taxation. To do so, we identify and estimate a DiD model of the effects of an exogenous income tax hike augmented in a way that integrates the hierarchical or mixed data structure. This procedure permits to identify correctly the average effect of an exogenous tax hikes on treated subjects and to estimate the tax effect *conditional on the cluster structure distinctive of this data generation process*. The procedure retains the great advantages of the DiD technique (that is, the correct identification of the effects of an *exogenous* income tax treatment) and, since it accounts for intra-cluster correlation of the error terms, it reduces the probability of Type I error (detecting a tax effect when this effect is not present). Our resulting mixed DiD tax model permits to evaluate the contribution of the between and within cluster variances to the total estimated variance of the labour response variables and to isolate the proportion of that variance that is generated across higher-level units (Regions in our case). Specifically, in our multilevel DiD tax model the regression intercepts –assumed to have particular distribution across the clusters and defined as dependent on some cofactors– are used to estimate different labour responses to income tax changes among treated and untreated taxpayers *once the nesting structure of the data generation process is accounted for*. Obtaining information about the between and within cluster variances of exogenous policy measures not only provides better information on the source of the overall variability of the response labour variables but may be relevant in a policy perspective.

To implement the above methodology, we use Italy as a case study. Italian regional and provincial authorities may use different local income tax rates and adopt different tax expenditure measures, not to mention subsidies and regulation. Income taxation is heterogeneous across Italian Regions/Provinces, and exogenous regional income tax shocks may affect, *inter alia*, the supply of labour services and employment (Bosco, Bosco, & Maranzano, 2025). We conduct the empirical study using a large and disaggregated Italian panel data containing provincial, regional and national tax and economic data and, to define the above empirical multi-level DiD model, we exploit a quasi-natural tax experiment that occurred in Italy at the end of 2006. The characteristics of the tax treatment (timing and treatment assignment) and the time pattern that response variables follow before the treatment, as well as the absence of confounding factors, have important implications for our empirical approach. The treatment affected a subset of five Italian Regions and their Provinces (treatment groups) from 2007 onwards and consisted in an exogenously imposed increase of their regional income tax rates with respect to the rates of the other Regions (control groups) which remained unchanged. Hence, the 2006/7 central government policy defines the framework of a quasi-natural tax experiment, involving the taxpayers living in the five (not-randomly-chosen) treated regions vs. the taxpayers of the remaining fifteen untreated control regions. As a result of the treatment design, the non-random part of our proposed mixed model will be consistent with the identification requirements needed for the efficient Two-

Way Fixed Effect (TWFE) estimation of the Average Treatment Effect upon Treated (ATET) in a panel data set under treatment homogeneity (Roth, Sant'Anna, Bilinski, & Poe, 2023, p. 2220).

We use Italian data collected at three levels: the provincial level (the lower level of the data set); the regional level in which each Province is nested; and the Macroregional defined as set of administratively defined aggregation of Regions (the higher level of the data set). These levels are defined according to both ISTAT and the Eurostat (2024) Nomenclature of territorial units for statistics (NUTS). The time dimension of our panel data set (2004–2021) permits to conduct a dynamic analysis of the tax effects upon treated units and to propose a line of investigation (is the tax effect transitory or time lasting?) not previously followed by the literature in a DiD framework which generally adopts this approach in "micro studies are based on research designs that allow for causal identification, but the approach only captures short-run effects and may miss important dynamic mechanisms" (Kleven, Kreiner, Larsen, & Søgaardt, 2025, p. 2850). We find that the extensive negative adjustments of various response variables measuring the supply of labour services of treated taxpayers are statistically significant, rapid, and sharp *but not long-time lasting*. We also find significant response differences across age and gender and compare our findings with the estimates of a pure (one level) DiD model for both male and female as well as young and mature workers. We find that controlling for cluster variability adds significance to the results and permits to obtain information about the time path of the ATET across treated units that otherwise would go unnoticed. In terms of policy, we stress that estimating the cluster variability of the labour response to taxation will help design "cluster related" effective tax strategies (possibly a not uniform tax policy over a national territory) and to correct those nationwide uniform income tax policies implicitly based on a "one size fits all" philosophy. Notice that the use of multilevel (hierarchical) structures in a DiD context with macro data is innovative with respect to previous research. It combines the advantages of a DiD approach to the causality relationship with the existing hierarchical administrative subdivision of our labour data set and its tendency to generate within-and-between group effects. As a result, this paper does not typify a direct application of some well-established framework to a new national dataset but shows how new theoretical insights/interpretations of the effects of taxes on labour supply may be obtained from a new methodological statistical perspective potentially applicable to other contexts/fields.

The remainder of the paper is structured as follows. Relevant literature on taxation and labour supply is presented in section 2 where empirical methods, hypotheses, and previous results are discussed. Section 2 also discusses the issue of workers heterogeneity in connection with the hierarchical structural of our data generation process. It also stresses the endogeneity problem plaguing previous analysis of income taxation and labour supply and motivates the adoption a multi-level DiD in the present hierarchical case. In section 3 we examine the main features of the Italian 2006/7 tax experiment that will be used to justify the DiD identification strategy adopted in the paper. The main statistical features of the Multi-level DiD tax model are presented and commented in section 4, which also contains a discussion of the relevance of the values of the Intra-Class Correlation Coefficient (ICC) as a measure of how much observations within a cluster/group are similar to each other, relative to the between clusters/groups observations. Section 5 contains a description of the data set with basic statistics (reported in the Appendix). Sections 6.1 and 6.2 present and discuss the main empirical findings of the DiD models based on distinct clustering identifications whereas Section 6.3 presents the results of a robustness analysis based on age and gender disaggregation and the discuss the similarities and dissimilarities of our results with those reported by previous literature. Section 7 briefly concludes.

2. Related literature on taxes and labour supply

Early research on labour supply responses to income tax variations initially adopted the intensive margin perspective (Blundell, 1992; Chetty, 2012; McClelland & Mok, 2012; Zodrow, 2025) and focused on the tax sensitivity of supply decisions for primary earners. Keane (2011; 2022), McClelland & Mok (2012), Stantcheva (2019), and Bosco, Bosco, & Maranzano (2025) discuss the main econometric problems facing early studies and subsequent developments as well as the estimated values of the labour supply elasticities (Keane, 2022).

Papers of Pencavel (1986), Killingsworth & Heckman (1986), Colombino & Del Boca (1990), Blundell & MaCurdy (1999) and others aimed to examine and interpret the generally observed offsetting of income and substitution effects attributable to income tax hikes and to incorporate in the empirical research design the intricate non-convexities of the individual budget constraints resulting from progressive income taxation and government transfers. With the notable exception of Hausman (1981), early results indicated that (i) the labour supply of primary earners shows little tax sensitivity whereas labour supply decisions of secondary earners were more sensitive to taxes and (ii) that adult male workers (married or unmarried) were less responsive than women.

The analysis of the household labour supply with more than one income earner was developed in papers of Aarson & French (2002), Blundell & Walker (1982), Blundlell & MaCurdy (1999), Bludner & Walker (1986), Blundell, Walker & Bourguignon (1988), Donni & Molina (2018), Hausman & Ruud (1984), and Sickles & Yazbeck (1998). Findings did not change the general picture emerging from the previously quoted studies: differentiated responses to tax hikes of primary earners (generally, male) and secondary earners (generally, women).

Early research did not contemplate the possibility that some individuals could face labour supply restrictions constraining the provision of their optimal hours of labour (intensive margin evaluation). Yet, constraints could be the result of some private factors (such as the time devoted to child-care and/or dependent relatives, the expenses for transportation and commuting, etc.) acting as fixed costs that might have a direct impact on labour supply decisions. When income taxation changes and the above factors are present, tax hikes may affect the very basic decision about labour participation by influencing the trade-off between work and non-work time allocation. Hence, like other kind of wage changes, income tax hikes may affect the dichotomous decision about labour market participation rather than just the quantity of work offered and possibly force some worker to entirely withdraw from the labour market (extensive marginal decision) or to switch to "unofficial" labour relations.

Microeconometric studies of labour supply (Heckman, 1993) recognised the distinction between the extensive and intensive margins, and studies analysing the differences between micro and macro responses of labour supply to tax reform (Blundell, Bozio, & Laroque, 2011) emphasised the above difference between the extensive and intensive margins. For example, Rogerson & Wallenius (2009), following the work of Prescott (2004), argue that the responsiveness of the extensive margin of labour supply to taxation plays a major role in explaining aggregate differences in total hours worked across countries. They show that an economy with fixed technology costs for firms and an inverted U-shape life-cycle productivity for workers can produce large aggregate extensive labour supply responses driven by employment changes occurring at either extreme of the working life. Typically, the elasticity at the extensive margin has been found to be somewhat larger than the elasticity at the intensive margin and, as female labour force participation increased over time almost everywhere, recent studies show that the labour supply elasticities of men and women have, to some extent, converged (Blau & Kahn, 2007; 2017). This result contradicts previous findings showing that the size of the wage elasticities at the two margins has been found to differ significantly by gender and age (Blundell & MaCurdy 1999) as well as family composition (Apps & Ray, 2009; Guner, Kaygusuz, &Ventura, 2012).

On the extensive margin for lower-income groups, Eissa and Liebman (1996) and Meyer and Rosenbaum (2001) show that in the USA the Earned Income Tax Credit (a refundable tax credit enacted in 1975 for low-to-moderate-income working individuals and couples, particularly those with children) has increased labour force participation (McClelland & Mok, 2012, 12). Other authors (e.g., Romer & Romer, 2010; Saez, Matsaganis, & Tsakloglou, 2012) show that for high-income earners there is some evidence that the efficiency costs of raising taxes on top-income taxpayers expressed in terms of labour supply and other margins may be offset by shifting in the timing or form/source of income acquisition (Auerbach & Siegel, 2000; Goolsbee, 2000). Other studies of labour response to tax changes analyse the potential adverse base effects of tax hikes such as tax avoidance, outright tax evasion, and a general reduction in economic activity (Piketty, Saez, & Stantcheva, 2014). They conclude that income tax can be responsible not only of a reduction of labour services supplied but also of a general contraction of the economic activity. As for self-employment, Heim (2010), Appelbaum & Katz (1986), and Kihlstrom & Laffont (1979) consider explicitly how stochastic shocks in demand or cost functions affect the decision to become self-employed. In those models, an asymmetric tax reduction, favouring the self-employed relative to other workers, increases self-employment.

When territorial aspects (differences in local income taxation) are accounted for, the analysis considers the tax-induced mobility of taxpayers and discusses whether income tax rates can be employed by local authorities to attract highly qualified foreigner taxpayers. The literature shows that local income taxes may affect local decisions (such as regional/local labour supply or residential location) and evaluates the possible consequences of these decisions on the level and geographical distribution of productive activities and on local technological spill over (Widmann, 2023). For instance, Akcigit, Baslandze, & Stantcheva (2016) find that "prolific" inventors migrate between countries in response to changes in personal top income tax rates and obtain an estimate of the elasticity of the number of foreign inventors (in a country) with respect to the top net-of-tax rate close to one. Moretti & Wilson (2017) examine the mobility responses of inventors to changes in personal top income tax rates across US states and find a corresponding elasticity with respect to the top net-of-tax rate of 1.8. In addition, since the market value of material goods, and its geographical distribution, may be affected by income taxation, other studies show that local income tax rates can generate tax externalities of various nature across local jurisdictions (Esteller-Moré & Solé-Ollé, 2002).

Finally, we may note that many of the above studies implicitly assume that the incidence of labour taxes is fully borne by workers. This approach makes sense under the neoclassical view that labour supply bears most of the incidence of labour taxes (on the contrary, classical economists such as Smith and Ricardo postulated that wage taxes increased gross-of-tax wages and reduced net profits). Yet, a bulk of available evidence shows that the neoclassical view of tax shifting is not always consistent with the labour data generation process, and that the estimated quantity responses are in fact a mix of demand and supply adjustments. In the extreme case, when the incidence is fully borne by firms, as in Saez, Schoefer, & Seim (2019) and Benzarti & Harju (2021), the net-of-tax wages should not change, workers should not respond to tax hikes and, consequently, the total labour response should be attributed to labour demand. Some papers have recognized the dual response of supply and demand and have provided evidence of a confounding role of firm responses (Chetty et al., 2011; Tazhitdinova, 2020; Gudgeon & Trenkle, 2024) to wage taxation, namely that incidence can be an important (albeit overlooked) factor affecting the estimates of structural labour supply tax elasticity.

Yet the question of who bears the income tax is not the only relevant tax conjecture. Following Benzarti (2024), there are three types of anomalies that apply to labour (and consumption) taxes. First, there is mounting evidence that questions the usefulness of using the relative magnitude of the supply and demand elasticities as a sufficient statistic for tax incidence. Second, there is some empirical evidence that tax incidence is asymmetric: market prices (and quantities) respond differently to increases and decreases in taxes. Third, statutory incidence appears to matter for economic incidence, which is an empirical finding that contradicts a fundamental implication of the *canonical* tax incidence model. Benzarti (2024) covers recent empirical evidence on the incidence of labour taxes by focusing on the evidence of payroll tax variations. There is a strong consensus among public and labour economists that the incidence of payroll taxes is likely borne by workers. The basis for this common wisdom is that labour demand is considered to be orders of magnitude more elastic than labour supply since competitive firms are expected to pay workers their marginal product of labour (or less) and thus should pass on any payroll taxes to workers. However, recent empirical evidence from the past two decades contradicts this view by showing that firms may actually bear most of the employer portion of payroll taxes and shows substantial employment effects of payroll tax cuts. Saez, Schoefer, & Seim (2019) study the consequences of a 16-percentage point cut (from 31% to 15%) of the employer-portion of the payroll tax in Sweden applied to employees younger than 26. The prediction of the canonical model is straightforward: because employees that are slightly older than 26 should be similar to the employees that are slightly younger, their marginal products of labour should be the same. Furthermore, since employers only care about the tax-inclusive wages (wages inclusive of all payroll taxes) and not about the posted wages (which are net of employer payroll taxes but inclusive of employee payroll taxes), a payroll tax cut should result in a decrease of the tax-inclusive wage. This would lead firms to raise posted wages until the tax-inclusive wages are equalized for similar workers, resulting in an increase in the posted and net-of-tax wages of the under-26 workers. In other words, the tax cut would be borne by workers, as is commonly believed. Instead, Saez, Schoefer, & Seim's (2019) findings show that the net-of-tax wages do not respond to the payroll tax cut, and, conversely, total wages decrease substantially. Therefore, the payroll tax cut is fully borne by firms, since it results in lower labour costs and does not lead to higher wages. Second, they estimate substantial employment effects, amounting to 2 to 3 percentage point employment increases for workers younger than 26. Hence, workers are very sensitive to changes in wages, which greatly affects the numbers of hours they want to work and their decision to work altogether. Conversely, firms are not very concerned with wages in their decision to hire workers. Instead, Saez, Schoefer & Seim (2019) interpret their results in a way that appears important in terms of clustered behaviour on the part of the firms: firms are subject to a "wage equity" constraint within firms and cannot pay a lower wage to younger workers, even though their marginal product of labour is lower than that of workers who are significantly older. When the payroll tax is cut, the wedge between labour costs and the marginal product of labour due to the wage equity constraints shrinks, resulting in a reduction in the inefficient unemployment of younger workers so that the effects of targeted payroll tax cuts spillover onto workers who were not originally targeted by them via collective firms' behaviour. Saez, Schoefer &Seim (2019) find that the payroll tax cut led to an increase in wages across clusters of firms that were more intensely treated by this cut (because they had a higher share of younger workers).

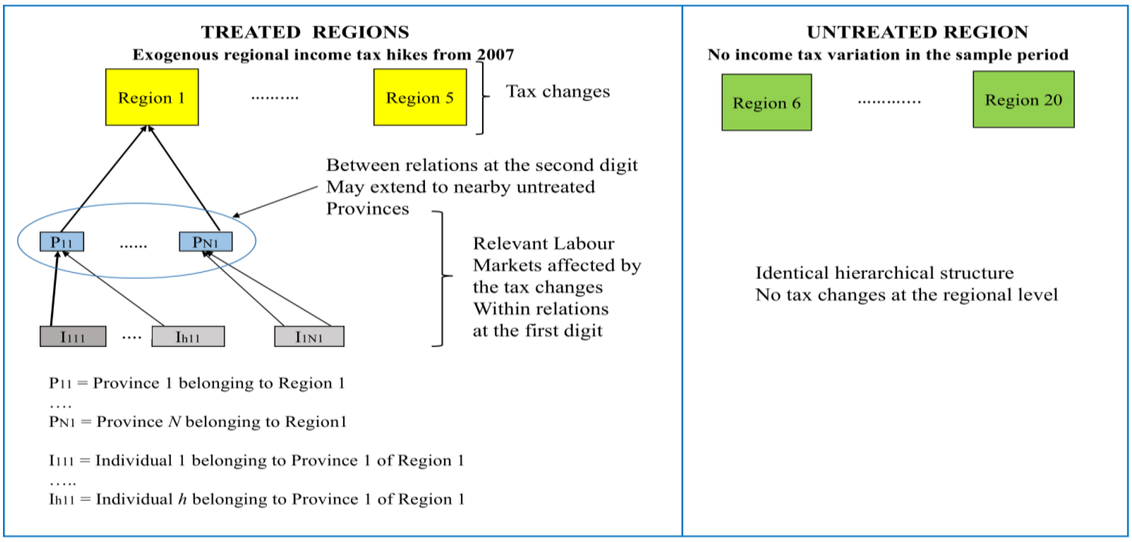
Summing up, while authors may differ on the magnitude of labour-supply elasticities, they largely agree on the sign. The consensus is that increasing tax rates usually reduces work effort. Considering the effect of a rise in a proportional tax, Meghir and Phillips (2010) write (p. 207): "in most cases this will lead to less work, but when the income effect dominates the substitution effect at high hours of work it may increase effort." Keane (2011) states the directionality of the effect without reservation, writing that (p. 963): "the use of labour income taxation to raise revenue causes people to work less." Here and elsewhere, empirical researchers may recognize the theoretical possibility that effort may increase with tax rates but view this as an empirical rarity rather than a regularity. Early studies of the relationship between labour supplied and wage per unit of labour have recognised that since wage is measured net of tax, and since the tax is progressive, the disposable wage may be correlated with the dependent variable and, hence, with the error term. Income tax progressivity makes things even more complicated when tax rates are included in the set of explicatory variables of labour offered. Kean (2011) and Stantcheva (2016) discuss the main econometric problems facing early cross-section OLS studies and subsequent developments and we refer readers to their discussion. Among the problems stressed by Keane (2011, p. 972 and following) and Stantcheva (2016) we stress those related to the endogeneity of the tax rates with progressive income taxation and the extensive vs. intensive margin responses. Yet, the endogeneity problem of wages and non-labour income arises from possible correlation with tastes for work Keane (2011, p. 972) but can also encompass the possible simultaneous relation between labour effort and income tax rates mentioned above for all those people who can adjust labour supply to reduce the tax burden. This has led scholar to abstain from seeking to estimate structural labour supply models but instead rely on a DiD approach (Keane, 2001, p. 1068). In Keane's terms, the idea of DiD is to compare the behaviour of a “treatment” group that was substantially affected by tax changes with that of a “control” group that was little affected. Yet, as noted by Benzarti (2024) whereas payroll taxes are well suited for quasi-experimental designs because they offer better variation (payroll tax exemptions on young workers, by cohort, past certain earning caps, etc.), similar sources of variation are seldom available for other labour taxes. Moreover, with a hierarchical nesting structure of labour supply macro data a “pooled” version of the DiD model is insufficient. Ignoring the hierarchical structure of the data when estimating the labour response to public policies means that the estimation strategy uses less information on the variability of the response variables than we really ought to use and does not guarantee the efficiency of the estimations. The problem is similar to the one noticed by Keane (2022) who emphasises that not only elasticities vary by age, education, gender and marital status but that when individuals differ in terms of human capital investment, the labour elasticity to taxes is higher than conventional wisdom suggests.

Therefore, in many cases the multiple sources of labour response variation may prove as important as the tax ATET estimation in itself. Then, when fitting a model of labour extensive or intensive reaction to exogenous changes in personal income taxation identifying correctly the causal relationship and avoiding the income tax endogeneity problem may not be sufficient. Properly considering the clustering feature of the data generation process is necessary to account for the correlations within units or groups and to estimate the multiple sources of variation simultaneously. In Section 5 we propose in an estimation strategy that deals with both issues and permits to obtain estimates that account for workers heterogeneity due to clusters effects, we present an augmented version of a DiD equation and model intercepts and coefficients in the random part of the model to accommodate different hypotheses of the process governing their variance. Before doing so, a short narrative of the Italian tax experiment is presented in Section 3 below.

3 The Italian tax experiment

In December 2006 the Italian central government forced five Italian regions (Abruzzo, Campania, Lazio, Molise, and Sicily) to raise the rates of two regional income taxes. They were ex-post “punished” for having increased by more than 7% the deficit of their regional health care with respect to the 2005 level. The other fifteen regions were not affected by the tax measure. The measure was adopted to force the punished regions to collect additional local resources and use them exclusively to reduce the deficits of their health system. The government determined the threshold for the inclusion in the treatment (the above 7%) at the end of 2006 and the measure was quite unexpected. The taxes in question are a business profit taxon (IRAP) and the Regional Surcharge on the National Income Tax. A third element of the treatment was related to regional tax expenditures. Italian regions were permitted to allow taxpayers’ deductions from the above regional surcharge payments as an alternative to subsidies, vouchers, service vouchers and other social support measures determined by their regional legislation in favour of some taxpayers. These powers were taken away from the five treated regions. As a result, they could not compensate the surcharge increase with tax deductions. For both taxations, the increase of the tax rates was compulsory, quantitatively relevant and finalized to collect additional resources for their regional health care systems. No other uses were permitted. At the end of 2006, the law introducing the new tax regimes for the above five regions indicated that the measures had to be adopted for a 3-year period, but the measures remained in force for the entire sample period used in this study. Then, once a region became treated in 2007, that region remained treated in the next periods (i.e., the Irreversibility of Treatment applies). More aspects of the tax treatment are discussed in Bosco, Bosco, & Maranzano, 2025). The following graph gives a synthesis of the nesting structure of the data and of the 2006/7 income tax treatment. Additional details are in the *Notes* below Figure1.

Figure 1. *The Italian tax treatment 2006/7 and the nesting structure using two levels*



Notes. Figure 1 illustrates the nesting structure of the Regional/Provincial Italian labour data. Notice that in the figure we describe a two-level structure in which provinces (NUTS-3) are nested within regions (NUTS-2). The core multilevel specification presented in Section 4 introduces an additional third level representing the Eurostat macro-regions (NUTS-1) in which regions and provinces are nested according to a time-constant one-to-multiple relationship. The tax treatment described in Section 3 was forcibly adopted by five treated Regions under a Central Government compulsory Disposition. The between relationship across Provinces indicated in the plot (the first digit indicates the Province and the second the Region) depends upon the common regional regulation and tax policy. At the bottom of the nesting structure, we have agents operating in provincial labour markets where within relations exist (the digit indicates individuals in each Province). As for the tax changes mentioned in left part, the tax hikes were the following: IRAP tax rate existing in 2006 (4.5%) and employed by the five treated regions was increased by almost a 1% and the actual tax rate for treated regional taxpayers became equal to 5.25%. From the tax year 2007 onwards, the five “treated” regions had to increase Regional Income tax surcharge by 0.5% the ongoing rate of 1.23% they applied as a surtax on the personal income tax base of their residents. The surcharge rate was further increased by an additional 0.3% and the net total increase was equal to 0.8% leading to an income Tax Surcharge rate of 2.03%. The exogenous and unexpected increases of Income taxation mainly fell on Individual Business and Self-Employed workers. Yet also the generality of treated regional taxpayers paying the national Income Tax were affected by the measure although less severely (they had to face only the increase of the regional personal income surcharge) unless in addition to paying the tax hike they had to forgo the above-mentioned tax deductions.

Hence, based on the realization of a *state of the world* (the deficit of their regional health care systems) determined when the treatment was not even under discussion, the treated taxpayers were exposed to an exogenously determined tax increase of the main regional direct taxes. Hence, the Italian tax experiment is a variant of a "design-based" experiment with (almost) random participation and permits conducting a "design-based inference" for DiD analysis where the randomness in the data does not come from a pure stochastic assignment of regions to treatment or from drawing regions from an infinite super-population of regions. In this connection, Roth, Sant'Anna, Bilinski, & Poe (2023, p. 2219) recall that methods that are valid from the canonical sampling-based view are typically also valid from the design-based view as well, with the recommendation of clustering standard errors at the level at which the treatment is independently assigned (the regional level in our case). This estimation strategy will be implemented in the nest sections.

4 Identification of a DID tax model with single and multilevel structure

In what follows we construct a multi-level DID estimation of the effect of the income tax treatment using a panel data set of labour supply and cofactors variables summarised in Table A1 (Appendix) which contains summary statistics of the response and cofactors as well as some stationarity tests.
To illustrate the estimated model, we first introduce the following notations:

$$y_{it} = \text{observed "labour response variable" in region } i \text{ at time } t$$

$$D1 = \begin{cases} 0 \text{ if the region is untreated} & (\text{irrespective of } t) \\ 1 \text{ if the region is treated} & (\text{irrespective of } t) \end{cases}$$

$$D2 = \begin{cases} 0 \text{ for years in which there was no treatment (irrespective of } i) \\ 1 \text{ for years in which there was treatment (irrespective of } i) \end{cases}$$

The total sample period is T = (1995… 2021). The No-Treatment period (dummy variable $D2$ = 0) is t = (1995 … 2006), and t = (2006 + 1 … 2021) is the Treatment period (dummy variable $D2$ = 1)

$\boldsymbol{X}$ = vector of cofactors/controls (to be specified in each equation)

As noted above, we define the target or response variable y in terms of potential outcomes and estimate the average effect of treatment on the treated units. This compares the potential outcomes with treatment to the potential outcomes with no treatment, in the treated group. Written mathematically, the estimated effect of the treatment is the β coefficient of panel data OLS equation in a single-level model (i.e., the baseline DiD regression model) is

$$y_{i,t} = Constant + [\alpha(D1) + \delta(D2) + \beta(D1 \times D2)] + \boldsymbol{X}_{i,t}\gamma + error_{i,t} \tag{1}$$

The crucial parameter to estimate is β. If we call R a binary income tax treatment indicator, we have $\hat{\beta} \equiv CATET = \mathbb{E}(y^1_{i,t} - y^0_{i,t} | X_{it}; R_{i,t})$ or $\hat{\beta} \equiv ATET = \mathbb{E}(y^1_{i,t} - y^0_{i,t} | R_{i,t})$ if cofactors are included or excluded. The Conditional Average Effect upon Treated (CATET) or the ATET will be interpreted as the mean effect (conditional or unconditional) of the "tax treatment" for those taxpayers who were compelled to participate in the central government program of income tax changes (i.e. the residents in the treated areas). Unless it is not explicitly specified, in all estimates the average treatment effect is estimated by adjusting for both cross sectional and time effects. Note that outcomes variables are independent on treatment conditional upon $\boldsymbol{X}$. We maintain that the *unconfoundedness* hypothesis holds. To correctly identify the parameters of interest in Equation (1), tax shocks need to be exogenous conditional on fixed effects and controls. Intuitively, this identifying assumption is that the national tax shock of 2007 is not favouring regions that are doing poorly relative to how fast they normally perform in terms of dependent variables. Then, the validity of comparing outcomes of regions having different distributions of the response variables relies on three key assumptions: (1) the national tax shock is exogenous, (2) targeted tax shocks are unrelated to any possible targeted level of the response variables, and (3) outcomes from untreated regions provide a reasonable counterfactual since the 2007 tax shock was absent. Since we control for region and year fixed effects in Equation (1), the first assumption maintains that Italian policy makers of the time were not systematically setting income tax policy to respond to idiosyncratic regional shocks other than the already mentioned budget unbalance of their health care systems.

Yet, being a single level model, the basic Equation (1) ignores the implications of both the within and between cluster variability of the data recorded at the Provincial level on labour responses to an income tax treatment. As it is illustrated in Fig. 1 above, provincial labour markets are nested into regional markets and so equation (1) cannot offer a satisfactory instrument for the ATET/CATET estimation because it ignores the two-level partition of the data and the hierarchical structure (Provinces nested into Regions) of the data set. Equation (1) needs to be transformed as it is explained in the following sections to incorporate and hierarchical nesting structure of the data.

4.1 The identification with a multilevel structure

Availability of provincial, regional, and macroregional data permits the identification of a multilevel DiD model of the reaction of response variables (measured at the provincial level) to exogenous income tax treatment operating at the regional level and affecting all the provinces of each region. This multilevel partition of the data introduces a hierarchical structure in the data set. In general, when data is collected from clusters (e.g., schools, administrative areas, hospitals, etc.) observations within clusters – the Italian Provinces (NUTS-3) within Regions (NUTS-2), in turn within Macro-regions (NUTS-1) in our case – are more likely to be similar. For example, workers from different regions might perform differently in a common test while the performances of workers of different provinces of the same region might have some similarities depending upon common practices, education, or professional training activities organised by their regions. Here, both the regions and the macro-regions are the clusters and test scores of workers' activities are provincial observations nested within regions. Consequently, when fitting a model of labour extensive or intensive reaction to changes in personal income taxation a researcher has to define an estimation strategy that not only solves the tax endogeneity problem by adopting the DiD approach but also allows to deal with the (additional and independent) clustering problem.

4.1.1 A three-level DID linear mixed-effect model for nested data

To deal with the hierarchical problem we introduce random variables for the intercept in a new version of Equation (1) where data are collected in each province of each region (treated and untreated) of each macro-region in each year (with and without treatment) and then estimate their variance to obtain indications on how the hierarchical structure of the data affects the results.

The resulting data structure of the model can be described as follows. Let i denote each province, j each region, h each macro-region, and t each year. We have $N = 107$ provinces ($i = 1, \dots, 107$) unevenly included in $J = 20$ regions ($j = 1, \dots, 20$ and n_j provinces for each region) that in turn belong to $H = 5$ macro-regions ($h = 1, \dots, 5$ and n_h regions for each macro-regions). Each provincial data is observed for the same T time interval [2004 … 2007, … 2020] where 2007 is the first year of the treatment (common to all treated Province/Regions/Macro-regions and in operation until the end of the sample data). Recall that between macro-regions and regions it holds a time-constant $1{:}\,n_j$ relationship, as well as a time-constant $1{:}\,n_h$ relationship holds between regions and provinces. Thus, for each given province we know to which region and macro-regions it belongs.

A three-level DID panel data model[4] can be represented according to a linear mixed representation such that the above Equation (1) is modified to have an intercept allowing for provincial (level-1) regional (level-2) and macro-regions (level-3) variability:

$$y_{ijh,t} = \gamma_{000} + \underbrace{\underbrace{[\alpha(D1) + \delta(D2) + \beta(D1 \times D2)]}_{\text{Tax Treatment Component}} + \underbrace{[\gamma_{01} W_{jh,t} + \gamma_{10} X_{ijh,t}]}_{\text{Cofactors}}}_{\text{Fixed Part}}$$

$$\underbrace{+u_{0h} + u_{0jh} + u_{0ijh} + r_{ijh,t}}_{\text{Random Part}} \qquad (2)$$

where,

- $y_{ijh,t}$ is the labour response variable recorded in Province i nested in Region j nested in Macro-Region h at time t

[4] In this paper we consider time series concerning socio-economic variables $y_{ijh,t}$ observed at the provincial level and thus belonging to a hierarchical structure such that provinces (NUTS-3) belong to regions (NUTS-2) that in turn belong to macro-regions (NUTS-1). Typically, a three-level structure refers to a multilevel structure in which individuals (level-1) are clustered into level-2 groups, which in turn belong to level-3 groups (e.g., students nested into classes nested into schools). Here, we do not consider the time as a proper index for the individuals. Rather, we consider the whole provincial time series as the level-1 observations, while provinces, regions and macro-regions serve as level-1, level-2, and level-3 clusters, respectively.

- $X_{ijh,t}$ is the vector of provincial level cofactors observed for Province i of Region j of Macro-region h in year t affecting the provincial response variables $y_{ijh,t}$
- $W_{jh,t}$ is the vector of regional level cofactors affecting the response variables $y_{ijh,t}$
- u_{0h}, u_{0jh} and u_{0ijh} are the random effects terms.
- $r_{ijh,t}$ is the idiosyncratic error term.

Notice that for all i, j, h, and t, we assume that both the random effects (REs) and the idiosyncratic error term $r_{ijh,t}$ are mutually independent following a Gaussian distribution having of parameters $u_{0h} \sim N(0, \sigma_h^2)$, $u_{0jh} \sim N(0, \sigma_j^2)$, $u_{0ijh} \sim N(0, \sigma_i^2)$, $r_{ijh,t} \sim N(0, \sigma_r^2)$. The level-specific variances are unknown parameters to be estimated through their joint distribution (for details see Sections 11 and 13 of Gelman, 2014).

Equation (4) shows the presence of a *fixed-effect part* and a *random part in* addition to the tax event component, which belongs to the fixed part of the model. Then, Equation (2) defines a mixed-effects models containing both fixed effects and random effects (Demidenko, 2013; Pinheiro & Bates, 2006).

The fixed-effects part of the model still contains the dummies needed for the estimation of the *ATET (*or *CATET* when cofactors are included) effect of the tax treatment at the provincial level (i.e., the interaction of the two dummies D1×D2). Then $\hat{\beta} \equiv CATET = \mathbb{E}\left(y_{ijh,t}^1 - y_{ijh,t}^0 \middle| X_{ijh,t}, W_{jh,t}; D_{i,t} = (D1 \times D2)\right)$ with $D_{i,t}$ = (0, 1). With Equation (2) we are postulating that a TWFE simple DID model with only level-1 individual cofactors cannot exhaustively explain the effect of the tax treatment. The reaction, for example, of labour supply to income taxation in a treated province (i.e. in a province belonging to a treated region) may be conditioned by idiosyncratic factors operating at both provincial and regional level and by panel fixed and time effects.

This part is responsible for the overall regression line and includes a common constant γ_{000} (say, the general *nationwide* intercept), that is analogous to the standard regression coefficients of the DiD regression of the initial Equation (1) and is estimated directly, plus cofactors recorded at both provincial and regional level. Specifically, we have province-specific covariates (i.e., ***X)*** and region-specific covariates (i.e., ***W),*** with the latter changing only across *j* and *t*, and not at the provincial level. This is important for our special case of policy analysis because the 2006 tax treatment applies at the group level (regional income taxation). In such cases, it is crucial for obtaining correct inference to recognize the cluster correlation (Wooldridge, 2010, Section 20.3.2) across units. In effect, if one has observables in the model measured at the group level (region) whether or not they change over time, it is necessary to assume that a province cannot belong to more than one group (Region) or has changed group during the sample period. Therefore, the DID fixed part of Equation (4) is correctly identified by a *No Multiple Membership* assumption, that is, during the sample period each first level unit (province) is continuously nested in the same second level unit (region). The assumption accords with the institutional framework of the Italian Province/Region relationship valid during our sample period.

The random effects part is shaped on the intrinsic hierarchical structure of the data and includes a province-specific random component u_{0ijh}, a region-specific random component u_{0jh} and a macro-regional random component u_{0h}. Each RE represents the variation of intercept associated with the corresponding unit (e.g., the i-th Province for u_{0ijh}) with respect to the national grand mean.

Notice that for each level all the units belonging to that level share the same variance (e.g. σ_i^2 is unique and common to all the provinces). This implies that our model is homoscedastic within level and heteroskedastic across levels. Moreover, the four variances are time constant, and can be specified as follows:

$$Var(y_{ijh,t}|W_{jh,t}, X_{ijh,t}) = Var(y_{i'j'h',t}|W_{j'h',t}, X_{i'j'h',t}) = \sigma_h^2 + \sigma_j^2 + \sigma_i^2 + \sigma_r^2 \quad \forall i \neq i', \forall j \neq j', \forall h \neq h'$$
$$Var(y_{jh,t}|W_{jh,t}, X_{ijh,t}) = Var(y_{j'h',t}|W_{j'h',t}, X_{i'j'h',t}) = \sigma_h^2 + \sigma_j^2 + \sigma_r^2 \qquad \forall j \neq j', \forall h \neq h'$$
$$Var(y_{h,t}|W_{jh,t}, X_{ijh,t}) = Var(y_{h',t}|W_{j'h',t}, X_{i'j'h',t}) = \sigma_h^2 + \sigma_r^2 \qquad \forall h \neq h'$$
$$Var(y_{ijh,t}|W_{jh,t}, X_{ijh,t}) \neq Var(y_{jh,t}|W_{jh,t}, X_{ijh,t}) \neq Var(y_{h,t}|W_{jh,t}, X_{ijh,t}) \quad \forall\, i, j, h$$

The motivation of the adopted version of the above DiD model is twofold. On the one hand, mixed models may provide more accurate standard errors and more reliable estimates of the treatment effect coefficients. On the other hand, the simultaneous estimation of the multiple sources of labour response variation may prove as important as the ATET estimation in itself. For that reason, we will specify different versions of Equation (4) for estimation purposes to obtain information about the way in which the response is affected by the different possible sources of variation exists across levels.

4.1.2 Specifications of the multi-level model useful for policy analysis

We can now consider three particular specifications of Equation (4) in which the random intercept and random slope linear mixed model are formulated in a way that permits policy-relevant interpretation of the outcomes.

First specification: Regional Clustering

We first start with a specification in which we allow only for region-specific random intercepts and for both provincial and regional cofactors. This is equivalent to impose in the random part of Equation (4) that $u_{0h} = u_{0ijh} = 0$ and all other parameters are different from zero. This means that we assume that the cofactors affect the response in a common "national" way whereas the average level of the response is allowed to vary across regions. Notice that here we are ignoring both the province-within-region-within-macroregion and the provinces-within-regions structure as only the provincial RE is included in the model.

Second specification: Provincial clustering

The second specification considers a province-specific clustering effect instead of the regional one. This is achieved by imposing that in Equation (4) $u_{0h} = u_{0jh} = 0$, i.e. that is the cofactors affect the response in a common "national" way, but the average level of the response varies among provinces. Notice that here we are ignoring the regions-within-macroregions structure as only the regional RE is included in the model.

Third specification: Macro Regional clustering

In the third specification we assume the existence of a macro-regional RE shifting the global intercept for each macro-regions while ignoring the provincial and regional hierarchical levels. That is, in Equation (4) we assume that $u_{0jh} = u_{0ijh} = 0$.

Fourth specification: Province-within-Regions clustering

The fourth specification assumes that $u_{0h} = 0$, that is, the cofactors affect the response in a common "national" way but the average level of the response is determined by the combination of a regional specific effect and a province-within-region effect evaluated by clustering the province of each region.

Eventually, one could also consider a restricted version of Equation (4) by imposing the constraint $u_{0h} = u_{0jh} = u_{0ijh} = 0$ and all other parameters are different than zero. In this case, Equation (4) reduces as the basic (or single-level) DID model with co-factors in Equation (1) and $\hat{\beta} \equiv CATET = \mathbb{E}\left(y^1_{ijh,t} - y^0_{ijh,t} \middle| X_{ijh,t}, W_{jh,t}, D_{i,t} = (D1 \times D2)\right)$ with $D_{i,t}$ = (0, 1) according to the occurrence of the treatment. In this case, we should conclude that the effect of the tax treatment on the response variables is immune of clustering elements. Furthermore, by excluding the level-2 vector $\boldsymbol{W}$, we are implicitly assuming that the reaction labour supply to income taxation is not conditioned by idiosyncratic factors operating at the regional level, but only by provincial cofactors directly affecting the conditional expected values of the response variables. Clearly, this version will represent the standard DiD model of the labour response to a tax treatment and will provide outcomes against which the results of the above specified cluster estimates can be compared.

Notice that the all the above specifications differ in terms of presence or absence of the multi-level REs and are nested sub-cases of Equation (4). To the test for the significance of REs requires suitable likelihood ratio testing (LRT) procedures, such as the implementation of LRT fitted with and without random intercept and calculate the log-likelihood of each model. The formula for likelihood ratio testing is given as $LRT = -2\log\left[\frac{L_a(\theta)}{L_b(\theta)}\right]$where the numerator is the log-likelihood of equations with fewer parameters (e.g., no random intercept parameter) and the denominator is the log Random Coefficient Model Log-likelihood of equations with greater parameters (e.g., with random intercept parameter). The null hypothesis is that model with fewer parameters is best while the alternate is in favour of a random intercept model or model with more parameters. Alternatively, the null hypothesis can be stated as H_0: $\sigma^2_{0h} = 0$ which means we can ignore the extra parameter. Then we will compare it to the χ^2 distribution where the degree of freedom is the number of extra parameters.

4.1.3. Variance decomposition and the Intra-Class Correlation Coefficient

When hierarchical samples come with a time component, there are at least two potential sources of correlation across observations: across time within the same individual (i.e., a generic NUTS-3 province) and across individuals within the same group (i.e., NUTS-2 regions in which each NUTS-3 province is institutionally nested). The two sources of correlation may also interact, that is, different units within the same group or cluster might be subject to unobserved shocks correlated across different periods. However, in our case, cluster identification does not change over time. Accordingly, we have a balanced panel data on each individual or unit, and each unit "naturally" belongs to a cluster.

Let us suppose that the response variable y_{ijh} is modelled according to Equation (2). We recall from the previously defined notation that, conditioning on the set of covariates, the total variance of the province-within-region-within-macroregion labour response variable is given by $Var(y_{ijh,t}|W_{jh,t}, X_{ijh,t}) = Var(y_{i'j'h',t}|W_{j'h',t}, X_{i'j'h',t}) = \sigma_h^2 + \sigma_j^2 + \sigma_i^2 + \sigma_r^2$. Then, we can define the (*residual)* Intraclass Correlation Coefficient (ICC)[5] as the linear correlation among pairs of (randomly sampled) observations within the same (randomly drawn) cluster controlling for the available set of cofactors.

The ICC can be used as a tool to determine whether a linear mixed model is even necessary; in particular it can be theoretically meaningful to understand how much of the overall variation in the response variable is explained simply by clustering. An ICC close to zero means that the REs at a given level is almost zero across all the groups, then the clustering structure is irrelevant, and one could have used ordinary linear regression models. In this case, it means the observations within clusters are no more similar than observations from different clusters and that clustering is useless. Conversely, the higher the ICC the stronger the similarity between response variables *within* the groups (i.e., the regions), then the hierarchical linear model is a more reliable method than standard regression analysis. In other words, a high ICC makes the multi-level DID even more justified.

In the framework of our administrative data, the ICC can be either computed with respect to the provincial, regional, and macro-regional level. Following Siddiqui et al. (1996, pp. 425-426) and Hox et al. (2017, pp. 20-22), in Table 1, we synthesize the level-specific ICCs that can be computed according to five specifications discussed in Section 4.1.2.

Table 1. *Clustering specifications and corresponding ICCs formulation*

Clustering specification	ICCs formulae
Regional	$ICC_{Reg} = \dfrac{\sigma_j^2}{\sigma_j^2 + \sigma_r^2}$
Provincial	$ICC_{Prov} = \dfrac{\sigma_i^2}{\sigma_i^2 + \sigma_r^2}$
Macro-regionals	$ICC_{Macro} = \dfrac{\sigma_h^2}{\sigma_h^2 + \sigma_r^2}$

[5] For technical details see Sections 2.2.3 and 4.5 of Hox et al. (2017) or Section 3.3 of Snijders & Bosker (2011)

Provinces-within-Regions	$ICC_{Reg} = \frac{\sigma_j^2}{\sigma_j^2 + \sigma_i^2 + \sigma_r^2}$ $ICC_{Prov-Reg} = \frac{\sigma_j^2 + \sigma_i^2}{\sigma_j^2 + \sigma_i^2 + \sigma_r^2}$
Provinces-within-Regions-within-Macroregions	$ICC_{Macro} = \frac{\sigma_h^2}{\sigma_h^2 + \sigma_j^2 + \sigma_i^2 + \sigma_r^2}$ $ICC_{Reg-Macro} = \frac{\sigma_h^2 + \sigma_j^2}{\sigma_h^2 + \sigma_j^2 + \sigma_i^2 + \sigma_r^2}$ $ICC_{Prov-Reg-Macro} = \frac{\sigma_h^2 + \sigma_j^2 + \sigma_i^2}{\sigma_h^2 + \sigma_j^2 + \sigma_i^2 + \sigma_r^2}$

Recall that, ATET or CATET is a *mean* measure of the effect of the treatment upon treated units. Then, a high ICC says that the deviation from this mean of the labour responses to income tax changes recorded in each province is conditioned by the region in which that the provinces is nested. This means that the response variables recorded in provinces belonging to treated regions tend to be closer or further to this common mean than provinces belonging to another treated region. Simple TWFE panel data DiD model with pooled data sets tends to obscure this information.

5 Summary statistics of the dataset

Multilevel models require that the grouping criterion is clear and that variables can be assigned unequivocally to their appropriate level. In our case group boundaries are not *fuzzy* (they are determined by the Italian Constitution) and variables do not generate cross-level interaction.

We estimate various version of Equation (4) using data of 107 Italian Provinces (NUTS-3 level), each nested in one of the 20 Regions (NUTS-2) and Macro-regions (NUTS-1) and recorded annually from 2004 to 2020. We also collected provincial and regional macroeconomic indicators on a yearly basis for the same time period. Data are provided by the Italian National Statistical Institute (https://www.istat.it/it/dati-analisi-e-prodotti/statistiche) and by the (Eurostat, 2024) regional database.

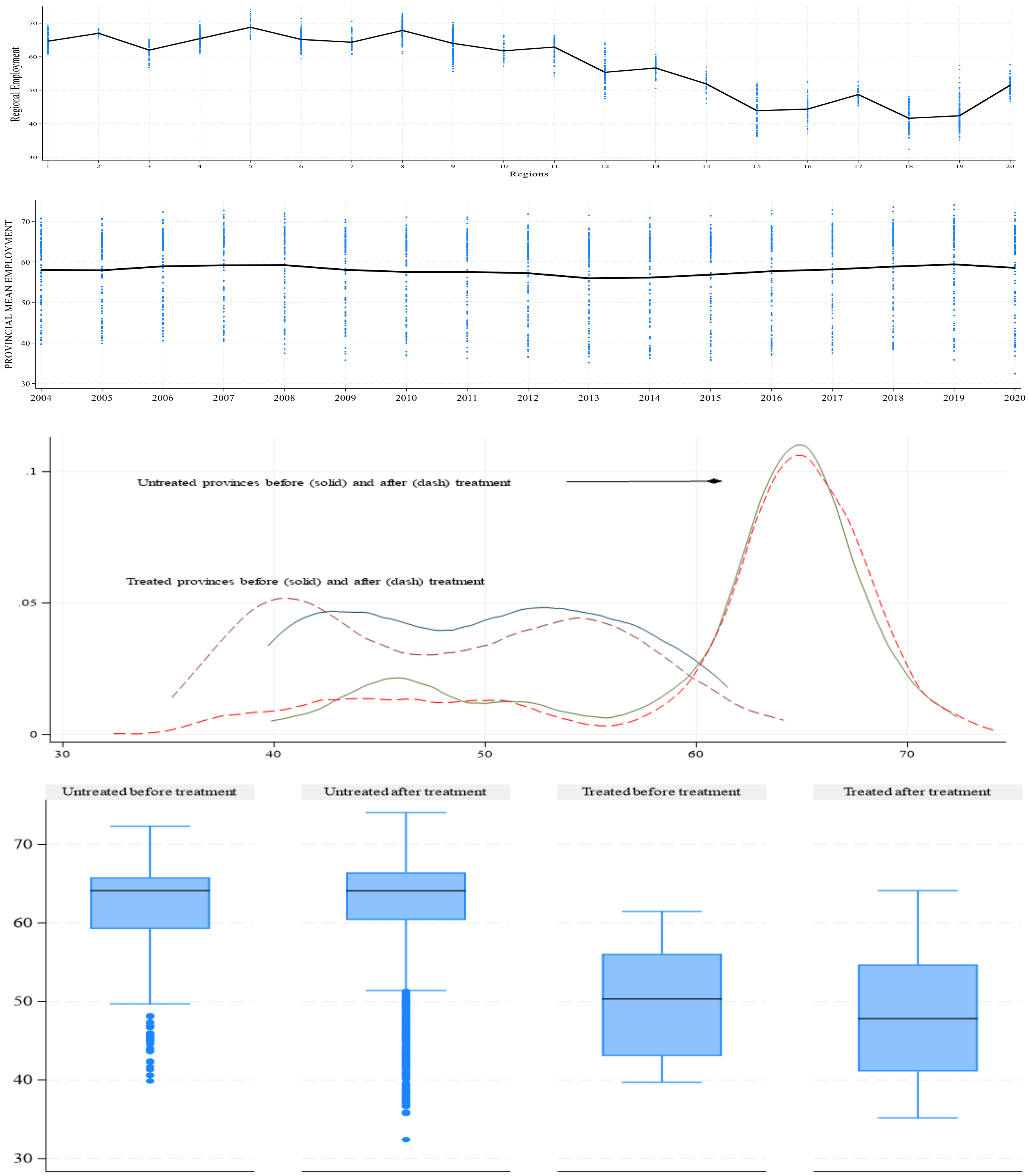
Our main response variable is the *Total Rate of Employment* computed as the percentage of 15-64 years-old population employed[6]. While no single variable can capture all the nuances existing in the labour market, we feel that using the employment rate of a province can help determine the economic health of that area and establish how well a business may perform or how easy it may be to get a job. In other terms, it allows a more complete understanding of the economic implications of income tax reform because it accounts for a flexible

[6] The ISTAT-Eurostat definition of the provincial employment rate (Eurostat, 2025) that we adopted omits individuals under 14 or over 64 years old, full-time students, people with limitations that prevent working, and institutionalized individuals. The rate takes the number of employed individuals and compares it to the total available labour force.

framework that admits interactions among consumption, investment and leisure choices over time. The other reason for using the Eurostat NUTS-3 *Rate of Employment* as a response variable is that it may come in different specification. We have the *Rate of Employment* recorded for the two genders and different age classes (Maranzano & Pelagatti, 2025). As a result, tax effects on employment can be evaluated for males and women of different age as well as for young and senior workers of the two genders by constructing specific *Employment Indexes*. To identify the tax effects on work incentives and consumption, Ziliak and Kniesner (2005, p. 778) use household-level data of labour supply defined as annual hours of work from all jobs on male heads of household. For those reasons, we prefer it to the unemployment rate, which measures the share of workers in the labour force who do not currently have a job (at least officially) but are actively looking for work.

Density plots and boxplots shown in Figure 2 below illustrate the time behaviour of the *Rate of Employment* before and after 2006 (the last year before treatment). From the plots we gather that the income tax treatment was administered to taxpayers living in regions with low level of employment.

Figure 2 *First panel Mean Employment Rate (age 15 – 64, Male and Female) in each region; Second panel Provincial Employment over time; Density plots and Boxplots*



Notes. Fig. 2 shows (first panel) the mean realizations of the response variable *Employment Rate* (age 15—64) in each (treated and untreated) region. Vertical points correspond to yearly realization of the variable and the solid black line shows the path of the mean values across regions. The second panel shows the provincial realizations of *Employment Rate* (age 15—64) from 2004 to 2020 at the provincial level. The vertical points are the actual realizations (103 observations) and the solid black lines indicate yearly mean values. The third plot shows the densities of the *Rate of Employment* recorded in treated provinces before and after the first year of the income tax change (treatment). The left shift of the density is confirmed by the reduction of the mean and median values clearly shown by the box-plot reproduced in the fourth plot. Notice the low outliers for the untreated provinces.

6 TWFE Estimates

We start with a single level DID model of the effect of income tax changes on the overall provincial *Employment Rate* (male and female, age 15-64). The causal effect is estimated through a panel TWFE specification with common treatment period identified by the basic DiD assumptions, individuals (i.e., provinces) fixed effects and time fixed effects in ordinary least squares estimation. The following Table 2 repost results of various version of the TWFE model to be intended as benchmark for the comparison with the next multilevel specifications. MOD1 is a pure ATET version without cofactors; MOD2 includes two province-specific cofactors (i.e., average per capita GDP and average per worker GVA); MOD3 augments model MOD2 with further region-specific cofactors (i.e., average per capita GDP and average per worker GVA at the regional level); and MOD4 adds a further control variable computed as the national annual average value of the employment rate and is included to incorporate the mean trend of the response variable (Mundlak, 1978).

Table 2. *Single-level DID ATET TWFE Estimations of various versions of Equation (1)*

Response Variable **RATE OF EMPLOYMENT**	**MOD1**	**MOD2**	**MOD3**	**MOD4**
INCOME TAX TREATMENT				
(CATET Treated vs Untreated)	**-1.72*** **	**-1.41*** **	**-1.32*** **	**-1.32*** **
YEARS				
2004	omitted	omitted	omitted	omitted
2005	-0.09	-0.20	-0.18	-0.18
2006 (last year before treatment)	.89***	0.52***	0.46**	0.66***
2007	1.56***	0.83***	0.72**	1.02***
2008	1.62***	0.86***	0.71**	0.87***
2009	.43*	0.04	-0.05	-0.07
2010	-0.11	-0.61**	-0.65**	-0.76
2011	-0.07	-0.72**	-0.77**	-0.87*
2012	-0.41	-0.99***	-1.06***	-1.14**
2013	-1.66***	-2.15***	-2.14***	-2.36***
2014	-1.47***	-1.93***	-1.89***	-2.10***
2015	-0.79**	-1.39***	-1.36***	-1.50**
2016	0.07	-0.77*	-0.83*	-0.85*
2017	0.50	-0.54	-0.65	-0.54
2018	1.26***	0.08	-0.06	0.13
2019	1.80***	0.31	0.08	0.54**
2020	0.91 ***	-0.12	-0.31	(omitted)
Constant	57.95***	52.10***	52.32***	60.71

COFACTORS				
Prov. Avg. GDP per capita (pps)		2.53×10^{-4}***	1.63×10^{-4}*	1.63×10^{-4}*
Prov. Avg. GVA per emp.		5.17×10^{-6}	3.48×10^{-5}	3.48×10^{-5}
Reg. Avg. GDP per capita (pps)			1.79×10^{-4}	1.79×10^{-4}
Reg. Avg. GVA per emp.			-7.5×10^{-5}	-7.5×10^{-5}
Reg. GDP growth rate			-0.01719	-0.01719
National Avg. Emp. Rate				-0.13783
Time Fixed Effects	√	√	√	√
Individual Fixed Effects	√	√	√	√
Observations	1,815	1,815	1,815	1,815
Parallel Trends Test				
F(1, 106)	0.37	0.28	0.29	0.29
	Prob > F = 0.544	Prob > F = 0.595	Prob > F = 0.5890	Prob > F = 0.589
Anticipation Effects				
Granger causality test				
F(2, 106)	0.31	0.19	0.18	0.18
	Prob > F = 0.735	Prob > F = 0.828	Prob > F = 0.837	Prob > F = 0.837

Notes. The null hypotheses of parallel trends[7] and No anticipatory effects[8] cannot be rejected at any level of significance. Interpretation of ATET/CATET is straightforward: the rate of employment is negatively affected by income tax increases in any specification of the model. MOD4 includes as cofactor the population grand mean of the response variable, National Mean Employment Rate. It is computed as $(107 \times 17)^{-1} \sum_{i=1}^{103} \sum_{t=2004}^{2020} y_{it}$. Including as a cofactor this "double demeaned" version of the response variable is motivated by the need to incorporate a long run trend of employment into a model such as (6) where time fixed effects are included, and therefore the long run changes in the general economic environment that have the same effect on all units are removed (Wooldridge, 2025). Then, since local employment is logically affected by the National Mean Employment, with the latter indicative of the national trend of employment, a factor incorporating the above general environment changes is needed to better define what affects local employment. Finally, when interpreting the estimated coefficient of the cofactors, the scale of measurement is to be considered. For example, as Provincial Avg. GDP PPS per capita is measured in euros per capita, the reported coefficient 2.53×10^{-4} means that if that the average Provincial GDP per capita would increase by 10000 euros, the rate of employment would increase by an 2.53% of its value. Legend: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

[7] To test for parallel trends, we define the model

$$y_{ij,t} = \gamma_{00} + [\alpha(D1) + \delta(D2) + \beta(D1 \times D2)] + \gamma_{10} X_{ij,t} + \psi_1[S_0 \times w_i \times YEAR] + \psi_2[S_1 \times w_i \times YEAR] + r_{ij,t}$$

where S_0 is a time dummy indicating pre-treatment years, S_1 indicates treatment years and w indicates either treated provinces (belonging to a treated region) ($i = 1$) or untreated provinces ($i = 0$). If the estimated $\psi_1 \neq 0$, then a difference in the slope of the time trend between treated and untreated units in pre-treatment years is statistically consistent with the data. Table 1 reports the Wald test of the H_0: $\psi_1 \neq 0$ vs $\psi_1 = 0$. The null is that of linear parallel trends ($\psi_1 = 0$). This means that the test statistics is obtained by dividing the maximum likelihood estimate (MLE) of the slope parameter ψ_1 by the estimate of its standard error. Under the above null, this ratio follows a standard normal distribution.

[8] The last rows of Table 2 report the results of a Granger-type causality test to assess whether treatment effects are observed prior to the treatment.

Results reported in Table 2 accords with the findings reported in the literature reviewed in Section 2 with respect to both micro estimates — typically using tax reforms as quasi-experiments — and macro estimates — typically based on structural estimations or calibrations. Both approaches find a negative effect of tax hikes, although the latter approach generally produces larger elasticities estimates[9]. Our estimates are obtained following a macro approach corrected to allow for causal identification and indicate that the exogenous increase of the income taxation affects negatively the rate of employment in any specification of the estimated model. Interestingly, cofactors have no impact of the response variable (with the exception per capita GDP recorded at the provincial level) but they affect the magnitude of the estimated ATET (or CATET) coefficients. In MOD1, for example, a 1% increase of the income tax rate leads to a 1.72% decrease of the ongoing mean employment rate in treated provinces whereas in MOD3 and MOD4 the effect reduces to a 1.3%. Province level cofactor (GDPs) is strongly statistically significant (for the interpretation of the magnitude, see the Notes below Table 2) and has the expected sign. As for regional level cofactors, GDP has effects analogous to the provincial GDP but its rate of growth has no statistically significant effect. Controlling for the National Mean Employment Rate has no consequences.

7 Multilevel DID estimates

Assume now that we model the provincial rate of employment following the hierarchical multilevel specifications presented in Section 4. Specifically, we consider the following five random-intercept specifications accounting for the nested NUTS hierarchy of the Italian administrative data:

- MixMOD1: random intercept model with regional clustering structure
- MixMOD2: random intercept model with provincial clustering structure
- MixMOD3: random intercept model with macro-regional clustering structure
- MixMOD4: random intercept model with province-within-regions clustering structure
- MixMOD5: random intercept model with province-within-regions-within-macroregions clustering structure.

All the specifications include the same set of control variables, namely the covariates used in MOD4 of the single-level specification.

[9] Even if much microeconometric evidence points to small elasticities, many economists contend that the true long-run effect of taxes is large due to dynamic effects (e.g., Prescott 2004). As recently stressed by Kleven, Kreiner, Larsen, & Søgaardt (2025) this debate relates to the stark difference between micro and macro elasticities of labour supply. Micro studies are based on research designs that allow for causal identification, but the approach only captures short-run effects and may miss important dynamic mechanisms. Macro studies are model-dependent and may be associated with specification bias, but they allow for potentially relevant dynamic responses.

Table 3. *DID ATET Estimations of the Mixed-effects ML regression with distinct clustering choices*

Response Variable RATE OF EMPLOYMENT	Id-Clusters identified by Institutional levels				
	MixMOD1 Id = Regions	**MixMOD2 Id = Provinces**	**MixMOD3 Id = Macro-regions**	**MixMOD4 Id = Provinces-within-Regions**	**MixMOD5 Id = Province-within-Regions-within-Macroregions**
			FIXED PART		
Treatment Effects					
D1	-7.82**	-9.62***	-1.75***	-7.90**	-0.62
D2	-0.03	-0.16	-1.51***	0.03	0.05
Tax Treatment = (D1×D2)	**-1.31*** **	**-1.24*** **	**-1.38** **	**-1.34*** **	**-1.36*** **
Cofactors Effects					
Provincial average per capita GDP (PPS)	4.93×10^{-4} ***	1.44×10^{-4} ***	5.99×10^{-4} ***	2.11×10^{-4} ***	2.14×10^{-4} ***
Provincial average per worker Gross Value Added	-2.69×10^{-4} ***	-0.03×10^{-4}	-2.1×10^{-4} ***	-0.25×10^{-4}	-0.24×10^{-4}
Regional average per capita GDP (PPS)	$-0.63x10^{-4}$	3.38×10^{-4} ***	-3.88×10^{-4} ***	1.95×10^{-4} ***	1.80×10^{-4} ***
Regional average per worker GVA	$-0.79x10^{-4}$	-2.09×10^{-4} ***	4.69×10^{-4} ***	-1.75×10^{-4} ***	-1.73×10^{-4} ***
Regional year-to-year GDP (PPS) growth rate	128.88×10^{-4}	228.34×10^{-4}	-12.13×10^{-4}	409.79×10^{-4}	253.92×10^{-4}
National average employment rate (15-64 age class)	0.45***	0.40***	0.44***	0.06	0.46 ***
Constant (γ_{00})	32.11***	35.59***	12.91**	25.51**	30.97***
			RANDOM PART		
var(Cons)	44.60 (14.31)	44.23 (6.48)	33.03 (21.16)	Var Reg. 45.60 (14.68) Var Prov. 4.23 (0.68)	Var. Macro 39.86 (29.14) Var Reg. 15.07 (5.76) Var Prov. 4.21 (0.67)

Var (Residual)	5.67 (0.19)	2.51 (0.09)	14.93 (0.50)	2.59 (0.09)	2.50 (0.09)
ICC	.903 (.028)	0.95 (0.01)	0.69 (0.14)	ICC Reg. 0.87 (0.04) ICC Prov. 0.95 (0.01)	ICC Macro 0.65 (0.18) ICC Reg. 0.89 (0.05) ICC Prov. 0.96 (0.02)
LR comparison test for linearity Prob >= chibar2	2184 0.0000	3154 0.0000	493 0.0000	3328 0.0000	3339 0.0000
AIC BIC	8424 8490	7453 7519	10115 10181	7281 7353	7273 7350
Number of Groups (Clusters)	21	107	5	21 + 107	5 + 21 + 107
Observations	1815	1815	1815	1815	1815

Notes. Standard Errors in parenthesis. The distribution of the LR test statistic of linear model is not the usual chi-squared with 1 degree of freedom but is instead a 50:50 mixture of a chi-squared with no degrees of freedom (that is, a point mass at zero) and a chi-squared with 1 degree of freedom. ICC and IC tests (AIC and BIC) are evaluated with respect to different DF according to the id variable adopted in each model variant. For example, DF when the id variable is Macro-Region (there are 5 main macro areas in Italy is only 5. Number of clusters are reported at the bottom of the Table. Legend: * p < 0.1; ** p < 0.05; *** p < 0.01.

As one can see all variants generate negative estimates of the treatment effect (on average we may say that income tax hikes reduce the rate of employment by 1.3 percentage points, in absolute value. The largest value of the estimated treatment is obtained when the data clustering is obtained in specification MixMOD3, which uses the five Italian macro areas (Est-North, West-North, Centre, South, and Isles), and the smallest estimate is generated by the MixMOD2 specification, which adopts the provincial clustering criterion. Contrary to the previous results obtained using the single-level specification, with the exception of the Regional year-to-year GDP (PPS) growth rate, cofactors are statistically significant. Altogether the results of the fixed part of the estimated model accord with those obtained using the single level version of the DiD model reported in Table 2 and the literature reviewed in Section 2.

The random part of Table 3 shows the estimated variance components. Recall that with this specification of the mixed model we only impose a random intercept. The term id (Identity of the group) means that we assume that the random effects of the first three version of the model can be either at the province level (as if each province formed a cluster) or at the regional level (each region forms a cluster) or at the macro-area level. MixedMod4 and MixedMod5 incorporate the hierarchical dimension of the estimation procedure and are used to generate estimates of the tax treatment assuming that the data generation process obeys reflect a vertical hierarchy going from the Provinces (level-1) to the corresponding Regions (level-2) to the corresponding Macroregions (level-3).

Both the reported LR tests compare the model with single-level ordinary linear regression. They are highly significant for our data, rejecting the null hypothesis of a non-hierarchical structure. Note that the ICCs never approach zero, meaning that the grouping is important as a source of explanation of variability. In MixMod1, MixMod2, and MixMod3, high value of ICC indicates that the share of explained variance that can be attributed to the groping level is very high compared to the residual unmodeled variability.

We report the ICCs for every level involved in the estimation. ICC changes as variables are added to the model. Yet, when the Provinces are used as the clustering level the estimated ICC is higher than the estimated ICC obtained using regional clustering and the Province level explains a larger part of the overall variance. Because of all the reasons discussed in Section 1 (e.g., culture and habits, and other similarities), within Provinces there is a high degree of similarity of taxpayers' responses to tax increases.

Finally, notice that the Akaike Information Criterion and Bayesian Information Criterion (BIC) are reported to compare the various version of the fitted model. The lowest BIC values (very similar among them) is obtained using MixedMod2 and MixedMod4 and MixedMod5 may indicate that these hierarchical structures of the data better describe the data generation process[10].

7.3 Robustness analysis: Age and gender specific analysis

To check the robustness of the results reported in Table 3 and, more importantly, to conduct disaggregated estimations of labour response we replicated the study using as response variable the *Rate of Employment* generated for different segments (Age and Gender) of the labour force. The purpose is to evaluate whether a relationship between data clustering and Age/Gender differences of taxpayers exists and, in the affirmative, whether young taxpayers (male and female) are more exposed to clustering effects than mature taxpayers (male and female). Results are generated by the estimation of same equation of Table 3 and are summarised in Table 4 below where the first column indicates the distinct classes of age and the second column indicates the distinct classes gender. With respect to the empirical specification presented in the above sections, we focus on the provincial and regional clustering and ignore the macro-regional level. Indeed, as already stressed, the third hierarchical level adds poor policy indications while preserving consistency with the other levels. To save space, we do not report the estimated control coefficients and other statistics but only focus on the estimated ATETs/CATET[11].

Table 4: *CATET estimates from Mixed-Effects DID panel model with Rate of Employment response (age and gender)*

Workers age class	Gender	Regional clustering		Provincial clustering		Provinces within Regions clustering	
		CATET	**ICC**	**CATET**	**ICC**	**CATET**	**ICC**
15 – 64	Total	-1.31***	0.89	-1.24***	0.95	-1.34***	ICC Reg. 0.87 ICC Prov. 0.95

[10] In section 1 we recalled empirical results of the empirical educational studies. Here we simply add that our results are in line with the main bulk of that literature. For example, we may adapt the conclusion of Siddiqui et al. (1996, pp. 425-426): "*Since the students within a classroom are more likely to be similar than the students within a school, the ICCs at the classroom level are always higher than or equal to the ICCs at the school level.*"

[11] Complete results are available upon request.

	Males	-1.74***	0.85	-1.53***	0.87	-1.77***	ICC Reg. 0.84 ICC Prov. 0.90
	Females	-0.99**	0.82	-0.91***	0.92	-1.03***	ICC Reg. 0.79 ICC Prov. 0.93
15 – 24	Total	+3.28***	0.89	+3.78***	0.84	+3.27***	ICC Reg. 0.89 ICC Prov. 0.91
	Males	+3.29***	0.85	+4.25***	0.68	+3.26***	ICC Reg. 0.84 ICC Prov. 0.87
	Females	+3.47***	0.83	+4.46***	0.56	+3.46***	ICC Reg. 0.83 ICC Prov. 0.85
25 – 34	Total	-1.02	0.92	-0.68	0.92	-1.04*	ICC Reg. 0.91 ICC Prov. 0.94
	Males	-3.26***	0.84	-2.53***	0.78	-3.30***	ICC Reg. 0.83 ICC Prov. 0.86
	Females	+0.82	0.89	+1.33*	0.89	+0.79	ICC Reg. 0.89 ICC Prov. 0.92
35 – 44	Total	-1.88***	0.89	-1.67***	0.92	-1.90***	ICC Reg. 0.88 ICC Prov. 0.93
	Males	-3.20***	0.83	-2.72***	0.78	-3.24***	ICC Reg. 0.83 ICC Prov. 0.86
	Females	-0.72	0.81	-0.42	0.87	-077	ICC Reg. 0.80 ICC Prov. 0.89
45 – 54	Total	-2.97***	0.75	-2.78***	0.83	-3.02***	ICC Reg. 0.74 ICC Prov. 0.85
	Males	-2.78***	0.72	-2.22***	0.60	-2.80	ICC Reg. 0.72 ICC Prov. 0.78
	Females	-3.35***	0.60	-3.26***	0.78	-3.44***	ICC Reg. 0.58 ICC Prov. 0.79
55 – 64	Total	-1.97***	0.86	-4.81***	0.84	-4.02***	ICC Reg. 0.90 ICC Prov. 0.91
	Males	-3.80***	0.88	-4.72***	0.79	-3.77***	ICC Reg. 0.88 ICC Prov. 0.89
	Females	-4.17***	0.87	-5.13***	0.80	-4.17***	ICC Reg. 0.87 ICC Prov. 0.89
20 – 64	Total	-4.03***	0.90	-1.88***	0.93	-2.00***	ICC Reg. 0.84 ICC Prov. 0.94
	Males	-2.61***	0.78	-2.33***	0.79	-2.64***	ICC Reg. 0.77 ICC Prov. 0.86
	Females	-1.45***	0.79	-1.37***	0.91	-1.50***	ICC Reg. 0.77 ICC Prov. 0.92
18 – 29	Total	+2.47***	0.92	+2.91***	0.92	+2.46***	ICC Reg. 0.92 ICC Prov. 0.94
	Males	+1.80*	0.88	+2.56***	0.84	+1.78**	ICC Reg. 0.88 ICC Prov. 0.90
	Females	+3.13***	0.91	+3.76***	0.87	+3.12***	ICC Reg. 0.90 ICC Prov. 0.92
Over 15	Total	-0.76***	0.91	-0.69***	0.95	-0.78***	ICC Reg. 0.89 ICC Prov. 0.96
	Males	-0.97**	0.88	-0.79***	0.90	-0.99***	ICC Reg. 0.87 ICC Prov. 0.93
	Females	-0.58*	0.85	-0.51**	0.92	-0.61***	ICC Reg. 0.83 ICC Prov. 0.94

Notes The CATET effects for various age/gender combination of provincial employment rate controlling for provincial average per capita GDP (PPS), provincial average per worker GVA, regional average per capita GDP (pps), regional average per worker GVA, regional year-to-year GDP (PPS) growth rate, and national average employment rate (15-64 age class). Results are obtained from MixMOD1 (regional clustering) and MixMod2 (provincial clustering) of Table 2. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. Complete estimation results and tests statistics for parallel trend and anticipation effects are available upon request.

The first three rows (i.e., referring to the 15-64 years old population) can be directly compared with the DiD results reported in Tables 2 and 3. Tax hikes reduces labour supply for the generality of taxpayers and no differences in sign exists between male and female workers. Yet, female labour supply reaction appears less intensive, and this finding is not in accord with much of the empirical literature reviewed in Section 2. A clustering effect emerges from any specification of the model as it is evident from the coefficients reported in columns 4, 6, and 8. The specification "Provinces within Regions" (last columns) assumes a hierarchical clustering (Provinces within each Region) and shows values of the ICC from which we may conclude that provincial clustering is more important than regional in explaining behavioural similarities across taxpayers of both genders. The same findings are obtained using other series of the Rate of Employment, namely Age ≥ 15 and $20 \leq$ Age ≤ 64. The difference between the latter and the previously studied largest cohort ($15 \leq$ Age ≤ 64) is that the latter includes workers with more educational qualifications (years of schooling).

Different results are obtained from the two Age cohorts between 25 and 44. Female reaction to income tax changes is not statistically significant or even positive. Women in the motherhood age interval, probably due to the labour search cost in which they may incur after motherhood and child caring, prefer not to quit their current occupation even if the opportunity costs of home activity are altered by income taxation. Results from the cohort $18 \leq$ Age ≤ 29 point in the same direction. The hierarchical clustering specification "Provinces within Regions" (last columns) produces values of the ICC from which we may conclude that, once again, provincial clustering is more important than regional in explaining behavioural similarities across taxpayers of both genders, although the provincial ICC is always larger for female taxpayers.

In general, male employment always responds negatively to income tax hikes in all the examined versions of the model with the relevant exception of the cohort $15 \leq$ Age ≤ 25 (workers potentially having an intermediate education and short occupational histories). Although this finding may be interpreted as evidence of a sort of tax induced replacement effect (young workers may substitute old workers who reduced their labour supply because of a tax hike), lifecycle considerations may be part of the explanation of the positive and statistically significant estimated coefficients in any version of the clustering hypothesis. As recently stressed by Kleven, Kreiner, Larsen, & Søgaardt (2025) the profile of earnings may be a step function with discrete changes at job switches such as occupation or firm switches. Expectations of earnings growth (possibly by promotions) over the lifecycle either through gains in the job-match component of wages or through mobility to firms with higher wage premia can justify the absence of a negative labour supply response on the part of male young workers. Moreover, one may argue that young and poorly qualified workers (male and female) cannot freely reduce effort and earnings within a given job cell. On the contrary, they may increase effort, if there is a prevalence of income over substitution effects after taxation or if they expect that their future larger earnings (often tied to job switches) are positively dynamically related to their current unchanged (or even increased) effort.

In general, the tax treatment effect decreases with age whereas ICC increases with age. The latter finding assesses that the degree of similarity within each Region (cluster) is likely to be smaller across young workers in spite of a likely higher probability of having similar educational backgrounds. On the contrary, the higher ICC generated by the models fitted with more mature workers' data shows that their post-tax-increase behaviour may depend by similar job qualifications and/or experiences resulting by the similarity of the common general environment). Tax effects on labour supply are to some extent affected by the environment similarities. Other estimation procedures could not permit obtaining similar findings. As for our specific tax treatment, one should recall that the tax hikes affect the Marginal Tax Rate and therefore affect proportionally more the workers with a large income tax base belonging to the last tax brackets. Since the income tax base is generally correlated with age, then senior workers, possibly at the end of the working career, may prefer to pay more taxes and receive a lower disposable income[12] to abandoning participation (for example by anticipating retirement). Hence, the same changes in the tax rates elicit more elastic response behaviour of young workers who may also underestimate the costs of searching for less taxed activities and overestimate their experience and reputation. On the contrary, fear of worse future labour market outcomes and opportunities (likely unemployment or future precarious working contracts), worries about the income risk of their families, specific cultural habits as well as other socio-economic factors, may induce mature people to accept reductions of net of tax wages without extensive or intensive work adjustments. Then, young people excessive optimism about occupation may produce effects possibly similar to the effects produced by salience. Finkelstein (2009), Waseem (2020,) and Taubinsky & Rees-Jones (2018) explore different taxpayers reactions based on salience and show that the efficiency costs of (indirect) taxation are amplified by differences in under-reaction across individuals and across tax rates. Yet, although inattention and imperfect optimization could be particularly important in the case of taxation (Chetty, Looney & Kroft, 2009 p.1145), lack of full optimization with respect to taxes may not be the exclusive explanation of the above age-contingent tax result.

In general, labour responses to tax hikes decreases with age possibly because workers above a certain age do not reduce participation for fear of foregoing all the benefits stemming from the social security scheme of the progressiveness of the entire tax system which cannot be compensated by the acceptance of unemployment subsidies. This effect may be reinforced by the fact that mature workers have comparatively higher earning profiles and face higher level of searching cost to obtain similar income in less taxed areas. If the tax system is progressive, the marginal tax rates that individuals face vary with earnings. As earnings vary over the lifetime of individuals, a progressive tax system implies that the marginal tax rates faced by workers also vary with age. Another reason comes from the tax/social security scheme targeted toward the old (see Gruber and Wise 1999).

---

[12] Another way to express the same concept may be that the Marginal Rate of Substitution between Income and Leisure is age dependent.

Models of age specific optimal income taxation are analysed by Lozachmeur (2006) who finds results partially similar to those reported above. As recently noted by Hummel (2025) many macro-empirical studies document a positive effect of income taxes on unemployment. However, when a distinction is made between marginal and average tax rates as in Lehmann et al. (2016), results show that average tax rates positively affect unemployment, whereas marginal tax rates have the opposite effect under the (plausible?) assumption of a given overall level of labour income taxation. Similarly, Manning (1993) finds that the marginal tax rate is negatively associated with unemployment. More indirect evidence comes from studies that analyse the effect of income taxes on wages. (see Hummel, 2025, for a discussion). The predictions from the model estimated in our paper are not entirely in consonance with the findings of those studies. The tax treatment analysed in this paper (basically, an asymmetric increase of a regional income surtax) is more similar to a marginal than an average rate increases and produces the reduction on the response labour variable.

8 Conclusions

While estimating the labour response to tax changes by DiD methods allows to identify a causal-effect relationship between labour and income taxes immune from the endogeneity problems affecting other empirical approaches, the use of mixed models enhance the DiD analysis by incorporating random effects for panel data, as well as data hierarchy and clustered designs. Our multilevel DiD versions of the income tax effects on labour response incorporate various mixed settings and allow to account for correlations within units or groups and to estimate multiple sources of variation simultaneously. In a nutshell, we obtain more reliable indications on the causal inference underlying the complex labour data generation processes.

We leverage on repeated measurements over time of labour response data and cofactors and employs mixed models to explicitly account for the non-independence of observations within the same group by including random effects, which helps to avoid biased standard errors and incorrect inference. Since we use hierarchical data (i.e., labour response of workers/taxpayers operating in Provinces which are in turn included in tax autonomous Regions, and above them in Macro-regions), our proposed DiD mixed models incorporate random effects for these clusters, acknowledging that observations within a cluster are more similar to each other than to observations in other clusters. Accordingly, we have been able to model individual variation/differences with respect to a baseline outcome (i.e., a pooled "single level" DiD model of the tax effect) and obtain more reliable estimates of the tax treatment effects. Our approach generates an improved design of the causal inference process by accounting for the hierarchical or repeated-measures nature of the labour response data. Then we trust that our mixed models provide a) more accurate standard errors and more reliable estimates of the treatment effect; b) more model flexibility which allows to capture the complex patterns of variation and relationships within the data; c) more efficiency. In short, our proposed method accommodates a nested hierarchical structure of the labour macro data into the framework of a DiD model and permits the estimation of the labour responses to exogenous income tax hikes taking a hierarchical nesting structure of the response variable into account. In addition, our procedure allows us to estimate the amount of variance of the response explained by each clustering level and permits to compare results from various nested and non-nested procedures. This generates more efficient estimations and permits to investigate issues that other approaches cannot explore using complete data pooling.

Using an Italian tax treatment experiment as a case study, we estimated the labour response to income tax changes accounting for both provincial, regional and macro-regional sources of variation in each local labour markets. We show that the empirical specification is robust to the inclusion on the random effects as the signs, the magnitudes and the significance of the estimated *CATET*s remain close to -1.30% (and strongly statistically significant), as well as the estimates for the control variables. A substantial difference with respect to the benchmark single-level DiD model is given by the high positive linear correlation estimated among units belonging to the same cluster and by the corresponding decomposition of the variance, which show a marked clustered nature which is completely ignored by the single-level model. A further robustness analysis was conducted assessing the policy response across various segments of the labour supply, showing that the negative response of both male and female working force to income tax hikes statistically significant and age dependent.

Declarations

- Conflict of interest/Competing interests: The authors have no competing interests to declare that are relevant to the content of this article.
- Funding: The authors did not receive support from any organization for the submitted work.
- Data availability and codes: All results presented in this paper can be reproduced using the R and Stata 18 software. Code and data can be provided upon request.
- Figures: All images included in the paper were created by the authors and do not require any publication permission.

References

1. **Akcigit, U., Baslandze, S., and S. Stantcheva (2016**). Taxation and the International Mobility of Inventors. American Economic Review, 106(10), 2930-2981. doi:10.1257/aer.20150237
2. **Appelbaum, E., and Katz, E. (1986)**. Measures of Risk Aversion and Comparative Statics of Industry Equilibrium. American Economic Review, 76(3), 524-529
3. **Apps P. & R. Rees (2009**). Public Economics and the Household, CUP, Cambridge UK
4. **Auerback A. & JM Siegel (2010)** Capital gains realizations of the rich and sophisticated. American economic review 90.2 pp. 276-282.
5. **Benzarti, Y. (2024)**, Tax Incidence Anomalies, NBER Working Paper No. w32819, https://ssrn.com/abstract=4922849
6. **Benzarti, Y., & J. Harju. 2021)**. Using payroll tax variation to unpack the black box of firm–level production. Journal of the European Economic Association 19 (5): pp. 2737–2764.
7. **Blau F. D. & L. M. Kahn (2007),** Changes in the Labor Supply Behavior of Married Women: 1980–2000. Journal of Labor Economics Volume 25, 3, pp. 393–438
8. **Blundell R., & T. MaCurdy (1999)**. Labor supply: A review of alternative approaches, in Ashenfelter, O. & D., Card (eds.), Handbook of Labor Economics, vol.3, chapter 27: pp. 1559–1695, North–Holland.
9. **Blundell R., A. Bozio, & G. Laroque (2011**), Extensive and Intensive Margins of Labour Supply: Working Hours in the US, UK and France. IZA DP No. 6051
10. **Blundell R., M. Costa Dias, C. Meghir , & J. Shaw (2016)**. Female Labor Supply, Human Capital, and Welfare Reform. Econometrica, Vol. 84, No. 5, pp. 1705–1753
11. **Blundell, R. & I. Walker (1982**). Modelling the joint determination of household labour supplies and commodity demands. The Economic Journal 92: pp. 351–64.
12. **Blundell, R. (1992). & I. Walker (1986).** A Life-Cycle Consistent Empirical Model of Family Labour Supply Using Cross-Section Data. Review of Economic Studies Volume 53, Issue 4, pp. 539–558, https://doi.org/10.2307/2297605
13. **Blundell, R. (1992).** Labor supply and taxation: A survey. Fiscal Studies 13: pp. 15–40
14. **Blundell, R., I. Walker, & F, Bourguignon (1988**). Labor supply incentives and the taxation of family income. Economic Policy, 3: pp. 133–61
15. **Bosco B., C.F. Bosco, & P. Maranzano (2025**), Labour responsiveness to income tax changes: empirical evidence from a DID analysis of an income tax treatment in Italy. Empirical Economics https://doi.org/10.1007/s00181-025-02748-7
16. **C. D. Romer & D. Romer (2010)** The Macroeconomic Effects of Tax Changes: Estimates Based on a New Measure of Fiscal Shocks. American Economic Review 00(3): pp. 763-801
17. **Chetty R., A. Looney, & K. Kroft (2009)** Salience and Taxation: Theory and Evidence. American Economic Review, 99(4), pp. 1145-1177
18. **Chetty, R. (2012).** Bounds on elasticities with optimization frictions: A synthesis of micro and macro evidence on labour supply. Econometrica 80 (3): pp. 969–1018.
19. **Colombino U. & D. Del Boca (1990).** The Effect of Taxes on Labor Supply in Italy. The Journal of Human Resources 25(3), pp. 390–414
20. **Demidenko, E.** (2013). Mixed models: theory and applications with R. John Wiley & Sons.
21. **Demidenko, E.** (2013). *Mixed models: theory and applications with R.* John Wiley & Sons.
22. **Donni, O. & J. A. Molina (2018**). Household Collective Models: Three Decades of Theoretical Contributions and Empirical Evidence. IZA Institute of Labor Economics, DP No. 11915
23. **Eissa N. & J. Liebman (1996)** Labor Supply Response to the Earned Income Tax Credit. The Quarterly Journal of Economics, Vol. 111, No. 2 pp. 605-637 https://doi.org/10.2307/2946689
24. **Esteller-Moré, A., and Solé-Ollé, A. (2002**). Tax Setting in a Federal System: The Case of Personal Income Taxation in Canada. International Tax and Public Finance, 9(3), 235-257. doi:10.1023/A:1016212110137
25. **Eurostat (2024).** NUTS - Nomenclature of territorial units for statistics. Retrieved December 17th, 2024 from https://ec.europa.eu/eurostat/web/nuts
26. **Eurostat (2025).** Employment (thousand persons) by NUTS 3 region. https://doi.org/https://doi.org/10.2908/NAMA_10R_3EMPERS
27. **Eurostat.** (2024). NUTS - Nomenclature of territorial units for statistics. Retrieved December 17th, 2024 from https://ec.europa.eu/eurostat/web/nuts

28. **Eurostat. (2024**). NUTS - Nomenclature of territorial units for statistics. Retrieved December 17th, 2024 from https://ec.europa.eu/eurostat/web/nuts
29. **Fay, R. E. I., & Herriot, R. A**. (1979). Estimates of Income for Small Places: An Application of James-Stein Procedures to Census Data. Journal of the American Statistical Association, 74(366a), 269-277. https://doi.org/10.1080/01621459.1979.10482505
30. **Finkelstein A. (2009**), E-ztax: Tax Salience and Tax Rates, The Quarterly Journal of Economics, Volume 124, 3, pp. 969–1010, https://doi.org/10.1162/qjec.2009.124.3.969
31. **Goolsbee A. (2000)** What Happens When You Tax the Rich? Evidence from Executive Compensation Journal of Political Economy, 108, 2 pp. ???
32. **Gudgeon, M., & S.Trenkle (2024),** The speed of earnings responses to taxation and the role of firm labour demand. Journal of Labour Economics 42 (3): https://doi.org/10.1086/723831
33. **Guner N., R. Kaygusuz & G. Ventura (2012).** Taxation and Household Labour Supply. The Review of Economic Studies, July 2012, Vol. 79, No. 3 (July 2012), pp. 1113–1149
34. **Hausman J. (1981),** Labor Supply, in H. Aaron & J. Pechman, eds., How taxes effect economic activity Washington, DC
35. **Hausman, J. & P. Ruud (1984).** Family labor supply with taxes. NBER Working Paper #1271
36. **Heckman J. T. (1993).** What has been learned about labour supply in the past twenty years? American Economic Review, Papers & Proceedings, 83, 2, pp. 116–121
37. **Heim D.** (2010) The Responsiveness of Self-Employment income to tax rate changes. Labour Economics, 17, pp. 940 –950
38. **Hox, J., Moerbeek, M., & Van de Schoot, R. (2017).** *Multilevel analysis: Techniques and applications*. Routledge.
39. **Hummel A. J**., (2025) Unemployment and tax design, Journal of Public Economics, Volume 246, 105359 https://doi.org/10.1016/j.jpubeco.2025.105359
40. **Keane M. (2011)** Labor Supply and Taxes: A Survey. Journal of Economic Literature 49, 4, pp. 961–1075
41. **Keane, M. P. (2022)**. Recent research on labor supply: Implications for tax and transfer policy. Labour Economics, vol. 77.
42. **Kihlstrom, R. E., & J- J. Laffont, (1979)** A General Equilibrium Entrepreneurial Theory of Firm Formation Based on Risk Aversion. Journal of Political Economy, 87(4), 719-748 doi:10.1086/260790
43. **Killingsworth M. R. & J.J. Heckman (1986**) Female Labor Supply: A Survey, Ashenfelter Orley C. and Richard Layard (Ed. s) Handbook of Labor Economics, vol.1, pp. 103-204, North-Holland
44. **Kleven H. K., Claus Kreiner, Kristian Larsen, & Jakob Søgaardt (2025)** Micro versus Macro Labor Supply Elasticities: The Role of Dynamic Returns to Effort. American Economic Review vol. 115, 9, pp. 2849–90
45. **Lozachmeur J-M. (2006**) Optimal Age Specific Income Taxation. Journal of Public Economic Theory, 8(4), pp. 697–711.
46. **Maranzano, P., & Pelagatti, M.** (2025). A Hodrick–Prescott filter with automatically selected breaks. Economic Modelling, 107132. https://doi.org/https://doi.org/10.1016/j.econmod.2025.107132
47. **McClelland R. & S. Mok (2012),** A Review of Recent Research on Labour Supply Elasticities. Congressional Budget Office, Working Papers Series 2012-12, Washington
48. **Meghir, C. & D. Phillips (2010),** Labour Supply and Taxes, in T Besley, R Blundell, M Gammie, & J Poterba (eds.), Dimensions of Tax Design: the Mirrlees Review. Oxford University Press, pp. 202–274.
49. **Mejía-Rodríguez, A. M., & Kyriakides, L.** (2022). What matters for student learning outcomes? A systematic review of studies exploring system-level factors of educational effectiveness. Review of Education, 10(3), e3374. https://doi.org/https://doi.org/10.1002/rev3.3374
50. **Meyer B. D. & D. T. Rosenbaum (2001**) Welfare, the Earned Income Tax Credit, and the Labor Supply of Single Mothers The Quarterly Journal of Economics, Volume 116, Issue 3, August 2001, Pages 1063–1114, https://doi.org/10.1162/00335530152466313
51. **Moretti, E., and Wilson, D. J. (2017**). The Effect of State Taxes on the Geographical Location of Top Earners: Evidence from Star Scientists. American economic review, 107(7), 1858-1903. doi:10.1257/aer.20150508
52. **Pencavel, J. (1986**). Labour supply of men, in Ashenfelter O. & R. Layard (eds.) Handbook of Labour Economics, Vol.1:5, Amsterdam: North Holland: pp. 3–102.

53. **Piketty T., E. Saez & S. Stantcheva (2014)** Optimal Taxation of Top Labor Incomes: A Tale of Three Elasticities American Economic Journal: Economic Policy vol. 6, no. 1, pp. 230–71)
54. **Pinheiro, J., & Bates, D. (2006)**. Mixed-effects models in S and S-PLUS. Springer science & business media.
55. **Pinheiro, J., & Bates, D.** (2006). *Mixed-effects models in S and S-PLUS*. Springer science & business media.
56. **Prescott, E.C. (2004)** Why Do Americans Work So Much More than Europeans? National Bureau of Economic Research. https://doi.org/10.3386/w10316
57. **Raudenbush, S. W., & A.S. Bryk (2002)** Hierarchical linear models: Applications and data analysis methods, Vol. 1. Thousand Oaks, CA: Sage.
58. **Rogerson R. & J. Wallenius (2009)**. Micro and macro elasticities in a life cycle model with taxes. Journal of Economic Theory, Vol. 144(6), pp. 2277-2292
59. **Roth J., P. H.C. Sant'Anna, A. Bilinski, & J. Poe** (2023), What's trending in difference-in-differences? A synthesis of the recent econometrics literature, Journal of Econometrics, 235, 2, pp. 2218–2244
60. **Saez E., M. Matsaganis , & P. Tsakloglou (2012).** Earnings determination and taxes: Evidence from a cohort-based payroll tax reform in Greece. The Quarterly Journal of Economics 127.1 pp. 493-533
61. **Saez E., B. Schoefer, & D. Seim (2019),** Payroll taxes, firm behavior, and rent sharing: evidence from a young workers' tax cut in Sweden. American Economic Review 109 (5): pp. 1717–1763.
62. **Sellström, E., & Bremberg, S**. (2006). Is there a "school effect" on pupil outcomes? A review of multilevel studies. Journal of Epidemiology and Community Health, 60(2), 149-155. https://doi.org/10.1136/jech.2005.036707
63. **Sickles, R. & A., Yazbeck (1998)**. On the dynamics of demand for leisure and the production of health. Journal of Business & Economic Statistics 16: pp. 187–197
64. **Siddiqui, O., Hedeker, D., Flay, B. R., & Hu, F. B**. (1996). Intraclass Correlation Estimates in a School-base Smoking Prevention Study: Outcome and Mediating Variables, by Sex and Ethnicity. American journal of epidemiology, 144(4), pp. 425-433. https://doi.org/10.1093/oxfordjournals.aje.a008945
65. **Snijders T. & R. Bosker (2012)** Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling, second edition. London: Sage Publishers
66. **Snijders, T. A., & Bosker, R. (2011)**. Multilevel analysis: An introduction to basic and advanced multilevel modelling. SAGE Publications Ltd
67. **Stantcheva S. (2016)** Labor Supply Responses to Taxation, https:// scholar.harvard.edu/ files/stantcheva/files/empirical_responses.pdf
68. **Taubinsky D. & A. Rees-Jones (2018)** Attention Variation and Welfare: Theory and Evidence from a Tax Salience Experiment. Review of Economic Studies 85, pp. 2462–2496.
69. **Tazhitdinova A. (2020)**. Do only tax incentives matter? Labour supply and demand responses to an unusually large and salient tax break. Journal of Public Economics 184: pp. 104–162
70. **Waseem M. (2020)** Does Cutting the Tax Rate To Zero Induce Behavior Different From Other Tax Cuts? Evidence from Pakistan. The Review of Economics and Statistics, 102(3), pp. 426–441
71. **Widmann R. (2023)**. Immigrant inventors and local income taxes: Evidence from Swiss municipalities. Journal of Public Economics, 219, 104822. doi:https: //doi.org/10.1016/ j.jpubeco.2023.104822
72. **Wooldridge J. M. (2010)**. Econometric analysis of cross section and panel data. MIT press, 2010.
73. **Wooldridge J. M. (2025).** Two-Way Fixed Effects, the Two-Way Mundlak Regression, and Difference-in-Differences Estimators, Empirical Econ (2025). https://doi.org/10.1007/s00181-025-02807-z
74. **Ziliak P. J. and Kniesner T. J. (2005).** The Effect of Income Taxation on Consumption and Labor Supply. Journal of Labor Economics, 23, 4, pp 769–795
75. **Zodrow G. (2025)** Advanced Introduction to Taxation, EE Elgar, Cheltenham, UK.

FONDAZIONE ENI ENRICO MATTEI WORKING PAPER SERIES

Our Working Papers are available on the Internet at the following address:
https://www.feem.it/pubblicazioni/feem-working-papers/

"NOTE DI LAVORO" PUBLISHED IN 2025

1. M. A. Marini, S. Nocito, Climate Activism Favors Pro-environmental Consumption
2. J. A. Fuinhas, A. Javed, D. Sciulli, E. Valentini, Skill-Biased Employment and the Stringency of Environmental Regulations in European Countries
3. A. Stringhi, S. Gil-Gallen, A. Albertazzi, The Enemy of My Enemy
4. A. Bastianin, X. Li, L. Shamsudin, Forecasting the Volatility of Energy Transition Metals
5. A. Bruni, Green Investment in the EU and the US: Markup Insights
6. M. Castellini, C. D'Alpaos, F. Fontini, M. Moretto, Optimal investment in an energy storage system
7. L. Mauro, F. Pigliaru, G. Carmeci, Government Size, Civic Capital and Economic Performance: An O-ring approach
8. A. Bastanin, C. F. Del Bo, L. Shamsudin, The geography of mining and its environmental impact in Europe
9. E. Cavallotti, I. Colantone, P. Stanig, F. Vona, Green Collars at the Voting Booth: Material Interest and Environmental Voting
10. L. Ciambezi, M. Guerini, M. Napoletano, A. Roventini, Accounting for the Multiple Sources of Inflation: an Agent-Based Model Investigation
11. L. Fontanelli, M. Guerini, R. Miniaci, A. Secchi, Predictive AI and productivity growth dynamics: evidence from French firms
12. A. Drigo, An Empirical Analysis of Environmental and Climate Inequalities across Italian census tracts
13. F.F. Frattini, F. Vona, F. Bontadini, I. Colantone, The Local Job Multipliers of Green Industrialization
14. A. Abatemarco, R. Dell'Anno, E. Lagomarsino, Measuring equity in environmental care: methodology and an application to air pollution
15. F. F. Frattini, Political Participation and Competition in Concurrent Elections: Evidence from Italy
16. M. Bonacina, M. Demir, A. Sileo, A. Zanoni, What Hinders Electric Vehicle Diffusion? Insights from a Neural Network Approach
17. C. A. Bollino, M. Galeotti, Native-borns and migrants do not contribute equally to domestic CO_2 emissions
18. D. Legrenzi, E. Ciola, D. Bazzana, Adaptation to climate-induced macrofinancial risks: top-down and bottom-up solutions
19. M. Bonacina, R. C. Tokong, Is Italy on Track? A Data-Driven Forecast for Road Transport Decarbonisation by 2030

Fondazione Eni Enrico Mattei
Corso Magenta 63, Milano – Italia

Tel. +39 02 403 36934

E-mail: letter@feem.it
www.feem.it

Fondazione Eni Enrico Mattei