Empirical Economics
https://doi.org/10.1007/s00181-025-02748-7

Labour responsiveness to income tax changes: empirical evidence from a DID analysis of an income tax treatment in Italy

Bruno Paolo Bosco[1] · Carlo Federico Bosco[2] · Paolo Maranzano[1,3]

Received: 19 July 2024 / Accepted: 31 March 2025

Abstract

This paper uses an Italian income tax treatment occurred in 2006/7 as a quasi-natural tax experiment and offers some fresh empirical evidence on how labour supply responds to exogenous income tax hikes. The specific features of the Italian tax measures, namely homogeneity and contemporaneity of the treatment across Italian regions, allow us to adopt the identification strategy of the two-way fixed effect (TWFE) panel data difference-in-differences (DID) model and to define the correct statistical framework of the study. We find that income tax hikes cause extensive negative adjustments of various response variables measuring the supply of labour services offered by treated taxpayers. DID estimated coefficients indicate that supply adjustments are significant, fast, and strong but not long-time lasting. Estimated time effects permit to evaluate the time impact of the tax measures and to conduct a dynamical policy evaluation of the tax hikes. We also show that treated families reduce in a similar manner their consumption with respect to families in the control groups and that analogous adjustment responses to income tax hikes characterise the growth of real per capita regional GDP. Altogether, estimates provide an overall view of the labour, consumption, and real growth effects of the income tax hikes. Results from DID estimations are further evaluated in comparison with the spatial–temporal patterns observed for every response variable in treated and untreated regions.

✉ Paolo Maranzano
paolo.maranzano@unimib.it

Bruno Paolo Bosco
bruno.bosco@unimib.it

Carlo Federico Bosco
fede.bosco28@gmail.com

1 Department of Economics, Management and Statistics (DEMS), University of Milano-Bicocca, Piazza dell'Ateneo Nuovo, 1, 20126 Milan, Italy

2 University of Pavia, Corso Strada Nuova, 65, 27100 Pavia, Italy

3 Fondazione Eni Enrico Mattei (FEEM), Corso Magenta, 63, 20124 Milan, Italy

Published online: 04 June 2025

Keywords Income taxation · Extensive labour supply change · TWFE panel data DID · Convergence tests · Taxation and regional growth

JEL Classification C10 · C18 · C21 · H2 · E2 · E32 · E62 · C23 · C26

1 Introduction, motivation, and content of the paper

Labour supply response to income taxes is a fundamental issue of public economics at both theoretical and empirical levels. Using Italy as a case study, we contribute to the existing empirical literature by presenting the results of a two-way fixed effect (TWFE) panel data difference-in-differences (DID) regression model of labour supply reaction to exogenous regional income tax hikes. Italy is not a federal country; yet she may provide a good case for studying the reaction of labour and other growth-related variables to regional income tax hikes. Since 1970 (the year of the creation of the Italian Ordinary Regions as an implementation *in slow motion* of the 1948 democratic Constitution of the Republic) and particularly in the last decades, Italian regions have accumulated large tax and expenditure powers. Regional authorities may use different local income tax rates and may adopt different tax expenditure measures. Hence, income tax heterogeneity exists countrywide, and exogenous regional income tax changes may affect the supply of labour services and regional growth paths. We study the regional differences in income taxation and their changes over time and investigate the impact of income tax hikes on labour supply and other growth-related variables.

To conduct the empirical study by means of a 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. They make the statistical framework of our analysis consistent with the identification requirements necessary for the efficient estimation of the Average Treatment Effect upon Treated (*ATET*) discussed by Roth et al. (2023, p. 2220) via DID under treatment homogeneity. The treatment affected a subset of Italian regions (treatment group) from 2007 onwards and consisted in an exogenously imposed increase in their regional income tax rates with respect to the rates existing in the other regions (control group) which remained unchanged. The new rates were imposed by the central government for reasons that we discuss later. Hence, the 2006/7 central government policy defines the framework of a quasi-natural tax experiment, fully described in Appendix 1, involving the taxpayers of five treated regions (not chosen at random) vs the taxpayers of the remaining fifteen untreated control regions. The goal of the estimation is to evaluate if exogenous differences in regional income taxation resulting from the central government income tax policy generated statistically significant responses of the labour supply of treated subjects. We also study the time adjustments of labour supply—and of other real variables such as the per capita regional income, its growth, and the family consumption—in the period following the income tax changes to evaluate how persistent the effect of income tax increase is over time. Results show that the negative adjustments of various response

variables measuring the supply of labour services of treated taxpayers are statistically significant, fast, and strong but not long-time lasting. Important for tax policy are the results concerning the response to tax hikes of *Self-Employed* people—possibly the taxpayers mostly affected by the 2006/7 tax treatment. Our DID estimations show that the share of *Self-Employed* workers is reduced by the increase in income taxation. Not surprisingly, we also find that treated families reduce in a similar manner their consumption. Analogous adjustment responses to tax hikes characterise the growth of regional per capita GDP. Altogether, our results show that exogenous income tax hikes affect the behaviour of workers and their labour supply. The result is robust to changes in the response variables and to the use of a larger data set. For instance, using other labour response variables (e.g. the annual *New VAT Certificates*, necessary for conducting self-employment activities) we obtain identical findings. As for the above-mentioned persistency of the *ATET* over time, we find that the tax effect is not long-time lasting. After treatment, adjustments of the labour supply to a long-run trend are observable, on average, 3 years after the introduction of the treatment. We also find that anticipatory effects are generally absent which is another valuable element for policy considerations.

The paper is structured as follows. After a brief Introduction presented in Sect. 1, the relevant literature is discussed in Sect. 2. The tax factors considered in this study are described in Sect. 3, together with the short narrative of the Italian 2006/7 tax experiment. Details on the latter are discussed in Appendix 1. Section 4 contains the discussion of the DID identification hypothesis in terms of a TWFE model as well as a statistical survey of the data with pre- and post-treatment plots of the candidate response variables. In Sect. 5, we present the empirical results, which are commented upon in comparison with the results of the existing international literature. In Sect. 6, we conduct a robustness analysis by using provincial, rather than regional, response variables measuring other indicators of labour activities and a larger data set. We employ an augmented version of the previous basic TWFE DID model which includes cofactors clustered at both provincial and regional levels to evaluate the sensitiveness of *ATET* to regional and provincial cofactors. Results support the findings of our previous estimated model that tax hikes have a negative but not time lasting impact on labour supply. Policy implications are also discussed. Section 7 presents some additional test procedures permitting to implement what we call a partial *falsification test* of the parallel trend assumption of the DID model. The test is based on convergence analysis for detecting the presence of parallel trends during the pre-treatment period. Section 8 briefly concludes. The spatial–temporal patterns observed for every response variable in treated and untreated regions are shown in Appendix 2.

2 Literature review

The empirical literature studying the relationships between income taxation and variables conducive to growth generally finds indications of heterogeneous aggregate effects of income tax changes. Zidar (2019, p. 1440). Some studies, for instance, provide evidence that lower-income households tend to have higher marginal propensities to consume (Dynan et al. 2004; Jappelli and Pistaferri 2010; Johnson et al. 2006;

McCarthy 1995; Parker 1999; Parker et al. 2013) and so they respond to tax policy by increasing consumption more than high-income groups. In analogous way, labour supply and other growth-related variables of different income groups may respond to income tax policy in an opposite way.

Focusing on labour supply, one should recall that the empirical literature studying the relationships between income taxation and labour supply is characterised by considerable controversy over how to estimate the responsiveness of labour supply to changes in wages and taxes. Keane (2011) and Stantcheva (2019) discuss the main econometric problems facing early studies and subsequent developments and the estimated values of the labour supply elasticities (Keane 2022). We refer the readers to their discussions. 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. Other authors (e.g. Romer and Romer 2010; Saez et al. 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 limited or can be offset by shifting in the timing or form/source of income acquisition (Auerbach and Siegel 2000; Goolsbee 2000). Ziliak and Kniesner (2005) find that consumption and worked hours are direct complements in utility and increase (both) with a compensated increase in net wage. 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 et al. 2014). They conclude that income tax can be responsible not only of a reduction in labour services supplied but also of a general contraction of the economic activity. As for self-employment (one of the response variables employed in this paper), Heim (2010), Appelbaum and Katz (1986), and Kihlstrom and 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. We are interested in estimating if a negative effect of income tax hikes (a reduction in self-employed people after the tax increase) is consistent with the data.

When territorial aspects 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 et al. (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 and 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 and Solé-Ollé, 2002).

Implementing the DID approach, Jakobsen et al. (2019) provide causal evidence of the impact of net-of-tax rate on wealth accumulation in Denmark, whereas Kleven et al. (2013) study the causal relationship between tax rates and a particular subset of the population of migrant workers (football-players called super stars). Baskaran (2021) presents results from a set of municipalities who, in North Rhine-Westphalia, increased their local property and business tax rates. He studies the revenue and base effects of local property and business tax hikes in a generalised DID design. His results suggest that the property tax hikes had a revenue elasticity of unity even in the long run. He concludes that tax changes had no adverse effects on property tax bases. For the business tax, he finds no significant effects on revenues and tax bases.

3 The Italian income tax treatment of 2006/7

The measures adopted by the Italian government at the end of 2006 are a good example of a possible exogenous income tax treatment. In December 2006, the Italian central government decided to impose an increase—automatic and compulsory—of two regional income tax rates, namely IRAP tax rates and regional Income Tax Surcharge to regions who had increased, with respect to 2005, the deficit of their regional health care systems above a certain percentage. The other regions were not affected by the tax measure.[1] The measure was adopted to force those regions to collect additional local resources and use them *exclusively* to reduce the deficits of their health system. The government finally decided that the percentage increase to be used as a threshold for the inclusion in the treatment group was the 7% of the previous year deficit. Five out of twenty Italian regions (Abruzzo, Campania, Lazio, Molise, and Sicily) fell into the treatment group. The content of the 2006 policy (described in details by Caruso and Dirindin (2019) is summarised below; a more complete description of the taxes involved in the treatment is shown in Appendix 1):

- IRAP tax rates. The ongoing standard IRAP tax rate existing in 2006 (4.5%) and employed by the five regions mentioned above was increased from the tax year 2007 onwards by almost a 1 percentage point. This implied that the actual tax rate for those regional taxpayers became equal to 5.25%. IRAP tax rates for taxpayers resident in other untreated regions were not modified.
- Regional Income Tax Surcharge. From the tax year 2007 onwards, the five treated regions had to increase by 0.5 percentage points the ongoing rate of 1.23% they applied to their residents as a surcharge rate on the national personal income tax (called IRPEF). Yet, since the five regions were under "controlled administration" by the central government, the surcharge rate was further increased by an additional

[1] If we call D ≡ Health Deficit = (Health expenditure – Health revenue)/Health expenditure, the criterion for eligibility was that $D_{it+1} \leq 1.07 \times D_{it}$. Those regions who did not respect that limit were assigned to the treated group. As discussed in the text, the number and identity of regions who would integrate the treatment group were ex ante uncertain. The idea was to punish regions who produced an "insufficient" fiscal effort in the year before the treatment and who had not reduced "appropriately" and autonomously the growth of their health deficit.

0.3 percentage point. Then, the net total increase was equal to 0.8 percentage points leading to an Income Tax Surcharge rate of 2.03%.

- A third element of the treatment is the following. Since the institution of the regional income surcharge, the Italian regions were permitted to allow deductions from the income surcharge payments instead of providing 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 the above taxes, the increase in the tax rates was compulsory, quantitatively relevant and finalised to collect some additional tax yield that the involved regions had to employ for the reduction in the prolonged budget unbalances of 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 (see below). Then, the condition of Irreversibility of Treatment applies to our case because once a region became treated in 2007 that region remained treated in the next periods.

The unexpected increase in IRAP Tax and Personal Income Surcharge mainly fell on Individual Business and Self-Employed Professionals (the main group of IRAP taxpayers within a region). Yet also the generality of treated regional taxpayers paying the national income tax was affected by the measure although less severely (they had to face only the increase in 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 realisation of a state of the word (the deficit of their regional health care systems) *determined when the measure was not in effect*, taxpayers, in the three-year period starting in 2007, were exposed to an exogenously determined tax increase of the main regional direct taxes.

Relevant for the applicability of our method is the circumstance that almost any Italian region could had been assigned to the treatment group. The health deficit was high and time increasing in the overwhelming majority of the regions, and it was completely unknown ex ante at what level the central government was going to set the above cut-off threshold determining a non-penalising annual tax increase. Some regions (e.g. Liguria) were expecting to be included in the treatment but remained out, whereas some other regions were expecting to remain out but were included (e.g. Lazio). The central government determined the above-mentioned 7% value in the last days of 2006 when the budget law for 2007 was finally approved by the Parliament and the 7% value arrived quite unexpected, according to national newspapers. As a result, at the end of 2006 each region had a positive probability of being eligible as a member of the treatment group based on her previous and irreversible behaviour. Therefore, in the Italian tax experiment 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. Actually, the Italian tax experiment is a variant of a "design-based" experiment with (almost) random participation based on past behaviour of treated and untreated units. Past behaviour qualified regional taxpayers

as future treated or untreated units and for this reason the inference drawn from our DID analysis can be seen as a "design-based inference". In this connection, Roth et al. (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).

Finally, it is worth noting, as a complication, that the design of the tax measure introduced with the 2006/7 treatment does not allow to exclude that some interactions in the response variables could be possible as well as special correlation in the treatment status among individuals.[2]

4 The DID identification

We follow Roth et al. (2023, p. 2220) and adopt the identification hypothesis of a TWFE panel data DID method to conduct the causal–effect study of the impact of a central government exogenous income tax policy. We employ response variables related to labour supply and other growth-related variables (such as family consumption) to evaluate the short- and long-run impact of income tax hikes taking both time and individual effects into account.[3] To illustrate the estimated model, we introduce the following notations:

$$y_{it} = \text{Observed response variable in region } i \text{ at time } t.$$

$$\boldsymbol{X}_{it} = \text{ matrix of cofactors/controls to be specified in each equation.}$$

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

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

The total sample period is $T = (1995\ldots\ 2021)$ with $t = (1995\ \ldots\ 2006)$ being the non-treatment period (dummy variable $D2 = 0$) and $t = (2007\ \ldots\ 2021)$ being the treatment period (dummy variable $D2 = 1$).

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,

[2] If spatial interaction among the taxpayers is present, the so-called stable unit treatment value assumption (STUVA) does not hold and identification of the *ATET* fails. Sun and Delgado (2024) present simulations of a DID that deals with spatial interactions and discuss the relevant literature.

[3] In the present case, migration motivated by tax minimising purposes had to be realised across regions (from treated to untreated regions) not municipalities, which may imply high moving costs. As a result, the estimation strategy may not consider the effects of the changes in the response variables induced by the inter-municipal migration motivated by municipal income tax differences as is it cleverly done by Rubolino and Giommoni (2023).

in the treated group. Written mathematically, the estimated effect of the treatment is theβ coefficient of panel data OLS equation:

$$y_{it} = \text{Constant} + \alpha \times D1 + \delta \times D2 + \beta \times (D1 \times D2) + \boldsymbol{X}_{it}\gamma + \varepsilon_{it} \qquad (1)$$

where ε_{it} is an idiosyncratic error term. The crucial parameter to estimate is β. If we call R a binary income tax treatment indicator, we have $\widehat{\beta} \equiv \text{CATET} = \mathbb{E}\left(y^1_{i,t} - y^0_{i,t} | X_{it}; R_{i,t}\right)$ or $\widehat{\beta} \equiv \text{ATET} = \mathbb{E}\left(y^1_{i,t} - y^0_{i,t} | R_{i,t}\right)$ if cofactors $\boldsymbol{X}$ are included or excluded. CATET (Conditional Average Effect upon Treated) or 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 otherwise specified, in all estimates the average treatment effect is estimated by adjusting for both cross-sectional and time effects. Notice that, given the design of the tax treatment, the response variables used in the estimations will all be independent on treatment conditional upon the $\boldsymbol{X}$ that will be used. Then in addition to the assumption of the parallel trend and the absence of anticipatory effects, our DID specification relies also on the *unconfoundedness* hypothesis (treatment participation is uncorrelated with the realisations of response variables). For Eq. (1) to identify the parameters of interest, 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 response 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) the targeted tax shocks are unrelated to any possible targeted level of the response variables, and (3) the outcomes from untreated regions provide a reasonable counterfactual. Since we control for region and year fixed effects in Eq. (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 budget unbalance of the health care systems mentioned in Sect. 2. Notice that the central government income tax policy introduced a contemporaneous treatment to a fixed group of regions (treatment effect homogeneity). Then, our TWFE DID model (1) does not make "forbidden comparisons" between already treated units (Roth et al. 2023, p. 2228).

Yet, the fact that assumptions (1), (2), and (3) permit to use TWFE regression specification for the estimation of the treatment parameter does not imply that other conditions for efficient parameter estimations of an OLS-based model are met to obtain asymptotically valid inference, particularly when the sample period is as large as ours. Stationarity and cross-sectional dependence are two issues. The former is not an absolute requirement for panel data analysis, but it is preferred in many cases. With a short time series of, for instance, up to 10 years, stationarity may not be critical. In that case, nonstationary of the data still allows using fixed effects (or first differences) models to control for unobserved heterogeneity and trends over time. Yet, with a sample like ours (27 years), it is necessary to conduct tests to check for the

presence of panel unit root because, in case of non-stationarity, the use of OLS-like procedures can produce invalid estimates. Granger and Newbold (1974) called such estimates "spurious regression" results: high R^2 values and high t-ratios yielding results with no real meaning. As it is shown in Table 1 (see Sect. 4.1), we reject the panel unit root hypothesis at usual significance levels and interpret the result as evidence that a statistically significant proportion of the units (both treated and untreated) are stationary.

4.1 Data description

We collected a set of regional ($N = 20$) potential response variables on a yearly basis from 1995 to 2021 (i.e. $T = 27$ years). Descriptive statistics of the response and control variables are given in Table 1 where stationarity and cross-sectional dependence tests are also reported. The relevance of these tests is discussed later. Data are divided between response variables:

1. Self-employment (measured as the share of self-employed people out of the total annual employment in the region). Data at the regional level are available from 2003 to 2020.
2. Number of New VAT Certificates: it measures the annual new flow of operators who in their activities require a VAT identification number. The tax authorities grant this certification (actually, a sort of specific tax number/code) to entities wanting to start some business activity. It approximates the net flow of annual business entities entering the respective professional markets, and it is computed as the Change of the annual stock of existing certificates = Existing Certificates at $t-1$ + New Solicited Certificates at t – (Expired Certificates at t – Cancelled Certificates at t). Data at the regional level are available from 2001 to 2021.
3. Family consumption in the logarithmic scale.
4. Regional per capita GDP, both measured in levels (€) and in Hodrick–Prescott (1997) filtered logarithmic scale.
5. Annual growth rate of regional per capita GDP.

and possible cofactors:

1. Internal Fixed Gross Investments for non-financial sectors: it consists of the acquisitions (net of sales) of fixed capital carried out by resident producers to which the increases in value of non-produced tangible goods are added;
2. Regional Value Added per worker (measure of labour productivity).

Descriptive statistics and tests are reported in Table 1.

Detailed comments and statistical tests conducted with the data of Table 1 are reported in Notes at the end of the table where we describe the results of panel data unit root and cross-sectional independence test. As stressed above, stationarity is an important issue in our panel data DID setting because we employ cross-sectional data observed for a lengthy period. Hence, we use cointegration tests to ensure that statistical relationships between trending variables are not spurious. In the Levin–Lin–Chu (LLC) unit root test (Levin et al. 2002), a common autoregressive parameter for all panels is assumed, so it does not allow for the possibility that some regional variables

Table 1 Descriptive statistics of response variables and cofactors (monetary variables are in 2015-chained values)

Variables	Obs	Mean	SD	Min	Max	Panel unit roots tests H_0: Panels contain unit roots		Panel Stationarity	Pesaran CD test for cross-sectional dependence			
20 Units /regions $T = 1995, \ldots, 2021$						Levin–Lin–Chu (LLC)	Harris–Tzavalis (HT)		Test statistics	*P* value	Average joint *T*	Mean abs (ρ)
Regional GDP (Euros)	540	8.35e + 09	8.24e + 09	4.25e + 08	3.86e + 10	$t^* = -2.429$ PV = 0.0076	$\rho = 0.6488$ PV = 0.0057	Yes, both tests	43.348	0.000	27.00	0.64
Per capita regional GDP (Euros)	540	27,699.96	7453.601	15,313	43,103	$t^* = -1.928$ PV = 0.0269	$\rho = 0.6669$ PV = 0.0210	Yes, both tests	53.49	0.000	27.00	0.75
Value Added per worker (Euros)	540	49,832.05	13,346.88	4253	76,483	$t^* = -2.314$ PV = 0.0103	$\rho = 0.6615$ PV = 0.0146	Yes, both tests	53.331	0.000	27.00	0.83
Gross Fixed Capital Formation GFCF (Euros)	540	1.62e + 10	1.56e + 10	8.45e + 08	8.07e + 10	$t^* = -2.594$ PV = 0.0047	$\rho = 0.7903$ PV = 0.9093	Yes LLC No HT	48.668	0.000	27.00	0.68
Self-employment (% of total employment) (2003–2020)	355	13.55634	2.32783	6.9	20.4	$t^* = -5.298$ PV = 0.000	$\rho = 0.995$ PV = 1	Yes LLC No HT	19.578	0.000	17.51	0.37

Table 1 (continued)

Variables	Obs	Mean	SD	Min	Max	Panel unit roots tests H_0: Panels contain unit roots		Panel Stationarity	Pesaran CD test for cross-sectional dependence			
20 Units /regions $T = 1995, \ldots, 2021$						Levin–Lin–Chu (LLC)	Harris–Tzavalis (HT)		Test statistics	P value	Average joint T	Mean abs (ρ)
New VAT Certificates (Units) (2001 – 2021)	421	25,853.42	22,523.57	879	90,191	$t^* = -4.879$ PV = 0.0000	$\rho = 0.6287$ PV = 0.1595	Yes LLC No HT	24.256	0.000	20.20	0.53
Family Consumption (Euros)	540	50,667.72	43,999.08	2593	204,864	$t^* = -3.0287$ PV = 0.0012	$\rho = 0.6649$ PV = 0.0185	Yes, both tests	43.009	0.000	27.00	0.60

Sources: ISTAT and Italian Treasury (Department of Finance). Mean, standard deviation (SD), min, and max are computed using the overall sample. All monetary values are expressed as values linked to the 2015 reference year: the linking provides a measure of the economic aggregate of interest in terms of volume, i.e. net of the underlying price dynamics. Note that the sum of the chained values of the components of an aggregate is not equal to the chained value of the aggregate itself. Under the null hypothesis of cross-sectional independence, the Pesaran CD test statistics is normally distributed, i.e. CD ~ $N(0,1)$, and p values close to zero indicate that data are correlated across panel groups. In the Levin–Lin–Chu (LLC) unit root test, a common autoregressive parameter for all panels is assumed, so it does not allow for the possibility that some regional variables contain unit roots while others do not. To mitigate the effects of possible cross-sectional correlation among time series, we removed the cross-sectional means from the series (demeaning). Also, we included for all the tests a linear trend term. The LLC test requires that the ratio of the number of panels to time periods tends to zero asymptotically. Hence, it is not well suited with relatively few periods. Hence, we also present the Harris–Tzavalis (HT) unit root test, which assumes that the number of panels tends to infinity while the number of periods is fixed. For both tests, the null hypothesis is H0: panels contain unit roots

contain unit roots while others do not. To mitigate the effects of possible cross-sectional correlation among the time series, we removed the cross-sectional means from the series (demeaning). Also, we included for all the tests a linear trend term. The LLC test requires that the ratio of the number of panels to time periods tend to zero asymptotically. Hence, it is not well suited with relatively few periods but fits our case well. We also present the Harris–Tzavalis (HT) unit root test (Harris and Tzavalis 1999), which assumes that the number of panels tends to infinity while the number of periods is fixed. For both tests, the null hypothesis is that the panels contain unit roots. For almost all variables, we are able to reject the null hypothesis.

The presence of parallel trend in pre-treatment periods for response variables is of great importance for DID identification since violation of parallel trend assumption will lead to biased estimation of the causal effect. For the response variables of interest, pre- and post-treatment period (i.e. end of 2006) trends are shown in Fig. 1. It is frequently assumed that the smaller the time period tested (for example a 2-year period), the more likely the assumption is to hold, but with a period of 27 years the possibility that treated and control groups have different outcome trends (which may generate time convergence) cannot be excluded.

Visual inspection of Fig. 1 shows that parallel variations prior to 2007 are a reasonable hypothesis for the response variables, and it is suggestive that the identification proposition of the DID is valid as far as parallel trend is concerned. Plots reproduced in Appendix 2 (Fig. 9) show the time series of index numbers with baseline 2006 = 100 for all available response variables. Abruzzo, Campania, Lazio, Molise, and Sicily are the five treated regions, while the non-treated regions are averaged across year (solid black lines). Both groups of plots (Fig. 1 and those reproduced in Appendix 2) indicate that the average change in the response variables for treated and untreated regions would had been the same in the absence of income tax hikes.

Yet, parallel trends cannot evaluated by confining ourselves to the *realm of optics*. In this paper, it will be subjected to statistical test based on an augmented version of Eq. (1) which is discussed at the beginning of the next section. The presence of anticipation effects will be also tested. All the tests are reported in Table 2 alongside with the ATET estimated coefficients.

Note that the reported descriptive statistics in Table 1 are computed on the overall sample, and they do provide information about the overall scale and the variability of the macroeconomic indicators under consideration. However, they do not provide any insight about the territorial heterogeneity typically affecting the Italian economic measurements (see, for example, De Philippis et al. 2022; Federici et al. 2023; Lo Cascio et al. 2019). In Appendix 2 (Figs. 5, 6 and 7), we show the evolution of the considered regional indicators for selected years along with the 1995–2020 average. For almost all indicators included in the sample, the charts show a pronounced regional heterogeneity, highlighting in many cases either the North–South gap (e.g. GDP per capita, Value Added, and employment), which persists over the decades, or clear patterns linked to urbanisation degree and population density (e.g. regional household consumption). Although territorial heterogeneity is not explicitly modelled within our DID framework by including distances between regions (Elhorst 2014) or geostatistical factors (Wikle, Zammit-Mangion, & Cressie, 2019), the cross-sectional (regional) fixed effects included in Eq. (1) can still mitigate the effect of spatial dependence by

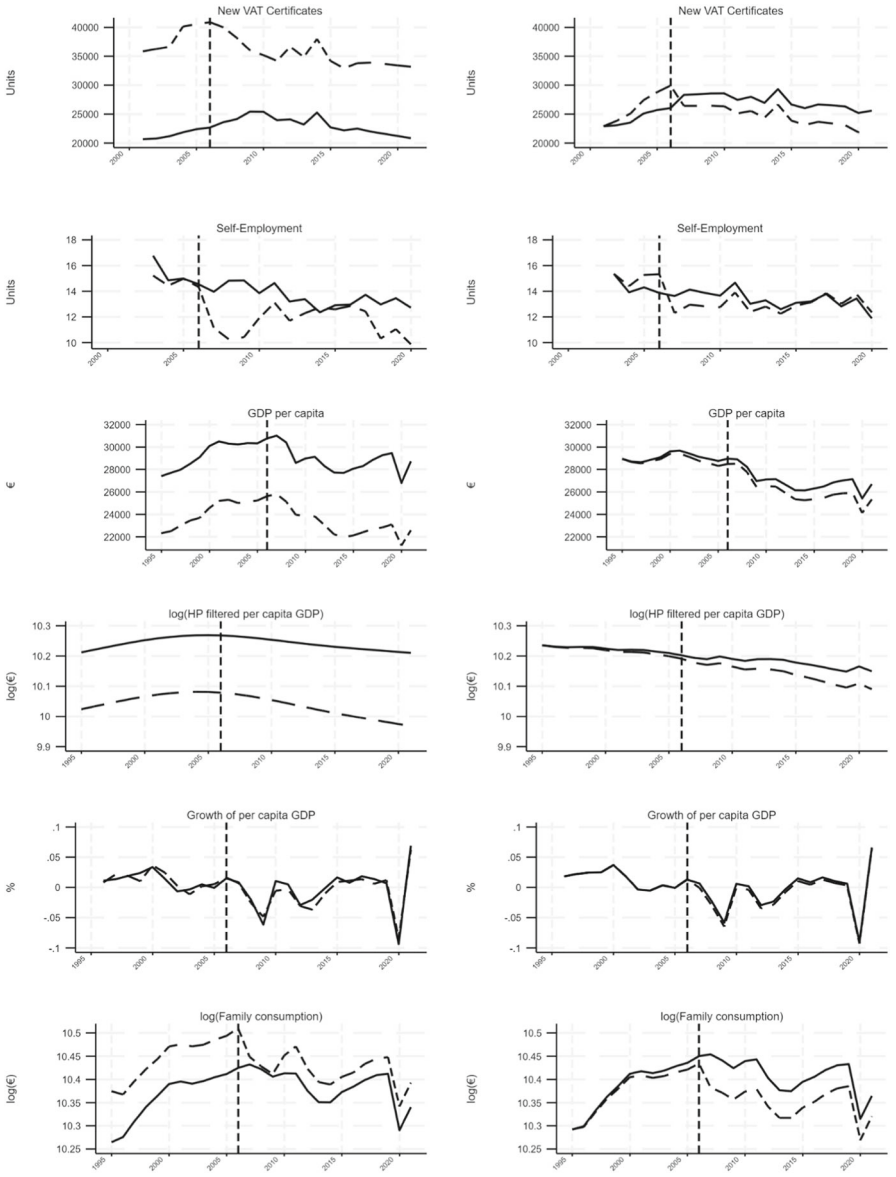

Fig. 1 Time behaviour of response variables before and after the 2006/7 tax treatment (Dashed lines = Treated regions; Continuous lines = Control regions). *Notes*. The plots permit a graphical diagnostic for parallel trend and show the time evolution of the response variables used in this study before and after the income tax treatment. 2006 is the last year before the treatment (vertical line). For each variable, the left plot consists of two lines showing the mean of the outcome over time for the treatment and the control groups. The right plot augments the DID model to include interactions of time with an indicator of treatment and plots the predicted values of this augmented model for the treatment and control groups. See also the maps of Appendix 2, which provide a detailed graphical analysis of the time–space evolution of the considered variables

Table 2 TWFE DID estimations of the Italian income tax treatment

Response variables	Self-employment	New VAT Certificates	Growth of per capita GDP	HP filtered per capita GDP (log values)	Per capita GDP	Family consumption (log values)
ATET or CATET Treated vs untreated regions	**− 1.165****	**− 4592.525****	**− 0.005***	**− 0.032*****	**− 629.19*****	**− 0.05****
Controls						
log(GFCF)	− 0.8	− 453.97	0.01	.088***	2241.21***	/
Log(Value Added per worker)	− 11.72***	− 13,460.289	0.03	0.24**	9867.1645**	/
Constant	160.11***	178,491.2	− 0.54*	5.67***	− 128,581.29***	10.29***
Time fixed effects	√	√	√	√	√	√
Region fixed effects	√	√	√	√	√	√
N	355	420	520	540	540	540
Ptrend test: H0 Linear parallel trend	$F(1, 19) = 2.30$ Prob > $F = 0.1455$ **H_0: cannot be rejected**	$F(1, 19) = 2.02$ Prob > $F = 0.1715$ **H_0: cannot be rejected**	$F(1, 19) = 0.02$ Prob > $F = 0.9030$ **H_0: cannot be rejected**	$F(1, 19) = 1.18$ Prob > $F = 0.2903$ **H_0cannot be rejected**	$F(1, 19) = 0.34$ Prob > $F = 0.5674$ **H_0cannot be rejected**	$F(1, 19) = 0.62$ Prob > $F = 0.4398$ **H0 cannot be rejected**
Granger type test: H0 no anticipation effects	$F(3, 19) = 1.50$ Prob > $F = 0.247$ **H_0cannot be rejected**	$F(5, 19) = 1.34$ Prob > $F = 0.29$ **H_0cannot be rejected**	$F(10, 19) = 9.99$ Prob > $F = 0.0000$ **Reject H_0**	$F(11, 19) = 11.29$ Prob > $F = 0.0000$ **Reject H_0**	$F(11, 19) = 7.29$ Prob > $F = 0.0001$ **Reject H_0**	$F(11, 19) = 7.12$ Prob > $F = 0.0001$ **Reject H_0**

Results from DID estimation of Eq. (1) with robust SE. H_0 in the Ptrend test is that pre-treatment linear trends are parallel. Granger type test is for anticipatory effects. H_0 is that there are no effects in anticipation of treatment. In the fourth and fifth columns, the dependent variable is the logged per capita income of the regional residents. We used the Hodrick–Prescott high-pass filter to separate the time series into trend and cyclical components. The trend component may contain a deterministic or a stochastic trend. The smoothing parameter determines the periods of the stochastic cycles that drive the stationary cyclical component. With the exception of the first, columns report CATET estimates obtained using Controls in addition to Time. ATET estimates (not reported) are very similar. More comments and test discussion are in Notes of Table 3. Data source: see previous table

capturing region-specific latent characteristics (Elhorst 2010, 2024) and preserving reliable estimates. In this sense, learning from examples like Merfeld (2019) and Muralidharan et al. (2023), a potential avenue for future investigations of Italian tax treatment could take into account the spatial structure between areas looking for spatial spill-overs and bordering effects of the policy.

As anticipated above, due to the panel macroeconomic nature of the dataset we also test for panel non-stationarity. Adding the cross-sectional dimension to the time series dimension offers an advantage in testing for nonstationary and cointegration since cross section increases the data set used in those tests, thus improving their power. However, the cross-sectional dimension also brings some new problems to our TWFE DID model, namely the existence of cross-sectional dependency that can bias usual panel data unit root test results in small samples. The Pesaran (2007) cross-sectional dependence test checks for this cross-sectional dependency issue. Our findings show, as expected, the presence of high correlation among regional variables across panels/regions. The importance of a priori check of the existence of unit roots in the panel data comes from the already known effect that the presence of unit roots in time series may cause a misinterpretation of estimated results. The unit root hypothesis is tested

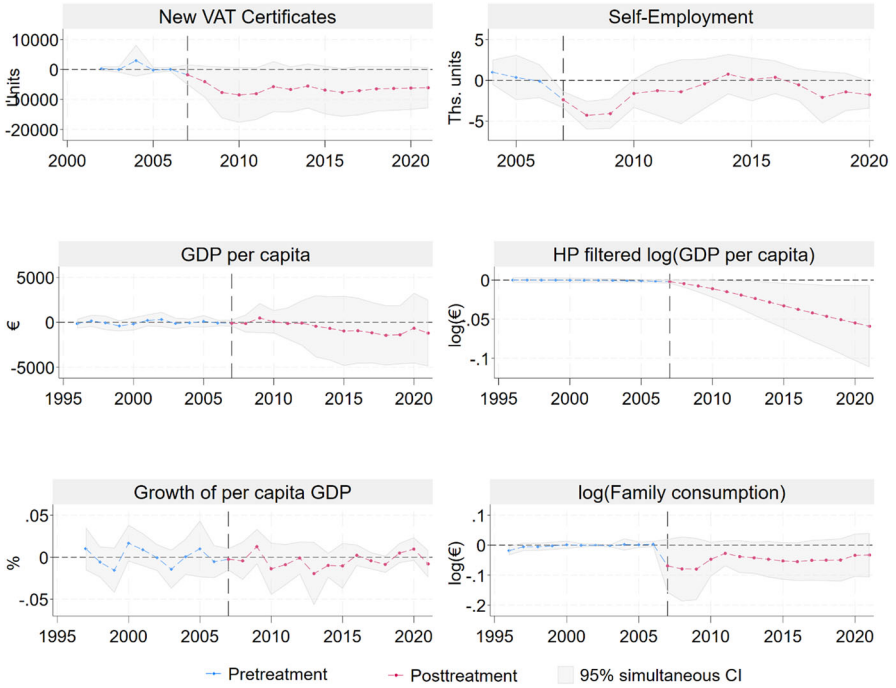

Fig. 2 Estimated ATET/CATET over time in treated regions for the response variables of interest. *Notes*. In the plots, the vertical line corresponds to the first year of treatment, and therefore, 2006 is the last year without treatment. The red lines indicate the estimated difference of each annual ATET and the 2006 value. As a result, the 2006 value of the parameter is zero. Details on how the above parameters are estimated are given in Notes in Table 3 where we report the results of the robustness analysis

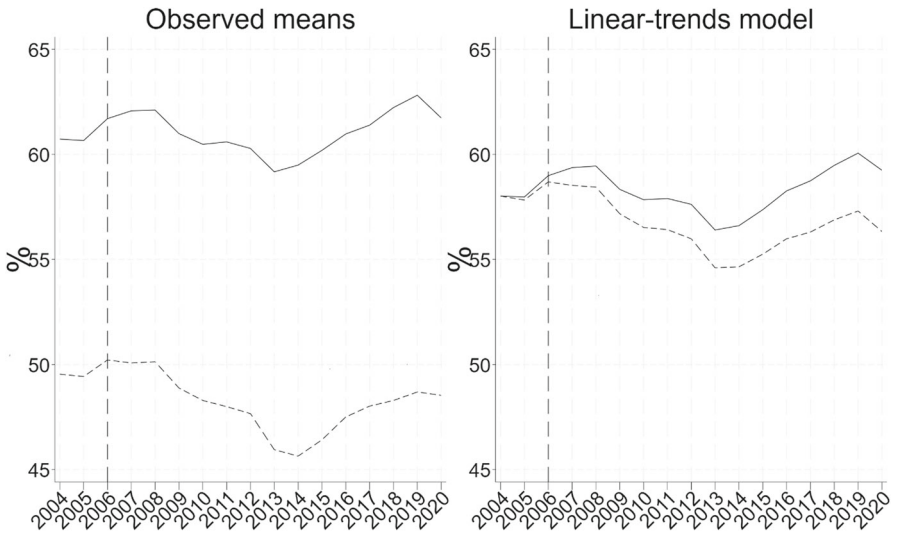

Fig. 3 Provincial rate of employment in Italy (age 15–64; male and female) by treatment group (Dashed lines = Treated provinces; Continuous lines = Control provinces). Graphical diagnostic for parallel trends and time duration of exposure to income tax treatment. Notes. The plot shows that before treatment period there was a parallel trend, and the post-treatment period shows an increase in the mean values of untreated units and a decrease in the mean values of the treated units

through the Levin et al. (2002) and Harris and Tzavalis (1999) unit root tests (i.e. for both, the null hypothesis is that the panels contain unit roots) accounting for linear trend and cross-correlation adjustments. For all the variables in the dataset, we reject the null that the panel contains unit roots and conclude that the series are trend stationary. Since in the OLS estimation of (1), panels are considered to be homogeneous, the opposite would be a drawback. The absence of unit roots is important for our DID analysis. In the most intuitive sense, stationarity does not mean that the series do not change over time, but simply that the way the series change does not itself change over time. If an exogenous event occurs (such as an exogenous tax change), changes in the response variables can be, loosely speaking, attributed to the influence of the event and not to some long-run property of the series.

5 Empirical results

TWFE panel DID results are presented in Table 2. The ultimate causal parameters of interest (ATET) are reported in the first row. Tests of parallel trend and anticipation effects are reported in the last rows and commented in Notes in the table. We refer the reader to Table 7 in Appendix for a detailed summary of the estimated coefficients, including those associated with temporal and cross-sectional fixed effects. The plots of the ATET evolution over the period following the tax shock are shown in Fig. 2 after the sections containing the discussion of the estimations of the tax effect on each response variable.

Before commenting the specific results for each response variable, we discuss the test for the presence of the parallel trend and anticipation effects. The last two rows of Table 2 report the estimated F-tests for parallel trend hypothesis and the test for absence of anticipation effects. The former is based on the results of an augmented linear DID model based on our Eq. (1) with additional terms that capture the differences in slopes between treated and controls. As for anticipation effects, Table 2 also reports the results of a Granger type test obtained from a linear DID model based once again on Eq. (1) augment in this case by interacting the dummy variable that marks treated observations with dummy variables for time periods prior to the treatment to capture any potential anticipatory treatment effects.

The augmented DID model employed for parallel trend test is the following

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

where S_0 is a time dummy indicating pre-treatment years, S_1 indicates treatment years, and w indicates either treated regions ($i = 1$) or untreated regions ($i = 0$). 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. In Table 1, we report the Wald test of the null hypothesis H_0: $\psi_1 \neq 0$ versus $\psi_1 = 0$. 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 of the slope parameter ψ_1 by the estimate of its standard error. Under the above null, this ratio follows a standard normal distribution.

As for the anticipation effects we adopt a procedure similar to a Granger model, but with a different parameterisation that includes lags in addition to leads. Recall that 2006 is the last year before treatment. Then, let m be the number of periods prior to 2006, $q \geq 0$ be the number of periods after 2006, and b is the baseline period. The fitted model is (using MOD1 of Table 2 as reference; no cofactors included)

$$y_{ij,t} = \gamma_i + \gamma_t + [\alpha(D1) + \delta(D2) + \beta(D1 \times D2)] + \sum_{\substack{k=m \\ k \neq 2006}}^{q} (B_{it}^k w_i)\tau_k + r_{ij,t}$$

where w_i has the same meaning as above (see fn. 4) and $B_{it}^k = (t_{it} = I_i + k)$ because we use all available leads and lags. Table 2 reports tests that coefficient referring to years prior 2007 are significantly different than zero.

The above Wald test for parallel trend and the Granger type test for anticipation effects are adopted in all the estimation procedures presented in the paper as it is recalled in Notes below each table.

Yet, one should recall that the Granger type test is less robust than the above parallel trend F-test since it requires more degrees of freedom because it estimates a higher number of parameters. Hence, the parallel trends test has higher statistical power. The results allow us to conclude (at any significance level) that the data generation process is consistent with the hypothesis of parallel trend but not always with the absence of anticipation effects hypothesis. The latter result may be informative for the

policy design of future tax treatments. In the following sections, the estimation results of each response variable are commented with greater emphasis on those related to labour supply.

5.1 Self-employment and New VAT Certificates

Individuals respond to the risk adjusted relative earnings opportunities, and therefore, labour income tax hikes should affect negatively their extensive and intensive labour supply. As for the duration of this effect, (Saez 2002, p. 1043) rightly emphasises that the shorter the time period upon which the income tax is assessed, the more relevant is the extensive effect of the tax changes. Table 2 shows the responsiveness to tax hikes of two extensive measures of the labour supply. The first is the annual level of *Self-Employment* (measured as the share of self-employed people out of the total annual employment in the region). The second (*New VAT Registration Certificates*) indicates the number of existing official Value-Added "certificates" (a sort of individual identification tax number for self-employed) that self-employed workers must possess to exercise legally their activities and cash the Value Added tax from their clients. It is a response variable measuring the (*official*) participation decision of a vast set of professionals offering labour services.

5.1.1 Self-employment

Table 2 shows that *Self-Employment* is affected by income tax hikes. Self-Employed people are probably the taxpayers mostly affected by the 2006/7 tax treatment since they will pay both taxes whose rates have increased in the five treated regions (see Appendix 1 for details). Our DID estimations show that the share of *Self-Employed* workers is reduced by the increase in income taxation. These results are in disagreement with previous literature (e.g. Heim 2010) and with Robson and Wren (1999) who state that higher tax rates lead individuals to take up self-employment to take advantage of the greater opportunity for tax avoidance and evasion that self-employment offer relative to paid employment (i.e. pure wage earners). According to this latter view, one should expect that income tax hikes increase rather than decrease *Self-Employment*. Our results indicate that this effect should be excluded in the Italian case. Yet, the estimated effect may indicate that reactions to an income tax increase may include tax evasion. Some self-employed taxpayers may simply enter the shadow economy as a result of the marginal income tax increase. On the other hand, the reduction in the share of self-employment may depend upon a reduction in low-middle income self-employed individuals who decided not to stay in business after the increase in the marginal income tax rate and simply quit working. Then the effects of income taxation on individual choices may differ from the effects of wealth taxation. Our results should be contrasted with those presented by Agrawal et al. (2025, p. 2) who analysed the effect of a wealth tax indicate that self-employed wealthy individuals are not significantly more likely to move to less taxed areas, because of tax reasons, than non-self-employed wealthy individuals.

Anticipatory effects on *Self-Employment* are excluded by the Granger-like test reported in Table 2. We interpret the test as the absence of anticipated avoidance behaviour on the part of self-employed taxpayers in the treated regions (a sign that the tax hikes arrived as a surprise) or as a clue that changing occupation in advance of the tax hikes was seen as either impossible or not profitable.

As for the duration of the tax effect over time, Fig. 2 (upper right plot) shows that the impact of the tax treatment on *Self-Employment* is statistically significant in the first year of the treatment and during the three years following the treatment, but the coefficients are no longer statistically significant afterwards. Then, the tax effect loses strength and significance as we move away from the initial year of the treatment. Still, results indicate that income tax hikes affect the level of self-employment in the treated areas although the effect last for a short time.

The significant negative effect of tax hikes on *Self-Employment* is also observable from the annual trends reported in the geographical maps reproduced in Appendix 2. We use them as additional elements useful to interpret the TWFE results. Figure 8 shows the absolute variation between 2006 and 2009 of *Self-Employment*, and Fig. 9 shows the time series of index numbers with base year 2006 (the last year prior to treatment). As for *Self-Employment*, in the five treated regions maps show a substantial reduction in this response variable (between 20 and 30 percentage points in the four years following the treatment according to the reported index numbers). On the contrary, in the untreated regions we observe on average an increase in recorded values.

5.1.2 New VAT (Registration) Certificates

The second labour response variable is the *New VAT Registration Certificates*. Operators who want to start legally a (new) business activity must obtain a VAT Certificate/identification number (similar to the tax code identification of general taxpayers) from the tax authorities. Hence, each year we observe the stock of existing certificates as well as the annual variation in the stock. In the estimations, we use this flow variable which measures the net increase/decrease in economic operators (plumbers, lawyers, mechanics, shop owners, engineers, architects, etc.) requiring and obtaining a formal VAT Certificate from the tax authorities, net of the number of certificates returned back to tax authorities or extinguished during each year. Clearly, employees are excluded from this set of taxpayers. In a DID model, a *CATET/ATET* parameter will correspond to the estimate of the effect of the income tax hikes on the (official) net flow of new labour supply provided in each region by professional business operators who need the above certification. Then, observing the post-treatment behaviour of this response variable and the difference between treated and untreated taxpayers one may obtain information on the effect of income tax increase on the participation decisions of that specific subset of workers/taxpayers. As it is stressed in Sect. 2, those who apply for a VAT Certificate/permission are subjected to the payment of *both* IRAP and IRPEF. Therefore, they represent the labourer category most affected by the tax treatment of 2007. Estimating the responsiveness of this variable to the tax treatment provides information about the (extensive) marginal reaction to taxation in the form of a participation decision, that is, to apply or not to apply for the certificate as a result of the tax change. As stressed by Saez (2002, p. 1048) for similar alternatives open

to taxpayers, this view is obviously a crude simplification of reality but captures the extensive margin labour supply decision of a class of workers. If one assumes away income effects and tax evasion, the decision to participate depends only on the difference between the (after tax) income with participation and unemployment transfers without participation.

The estimation of the effect of the tax treatment indicates that an *average* decrease in the new applications for the VAT Certificates roughly corresponds to 4600 units (about the 6% of the average 2006 value), which is the estimate of the average annual reduction in new (official) operators in the treated regions with respect to the untreated ones. This means that the tax treatment has discouraged new professionals to enter the relevant markets in treated regions or has led some of the already active operators to leave the (official) markets.

Once again, the test reported in the last rows of Table 2 suggests that we cannot reject the null that trends are parallel and that anticipatory effects are absent even for this variable. Recall that operators (old and new) requiring a VAT Certificate/Registration to carry a business are the main IRAP taxpayers, and the result seems to indicate that the tax base of IRAP is rigid to anticipatory adjustments.

5.2 Per capita income growth

We extend the DID analysis of the income tax hikes to per capita GDP and Consumption. The results are reported in the third, fourth, fifth, and sixth columns of Table 2. Three income response variables are considered: *annual growth of per capita regional GDP*, and *per capita regional GDP* in both levels and logarithmic scale (the latter is Hodrick–Prescott filtered), which proxies the income tax base for the generality of taxpayers. For the log per capita GDP, we use the Hodrick–Prescott high-pass filter to separate the time series into trend and cyclical components. When income is the response variables, DID may require cofactors. Our cofactors are Gross Fixed Capital Formation (GFCF, proxy of investments) and Value Added per regional Worker (proxy of the average labour productivity). Estimates show that personal income growth is strongly negatively affected by income tax increase and this result accords with the previous extensive labour supply contraction shown in the first three columns of Table 2. As for *per capita Regional GDP* in levels, results show that it is negatively affected by the treatment, but anticipated effects cannot be excluded, which implies that some taxpayers in treated regions may have anticipated the tax treatment and have adopted tax-reducing strategies before 2007. Analogous negative effect is obtained for the response represented by the log of the annual rate of per capita GDP growth.

We may contrast the above results with Zidar (2019, p. 1423). A clear similarity emerges between our response results and his estimates, although, in our case, estimates indicate a more pronounced decline of the response variables in treated regions. The exogenous increase in both taxes (the regional Income Surcharge and the increase in the business quasi-income tax IRAP), taken in isolation, apparently reduces the post-treatment regional GDP levels and growth over time. Clearly this drastic conclusion should be moderated by recalling that the effects on income and growth, via the

local multiplier effects, should be evaluated by weighting the effects of both taxes, on the one hand, and public expenditures on the other (Ramey 2011a, b).

The graphical trend (Figs. 8 and 9 in Appendix 2) of the per capita GDP shows a marked generalised reduction following the introduction of the tax treatment. While the average value in the untreated regions shows a pro-cyclical pattern with quick alternations of growth and stagnation, Sicily, Campania, and Molise (three of the treated regions) follow a smooth process of GDP loss that is not reabsorbed in the following years. The general conclusion is that the GDP response variables negatively react to the income tax treatment.

5.3 Family consumption

Results obtained for labour supply and GDP growth might indicate that tax hikes should negatively affect also other labour and growth-related exogenous variable. Table 2 also reports results for the response variable represented by family consumption. We observe that estimated parameters (last column of Table 2) show a negative value of the overall consumption elasticity in the treated region to be approximately 0.5%. The 2007 tax treatment clearly made the (national and regional) income taxation more progressive in treated regions. Thus, by changing the disposable income distribution in treated regions, a more redistributive taxation may have changed the relative aggregate demand for goods and services with possible decrease in consumption of taxpayers of the higher taxable brackets. This in turn may have led to changes in the demand for workers in various sectors, particularly those offering more income elastic goods. With non-homothetic consumption preferences, when we allow consumption to adjust (i.e. when income effects are at work), any change in progressivity differentially affects the labour supply of skilled and unskilled workers because the compositions of agents' consumption baskets change with their disposable income. If this is the case, the marginal utility of income may vary across income groups, distinctively affecting the incentives to work of each group (Oni 2023, p. 4). Consumption results seem consistent with those obtained for labour supply and may indicate that the permanent income hypothesis is not consistent with the data[4] and that the tax shock has a strong effect during the first years following the treatment and then loses intensity and significance (at the 95%).

Similar deductions can be made from the observed trends in household consumption exposed in Appendix 2 (Figures 8 and 9), which show that the reduction in consumption in the treated regions was stronger than in the control group between 2008 and 2011, being partially reabsorbed by some regions (yet, Lazio, Sicily and Campania did not catch up). This may suggest that durable consumption might be affected by the treatment at the beginning of the post-treatment period and then, since current consumption may be measured with error, its point estimates may be less reliable than other response variables such as self-employment. However, the results are somewhat

[4] This is so because consumption should not respond to *predictable* fluctuations in income (see for example, Souleles 1999, p. 947, who finds significant evidence of excess sensitivity in the response of households' consumption to their income tax refunds). Nguyen et al. (2021, p. 456 and Fig. 4) also find a significant sensitivity of consumption to general level of taxation.

not clear-cut for no distinction is made for durable goods consumption vs current consumption and, above all, for different income groups.

5.4 A policy interpretation and the evaluation of the tax effects over time

Altogether, our results show that income tax hikes reduce labour response over time. As in Wen and Gordon (2014) where in a *probit* model the tax convexity variable and the net-of-tax income difference between self- and paid employment have the predicted (negative and statistically significant) effect on self-employment, in our DID estimation *Self-Employment* negatively reacts to income tax increase. Then tax hikes are perceived as disincentive to entry the legal self-employment professional markets and produce negative effects at the extensive margin by reducing the number of individual professionals or the number of new companies of entrepreneurs who plan to exploit some business opportunities (Aghion and Howitt 1990). Our DID results accord with those obtained by some recent studies in which income/profit taxation is a determinant of the incorporation decision (see Barrios et al. 2009; Djankov et al. 2010) and represents an instrument to support business conditions (Da Rin et al. 2011).

Graphical trends shown in Appendix 2 support our TWFE DID estimates. The index numbers, in particular, clearly show that the pattern of the treated group and of the control group is widely divergent and that the regions affected by the income tax hikes have experienced large reductions (around 20% in 2009) in the number of VAT Certificates compared to the rest of Italy.

As for the effect of the tax treatment over time, the upper plots of Fig. 2 show that the effect of the tax treatment on the supply of professional services and the log of family consumption are strong and statistically significant (slightly at the 95% all over the sample). Yet, the difference between the estimated ATET in the years following the introduction of the treatment and the ATET estimated for 2007 (red lines) is negative. This means that after the introduction of the treatment there is a reduction in the "intensity" of the effects of the tax hikes. Labour supply and family consumption somehow readjust towards long-run trend values. On the contrary, the plots of the GDP per capita and its growth show that the above difference between estimated ATETs is not significantly different than zero (at 95%). In the latter case, the tax effects are more persistent.

Finally, some observations about the policy implications of our results are in order. DID analysis of income tax changes shows that tax reforms may not generate complete instantaneous modifications of existing labour supply. Modifications may be distributed over several years following the tax changes and the time adjustments of the labour response may be affected by the modifications of the general economic conditions occurring during adjustment years and/or by the effects of other taxes such as consumption and property taxes. Labour market reforms intervening during the time adjustments of the labour response following to tax changes can increase or reduce labour rigidity or can hinder firms' ability to adjust their workforce for new technologies or innovative processes. This should be carefully considered by policy makers. In the case of the 2006/7 Italian treatment, for instance, the tax reform was followed by

the income contraction triggered by 2008 world financial crisis that negatively affected the response variables studied in this paper.

6 A robustness analysis

We replicate the TWFE panel DID estimation using *Rate of Employment* (male and female, age from 15 to 64 years, Eurostat NUTS-3 data) recorded at the provincial level in treated and untreated regions from 2004 to 2021. Altogether, we use data of 103 provinces distributed among 20 regions: 5 treated and 15 untreated. The model is identified by the same basic TWFE DID assumptions of Eq. (1), plus the requirement that each province belonged to the same region at the beginning and at the end of the sample period 2004–2020 (no province abandoned the original region). We also include cofactors relative to both provincial and regional levels. Equation (1) rewrites as

$$y_{ijt} = \beta_0 \underbrace{+[\alpha(D1) + \delta(D2) + \beta(D1 \times D2)]}_{\text{Tax Treatment Component}} + \sum_{i=1}^{P} \beta_i X_{ijt} + \sum_{j=1}^{R} \beta_j W_{jt} + u_{ijt} \quad (2)$$

where $t = (2004, \ldots, 2021)$ and X_{ijt} is the year value of each of the possible P cofactors existing at the *provincial* level i (when i belongs to region j) and W_{jt} is the year value of each possible R cofactor existing at the *regional* level j and associated with the provinces nested in that region (for each $i \in j$). Finally, u_{ijt} is an error term normally distributed. Clearly, the Tax Treatment Component is the same as in Eq. (1).

All estimated versions reported below include both unit fixed effects and time fixed effects in ordinary least squares estimation. Table 3 reposts results of various version of the model.

MOD1 is a no-cofactors (pure *ATET*) model; MOD2 includes provincial cofactors only; MOD3 has cofactors of both provincial and regional level; and MOD4 includes a variable computed as the annual average value of the employment rate and is included to incorporate the national mean trend of the response variable.

Results reported in Table 3 indicate that the exogenous increase in the income taxation affects negatively the rate of employment in any specification of the estimated model. As expected, the presence of cofactors affects the magnitude of the estimated A*TET* (or *CATET*) coefficients. In MOD1 for example, a one point percentage increase in the income tax rate leads to a 1.7% 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 Notes in 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. Graphical illustration of the existence of parallel trends is reproduced in Fig. 3. The plot supports the results of the tests reported in the last rows of Table 2 and those commented about in Notes in the table and in the footnote. Yet,

Table 3 DID ATET TWFE estimations of various versions of Eq. (2)

Response Variable RATE OF EMPLOYMENT	MOD1	MOD2	MOD3	MOD4
INCOME TAX TREATMENT				
(*CATET* Treated vs. Untreated)	**− 1.74*****	**− 1.69*****	**− 1.35*****	**− 1.35*****
Constant	58.01***	56.73***	50.45***	− 77.12
Provincial GDP PPS		.00009**	.00008**	.00008**
Regional GDP PPS			5.167e−10***	5.167e−10***
Regional Growth Rate			0.53	0.53
National Mean Employment Rate				2.20**
Time Fixed Effects	√	√	√	√
Individual Fixed Effects	√	√	√	√
Observations	1751	1751	1648	1648
Parallel Trends Test $F(1, 102)$	0.42 Prob > F = 0.516	0.47 Prob > F = 0.4942	0.44 Prob > F = 0.5077	0.44 Prob > F = 0.5077
Anticipation Effects Granger causality test $F(2, 102)$	0.32 Prob > F = 0.725	0.34 Prob > F = 0.7115	0.43 Prob > F = 0.514	0.43 Prob > F = 0.514

The null hypothesis of both Parallel Trends test and No Anticipatory Effects test cannot be rejected at any level of significance. The interpretation of ATET/CATET is the following: 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 $(103 \times 17)^{-1} \sum_{i=1}^{103} \sum_{t=2004}^{2020} y_{it}$. 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 2021). 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, since Provincial GDP PPS is measured in millions of euros, the reported coefficient .00008323 means that if that Provincial GDP increased by 1.000.000 euros, the rate of employment would increase by an 8% of its value. Legend: $*p < 0.1$; $**p < 0.05$; $***p < 0.01$

both procedures lead us to conclude that the pre-treatment parallel trend assumption is consistent with the data.

7 Additional statistical tests of parallel trends

The presence of different trends between treated and untreated units may jeopardise the parallel trend assumption of the DID. Given the macroeconomic nature of our regional data set, the presence of converging trends of some response variables cannot be excluded a priori. When the time period tested is short (for example a 2-year period), the assumption is the more likely to hold. Yet with a period of 27 years the possibility that treated and control groups have different outcome trends (which may generate time convergence or time divergence among them) cannot be excluded. Then, in addition to the tests reported in the last rows of Table 2 and the information provided by the visual inspection of the plots of Fig. 1, in this section we propose *a falsification test* that *at least* excludes that the response variables of treated and untreated units follow some *converging trend*. Time convergence, i.e. converging trends of the response variables for treated and control groups (employment, labour supply, consumption, etc.), would then be at odds with the hypothesis of parallel trends. This is so because the change (or "path") in response variables over time that regions in the treated group would have experienced if they had not been included in the treatment may not be the same as the path of response variables that regions in the control group actually experienced. In the case of a quasi-natural experiment where assignment is possibly not entirely random, the treatment assignment must be mean independent of factors that affect the trend in the response variables. Roth et al. (2023, p. 2222) clarify that if the bias for selecting into the treatment is the same each year, then parallel trends allow for selection bias. Excluding converging trends is not synonymous of validating parallel trends, however. Trends can be non-parallel because they might *diverge*. Also discovering that the trends diverge would be a sign that the treatment assignment may not be mean-independent of factors that affect the trend in the response variables and that there is absence of parallel trends. Hence, discovering that response variables do no converge is not necessarily a sign of parallel trends.

Bearing in mind the above limitation, we propose an application of the *log*(t) convergence *test* of Phillips and Sul (2007, 2009), which proposed the "*log*(*t*)" regression test approach to test for the convergence hypothesis based on a nonlinear time-varying factor model. The proposed approach has the following merits: first, it accommodates heterogeneous agent behaviour and evolution in that behaviour. Second, the proposed test does not impose any particular assumptions concerning trend stationarity or stochastic non-stationarity, and consequently, it is robust to the stationarity property of the series. In other words, it could be used whatever the unit root results reported in Table 1. Yet, to make the test more adherent to the time span of the Italian tax treatment we apply separately the test to the entire sample period (1995–2021), to the pre- and post-treatment periods, and to the two subsets of data (treated and control units) separately.

Results of *log*(*t*) tests are reported in Tables 4, 5, and 6.

Results show that during the pre-treatment period the hypothesis of convergence is not consistent with the data and this finding supports the DID parallel trend hypothesis. In one case (i.e. *per capita GDP growth*), the null hypothesis of convergence cannot be rejected at usual significance levels for the entire time period and for the post-treatment period (this contradicts the Ptrend test reported in Table 2 for that variable), but no convergence is found for the pre-treatment period.

Obtaining the rejection of the null of convergence of the response variables when pooling treated and control observations together just indicates that the series do not converge. However, it does not exclude that trends could be not parallel because they *diverge*. Yet, a different result (not rejecting the null hypothesis of convergence) would be a clear contradiction of the parallel trend hypothesis and that is why we propose

Table 4 Phillips–Sul log(*t*) tests of convergence

Response Variables (treated and untreated regions)	1995–2021	1995–2006	2007–2021
Self-Employment (2003–2020)	− 0.8210 (− 3.0030) Rejected	− 2.5351 (− 105.3844) Rejected	− 1.3775 (− 5.4437) Rejected
New VAT Certificates (2001–2021)	− 0.7717 (− 29.1923) Rejected	− 1.4358 (− 12.5197) Rejected	− 0.9361 (− 55.5262) Rejected
Growth of Regional GDP p–c	2.7964 (1.85) **H_0Not Rejected**	− 4.6937 (− 2.0528) Rejected	3.3804 (2.1556) **H_0Not rejected**
HP filtered logarithm of per capita regional GDP	− 0.8112 (− 253.81) Rejected	− 0.9415 (− 40.8417) Rejected	− 0.9571 (− 53.9166) Rejected
Regional per capita GDP	− 0.8233 (− 23.3134) Rejected	− 0.9610 (− 26.2895) Rejected	− 0.8512 (− 32.0493) Rejected
Logarithm of Family Consumption	− 0.8788 (− 82.5425) Rejected	− 1.0069 (− 53.7084) Rejected	− 0.9222 (− 37.5706) Rejected

The table shows the results of a log(*t*) test for overall convergence of the response variables used in the DID analysis (see Table 1) proposed by Phillips and Sul (2007, 2009) in order to investigate the presence of convergence (the null hypothesis is that there is convergence) by adopting the Andrews estimator of long-run variance (fixed or adaptive bandwidth of the kernel) based on Quadratic Spectral kernel. Following Phillips and Sul (2007, 2009), we use the recommended value of $r = 30\%$ corresponding to the share of values to be discarded when computing the statistics. If the *t* test of log(*t*) test is smaller than − 1.65, then the null hypothesis of convergence of the variable is rejected at 5% significance. T-statistics are in parenthesis. Notice that, due to the low data availability of New VAT Certificates and Self-Employment, even if computation of the test for both the pre-treatment and post-treatment periods is feasible, asymptotic properties could be unsatisfied. Results show that during the pre-treatment period the hypothesis of convergence is not consistent with the data and this finding does not contradict the DID parallel trend hypothesis. In one case (growth of regional per capita GDP), post-treatment period and the entire period show convergence

Table 5 Phillips–Sul log(t) tests of convergence for untreated regions only

Response Variables (untreated regions only)	1995–2021	1995–2006	2007–2021
Self-Employment (2003–2020)	− 1.2245 (− 6.6226) Rejected	− 2.6070 (− 18.1581) Rejected	− 1.0843 (− 4.5208) Rejected
New VAT Certificates (2001–2021)	− 0.7446 (− 30.1970) Rejected	− 1.3935 (− 14.0926) Rejected	− 0.9188 (− 48.3598) Rejected
Growth of Regional GDP p–c	2.2234 (1.4464) **H_0not rejected**	− 4.6663 (− 2.2425) Rejected	2.8482 (1.2841) **H_0not rejected**
HP filtered logarithm of per capita regional GDP	− 0.7526 (− 153.3184) Rejected	− 0.9037 (− 37.4473) Rejected	− 0.9260 (− 50.2768) Rejected
Regional per capita GDP	− 0.7894 (− 18.6610) Rejected	− 0.9299 (− 23.6503) Rejected	− 0.8154 (− 34.8615) Rejected
Logarithm of Family Consumption	− 0.8783 (− 78.1308) Rejected	− 1.0022 (− 54.3171) Rejected	− 0.9188 (− 40.0264) Rejected

The table shows the result of log(t) test of convergence for the untreated regions only. T-statistics are in parenthesis. Notice that, due to the low data availability of New VAT Certificates and Self-Employment, even if computation of the test for both the pre-treatment and post-treatment periods is feasible, asymptotic properties could be unsatisfied. Results show consistency with those related to the full sample, that is, they do not contradict the DID parallel trend hypothesis except for the case of the growth of regional pe capita GDP, for which convergence is found in post-treatment and entire periods

this test (which gives possible indeterminate outcomes) in addition to those reported in Table 2.

8 Conclusions

We have analysed the effects of the Italian binary income tax treatment that was adopted at the end of 2006 and remained on afterwards. It affected taxpayers in a treated group of regions with no variation in treatment timing and assignment. TWFE panel DID results provide robust evidence that labour supply and other regional growth-related variables are responsive to income tax hikes. Robustness analysis conducted by using a multi-level TWFE DID panel, which includes provincial and regional level cofactors, supports the results.

Under the conditions discussed in this paper, the DID analysis of income tax changes has two main advantages with respect to other approaches. The tax treatment represents a truly exogenous independent variable, and the estimates of the ATET over time (i.e. during treatment) show that tax reforms may not generate complete instantaneous modifications of the existing labour supply. We find that income tax hikes cause

Table 6 Phillips–Sul log(t) tests of convergence for treated regions only

Response Variables (treated regions only)	1995–2021	1995–2006	2007–2021
Self-Employment (2003–2020)	− 2.1374 (− 2.3277) Rejected	− 2.0493 (− 1.5835) **H_0 not rejected**	− 0.8367 (− 0.7607) **H_0 not rejected**
New VAT Certificates (2001–2021)	− 0.8244 (− 62.8409) Rejected	− 1.4616 (− 11.7273) Rejected	− 0.9373 (− 35.9240) Rejected
Growth of Regional GDP p–c	3.2923 (2.1344) **H_0 not rejected**	− 3.8758 (− 1.6369) **H_0 not rejected**	1.4841 (0.8986) **H_0 not rejected**
HP filtered logarithm of per capita regional GDP	− 0.7645 (− 44.4023) Rejected	− 1.0604 (− 44.7164) Rejected	− 0.8736 (− 35.9767) Rejected
Regional per capita GDP	− 0.6442 (− 13.5016) Rejected	− 1.2055 (− 24.0175) Rejected	− 0.7757 (− 19.9970) Rejected
Logarithm of Family Consumption	− 0.8118 (− 82.0494) Rejected	− 1.0345 (− 51.2356) Rejected	− 0.9468 (− 33.0851) Rejected

The table shows the result of log(t) test of convergence for the treated regions only. T-statistics are in parenthesis. Notice that, due to the low data availability of New VAT Certificates and Self-Employment, even if computation of the test for both the pre-treatment and post-treatment periods is feasible, asymptotic properties could be unsatisfied. The table shows that the convergence hypothesis among untreated regions seems consistent with the data only for Self-Employment and the growth rate of per capita GDP during the pre-treatment period

extensive negative adjustments of various response variables measuring the supply of labour services offered by treated taxpayers. DID estimated coefficients indicate that supply adjustments are significant, fast, and strong but not long-time lasting. Labour supply modifications are distributed over several years following the tax changes, and the time adjustments of the response may be affected by the modifications of the general economic conditions occurring during years following the tax treatment and/or by the effects of other taxes such as consumption and property taxes. Our findings are consistent with the evidence presented on labour force participation by previous literature which indicates that the supply of professional self-employed person is adversely affected by the tax increase. Yet the DID analysis shows that adjustments of the labour supply of that segment of working population is observable for three years after the treatment. Similar results, although less pronounced, are obtained for other labour response variables. Results obtained for the response variable *New VAT (Registration) Certificates* indicate that an exogenous modification of regional taxation has robust extensive negative repercussion on the supply of professional services. Tax hikes may possibly delate or cancels the decisions of potential operators to enter the official markets or tempt some of them to enter the underground economy. Similar findings are

obtained for family consumption. Estimates show a negative reaction of consumption to income tax hikes with post-tax temporal adjustments similar to those characterising labour supply.

We found that the presence of anticipatory effects is not consistent with the data, and this may be relevant in terms of evaluation of tax policy measures. The absence of anticipatory effects somehow implies that the specific tax bases of the two income taxes analysed in this paper may be more rigid to anticipatory adjustments on the part of taxpayers than the tax base given by the general personal income tax.

Finally, our DID estimates of the ATET of the income tax increase on regional per capita GDP and its growth indicate that they react negatively to a heavier income taxation. Results of Table 7 in Appendix 2, *where time effects are reported,* indicate that the ATET value of the tax treatment on the per capita GDP is 630 euros which is interpretable as the average reduction—across time and treated regions—of the net-of-tax personal disposable income. Per capita personal income growth is negatively affected by an income tax increase, too. Both results are consistent with the estimated contraction of labour supply and family consumption. Yet, estimates show that the negative effects of the tax increase on per capita GDP growth loses significance at the end of the sample period. This finding shows that per capita GDP is more sensitive to long-term than to short-term fluctuations, including income tax hike. We obtained the opposite result for labour supply. In the long run, the dynamic of regional per capita income and its growth is affected by a larger set of factors, such as those entering the local multiplier of the public expenditure financed with the additional tax yield. Then, long-run effects of income tax increase should be evaluated by weighting the effects of taxes, on the one hand, and public expenditures and cyclical components on the other.

Appendix 1: The 2006/7 regional income tax policy in Italy[5]

The appendix provides a brief description of the two Italian income taxes involved in the tax treatment as well as a description of the treatment.

IRAP

From 1997, Italian corporate entities and individual natural persons generating income in the form of business income (*Reddito d'impresa*) as well as non-resident companies, but only on Italian source income, are subject to a regional production tax, IRAP. IRAP taxpayers include those carrying out business activities and self-employment, operating both individually and in partnership, private non-commercial entities as well as public administrations and other public bodies.

IRAP taxes the production activity, the exchange of goods and the supply of services. In this paper, IRAP is viewed as an Income Tax *in disguise* because it is charged on the value of net production resulting from the business pursued within the region. For commercial business and professionals, state administrations, regions, provinces,

[5] This appendix relies on Bosco et al. (2024).

municipalities, and generally for public administrations whose main object is not commercial business, the basis of assessment is the value of remunerations. The 90% of its regional yield is destined to the financing of the health system in each regional (in addition to central transfers and other funds). As a rule, IRAP tax rates are fixed by the regional governments within an interval defined by the national law that leaves some degree of autonomy to the regional authorities. Hence, although the tax structure is completely defied by the parliamentary law of 1997, IRAP tax rates may vary across regions and time.

The regional income tax surcharge

The regional income tax surcharge was introduced in the Italian tax system in December 1997 (D.Lgs. number 446/1997), and it is applied from the 1998 tax year. The regional surcharge is a regional tax, whose rate is added on top of those rates taxing the income of natural persons at the national level (IRPEF). Then, it falls on the taxable income as it is defined for IRPEF purposes net of any applicable tax deductions and credits for incomes produced abroad. Tax rates vary from region to region and may increase in relation to the taxpayer's income bracket. The ordinary regional income tax rate ranges from 1.23 to 3.33%. In the event that the region decides not to adopt a single rate but a plurality of differentiated rates, the latter must be structured exclusively in relation to the same income brackets established for the IRPEF tax, and then, tax rates must be increasing according to each bracket. The regions can arrange tax deductions in favour of the family by increasing those already existing for national IRPEF purposes and can also adopt other economic support measures in favour of IRPEF subjects. During the period of the present study, the regional surcharge rates varied greatly between regions. In 2000, the surtax rate was determined for the entire Italian territory by the national law at the level of 0.9 but each regions had the power to increase that rate up to 1.4. In that year, only Lombardy, Marche, and Veneto applied the additional taxation on a true progressive scale. At the end of our period (2021), other regions, such as Abruzzo, Basilicata, Calabria ($t = 1.73$), Campania ($t = 2.03$), Sardinia ($t = 1.23$), Sicily ($t = 1.23$), Valle d'Aosta ($t = 1.23$), and Veneto ($t = 1.23$), maintained (or switched to) the proportional (constant tax rate) surcharge tax (see the dedicated webpage from the Italian Ministry of Economy and Finance or MEF: https://www1.finanze.gov.it/finanze2/dipartimentopolitichefiscali/fiscalitalocale/addregirpef/sceltaregione.htm, accessed on August 27, 2024).

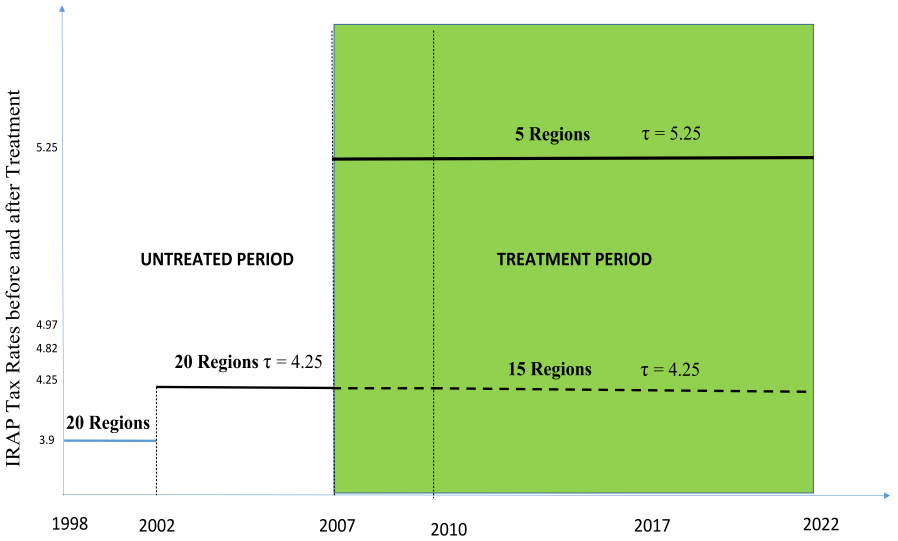

Fig. 4 Regional tax rates used in the estimations. IRAP tax rates before and after the Italian Tax Treatment at the end of 2006. The plot describes how the tax rates of IRAP have been modified by the Italian Budget Law of 2006 (Law 27/12/2006, n. 296, Disposizioni per la formazione del bilancio annuale e pluriennale dello Stato; legge finanziaria 2007). Recall that, in addition to the IRAP measures, the ordinary standard rate of 0.9% that regions applied as a Surcharge on the personal income tax base of their residents was increased in treated Regions by a 0.5% plus a further 0.3%. As a result, in treated regions the ordinary Surcharge rate reached the value of 2.03%. Treated regions are Abruzzo, Campania, Lazio, Molise, and Sicily

Figure 4 gives a synthetic view of the Italian Income Tax Treatment introduced at the end of 2006. From 2007, taxpayers in treated regions faced a uniform increase in the IRAP and Regional Surcharge rates. Uniformity means that the statutory rates were increased by the same percentage for all treated taxpayers without differences related the value of the tax base for bottomand top-income groups.

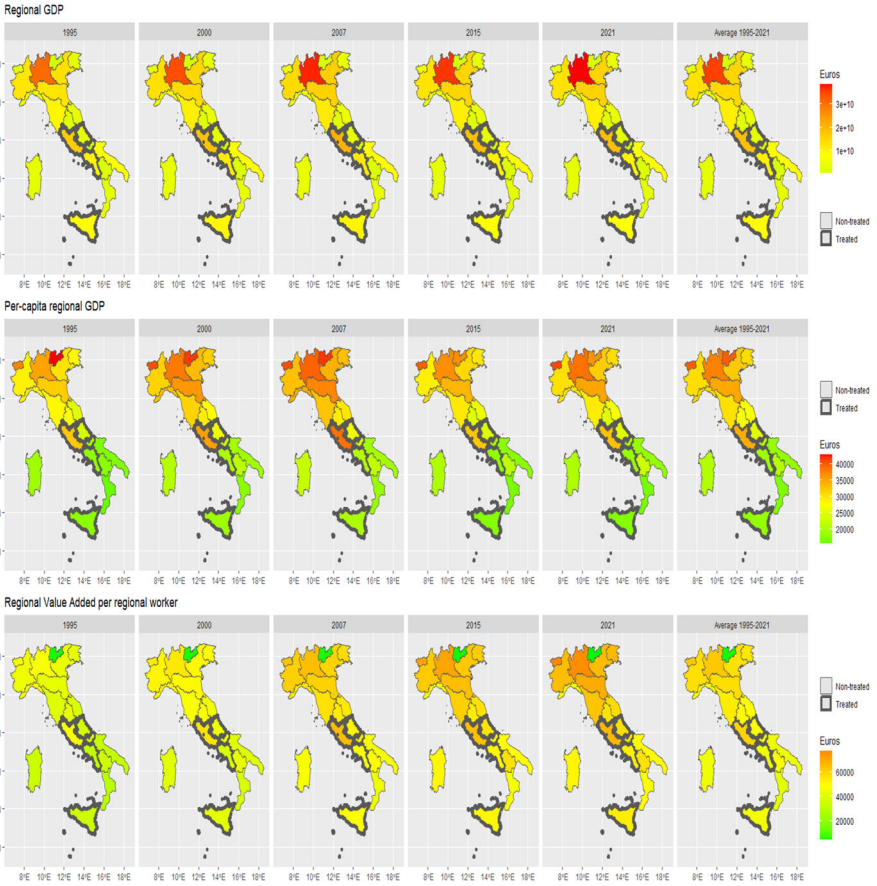

Fig. 5 Observed values of overall GDP, per capita GDP, and Value Added per worker across the Italian regions from 1995 to 2021. The last column depicts the regional average value 1995–2021. Regions with a marked border are those treated (Lazio, Campania, Abruzzo, Molise, and Sicily)

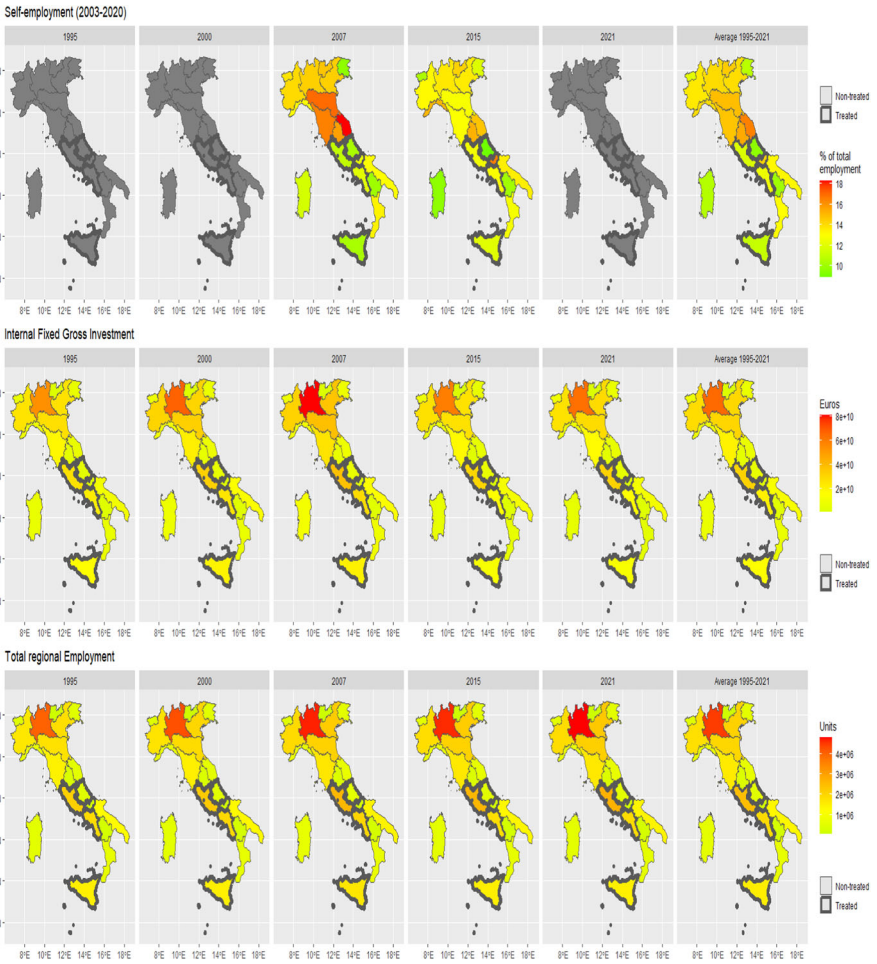

Fig. 6 Observed values of self-employment, internal fixed gross investment, and total employment across the Italian regions from 1995 to 2021. The last column depicts the regional average value 1995–2021. Regions with a marked border are those treated (Lazio, Campania, Abruzzo, Molise, and Sicily)

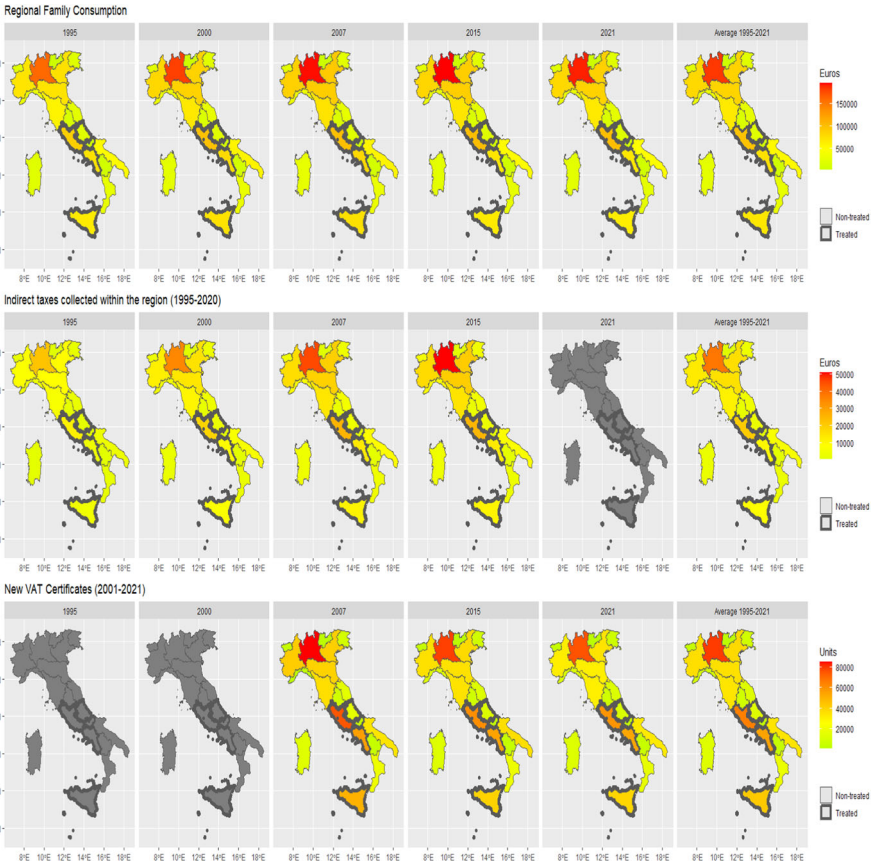

Fig. 7 Observed values of family consumption, indirect taxes collected, and New VAT Certificates across the Italian regions from 1995 to 2021. The last column depicts the regional average value 1995–2021. Regions with a marked border are those treated (Lazio, Campania, Abruzzo, Molise, and Sicily)

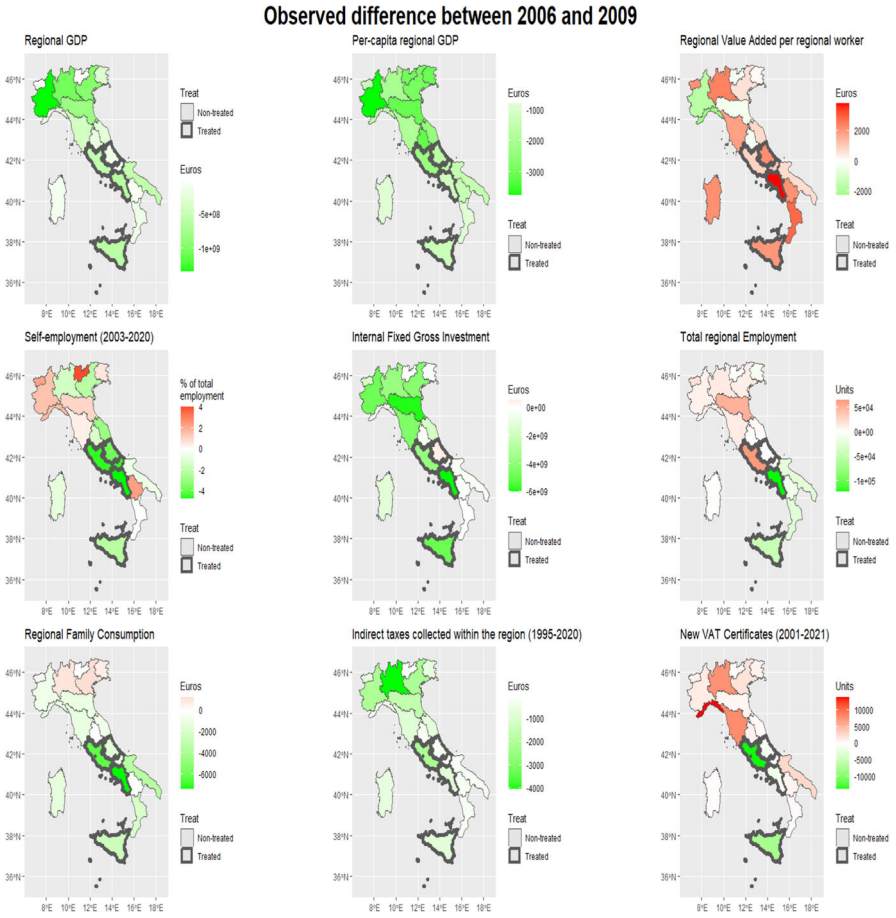

Fig. 8 Absolute variation from 2006 to 2009 for the available macroeconomic indicators at the regional level. The five treated regions are highlighted using bold regional borders

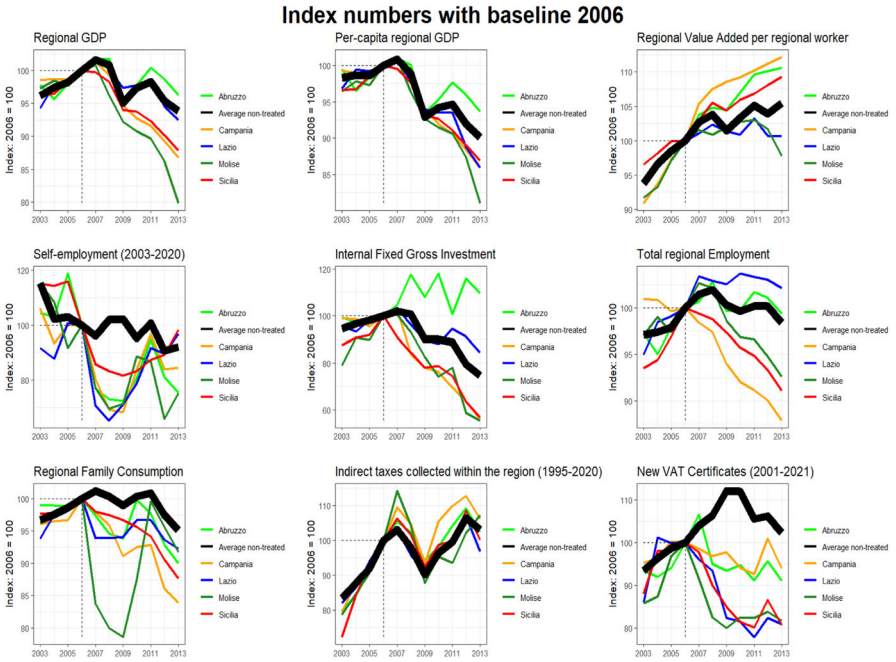

Fig. 9 Time series of index numbers with baseline 2006 = 100 for the available macroeconomic indicators. Estimated values represent the percentage variation for a given year from the baseline year (2006). Abruzzo, Campania, Lazio, Molise, and Sicilia are the five treated regions, while the non-treated regions are averaged across year (solid black lines)

Appendix 2: Additional regional plots

This appendix provides additional illustrative figures obtained using available data. They depict the regional distribution of the macroeconomic indicators from 1995 to 2021 employed in the paper (Table 7).

Table 7 DID estimations of the Italian income tax treatment (time effects reported)

Response Variables	Self-Employment	New VAT Certificates	Growth of per capita GDP	HP filtered per capita GDP (log values)	Per capita GDP	Family Consumption (log values)
ATET or CATET						
Treated vs. untreated regions	− 1.165**	− 4592.525**	− 0.005*	− 0.032***	− 629.19***	− 0.05**
1996				− 0.004	− 259.3	0.006*
1997			0.003	− 0.007	− 33	0.039***
1998			0.006	− 0.006	− 11	0.068***
1999			0.006	− 0.007	93.66	0.092***
2000			0.019***	− 0.012	601.68	0.118***
2001		(base)	0	− 0.018	656.04	0.123***
2002		362.62*	− 0.021***	− 0.017	382.62	0.118***
2003		1009.58**	− 0.024**	− 0.018	73.91	0.124***
2004	− 1.30***	2759.96***	− 0.015**	− 0.024	− 107.4	0.133***
2005	− 0.79	3532.96***	− 0.0188**	− 0.029	− 312.69	0.140***
2006	− 1.10**	4066.77***	− 0.005	− 0.036	− 116.92	0.154***
2007	− 1.73***	6032.48***	− 0.0121	− 0.038	− 2.13	0.156***
2008	− 1.20**	6118.95***	− 0.042***	− 0.043	− 687.6	0.144***
2009	− 1.44***	6257.96**	− 0.077***	− 0.035	− 1955.64	0.128***
2010	− 1.61***	6245.37**	− 0.012	− 0.044	− 1832.65	0.143***
2011	− 0.58	5098.01*	− 0.0163*	− 0.051	− 1823.48	0.148**
2012	− 2.20***	5637.86*	− 0.047***	− 0.045	− 2314.24	0.108***

Table 7 (continued)

Response Variables	Self-Employment	New VAT Certificates	Growth of per capita GDP	HP filtered per capita GDP (log values)	Per capita GDP	Family Consumption (log values)
2013	− 1.90**	4588.62	− 0.042***	− 0.046	− 2823.70*	0.082***
2014	− 2.55***	6938.05**	− 0.021**	− 0.049	− 2855.40*	0.081***
2015	− 2.02**	4246.26	− 0.003	− 0.059	− 2712.25*	0.101***
2016	− 1.87**	3578.83	− 0.009	− 0.066*	− 2543.81	0.112***
2017	− 1.26	4181.6	− 0.0013	− 0.074*	− 2204.57	0.127***
2018	− 2.18**	3998.34	− 0.008	− 0.084**	− 2043.94	0.138***
2019	− 1.55*	3778.65	− 0.011	− 0.092**	− 1955.1	0.141**
2020	− 3.10***	2645.1	− 0.11***	− 0.076**	− 3679.24**	0.023
2021		3013.83	0.049***	− 0.092**	− 2413.5	0.074**
log(GFCF)	− 0.8	− 453.97	0.01	.088***	2241.21***	
Log(Value Added per worker)	− 11.72***	− 13,460.289	0.03	0.24**	9867.1645**	
Constant	160.11***	178,491.2	− 0.54*	5.67***	− 128,581.29***	10.29***
N	355	420	520	540	540	540
Ptrend Test: H0 Linear parallel trend	$F(1, 19) = 2.30$ Prob > $F = 0.1455$ **H_0: cannot be rejected**	$F(1, 19) = 2.02$ Prob > $F = 0.1715$ **H_0: cannot be rejected**	$F(1, 19) = 0.02$ Prob > $F = 0.9030$ **H_0: cannot be rejected**	$F(1, 19) = 1.18$ Prob > $F = 0.2903$ **H_0cannot be rejected**	$F(1, 19) = 0.34$ Prob > $F = 0.5674$ **H_0cannot be rejected**	$F(1, 19) = 0.62$ Prob > $F = 0.4398$ **H0 cannot be rejected**
Granger Type Test: H0 No Anticipation effects	$F(3, 19) = 1.50$ Prob > $F = 0.247$ **H_0cannot be rejected**	$F(5, 19) = 1.34$ Prob > $F = 0.29$ **H_0cannot be rejected**	$F(10, 19) = 9.99$ Prob > $F = 0.0000$ **Reject H_0**	$F(11, 19) = 11.29$ Prob > $F = 0.0000$ **Reject H_0**	$F(11, 19) = 7.29$ Prob > $F = 0.0001$ **Reject H_0**	$F(11, 19) = 7.12$ Prob > $F = 0.0001$ **Reject H_0**

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00181-025-02748-7.

Acknowledgements This paper is part of a research project on income taxation and growth in Italy. We owe the idea of using the tax experiment of 2007 to Alessandro Santoro of the University of Milano-Bicocca. Enza Caruso and Giuseppe Pisauro kindly provided details about the Italian budget law of December 2006 (containing the tax rules used in this paper). A previous version of the paper was presented at seminars held at Bicocca University and at RCEA, International Conference in Economics, Econometrics and Finance, Brunel University London, May 2024. With the usual disclaimers, we thank seminars' participants and particularly Matteo Pelagatti, Giovanni Urga, and Marios Zachariadis for helpful comments. We also thank two anonymous referees for their observations and suggestions on a previous version. We are the only responsible for any remaining errors and mistakes.

Funding Open access funding provided by Università degli Studi di Milano - Bicocca within the CRUI-CARE Agreement. The authors did not receive support from any organisation for the submitted work.

Data availability and codes All results presented in this paper can be reproduced using the R and Stata 18 software. The codes were developed entirely by the authors. Attached to the submission files, we attach the complete set of dataset and scripts. For the reproducibility purposes, all scripts and the data are made available for the public on the following GitHub folder: https://github.com/PaoloMaranzano/BoscoMaranzano_DiD07RegionsIta.git.

Declarations

Conflict of interest The authors have no conflict of interest to declare that are relevant to the content of this article.

References

Aghion P, Howitt P (1990) A model of growth through creative destruction. Natl Bureau Econ Res Work Paper Ser 32:23. https://doi.org/10.3386/w3223

Agrawal DR, Foremny D, Martínez-Toledano C (2025) Wealth tax mobility and tax coordination. Am Econ J Appl Econ 17(1):402–430. https://doi.org/10.1257/app.20220615

Akcigit U, Baslandze S, Stantcheva S (2016) Taxation and the international mobility of inventors. Am Econ Rev 106(10):2930–2981. https://doi.org/10.1257/aer.20150237

Appelbaum E, Katz E (1986) Measures of risk aversion and comparative statics of industry equilibrium. Am Econ Rev 76(3):524–529

Auerbach AJ, Siegel JM (2000) Capital-gains realizations of the rich and sophisticated. Am Econ Rev 90(2):276–282. https://doi.org/10.1257/aer.90.2.276

Barrios S, Huizinga H, Laeven L, Nicodème G (2009) International taxation and multinational firm location decisions. Retrieved from https://EconPapers.repec.org/RePEc:euf:ecopap:0356

Baskaran T (2021) The revenue and base effects of local tax hikes: evidence from a quasi-experiment. Int Tax Public Finance 28(6):1472–1518. https://doi.org/10.1007/s10797-021-09657-2

Bosco B, Bosco CF, Maranzano P (2024) Income taxation and labour response. Empirical evidence from a DID analysis of an income tax treatment in Italy. FEEM Working Paper No. 16.2024. Available at SSRN: https://ssrn.com/abstract=4882611. https://doi.org/10.2139/ssrn.4882611

Caruso E, Dirindin N (2019) Salute ed Economia-Questioni di economia e politica sanitaria. Bologna, Italy, Il Mulino. https://www.mulino.it/isbn/9788815279002
Da Rin M, Di Giacomo M, Sembenelli A (2011) Entrepreneurship, firm entry, and the taxation of corporate income: evidence from Europe. J Public Econ 95(9):1048–1066. https://doi.org/10.1016/j.jpubeco.2010.06.010
De Philippis M, Locatelli A, Papini G, Torrini R (2022) La crescita dell'economia italiana e il divario Nord-Sud: trend storici e prospettive alla luce dei recenti scenari demografici [Italian economic growth and the North-South gap: Historical trends and future projections in light of the recent demographic scenarios]. *Bank* of Italy Occasional Paper(683)
Djankov S, Ganser T, McLiesh C, Ramalho R, Shleifer A (2010) The effect of corporate taxes on investment and entrepreneurship. Am Econ J Macroecon 2(3):31–64. https://doi.org/10.1257/mac.2.3.31
Dynan KE, Skinner J, Zeldes SP (2004) Do the rich save more? J Polit Econ 112(2):397–444. https://doi.org/10.1086/381475
Eissa N, Liebman JB (1996) Labor supply response to the earned income tax credit. Q J Econ 111(2):605–637. https://doi.org/10.2307/2946689
Elhorst JP (2010) Applied spatial econometrics: raising the bar. Spat Econ Anal 5(1):9–28. https://doi.org/10.1080/17421770903541772
Elhorst JP (2014) Spatial econometrics: from cross-sectional data to spatial panels, vol 479. Springer, Berlin
Elhorst JP (2024) Raising the bar in spatial economic analysis: two laws of spatial economic modelling. Spat Econ Anal. https://doi.org/10.1080/17421772.2024.2334845
Esteller-Moré A, Solé-Ollé A (2002) Tax setting in a federal system: the case of personal income taxation in Canada. Int Tax Public Finance 9(3):235–257. https://doi.org/10.1023/A:1016212110137
Federici D, Ferrante F, Parisi V (2023) The educational roots of persistent territorial disparities: the case of Italy. GeoJournal 88(2):2063–2082. https://doi.org/10.1007/s10708-022-10740-2
Goolsbee A (2000) What happens when you tax the rich? Evidence from executive compensation. J Polit Econ 108(2):352–378. https://doi.org/10.1086/262122
Granger CW, Newbold P (1974) Spurious regressions in econometrics. J Econom 2(2):111–120
Harris RDF, Tzavalis E (1999) Inference for unit roots in dynamic panels where the time dimension is fixed. J Econom 91(2):201–226. https://doi.org/10.1016/S0304-4076(98)00076-1
Heim BT (2010) The responsiveness of self-employment income to tax rate changes. Labour Econ 17(6):940–950. https://doi.org/10.1016/j.labeco.2010.02.010
Hodrick RJ, Prescott EC (1997) Postwar U.S. business cycles: an empirical investigation. J Money Credit Bank 29(1):1–16. https://doi.org/10.2307/2953682
Jakobsen K, Jakobsen K, Kleven H, Zucman G (2019) Wealth taxation and wealth accumulation: theory and evidence from Denmark. Q J Econ 135(1):329–388. https://doi.org/10.1093/qje/qjz032
Jappelli T, Pistaferri L (2010) The Consumption response to income changes. Annu Rev Econ 2(1):479–506. https://doi.org/10.1146/annurev.economics.050708.142933
Johnson DS, Parker JA, Souleles NS (2006) Household expenditure and the income tax rebates of 2001. Am Econ Rev 96(5):1589–1610. https://doi.org/10.1257/aer.96.5.1589
Keane MP (2011) Labor supply and taxes: a survey. J Econ Lit 49(4):961–1075. https://doi.org/10.1257/jel.49.4.961
Keane MP (2022) Recent research on labor supply: implications for tax and transfer policy. Labour Econ 77:102026. https://doi.org/10.1016/j.labeco.2021.102026
Kihlstrom RE, Laffont J-J (1979) A general equilibrium entrepreneurial theory of firm formation based on risk aversion. J Polit Econ 87(4):719–748. https://doi.org/10.1086/260790
Kleven HJ, Landais C, Saez E (2013) Taxation and international migration of superstars: evidence from the European football market. Am Econ Rev 103(5):1892–1924. https://doi.org/10.1257/aer.103.5.1892
Levin A, Lin C-F, Chu C-SJ (2002) Unit root tests in panel data: asymptotic and finite-sample properties. J Econom 108(1):1–24
Lo Cascio I, Mazzola F, Epifanio R (2019) Territorial determinants and NUTS 3 regional performance: a spatial analysis for Italy across the crisis. Pap Reg Sci 98(2):641–678. https://doi.org/10.1111/pirs.12372
McCarthy J (1995) Imperfect insurance and differing propensities to consume across households. J Monet Econ 36(2):301–327. https://doi.org/10.1016/0304-3932(95)01214-1
Merfeld JD (2019) Spatially heterogeneous effects of a public works program. J Dev Econ 136:151–167. https://doi.org/10.1016/j.jdeveco.2018.10.007

Meyer BD, Rosenbaum DT (2001) Welfare, the earned income tax credit, and the labor supply of single mothers. Q J Econ 116(3):1063–1114. https://doi.org/10.1162/00335530152466313
Moretti E, Wilson DJ (2017) The effect of state taxes on the geographical location of top earners: evidence from star scientists. Am Econ Rev 107(7):1858–1903. https://doi.org/10.1257/aer.20150508
Muralidharan K, Niehaus P, Sukhtankar S (2023) General equilibrium effects of (improving) public employment programs: experimental evidence from India. Econometrica 91(4):1261–1295. https://doi.org/10.3982/ECTA18181
Nguyen ADM, Onnis L, Rossi R (2021) The macroeconomic effects of income and consumption tax changes. Am Econ J Econ Pol 13(2):439–466. https://doi.org/10.1257/pol.20170241
Oni MH (2023) Progressive income taxation and consumption baskets of rich and poor. J Econ Dyn Control 157:104758. https://doi.org/10.1016/j.jedc.2023.104758
Parker JA (1999) The reaction of household consumption to predictable changes in social security taxes. Am Econ Rev 89(4):959–973. https://doi.org/10.1257/aer.89.4.959
Parker JA, Souleles NS, Johnson DS, McClelland R (2013) Consumer spending and the economic stimulus payments of 2008. Am Econ Rev 103(6):2530–2553. https://doi.org/10.1257/aer.103.6.2530
Pesaran MH (2007) A simple panel unit root test in the presence of cross-section dependence. J Appl Econom 22(2):265–312
Phillips PCB, Sul D (2007) Transition modeling and econometric convergence tests. Econometrica 75(6):1771–1855. https://doi.org/10.1111/j.1468-0262.2007.00811.x
Phillips PCB, Sul D (2009) Economic transition and growth. J Appl Econom 24(7):1153–1185. https://doi.org/10.1002/jae.1080
Piketty T, Saez E, Stantcheva S (2014) Optimal taxation of top labor incomes: a tale of three elasticities. Am Econ J Econ Policy 6(1):230–271. https://doi.org/10.1257/pol.6.1.230
Ramey VA (2011a) Can government purchases stimulate the economy? J Econ Lit 49(3):673–685. https://doi.org/10.1257/jel.49.3.673
Ramey VA (2011b) Identifying government spending shocks: it's all in the timing. Q J Econ 126(1):1–50. https://doi.org/10.1093/qje/qjq008
Robson MT, Wren C (1999) Marginal and average tax rates and the incentive for self-employment. South Econ J 65(4):757–773. https://doi.org/10.1002/j.2325-8012.1999.tb00198.x
Romer CD, Romer DH (2010) The macroeconomic effects of tax changes: estimates based on a new measure of fiscal shocks. Am Econ Rev 100(3):763–801. https://doi.org/10.1257/aer.100.3.763
Roth J, SantAnna PHC, Bilinski A, Poe J (2023) What's trending in difference-in-differences? A synthesis of the recent econometrics literature. J Econom 235(2):2218–2244. https://doi.org/10.1016/j.jeconom.2023.03.008
Rubolino E, Giommoni T (2023) Taxation and mobility: evidence from tax decentralization in italy. CESifo Working Paper No. 10655, Available at SSRN: https://ssrn.com/abstract=4583745. https://doi.org/10.2139/ssrn.4583745
Saez E (2002) Optimal income transfer programs: intensive versus extensive labor supply responses. Q J Econ 117(3):1039–1073. https://doi.org/10.1162/003355302760193959
Saez E, Matsaganis M, Tsakloglou P (2012) Earnings determination and taxes: evidence from a cohort-based payroll tax reform in Greece. Q J Econ 127(1):493–533. https://doi.org/10.1093/qje/qjr052
Souleles NS (1999) The response of household consumption to income tax refunds. Am Econ Rev 89(4):947–958. https://doi.org/10.1257/aer.89.4.947
Stantcheva S (2019) Lecture 4: Labor Supply Responses to Taxation—Spring 2019. Retrieved from https://scholar.harvard.edu/files/stantcheva/files/empirical_responses_4.pdf
Sun S, Delgado MS (2024) Local spatial difference-in-differences models: treatment correlations, response interactions, and expanded local models. Empir Econ. https://doi.org/10.1007/s00181-024-02610-2
Wen J-F, Gordon DV (2014) An empirical model of tax convexity and self-employment. Rev Econ Stat 96(3):471–482. https://doi.org/10.1162/REST_a_00388
Widmann R (2023) Immigrant inventors and local income taxes: evidence from Swiss municipalities. J Public Econ 219:104822. https://doi.org/10.1016/j.jpubeco.2023.104822
Wikle CK, Zammit-Mangion A, Cressie N (2019) Spatio-temporal Statistics with R. Chapman and Hall/CRC, Cambridge
Wooldridge JM (2021) Two-way fixed effects, the two-way mundlak regression, and difference-in-differences estimators. Available at SSRN 3906345. https://doi.org/10.2139/ssrn.3906345
Zidar O (2019) Tax cuts for whom? Heterogeneous effects of income tax changes on growth and employment. J Polit Econ 127(3):1437–1472. https://doi.org/10.1086/701424

Ziliak JP, Kniesner TJ (2005) The effect of income taxation on consumption and labor supply. J Law Econ 23(4):769–796. https://doi.org/10.1086/491611