Journal of Business Ethics (2021) 170:471–496
https://doi.org/10.1007/s10551-019-04393-4

ORIGINAL PAPER

The Effects of Immigration on Labour Tax Avoidance: An Empirical Spatial Analysis

Diego Ravenda[1] · **Maika M. Valencia-Silva**[2] · **Josep M. Argiles-Bosch**[3] · **Josep García-Blandón**[4]

Received: 1 April 2019 / Accepted: 9 December 2019 / Published online: 20 December 2019

Abstract

We investigate whether the geographic concentration of non-EU immigrants in the various Italian provinces affects labour tax avoidance (LTAV) practices adopted by firms located in the same provinces, as well as in the neighbouring provinces, and operating in construction and agriculture industries that mostly employ immigrants in Italy. For this purpose, we develop a LTAV proxy based on the financial accounting information of a sample of 993,606 firm-years, disseminated throughout the 108 Italian provinces, over the period 2008–2016. Our results, based on a Spatial Durbin Model panel regression, reveal a statistically significant positive association between the concentration of non-EU immigrants and LTAV at province level, as well as the presence of spillover effects among neighbouring provinces. Our findings are robust to several additional analyses, including instrumental variable estimations. Our study provides empirical support to previous structuralist or marginalization theories holding that socioeconomically marginalized groups, such as non-EU immigrants, are more likely to be involved in labour exploitation practices, which could underlie our LTAV outcomes. Furthermore, it supports the need for tax authorities to strengthen labour inspections, coordinated at national level, especially in those contexts where non-EU immigrants are mostly employed. On the other hand, a greater social integration, assistance, and recognition of rights of immigrants may help to alleviate their situation of weakness that makes them more vulnerable to LTAV practices. Finally, tackling LTAV, associated with the underemployment of immigrants, may prevent its negative effects for society arising from the reduction of public resources to sustain the social welfare and finance public goods and services.

Keywords Immigration · Labour tax avoidance · Spatial analysis

✉ Diego Ravenda
dravenda@hotmail.com; d.ravenda@tbs-education.es

Maika M. Valencia-Silva
maika.melina.valencia.silva@campus.eae.es

Josep M. Argiles-Bosch
josep.argiles@ub.edu

Josep García-Blandón
josep.garcia@iqs.url.edu

1 TBS Business School, Campus Barcelona, C/Trafalgar, 10, 08010 Barcelona, Spain

2 EAE Business School, Campus Barcelona, C/Tarragona, 110, 08015 Barcelona, Spain

3 University of Barcelona, Facultat d'Economia i Empresa, Av. Diagonal, 690, 08034 Barcelona, Spain

4 IQS School of Management, Universidad Ramón Llull, Via Augusta, 390, 08017 Barcelona, Spain

Abbreviations

AbSSCs	Abnormal social security contributions
ISTAT	Italian Institute of Statistics
LTAV	Labour tax avoidance
NSSCs	Normal social security contributions
SAR	Spatial autoregressive model
SDM	Spatial durbin model
SEM	Spatial error model
SSCs	Social security contributions
UDW	Undeclared work

Introduction

One of the most relevant demographic and socioeconomic trend worldwide during the last decades has been the steady increase of the foreign-born population, especially within developed countries (Longhi et al. 2010a). In this regard, recent estimates indicate that in 2015 about 244 million people were international migrants in the world, resulting in an increase of more than 40% since 2000 (United Nations Department of Economic and Social Affairs (UNDESA) 2017). This provides further evidence of the ongoing intensive migratory flows fuelled by widespread economic inequalities between countries, which push migrants from the most disadvantaged territories to seek new economic opportunities, and by the need to escape from armed conflicts, political instability, and persecutions (Theodore et al. 2018). In this line, in Italy, the non-EU resident immigrants amounted to 3.7 million (6.26% of resident population) in 2017, with an increase of 42% since 2008, when they represented 4.37% of resident population (Italian Institute of Statistics (ISTAT) 2018). Indeed, the geographic position of Italy, at the southern boundary of the European Union (EU) and at the crossroads of several Mediterranean migration pathways, makes it a natural bridge for the entry of migrants from North Africa and the Middle East into the European economy in general (Harney 2011; Triandafyllidou and Maroukis 2012). It is no accident that, after Spain, Italy is the European state that has received the largest number of immigrants over the past 25 years, mainly from developing countries and Eastern Europe (Fullin and Reyneri 2011). Furthermore, Italy is among the EU countries that over the period 2013–2015 have been most affected by the unprecedented inflow of refugees, asylum seekers and other undocumented migrants (Constant and Zimmermann 2016; Dustmann et al. 2017).

This increased relevance of the phenomenon of immigration and its socioeconomic effects have intensified in recent times the concerns of policy-makers and local populations on the issue of the integration of immigrants in the socio-economic context of the host countries and specifically in their labour market (Longhi et al. 2010b). In this respect, prior research documents that, in several price-competitive sectors with highly wavering demand, employers, willing to violate immigration and labour regulations, resort to undeclared immigrant workers and their exploitation to minimize labour costs (Maroukis et al. 2011; Theodore et al. 2018; Yea 2017). Indeed, the scarce employment options due to their restricted or absent labour rights, the lack of information about their rights, the limited language skills, the non-recognition of qualifications and work experiences achieved in other countries, as well as other forms of discrimination may lead immigrants to accept substandard employment within the informal economy or more precarious, insecure and illegal working conditions, especially in sectors characterized by low-skilled jobs, mostly unattractive to nationals (Annisette and Trivedi 2013; Cappelen and Muriaas 2018; Lewis et al. 2015; Strauss and McGrath 2017). Therefore, for several immigrants the subjection to a state of severe labour exploitation in the formal or informal economy may represent the only viable livelihood option, at least for a certain period, while establishing themselves within a host society (Lewis et al. 2015; Pajnik 2016). In this regard, the International Labour Organization (ILO) (2013) underlines how their precarious legal status and engagement in non-standard and undeclared work make immigrant workers more vulnerable to extreme forms of labour exploitation such as forced or unfree labour, defined as: "*all work or service which is exacted from any person under the menace of any penalty and for which the said person has not offered himself voluntarily*".

Despite the current social relevance of the above issues, empirical studies, aiming to unveil the effects of immigration and its regulation on labour market practices starting from data at microeconomic level, are relatively scarce (Borjas 2017; Di Porto et al. 2018; Monras et al. 2018; Yea 2017). Hence, to address this research gap, in this paper we aim to assess whether the geographic concentration of non-EU immigrants[1] in the various Italian provinces significantly influences the labour tax avoidance (LTAV) practices adopted by firms located not only in the same provinces, but also in the neighbouring provinces, because of the presence of spatial spillover effects. It is essential to clarify that we broadly define LTAV as the reduction of firm's explicit labour tax liability through specific procedures, regardless of their legality. Indeed, similar to prior definitions of income tax avoidance (Donohoe 2015; Lanis et al. 2018), we include within LTAV a continuum of labour tax planning strategies, spanning from relatively benign strategies envisioned by tax policies on the left to extremely aggressive or illegal strategies on the right. Importantly, we include in the labour tax definition all social security contributions (SSCs) and other insurances, computed on gross salaries of all workers, that the employers are legally required to withhold and pay to tax authorities to support the social protection of their employees (Ravenda et al. 2015).

Tax avoidance procedures may be unquestionably illegal as in the case of the employment of undeclared workers. However, when their legality cannot be clearly assessed or questioned, they may involve violations of the spirit of the law, generally considered as unethical, socially irresponsible and even representing a sustainability issue (Bird and Davis-Nozemack 2018; Lanis and Richardson 2015; Ylönen and Laine 2015). Specifically, labour tax may be avoided by abusing of subcontracted workforce, self-employed people or other forms of precarious, and in general non-standard employment arrangements, aiming to circumvent the social security regulations, when the working relationship should be regulated as standard subordinate employment according

[1] In our study, according to the official statistics, we consider an immigrant any resident with non-EU nationality, namely citizens of countries that do not belong either to the EU or the European economic area.

to the labour law (EC 2014; Pfau-Effinger 2009). Hence, LTAV is one of the primary objectives as well as the natural effect of the employment of undeclared work (UDW) and other labour exploitation practices.

We adopt a measure of LTAV based on some related accounting information included in the publicly available financial statements of the employing firms. More specifically, our LTAV proxy is based on the abnormal values of the ratio of SSCs paid to lagged total assets of 993,606 firm-years, disseminated throughout the 108 Italian provinces over the period 2008–2016, in construction and agriculture industries. We specifically focus on construction and agriculture given that, on the one hand, they are among the industries with the highest employment of non-EU immigrants in Italy and other EU countries (Corrado 2011; Directorate General of Immigration and Integration Policies 2018; Pajnik 2016; Prosser 2016; Strauss and McGrath 2017), and, on the other hand, they experience higher rates of UDW and other LTAV practices, compared to other industries (Buehn 2012; Trinci 2006; Williams and Nadin 2012). In addition, the effects of recent labour reforms in several European countries, including Italy,[2] aiming to bring greater flexibility to the labour market, with the consequent relaxation of the employment social protection, have particularly affected these industries and the involved migrant workers (Pajnik 2016). Importantly, our LTAV proxy may reflect not only illegal practices, but also a strategic use of the legal tools available to relieve the labour tax burden. However, we assume that, due to our research design that considers the peculiarities of each industry and year, the illegal forms of LTAV such as UDW are more likely to be the primary driver of the extremely abnormal values taken by our LTAV proxy. Indeed, the room to legally relieve labour tax is quite limited, quickly exhausted and UDW is the primary illegal means commonly employed to evade labour tax (Feld and Schneider 2010; Williams and Nadin 2012). In addition, although our measure of LTAV cannot capture all informal economic activity (e.g. unregistered firms are excluded), it may provide evidence of the relationship between non-EU immigration and LTAV within its validity boundaries and the results may be extrapolated to the general economic context.

In terms of methodology, we adopt a two-step regression procedure aiming to aggregate firm-level LTAV measures at province level in the first step and to estimate a Spatial Durbin Model (SDM) regression (LeSage and Pace 2009), across the 108 provinces for 9 years (2008–2016), in the second step. In particular, the usage of a SDM panel fixed-effect regression allows accounting for spatial interdependence among province-level observations that, if unaddressed, may bias the estimations, and specifically unveiling not only the effect of non-EU immigrant concentration in a province on LTAV in the same province (direct effects), but also the effect of non-EU immigrant concentration in a province on LTAV in the neighbouring provinces (indirect or spillover effects). In this vein, it is plausible to assume that immigrants resident in a province may move to the neighbouring provinces for work within an affordable distance limit and that, in general, a province is influenced by its neighbouring provinces in several economic, demographic and social aspects (Bastida et al. 2013).

Overall, our results support our hypothesis on the positive association between non-EU immigrant concentration and LTAV at province level and reveal the presence of spillover effects among neighbouring provinces. Our findings are robust to several additional analyses, including instrumental variable estimations to account for any endogeneity that may arise from reverse causality or correlated omitted variable bias. Hence, our results may provide empirical support to previous structuralist or marginalization theories (Cappelen and Muriaas 2018; Williams and Horodnic 2015), holding that spatially and socioeconomically marginalized groups, such as non-EU immigrants, are more likely to be involved in UDW and/or other labour exploitation practices, which could underlie our LTAV outcomes. Furthermore, our findings may suggest that labour market competition, caused by increased immigration, may negatively affect working conditions and enhance LTAV also for low-skilled/paid national workers, mostly employed in agriculture and construction industries.

Previous studies document the tendency of immigrants to be underemployed in the informal economy of the host countries, using case studies, interviews, surveys, and macroeconomic statistics (Bohn and Owens 2012; Borjas 2017; Cappelen and Muriaas 2018; Pajnik 2016; Theodore et al. 2018; Yea 2017). In this research context, our study is, to our knowledge, the first attempt to provide empirical evidence of the impact of immigration on LTAV, the logical effect of UDW and other labour exploitative practice, by starting from firm-level accounting information to carry out a spatial econometric analysis. Specifically, the spatial analysis based on the SDM model provides new evidence on spillover effects across neighbouring provinces due to the presence of spatial clusters in LTAV practices and the higher mobility of immigrants for work reasons, relative to natives, that Di Porto et al. (2018) only assume to explain their findings on the impact of the regularization of migrant workers on firm regular employment in Italy. Therefore, our results reveal that spatially clustered determinants of LTAV may lead to an emulation behaviour of labour-intensive neighbouring firms that resort to LTAV, rather than to technology, to compete locally and globally through the reduction of labour costs. This context may produce unfair competition for firms not

[2] The most recent labour market reform in Italy, the so-called Jobs Act, was enacted by the Renzi government in 2014.

engaging in LTAV practices. Therefore, a stricter enforcement of labour regulation in specific provinces may spill over its beneficial effects to the neighbouring province. In addition, our paper contributes to the literature given that it empirically confirms that, at least in certain industries dominated by low-skilled jobs, non-EU immigration may provide opportunities for LTAV practices, including UDW. These effects highlight the need for tax authorities to strengthen controls and labour inspections, in a coordinated manner throughout the national territory, especially in those contexts where non-EU immigrants are mostly employed. Finally, a greater social integration and recognition of rights of immigrants may help to alleviate their situation of weakness that makes them more vulnerable to labour exploitation practices. The alternative would be to allow LTAV practices to flourish, with the consequent negative effects for society in terms of reduction of public resources to sustain the social welfare and finance public goods and services.

The remainder of the paper proceeds as follows: "Working Conditions of Immigrants within the Italian Context" section examines the working conditions of immigrants in Italy; "Theoretical Research Background and Hypothesis" section reviews the research theories supporting the main hypothesis; Sect. "Research Design" describes the research design and sample data; "Regression Results and Analyses" section presents empirical results; "Conclusions and Discussion" section includes concluding remarks.

Working Conditions of Immigrants Within the Italian Context

Several previous studies examine the working conditions of immigrants, especially non-EU citizens, within the Italian context. In particular, several scholars assert that Italy is an attractive transit or settlement country for non-EU migrants not only for its proximity to the hotspots of North Africa and the Middle East, but also for the relatively large informal economy that provides employment opportunities for undocumented immigrants, especially in Southern Italian regions (Corrado 2011; Fullin and Reyneri 2011; Harney 2011; Triandafyllidou and Maroukis 2012). However, although the occurrence of some clandestine entries along Italy's extensive coastline, most of the non-EU immigrants enter Italy legally documented, as refugees or asylum seekers, and subsequently they over-stay their visa or breach its conditions by working (Dustmann et al. 2017; Harney 2011). Indeed, in Italy immigrants applying for asylum are not allowed to legally work for the first 6 months following their application or before their claim is positively evaluated by the immigration authorities (Constant and Zimmermann 2016; Dustmann et al. 2017). As this evaluation process may take far more than 6 months (Dustmann et al. 2017), in the meantime, several asylum seekers are absorbed in the underground economy, where they can find additional financial support to the modest allowance (pocket money) they receive from the government (Harney 2011). In addition, as most of asylum applications end up being denied (Commissione Nazionale per il Diritto di Asilo 2018; Seifert and Valente 2018), working informally represents the only available option for the significant proportion of immigrants that, after the asylum denial or the loss of their temporary permit, decide to remain in the country illegally (Hatton et al. 2017). Corrado (2011) suggests that, in Southern European Mediterranean countries, the seasonality and high labour-intensity of dominant economic sectors (e.g. agriculture, construction, and tourism) lead to a demand for a flexible, less qualified and poorly paid labour force which "escapes the regulated nature of unionized, formal sector employment and is available only when needed by employers". Hence, Italian agriculture, mostly represented by medium and small-sized farms, highly seasonal and widely exposed to global competition, greatly relies, for its subsistence, on the underemployment of cheap and undeclared labour force, typically consisting of irregular immigrants, asylum seekers, and refugees (Corrado 2011; Maroukis et al. 2011).

Finally, it is worth mentioning that a severely exploitative labour practice, mostly involving immigrants, widespread in the Italian agriculture and construction sectors and carried out by Italian Mafias is the so-called *Caporalato*. More specifically, *Caporalato* is a crime provided for by the Italian penal code (article 603-bis), consisting in the illicit brokering and exploitation of workforce. Specifically, illegal labour brokers called *Caporali*, often associated with Mafia organizations, hire, on behalf of farmers or builders, migrant workers to be illegally exploited and retain, as compensation, about half of the daily salary of the workers, as well as charging them for additional service fees (Flai-Cgil 2014; Seifert and Valente 2018).

In summary, our overview of the working conditions of non-EU immigrants in Italy provides concordant arguments that may support a research hypothesis on the existence of a significant positive association between the spatial presence of non-EU immigrants and LTAV practices adopted by firms of industries that mostly employ immigrants.

Theoretical Research Background and Hypothesis

In addition to some obvious conclusions that can be drawn from the analysis of the specific Italian context, other theories, suggested in prior research, may support our hypothesis on the role of non-EU immigration in fostering LTAV practices. In this regard, based on 74 semi-structured interviews conducted with Polish labour migrants in Norway, Cappelen and Muriaas (2018) show that the involvement of immigrants in insecure, precarious and undeclared work is mainly triggered by a combination of voluntary exit from the formal labour market, to achieve higher net income, as well as structures, such as the immigrants' social life (e.g. lack of social networks and integration within the native community) or their work life (e.g. difficulties in getting legally declared work), that make it more likely for this type of workers to be forced to accept these working conditions. Nonetheless, the authors consider the influence of external societal structure more determinant and then call for more research on how to best integrate labour migrants into the civil society of the host country. In summary, the authors suggest that both the structuralist and the individualistic neoliberal perspectives are applicable to explain the UDW of Polish labour migrants in Norway.

Indeed, in structuralist theories, UDW is mainly driven by *poverty escape* (Pfau-Effinger 2009) and survival motivations of marginalized populations such as the immigrants. Specifically, these population groups are necessarily excluded from the formal labour market and related government benefits according to the logic of the modern globalized capitalism, which leads employers to reduce labour costs through labour exploitation practices, including informal waged work and dependent or false self-employment, a form of work which is largely unregulated, underpaid, precarious and insecure (Adom 2014; Cappelen and Muriaas 2018; Williams and Round 2010). UDW is also viewed to be a "direct outcome of the demise of the intended full-employment/comprehensive formal welfare state regime characteristic of the Fordist and socialist era" (Williams and Round 2010). On the other hand, according to neoliberal theories, UDW is an outcome of people voluntarily exit the formal labour market to achieve more autonomy, flexibility, better remuneration, and avoidance of taxes and inefficient labour over-regulation (Cappelen and Muriaas 2018; Williams and Round 2010). Hence, participants in UDW are seen as microentrepreneurs choosing to operate off-the-books and outside the law in order to avoid the costs of market over-regulation and establish a real free market (Williams and Round 2010). Finally, in more recent years, post-structuralist theories suggest that UDW is the result of voluntary exit rather than exclusion, although the decision, conducted for and by kin, neighbours, friends and acquaintances, is mostly driven by social and redistributive rationales rather than by financial gain purposes (Williams and Round 2010). In addition, UDW is seen as a way to escape the exploitation of workers in the neoliberal global economic system and the corruption and bribes that can be part and parcel of the formal economy (Adom 2014; Biles 2009).

The relevance of these theories depends on the considered population group and socioeconomic context. In this vein, some scholars argue that the structuralist perspective, supporting "forced exclusion", is more applicable to waged undeclared work of relatively deprived populations, whereas the neoliberal perspective, supporting "voluntary exit", is more applicable to own-account informal workers that are relatively more affluent (Gurtoo and Williams 2009; Williams and Round 2010). In this regard, two contrasting perspectives on the socioeconomic and spatial variations in UDW are prevalent in the literature, namely the marginalization and reinforcement theories. Specifically, the dominant marginalization theory holds that informal work mostly involves low-paid, insecure, unregulated and low-qualified jobs carried out by spatially and socioeconomically marginalized people with fewer opportunities in the labour market, including immigrants and less affluent population groups, to cope with poverty (Beręsewicz and Nikulin 2018; Williams and Horodnic 2015). In this line, previous studies find that marginalized and low-skilled immigrants are more likely to be underemployed in the informal economy of the host countries especially in low-skilled labour-intensive industries (Bohn and Owens 2012; Theodore et al. 2018; Venkatesh and Fiola 2006).

On the other hand, the more recently developed reinforcement theory assert that the involvement in UDW is lower among marginalized populations, implying that the informal economy enhances the socioeconomic and spatial disparities produced by the formal economy (Williams and Horodnic 2015). In this respect, Williams & Nadin (2014) find that, in East-Central Europe and Western European nations, the marginalization and reinforcement perspectives co-exist given that marginalized groups, such as the unemployed, are more likely to be involved in UDW but gain significantly less and are more vulnerable to labour exploitation than those working undeclared as a complement to declared jobs. On the other hand, based on surveys conducted in various EU countries, other studies find that the marginalization thesis may only be valid for some marginalized populations but not for others (Williams and Horodnic 2015). These conflicting results highlight the need of a more nuanced interpretation of the marginalization thesis that should consider the socioeconomic context, the industry, as well as the peculiarities of the population group under analysis.

Specifically, in our study we consider that a structuralist perspective may be applicable to the non-EU immigrants in their employment in the agriculture and construction industries in Italy. Indeed, their previously described marginalized status, in terms of labour and social rights, economic conditions, and social integration, may make them vulnerable and forced victims of a capitalist exploitation, aiming at reducing labour cost, including related taxation, and enhancing competitiveness of the employing firms in the global market.

Finally, restrictive migration regimes, aiming to reduce immigrant rights in the host country, may even be used by employers and governments to undermine wages, terms and rights of all workers broadly (Strauss and McGrath 2017). Indeed, immigrant and national workers cannot remain conceptually and spatially compartmentalized from one another (Strauss and McGrath 2017). Hence, a higher presence of immigrants within the local workforce may also affect the labour practices, including LTAV, for national workers that may need to compete in the labour market with less demanding and more easily exploitable immigrants (Bohn 2010). In this regard, prior research on immigration in EU countries finds that immigration can negatively impact the working conditions of previous immigrants and low-paid/skilled native workers, that are close substitutes for immigrants, especially in sectors such as agriculture and construction (D'Amuri et al. 2010; Dustmann et al. 2013; Prosser 2016). In summary, all our previous arguments lead the following main hypothesis of our study:

Hypothesis *Ceteris paribus*, the concentration of non-EU immigrants is positively associated with the intensity of LTAV, across Italian provinces.

Research Design

Data and Sample Selection

To estimate our main LTAV proxy we use annual accounting data of all private firms located in the 108 Italian provinces and available on the AIDA database[3] over the period 2007–2016. The period 2007–2016 is constrained both by the availability of accounting data in AIDA, that are limited to a 10-year history,[4] and by the availability of data on non-EU immigration in Italy,[5] needed for our analysis, restricted to the 2007–2017 period. Consistent with the scope of our study, the final sample is reduced to Construction (NACE[6] codes: 41, 42, 43) and Agriculture (NACE code: 01) industries and finally consists of 167,920 firms and 993,606 firm-years. It should be noted that the fiscal year 2007 observations are lost in the analysis given that, to compute several variables needed for the estimations, we include 1 year lagged data. Panel A of Table 1 (Sample composition) summarizes the distributions of our sample firm-years by industry and Italian region.[7] Furthermore, the table classifies the Italian regions into their higher first-level NUTS[8] (North West; North East; Centre; South; Islands) and indicates the provinces included in each of the 20 Italian regions. On the other hand, Panel B of Table 1 shows the distribution of sample firm-years by province by ordering provinces in decreasing order of number of firm-years hosted.

Interestingly, 89.22% of firm-years belong to Construction industry, whereas only 10.78% belongs to Agriculture industry. The predominance of Construction in our sample should be considered in assessing the relevance of our study outcomes for policy-makers and tax authorities as well as when extrapolating the results to the general economic context. Furthermore, northern Italian regions (North West and North East) host the highest number of firm-years (40.74%), compared to the centre regions (26.07%) and *Mezzogiorno*[9] (South and Islands) regions (33.19%). This is consistent with the traditional greater economic development and performance of Northern Italy compared to the rest of Italy (Jayet et al. 2010). Finally, Rome (127,341), Milan (56,932) and Naples (44,430) are the provinces that host the highest number of firm-years, with Rome significantly outrunning the others, consistent with their greater population and density, whereas Biella (1325), Carbonia-Iglesias (1178), and Medio Campidano (775) are the provinces with the lowest number of firm-years.

Measure of LTAV and Descriptive Statistics

In Italy, a social security statutory flat rate, ranging from approximately 29% to 32% of each employee gross remuneration (payroll costs), is charged to the employer as SSCs.[10] The social security tax burden and the related policies are the same across the Italian territory and the actual rate depends on: the nature of the activity performed by the company, the number of employees of the company, the

[3] AIDA is a database managed by Italian Bureau Van Dijk, which includes financial statements and other relevant details of 1 million companies in Italy, with up to 10 years of history.

[4] We extracted the accounting data from AIDA over the first 5 months of 2018, when accounting data for fiscal year 2017 were not available yet.

[5] Data on immigration in Italy are provided by the Italian Institute of Statistics (ISTAT) and publicly available on: http://stra-dati.istat.it/.

[6] NACE (for the French term: nomenclature statistique des activités économiques dans la Communauté européenne) is the industry standard classification system used in the European Union. The current version is revision 2 and was established by Regulation (EC) No 1893/2006.

[7] The regions of Italy are the first-level administrative divisions of Italy, constituting its second NUTS (Nomenclature of Territorial Units for Statistics) administrative level. Each of the 20 regions is divided into provinces.

[8] NUTS (Nomenclature of Territorial Units for Statistics) is a geocode standard, developed by the European Union, for referencing the administrative divisions of EU countries for statistical purposes.

[9] *Mezzogiorno* or *Meridione d'Italia* is an economic macro-region traditionally comprising the territories of the former Kingdom of the two Sicilies (all the southern section of the Italian Peninsula and Sicily) as well as the island of Sardinia.

[10] The reference legislation on social security contributions, including their computation rules and settlement, includes law no 335 of August 8th, 1995 and other following circulars of INPS (the national social security institute).

Table 1 Sample composition

Panel A: distribution of sample firm-years by Italian region and industry for the period 2008–2016

Regions (Provinces)	Agriculture		Construction		Total	
	Firm-years	%	Firm-years	%	Firm-years	%
North West						
Lombardy (Bergamo; Brescia; Como; Cremona; Lecco; Lodi; Mantua; Milan; Monza e della Brianza; Pavia; Sondrio; Varese)	9882	0.99	147,018	14.80	156,900	15.79
Piedmont (Alessandria; Asti; Biella; Cuneo; Novara; Turin; Verbano-Cusio-Ossola; Vercelli)	4448	0.45	41,068	4.13	45,516	4.58
Liguria (Genova; Imperia; La Spezia; Savona)	767	0.08	16,490	1.66	17,257	1.74
Aosta Valley (Valle d'Aosta/Vallée d'Aoste)	137	0.01	2245	0.23	2382	0.24
North East						
Veneto (Belluno; Padua; Rovigo; Treviso; Venice; Verona; Vicenza)	7898	0.79	68,979	6.94	76,877	7.74
Emilia-Romagna (Bologna; Ferrara; Forlì-Cesena; Modena; Parma; Piacenza; Ravenna; Reggio nell'Emilia; Rimini)	8633	0.87	64,199	6.46	72,832	7.33
Trentino-South Tyrol (Bolzano/Bozen; Trento)	1834	0.18	15,611	1.57	17,445	1.76
Friuli-Venezia Giulia (Gorizia; Pordenone; Trieste; Udine)	2023	0.20	13,573	1.37	15,596	1.57
Centre						
Lazio (Frosinone; Latina; Rieti; Rome; Viterbo)	11,533	1.16	146,767	14.77	158,300	15.93
Tuscany (Arezzo; Florence; Grosseto; Livorno; Lucca; Massa-Carrara; Pisa; Pistoia; Prato; Siena)	10,224	1.03	49,126	4.94	59,350	5.97
Marche (Ancona; Ascoli Piceno; Fermo; Macerata; Pesaro e Urbino)	2564	0.26	23,457	2.36	26,021	2.62
Umbria (Perugia; Terni)	2621	0.26	12,728	1.28	15,349	1.54
South						
Campania (Avellino; Benevento; Caserta; Naples; Salerno)	8660	0.87	88,381	8.89	97,041	9.77
Apulia (Bari; Barletta-Andria-Trani; Brindisi; Foggia; Lecce; Taranto)	11,502	1.16	54,957	5.53	66,459	6.69
Abruzzo (Chieti; L'Aquila; Pescara; Teramo)	2078	0.21	25,211	2.54	27,289	2.75
Calabria (Catanzaro; Cosenza; Crotone; Reggio di Calabria; Vibo Valentia)	3979	0.40	21,942	2.21	25,921	2.61
Basilicata (Matera; Potenza)	1994	0.20	8669	0.87	10,663	1.07
Molise (Campobasso; Isernia)	682	0.07	4417	0.44	5099	0.51
Islands						
Sicily (Agrigento; Caltanissetta; Catania; Enna; Messina; Palermo; Ragusa; Syracuse; Trapani)	12,366	1.24	56,873	5.72	69,239	6.97
Sardinia (Cagliari; Carbonia-Iglesias; Medio Campidano; Nuoro; Oristano; Sassari)	3269	0.33	24,801	2.50	28,070	2.83
Total	107,094	10.78	886,512	89.22	993,606	100

Panel B: Distribution of sample firm-years by province for the period 2008–2016

Province	Firm-years	Province	Firm-years	Province	Firm-years	Province	Firm-years
Rome	127,341	Cagliari	10,436	Trapani	5891	Rovigo	3643
Milan	56,932	Frosinone	10,355	Viterbo	5852	Lecco	3628
Naples	44,430	Messina	10,046	Forlì-Cesena	5824	Cremona	3605
Bergamo	24,718	Parma	9306	Alessandria	5822	Massa-Carrara	3376
Turin	22,084	Reggio nell'Emilia	9251	Barletta-Andria-Trani	5688	Campobasso	3330
Bari	22,016	Trento	9092	Lucca	5609	La Spezia	3037
Brescia	21,112	Bolzano/Bozen	8353	Syracuse	5581	Sondrio	2957
Caserta	20,984	Udine	8122	Mantua	5561	Nuoro	2801
Salerno	19,597	Genova	7881	Ravenna	5407	Rieti	2692
Verona	16,941	Taranto	7718	Benevento	5386	Lodi	2680
Catania	16,846	Pisa	7405	Catanzaro	4980	Asti	2382
Bologna	16,265	L'Aquila	7335	Grosseto	4929	Valle d'Aosta	2382
Padua	14,439	Ancona	7142	Macerata	4870	Fermo	2314

Table 1 (continued)

Panel B: Distribution of sample firm-years by province for the period 2008–2016

Province	Firm-years	Province	Firm-years	Province	Firm-years	Province	Firm-years
Foggia	13,535	Pesaro e Urbino	6951	Rimini	4775	Belluno	2260
Venice	13,275	Agrigento	6886	Ascoli Piceno	4744	Imperia	2257
Modena	13,266	Chieti	6817	Piacenza	4577	Vibo Valentia	2221
Treviso	13,250	Cuneo	6816	Prato	4542	Crotone	2065
Florence	13,175	Potenza	6710	Reggio di Calabria	4448	Trieste	2065
Vicenza	13,069	Teramo	6699	Caltanissetta	4209	Isernia	1769
Monza	12,355	Avellino	6644	Livorno	4177	Oristano	1591
Cosenza	12,207	Como	6641	Ferrara	4161	Verbano-Cusio-Ossola	1560
Palermo	12,187	Brindisi	6519	Savona	4082	Vercelli	1558
Latina	12,060	Siena	6499	Novara	3969	Enna	1506
Perugia	11,543	Pescara	6438	Matera	3953	Gorizia	1468
Sassari	11,289	Ragusa	6087	Pordenone	3941	Biella	1325
Lecce	10,983	Arezzo	5959	Terni	3806	Carbonia-Iglesias	1178
Varese	10,768	Pavia	5943	Pistoia	3679	Medio Campidano	775

AIDA database, 2018. Agriculture industry includes firms with NACE code 01 (Crop and animal production, hunting and related service activities); Construction industry includes firms with NACE codes: 41(Construction of buildings), 42 (Civil engineering), and 43 (Specialized construction activities)

legal form of the company, and the employee's position, legal status and type of labour contract. Furthermore, some remuneration concepts are completely or partially excluded from the social security tax base,[11] namely fringe benefits, meal, travel and transfer allowances, proceeds received as compensation for damages, disbursements for education and training for employees, among others. Within this legal framework, employers may opportunistically and even fraudulently reduce the social security tax base below the reported employee gross remuneration to avoid SSCs. For example, they may rearrange regular taxable salaries with some of the above-mentioned kinds of compensation that are partially or totally exempt from SSCs under the legislation in force. In these scenarios, the variability of the effective rate of SSCs to gross salaries, reported in the income statement according to Italian accounting regulation,[12] may provide evidence of LTAV across firms, similar to the effective rate of income taxes to pre-tax income, widely used to measure income tax avoidance in previous research (Lanis and Richardson 2012; Platikanova 2015). Nonetheless, a LTAV proxy based on the effective rate of SSCs to gross salaries may provide biased LTAV results as it may be significantly affected by factors, possibly unrelated to LTAV, such as industry peculiarities, firm size and capital intensity, year-specific macroeconomic and regulatory conditions. More importantly, this proxy cannot signal LTAV through the underreporting of salaries for undeclared workers.

To address these concerns, we develop a measure of LTAV based on the ratio of SSCs paid[13] to lagged total assets. More specifically, we follow the intuition of Seifert and Valente (2018) who assume that illegal employment (UDW) displacing legal workforce may lead to underreported labour input and overreported labour productivity. Specifically, they find that the 2011 non-EU migrant wave in southern Italy caused a statistically significant increase of labour productivity of around 11% in 2011 and 2012 in vineyard farms of Sicily and Apulia regions. They show that this effect corresponds to around 10 million hours irregularly worked in the treated regions in each year, or around 21,000 full-time employees. Similarly, we assume that UDW may lead to abnormally low reported payroll costs (labour input) relative to sale revenues given that a part of the worked hours is undeclared and paid in black [envelope wages (Williams and Horodnic 2015)]. Hence, our LTAV proxy is abnormal level of the ratio SSCs to lagged assets (*AbSSCs*), computed as the residuals of Eq. (2) model, simultaneously estimated with Eq. (1) for each of the 36 two-digit NACE

[11] The social security tax base is defined by the Legislative Decree n. 314 of 1997.

[12] Italian accounting regulation for private companies is based on the Italian Civil Code (articles from 2423 to 2429), compliant with 2013/34/UE Directive, and accounting standards issued by Organismo Italiano di Contabilità (Italian Accounting Standard Setter).

[13] Most of SSCs reported as expenses in the income statement are likely to be fully paid given that Italian social security regulation obliges the employer to pay them within the 16th day of the month following the last salary payment period.

industry-year,[14] using a cross-sectional two-stage least square procedure (Cameron and Trivedi 2010). Specifically, the predicted dependent variable of the Eq. (1) is included as covariate in Eq. (2).

$$\frac{PAYR_{i,t}}{ln(TA_{i,t-1})} = \beta_0 + \beta_1 \frac{1}{ln(TA_{i,t-1})} + \beta_2 \frac{SALES_{i,t}}{ln(TA_{i,t-1})} + \beta_3 \frac{\Delta SALES_{i,t}}{ln(TA_{i,t-1})} + \beta_4 \frac{\Delta INV_{i,t}}{ln(TA_{i,t-1})} + \varepsilon_{i,t} \quad (1)$$

$$\frac{SSC_{i,t}}{ln(TA_{i,t-1})} = \beta_0 + \beta_1 \frac{1}{ln(TA_{i,t-1})} + \beta_2 \frac{PAYR_{i,t}}{ln(TA_{i,t-1})} + \varepsilon_{i,t} \quad (2)$$

where the subscripts i and t refer to an individual firm and year, respectively; $SSC_{i,t}$ is expenses for SSCs; $ln(TA_{i,t-1})$ is the natural logarithm of total assets[15]; $SALES_{i,t}$ is the net sales; $\Delta SALES_{i,t}$ is the change in net sales from year $t-1$ to t ($SALES_{i,t}$–$SALES_{i,t-1}$); $\Delta INV_{i,t}$ is change in finished product and work-in-process inventories from year $t-1$ to t[16]; and $PAYR_{i,t}$ is total payroll costs, excluding SSCs. Hence, *AbSSC* is the difference between reported SSC_t (deflated by $ln(TA_{t-1})$) and normal SSCs (NSSCs) corresponding to the fitted values of Eq. (2). Our estimation model in Eq. (1) is consistent with models adopted in several prior accounting studies to estimate normal and abnormal production costs and discretionary expenses (Cai et al. 2018; Hong and Andersen 2011; Ravenda et al. 2019). Furthermore, our proxy may resemble that proposed by Badertscher et al. (2017), to measure income tax avoidance, which is based on the abnormal values of the ratio of income taxes paid to lagged total assets to account for tax avoidance carried out through the underreporting of the accounting income as well as of the taxable income. Nonetheless the estimation procedure and the predictors of their regression model are completely different from those of our LTAV model given that they are more tailored to the peculiarities of corporate income tax.

Finally, we assume that firms engaging more actively in LTAV practices are more likely to exhibit lower and negative values of *AbSSC*, and vice versa. Indeed, lower *AbSSC* may arise from lower SSCs relative to reported payroll costs, as result of a strategic reduction of the tax base, and/or from higher predicted payroll costs, based on Eq. (1), compared to actual payroll costs, which may provide evidence of their underreporting due to the employment of undeclared workers.

Table 2 reports descriptive statistics for the variables included in the Eq. (3) regression model that is estimated cross-sectionally for each year of the period 2008–2016 and whose coefficients on province dummy variables are used as province-level LTAV measures. Variable values are showed for the years 2008 and 2016 as well as the total period 2008–2016. In addition, we carry out the non-parametric comparison tests to determine if the variables significantly differ between 2008 and 2016 (Wilcoxon test) as well as throughout the whole period 2008–2016 (Friedman test[17]). All continuous variables are winsorized at the top and bottom 1% of their distributions to avoid the influence of outliers.

As expected, the mean of variable *AbSSCs* is very close to 0 in each year of the period 2008–2016,[18] consistent with its cross-sectional estimation for each industry-year [see Eq. (2)]. More importantly, the results of Wilcoxon and Friedman tests show that the distribution of *AbSSCs*, and then its median, significantly changes over the examined period, providing evidence of a longitudinal variability in LTAV practices. Specifically, the lower negative median in 2008 compared to 2016 may suggest more widespread LTAV practices across firms in the former year, which may be associated with the start of the global economic downturn.[19] As regards the control variables, the comparison tests also show their significant variability over the analysed period.

[14] We repeat our estimations using three-digit NACE rather than two-digit NACE and the results obtained are qualitatively analogous to those presented.

[15] We deflate all variables by natural logarithm of lagged total assets to address the nonlinearity of the model. An untabulated analysis of residuals shows that this expedient significantly improves the explanatory power of the model.

[16] We include this variable to exclude inventory adjustments from the possible determinants of the regression residuals ultimately affecting our LTAV measure.

[17] We specifically apply the Skillings-Mack (SM) test (Skillings and Mack 1981), which is a generalization of the Friedman test in the presence of missing data. This test may be suitable for our analysis given that several firms do not appear in the observations of all years of the period 2008–2016.

[18] Untabulated *t*tests show that variable *AbSSCs* is not significantly different from 0 in any year of the period 2008–2016.

[19] In 2008, Italian GDP dropped by 1.05% (The World Bank 2018).

Table 2 Descriptive statistics and comparisons of firm-level variables over time

Variables	2008			2016			Total period 2008–2016			Tests	
	Mean	Median	Std	Mean	Median	Std	Mean	Median	Std	Wilcoxon (2016 vs. 2008)	Friedman
Dependent variable											
AbSSCs	0.000	−0.189	2.570	0.000	−0.079	2.633	0.000	−0.131	2.481	***	***
Control variables											
SIZE	6.180	6.303	1.807	6.159	6.240	1.747	6.202	6.301	1.763	***	***
AGE	4.751	2.000	11.331	12.670	10.000	11.042	8.223	5.000	11.563	***	***
LEVER	0.729	0.822	0.305	0.704	0.767	0.304	0.722	0.802	0.302	***	***
CAPINT	0.222	0.079	0.296	0.203	0.064	0.284	0.214	0.070	0.292	***	***
ROA	0.001	0.001	0.110	0.009	0.003	0.120	0.002	0.001	0.111	***	***
LOSS	0.153	0.000	0.360	0.163	0.000	0.369	0.176	0.000	0.380	***	***
GROW	0.530	0.062	1.685	0.255	0.003	1.225	0.298	0.008	1.319	***	***
DAC	0.001	−0.039	0.368	0.000	0.000	0.270	0.001	−0.007	0.296	***	***
AbMATL	−0.500	−8.663	47.765	0.173	−3.319	35.426	−0.054	−4.291	38.803	***	***
AbSERV	−0.589	−5.289	38.628	−0.125	−1.603	29.813	−0.362	−2.681	32.586	***	***
CASHTA	0.107	0.025	0.185	0.117	0.036	0.179	0.105	0.026	0.177	***	***
ETR	0.000	0.076	0.350	0.000	0.068	0.336	0.000	0.070	0.354		***
SD_ROA	0.355	0.111	0.671	0.445	0.185	0.663	0.430	0.177	0.663	***	***
INVENTA	0.311	0.111	0.362	0.265	0.063	0.346	0.297	0.091	0.360	***	***
Number obs.	85,982			116,534			993,606				

The sample full period spans 2008–2016

*, ** and *** denote significance levels at 10%, 5% and 1%, respectively, based on a two-tailed Wilcoxon rank-sum test for the differences in medians between variables in 2008 and variables in 2016, and a two-tailed Friedman test for the differences among variable annual distributions over the whole period 2008–2016

Table 3 presents descriptive statistics of variable *AbSSCs* by Italian region, classified into their higher first-level NUTS (North West; North East; Centre; South; Islands), to produce a first overview of the spatial distribution of LTAV practices across the Italian territory.

It is noteworthy that the means of variable *AbSSCs* by region are all significantly ($p < 0.01$) different from 0, based on two-tailed *t* test, except for region Tuscany. Furthermore, the means are all positive for northern Italian regions, whereas they are all negative for southern Italian regions and islands. These results provide preliminary evidence of the spatial heterogeneity of LTAV across the Italian regions over the period 2008–2016 and, specifically, suggest that LTAV may be on average more intensive in southern Italy, including islands, compared to northern Italy. These outcomes confirm previous studies (Confcommercio Studies Office 2017) suggesting that informal labour is more widespread in southern Italian regions, consistent with the historical dualism between northern and southern Italy in terms of socioeconomic development (Jayet et al. 2010).

Figures 1 and 2 show the maps of province-level LTAV distribution for 2008 and 2016, respectively, based on the results of Eq. (3) regression estimations.

Interestingly, in both years, LTAV is more intensive (lower *AbSSCs*) in southern Italian provinces relative to northern Italian provinces, consistent with the descriptive statistics of variable *AbSSCs* by Italian region shown in previous Table 3. Nonetheless, the picture also varies considerably among northern Italian provinces.

The maps in Figs. 3 and 4 show the spatial distribution by province of the non-EU immigrant worker concentration in agriculture and construction, computed as the ratio of non-EU immigrants employed in construction and agriculture industries to the total people employed in the same industries, in percentage.[20]

Interestingly, the higher non-EU immigrant worker concentration (Figs. 3, 4) and the higher LTAV intensity (Figs. 1, 2) in the Central Italian provinces relative to the Northern Italian provinces may support a positive association between the two variables, consistent with the hypothesis of our study. On the other hand, the higher LTAV intensity in Southern Italy, relative to Central and Northern Italy, is not generally associated with a higher non-EU immigrant

[20] We elaborate the ratio based on data provided by ISTAT.

Table 3 Descriptive statistics of LTAV proxy (*AbSSCs)* by Italian region

Regions (Provinces)	*AbSSCs*				
	N	Mean	Median	Std	*t* test
North West					
Lombardy (Bergamo; Brescia; Como; Cremona; Lecco; Lodi; Mantua; Milan; Monza e della Brianza; Pavia; Sondrio; Varese)	156,900	0.188	−0.129	2.524	***
Piedmont (Alessandria; Asti; Biella; Cuneo; Novara; Turin; Verbano-Cusio-Ossola; Vercelli)	45,516	0.380	−0.121	3.146	***
Liguria (Genova; Imperia; La Spezia; Savona)	17,257	0.288	−0.127	2.892	***
Aosta Valley (Valle d'Aosta/Vallée d'Aoste)	2,382	1.166	−0.079	5.863	***
North East					
Veneto (Belluno; Padua; Rovigo; Treviso; Venice; Verona; Vicenza)	76,877	0.187	−0.128	2.389	***
Emilia-Romagna (Bologna; Ferrara; Forlì-Cesena; Modena; Parma; Piacenza; Ravenna; Reggio nell'Emilia; Rimini)	72,832	0.120	−0.137	2.307	***
Trentino-South Tyrol (Bolzano/Bozen; Trento)	17,445	0.228	−0.116	3.434	***
Friuli-Venezia Giulia (Gorizia; Pordenone; Trieste; Udine)	15,596	0.099	−0.128	2.453	***
Centre					
Lazio (Frosinone; Latina; Rieti; Rome; Viterbo)	158,300	−0.050	−0.133	2.335	***
Tuscany (Arezzo; Florence; Grosseto; Livorno; Lucca; Massa-Carrara; Pisa; Pistoia; Prato; Siena)	59,350	0.012	−0.112	2.503	
Marche (Ancona; Ascoli Piceno; Fermo; Macerata; Pesaro e Urbino)	26,021	−0.052	−0.152	1.712	***
Umbria (Perugia; Terni)	15,349	0.072	−0.121	2.501	***
South					
Campania (Avellino; Benevento; Caserta; Naples; Salerno)	97,041	−0.174	−0.130	2.284	***
Apulia (Bari; Barletta-Andria-Trani; Brindisi; Foggia; Lecce; Taranto)	66,459	−0.234	−0.121	2.613	***
Abruzzo (Chieti; L'Aquila; Pescara; Teramo)	27,289	−0.133	−0.157	2.167	***
Calabria (Catanzaro; Cosenza; Crotone; Reggio di Calabria; Vibo Valentia)	25,921	−0.275	−0.154	2.459	***
Basilicata (Matera; Potenza)	10,663	−0.086	−0.060	2.332	***
Molise (Campobasso; Isernia)	5,099	−0.257	−0.133	2.287	***
Islands					
Sicily (Agrigento; Caltanissetta; Catania; Enna; Messina; Palermo; Ragusa; Syracuse; Trapani)	69,239	−0.343	−0.142	2.392	***
Sardinia (Cagliari; Carbonia-Iglesias; Medio Campidano; Nuoro; Oristano; Sassari)	28,070	−0.236	−0.156	2.234	***

The sample full period spans 2008–2016

*, ** and *** denote significance levels at 10%, 5% and 1%, respectively, based on a two-tailed *t* test for difference from 0 of the means of *AbSSCs* by region

worker concentration in the former relative to the latter. This apparently contradictory result may arise from the fact that the non-EU immigrant worker concentration only considers employees regularly declared. Nonetheless, the higher number of undeclared immigrant workers in Southern Italy may be an important determinant of the higher LTAV in this territory relative to Central and Northern Italy. Furthermore, this analysis, based solely on maps, is univariate and purely descriptive as it does not consider the spatial effects of other important determinants for which we control in the fixed-effect SDM regression that allows examining the effects within the same province throughout the period 2008–2016, while removing any unobserved heterogeneity that is constant over time and may bias the coefficients (Cameron and Trivedi 2010). It should be mentioned that the average non-EU immigrant worker concentration in agriculture and construction by province increases by 49.48%, from 17.13 to 25.61%, over the period 2008–2016, confirming the growing relevance of the non-EU immigration phenomenon in Italy. In addition, the previous maps provide a visual confirmation of the presence of spatial clusters at province level in non-EU immigrant concentrations and LTAV practices. These clusters may lead to a spatial autocorrelation in our data that can be specifically addressed by using an SDM regression for our hypothesis-related estimations.

Finally, Table 4 displays the variance inflation factor (VIF) for all explanatory variables included in the final Eq. (4) regression model as well as Pearson correlations between the same variables. The mean VIF for the full model is 1.83 with individual variable VIFs ranging from 1.02 to 2.84, which is far below the value of 10, a generally accepted maximum threshold to rule out multicollinearity issues in the model (Cameron and Trivedi 2010). These VIF results may also relieve some multicollinearity concerns

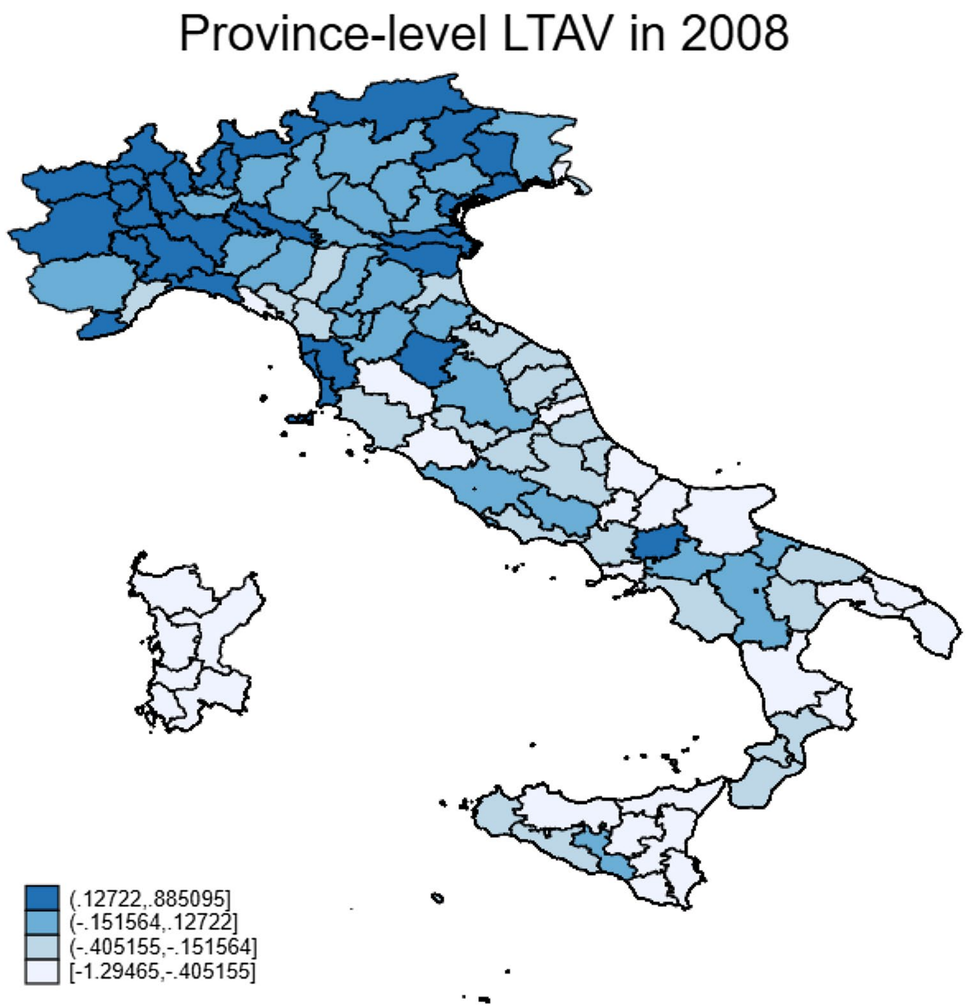

Fig. 1 Spatial distribution of province-level LTAV across Italy in 2008

arising from relatively high correlations coefficients between *UNEMPL* and *IMMIGR* (−0.706) and *UNEMPL* and *HGRSAL* (−0.710).

Hypothesis-Related Spatial Regression Model

Our LTAV proxy based on *AbSSCs* is initially estimated at firm-level. However, to test our main hypothesis, we need to build a measure of LTAV at province level to be regressed on our measure of non-EU immigrant concentration, available for each province, as well as on other province-level macroeconomic control variables that may spatially explain LTAV. Indeed, previous studies (Moulton 1990; Okkerse 2008) show that a regression model including individual-level variables jointly with regional-level variables may be misspecified and bias downward the standard errors of the variables measuring regional characteristics. Furthermore, such a comprehensive regression model could not account for spatial effects. Therefore, following prior research (Easton 2001; Fairlie and Meyer 2003; Gavosto et al. 1999), we adopt a two-step estimation approach to aggregate firm-level LTAV measures at province level. Specifically, in the first step, we run a cross-section regression for each year of the period 2008–2016 with a basic set of firm-level control variables, that previous studies show to be associated with tax avoidance practices within firms (Kim and Zhang 2016; Lanis and Richardson 2012, 2015; Ravenda et al. 2015), and a full set of 107 regional dummies (*PROVINCE*) by province and industry dummies (*INDUSTRY*) by three-digit NACE codes. We omit the province of Rome[21] as a base regional dummy in the model. Hence, the coefficients on province dummy variables provide an average measure of LTAV for that province relative to the province of Rome, corrected for differences in the firm group composition among provinces. Therefore, we estimate the following Eq. (3) model, whose control variables (*CONTROLS*) are defined in the "Appendix":

[21] We repeat the analysis by omitting Milan or Florence and the results obtained are qualitatively analogous to those presented.

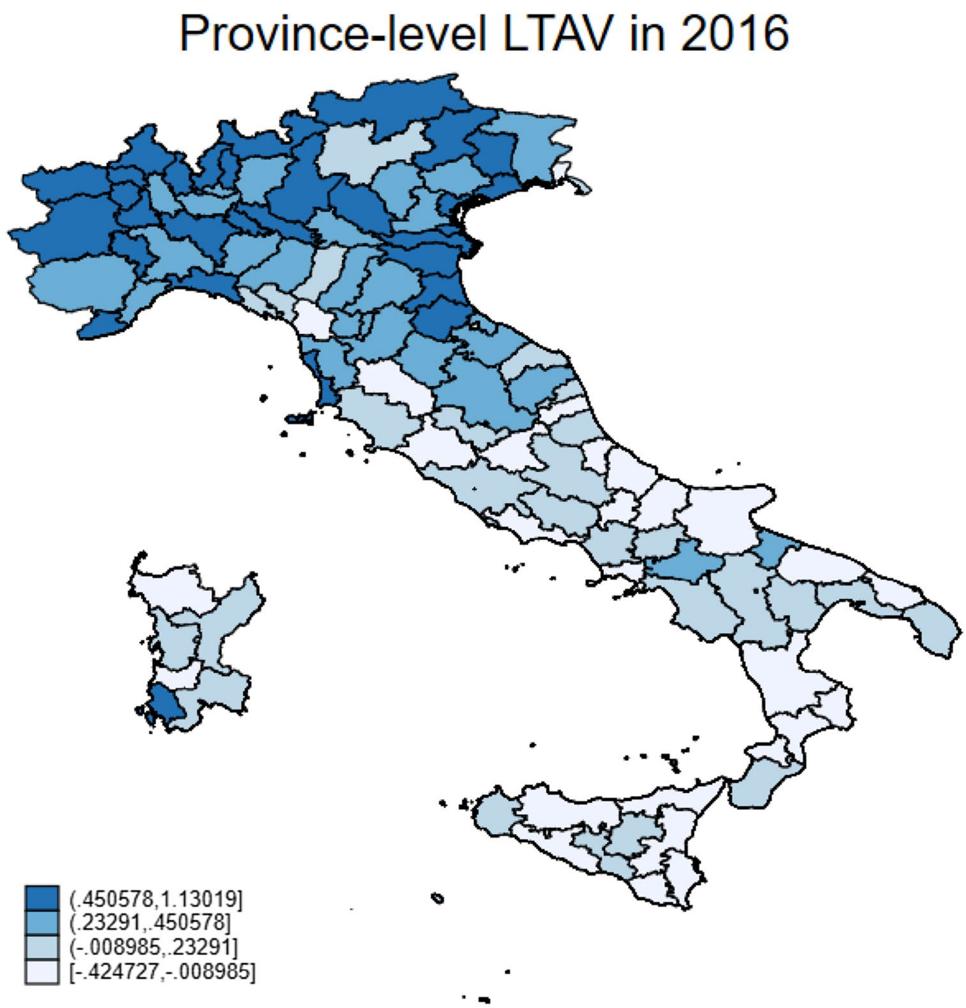

Fig. 2 Spatial distribution of province-level LTAV across Italy in 2016

$$AbSSCs_{i,t} = \beta_0 + \sum_r \beta_r PROVINCE^r_{i,t} + \sum_k \beta_k CONTROLS^k_{i,t} + \sum_s \beta_s INDUSTRY^s_{i,t} + \varepsilon_{i,t} \tag{3}$$

Subsequently, in the second step, we regress the estimated coefficients of the province dummies on a measure of non-EU immigrant concentration (*IMMIGR*) for each province and year, representing our hypothesis-related independent variable, and on province-level controls that may be associated with LTAV such as unemployment rate, population density, GDP growth, reported crimes, hourly gross wage.

Importantly, to account for spatial interdependence among province-level observations, that, if unaddressed, may bias the estimations (Anselin 2010), we adopt a Spatial Durbin Model (SDM) panel fixed-effect regression (Elhorst 2014a; LeSage and Pace 2009). SDM is a global spillover specification. This means that changes in one region spill over into not only the neighbouring regions, but also the neighbours of the neighbours, and so on, such that "a new long-run steady state equilibrium arises" (LeSage 2014). Therefore, this approach allows us to exploit and capture both the effect of non-EU immigrant concentration in a province on LTAV in the same province (direct effects), and the effect of non-EU immigrant concentration in a province on LTAV in the neighbouring provinces (indirect or spillover effects). Specifically, SDM, introduced by LeSage and Pace (2009), includes spatial lags of both the dependent variables and explanatory variables. Spatially lagged variables contain for each regional-level observation the weighted sum of the corresponding variable values of neighbouring regions and they are practically computed by multiplying each variable by a spatial weight matrix (W). W is a diagonal matrix of dimension $n \times n$, where n is the number of observations, and each observation represents a location. Non-zero elements in the i, j row and column positions of the matrix W,

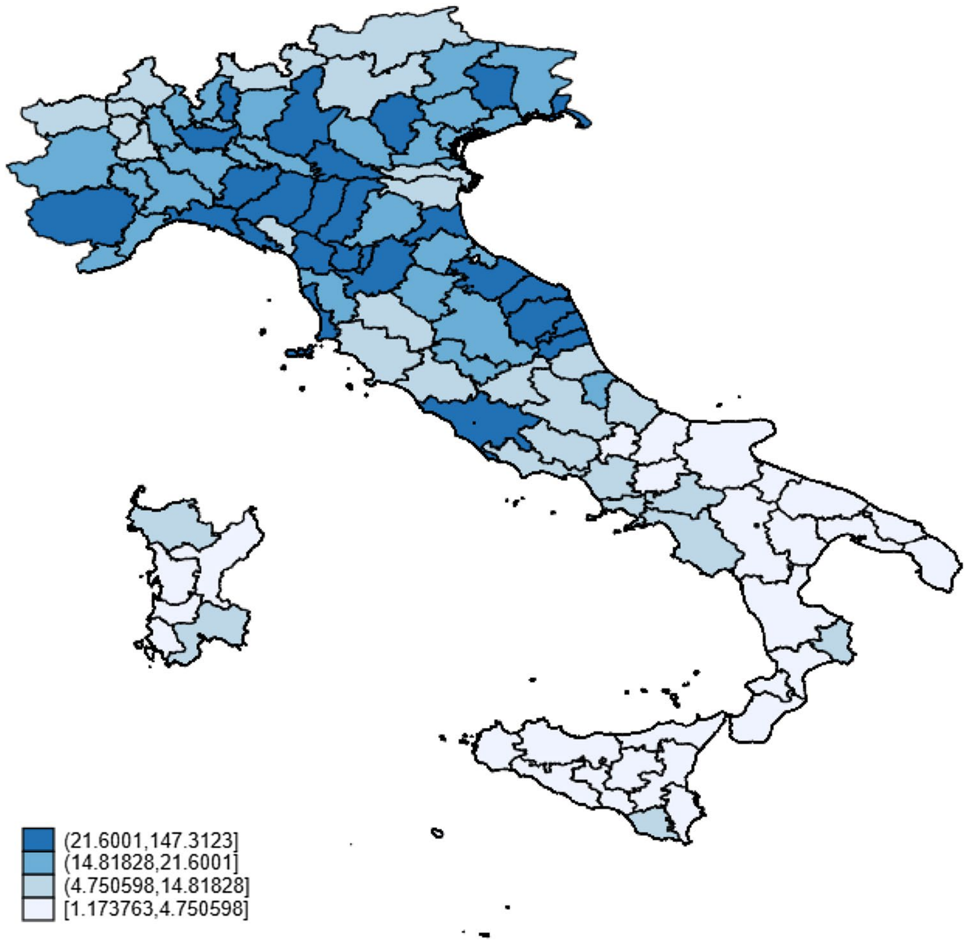

Fig. 3 Spatial distribution of non-EU immigrant worker concentration in construction and agriculture industries in 2008

based on distance metrics, indicate that region/observation *j* is a neighbour to *i* (LeSage 2014).

LeSage and Pace (2009) assert that SDM offers several advantages over other spatial regression models that, for example, only include a spatial autoregressive process in the error term [spatial error model (SEM)] or a spatially lagged dependent variable as an additional explanatory variable (spatial autoregressive model (SAR)). Specifically, SDM produces unbiased coefficient estimates even when the true data-generating process (DGP) is simply a SAR or a SEM. Therefore, in the presence of uncertainty about the form of spatial dependence in the underlying DGP, SDM is always the best option. Furthermore, SDM does not impose any prior restrictions on the magnitude of spillover effects, which can also be different for different explanatory variables (Elhorst 2014a). Finally, SDM is preferable over alternative spatial regression models given that ignoring spatial dependence in the dependent variable and/or in the independent variables, if present, will lead to biased and inconsistent coefficient estimates for the variables included in the regression equation. In contrast, ignoring spatial dependence in the disturbances will only cause a loss of efficiency (Elhorst 2010).

We consider that SDM methodology may be appropriate for our study due to the importance of immigration networks, the plausible assumption that immigrants resident in a province may commute to the neighbouring provinces within certain distance limits, and the fact that a province may be influenced by its neighbouring provinces in several economic, demographic and social aspects, including LTAV practices (Bastida et al. 2013). For example, spatial clusters in terms of LTAV practices may arise from an emulation behaviour of the neighbour adopted by firms engaging in LTAV to compete through the reduction of labour costs. Therefore, it is very likely the presence of spatial spillover effects across provinces in terms of immigrant concentration impact on LTAV, as well as the existence of a strong spatial autocorrelation of non-EU immigrant concentrations and LTAV practices across regions, so that we can expect random terms of our regression model to exhibit spatial autocorrelation (Jayet et al. 2010). In this respect, the validity of our assumptions justifying the adoption of panel fixed-effect SDM for our analysis is supported by several statistical tests, on the presence of either a spatially lagged dependent variable

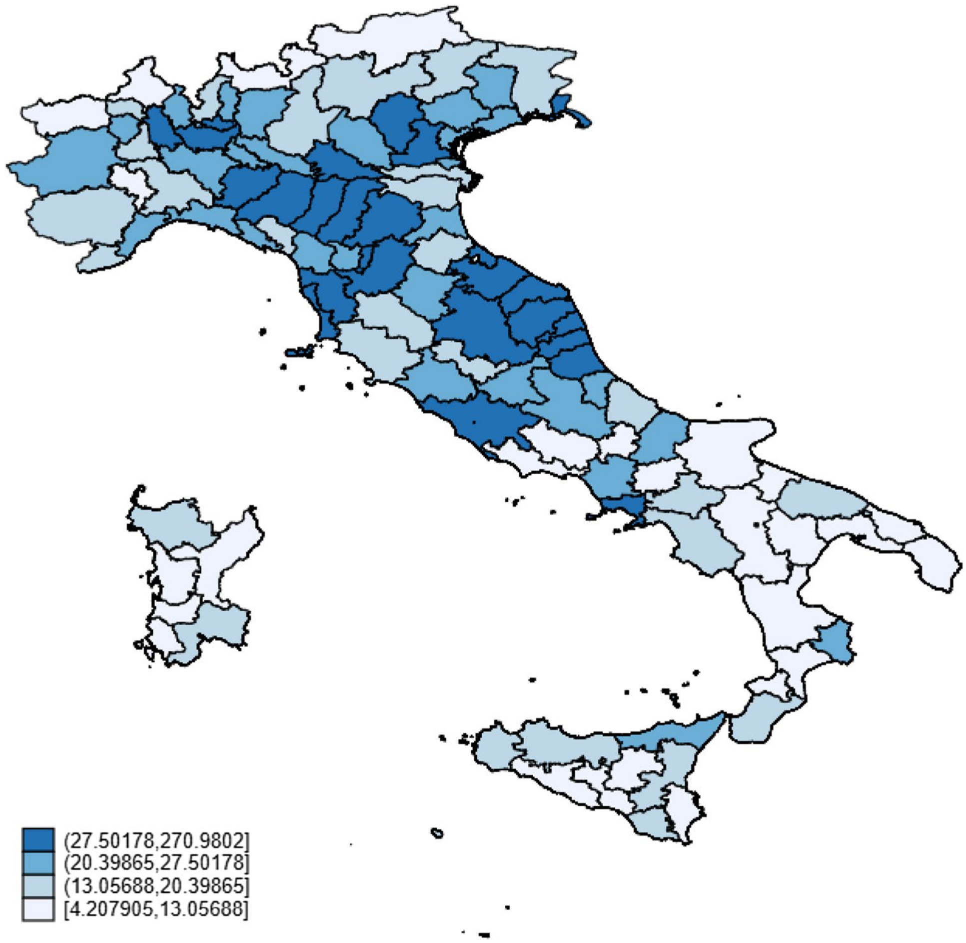

Fig. 4 Spatial distribution of non-EU immigrant worker concentration in construction and agriculture industries in 2016

Table 4 Variable VIF and Pearson correlations

Variables	VIF	*IMMIGR*		*DENSITY*		*CRIME*		*UNEMPL*		*HGRSAL*		Δ*GDP*
IMMIGR	2.49	1										
DENSITY	1.14	0.212	***	1								
CRIME	1.34	0.453	***	0.301	***	1						
UNEMPL	2.84	−0.706	***	−0.045		−0.244	***	1				
HGRSAL	2.14	0.586	***	0.146	***	0.242	***	−0.710	***	1		
Δ*GDP*	1.02	0.039		0.018		−0.008		−0.074	**	0.134	***	1
Mean VIF	1.83											

*, ** and *** denote significance levels at 10%, 5% and 1%, respectively, based on a two-tailed test. Variables are defined in the "Appendix"

and/or spatially lagged residuals, that we present in the results section of this paper.

In summary, our second step SDM regression, used to test our hypothesis, is the following:

$$LTAV\_PROV_{i,t} = \rho WLTAV\_PROV_{i,t} + \beta_1 IMMIGR_{i,t} + \sum_k \beta_k CONTROLS^k_{i,t} + \theta_1 WIMMIGR_{i,t} + \sum_k \theta_k WCONTROLS^k_{i,t} + u_i + v_{i,t} \tag{4}$$

where LTAV_PROV$_{i,t}$ is LTAV, in terms of *AbSSCs*, at province i level in year t, measured as the estimated coefficients on province dummies in Eq. (3); W is the spatial weight matrix, with elements equal to the reciprocal of distance

Table 5 Regression estimations of normal and abnormal SSCs

Variables	$SSC_{i,t}/ln(TA_{i,t-1})$								
	Total sample			Construction			Agriculture		
	Coef.	*t* stat	*p* val.	Coef.	*t* stat	*p* val.	Coef.	*t* stat	*p* val.
$1/ln(TA_{i,t-1})$	2.345	4.90	0.000	3.025	5.20	0.000	0.304	3.49	0.008
$[PAYR_{i,t}/ln(TA_{i,t-1})]$	0.372	31.89	0.000	0.410	89.20	0.000	0.258	152.60	0.000
Intercept	− 1.302	− 6.34	0.000	− 1.559	− 6.09	0.000	− 0.529	− 30.28	0.000
Mean R^2	0.573			0.616			0.444		
Mean *F*	616		0.000	5,277		0.000	69,952		0.000
Mean obs.	27,600			32,833			11,901		
Total obs.	993,606			886,494			107,112		
N. Industry-years	36			27			9		

The *p* values are two-tailed. The coefficients and R^2 are the mean values of coefficients and R^2 of cross-sectional estimations across 36 two-digit NACE industry-years. The *t* statistics are calculated using the standard error of the related mean coefficient across industry-years. $ln(TA_{i,t-1})$ is the natural logarithm of lagged total assets; $SSC_{i,t}$ is social security contribution expenses; $[PAYR_{i,t}/ln(TA_{t-1})]$ is predicted payroll costs deflated by $ln(TA_{i,t-1})$ resulting from the first-stage regression in Eq. (1). Agriculture industry includes firms with NACE code 01 (Crop and animal production, hunting and related service activities); Construction industry includes firms with NACE codes: 41(Construction of buildings), 42 (Civil engineering), and 43 (Specialized construction activities)

between provinces (before normalization),[22] by which covariates are premultiplied to compute their spatially lagged version; *IMMIGR*, the independent variable of interest, is the non-EU immigrant concentration, computed as the fraction of non-EU residents per 1000 residents in each province and year, restricted to the working-age population (between 18 and 59 years of age)[23]; u_i are unobserved province fixed-effects arising from the panel data nature of our sample. The rest of province-level control variables (*CONTROLS*) are defined in the "Appendix". It is noteworthy that before estimating the Eq. (4) model both *IMMIGR* and *CONTROLS* variables are spatially differentiated from the reference province of Rome[24] for each year, consistent with the reference to Rome of coefficients on province dummies.

Regression Results and Analyses

Estimations of Firm-Level and Province-Level LTAV

Table 5 shows the results of Eq. (2) regression estimations, whose fitted values are NSSCs and residuals (*AbSSCs*) are used as firm-level LTAV proxy. Based on the Fama and MacBeth's (1973) methodology, the reported coefficients and R^2 are mean values of cross-sectional estimations across 36 two-digit NACE industry-years. Furthermore, the significance levels of the coefficients are estimated using the standard errors of the coefficients across industry-years. Given the industry heterogeneity of the sample, we also report the results separately for construction and agriculture. Finally, to mitigate the influence of outliers, all variables of Eqs. (1) and (2) are winsorized at the top and bottom 1 percent of their distributions, before running the estimations.

It should be noted that all the estimated regressions are significant at the 0.01 level according to the F tests. In addition, the average coefficient on variable $[PAYR_{i,t}/ln(TA_{i,t-1})]$, the fitted value of Eq. (1) model, is positive and significant ($p < 0.01$), as expected. More importantly, an average R^2 of 0.573 indicates that the explanatory power of the model is very satisfactory. Specifically, untabulated values of R^2, across the 36 estimations, range from a minimum of 0.436 for NACE code 01 (Crop and animal production, hunting and related service activities) and year 2008 to a maximum of 0.753 for NACE code 42 (Civil engineering) and year 2008. Indeed, this model fit significantly improves the R^2 of 0.29 recorded by the different regression model used by Ravenda et al. (2015) in their first attempt to estimate the abnormal level of SSCs as a measure of LTAV. Furthermore, it outperforms the goodness of fit of other regression models adopted to estimate income tax avoidance through abnormal book-tax differences (Desai and Dharmapala 2009; Kim and Zhang 2016) and abnormal cash taxes paid to lagged total assets (Badertscher et al. 2017), whose R^2 are below 0.30. Finally, the estimation results by industry show that mean R^2

[22] We adopt a threshold distance of 57.14 km, beyond which the elements of *W* are set to 0. This is the best threshold distance based on the results of a Lagrange multiplier test. *W* is spectrally normalized so that its largest eigenvalue is 1. The choice of the spatial weight matrix is justified based on a theoretical argument on the mobility pattern and possibilities of immigrants.

[23] This age restriction is also motivated by the related data availability from the Italian Institute of Statistics (ISTAT).

[24] Using Milan or Florence as a reference province, rather than Rome, leads to qualitatively similar results to those presented in this study.

Table 6 Regression estimations of province-level LTAV measures

Variables	*AbSSCs*			
	Pred. sign	Coef.	*t* stat	*p* val.
SIZE	–	−0.041	−13.26	0.000
AGE	?	0.006	6.66	0.000
LEVER	–	−0.102	−5.27	0.001
CAPINT	+	−0.161	−6.58	0.000
ROA	?	0.017	0.41	0.695
LOSS	?	−0.083	−6.80	0.000
GROW	+	0.006	1.00	0.349
DAC	+	−0.050	−5.20	0.001
AbMATL	–	−0.001	−4.46	0.002
AbSERV	–	0.000	−1.33	0.221
CASHTA	–	−0.063	−3.83	0.005
ETR	–	−0.106	−5.93	0.000
SD_ROA	–	−0.024	−4.84	0.001
INVENTA	?	−0.066	−3.34	0.010
PROVINCE (dummies)		Yes		
INDUSTRY (dummies)		Yes		
Mean R^2		0.041		
Mean F		492		0.000
Mean observations		110,401		
Total observations		993,606		
Number of years		9		

The *p* values are two-tailed. The coefficients and R^2 are the mean values of coefficients and R^2 of cross-sectional estimations across 9 years. The *t* statistics are calculated using the standard error of the related mean coefficient across years. Variables are defined in the "Appendix"

for construction (0.616) is higher than mean R^2 for agriculture (0.444), providing evidence that the unexplained variation of paid SSCs, which may also be attributed to unobserved LTAV practices, is greater in the agriculture industry.

Table 6 presents the results of Eq. (3) regression estimations following the Fama and MacBeth's (1973) procedure. Specifically, the reported coefficients and R^2 are mean values of cross-sectional estimations across the 9 years of the period 2008–2016. Therefore, the significance levels of the coefficients are computed using the standard errors of the coefficients across years.

It should be noted that all the estimated regressions are significant at the 0.01 level according to the *F* tests. As regards the control variables, most of them are significant at conventional levels ($p<0.05$), except variables *ROA*, *GROW*, and *AbSERV*. Furthermore, their sign is mostly consistent with our predictions, made based on previous studies on labour and income tax avoidance (Kim and Zhang 2016; Lanis and Richardson 2012, 2015; Ravenda et al. 2015), with the relevant exception of variable *CAPINT* (capital intensity). Specifically, its negative sign suggests that more capital-intensive firms are more likely to engage in LTAV. These firms may also be more indebted and incur higher interest expenses as well as higher depreciation expenses. Therefore, they could underreport payroll costs to avoid payments of *SSCs*, without significantly increasing the accounting income that in Italy is the basis for the computation of the taxable income[25] (Gavana et al. 2013). It should be mentioned that, as almost all the firms in our sample are not listed on the stock exchange, their tax minimization incentives, through the underreporting of earnings, may prevail over capital market considerations that are commonly more relevant for listed firms and may, conversely, lead to upward manage earnings (Coppens and Peek 2005; Marques et al. 2011).

Hypothesis-Related Regression Results

Before estimating the Eq. (4) SDM regression, we present in Table 7 the estimation results of panel data regressions including simultaneously the firm-level variables of Eq. (3) and the regional-level variables of Eq. (4),[26] with standard errors clustered at the province level. Although these estimations cannot explicitly reveal any spatial spillover effect, they may provide a first indication of the relationship between firm-level LTAV and the concentration of non-EU immigrants in each province.

In the estimation 1, we include firm fixed-effects and omit *PROVINCE and INDUSTRY* dummy variables that are dropped by the firm fixed-effect estimation for being time-invariant within each firm. Conversely, in the estimation 2, we omit firm fixed-effects and include *PROVINCE* and *INDUSTRY* dummy variables. Interestingly, in both estimations the coefficient on the hypothesis-related variable *IMMIGR* is negative and significant ($p<0.01$), suggesting that the concentration of non-EU immigrants in each province and year is positively associated with the intensity of LTAV of the firms located in the same province in the same year. By controlling for industry, province and year in the estimation 2, we may also account for potential effects of differences in the employer's perception of labour inspection risks that may affect LTAV propensity. Indeed, labour inspection plans are mostly defined by public authorities every year at industry and province levels (Di Porto et al. 2018).

[25] The Italian Tax Code (Presidential decree 22, December 1986) sets the derivation principle in the Article 83, stating that taxable income is computed based on the accounting income that should only be adjusted, when accounting standards differ from tax rules.

[26] Regional-level variables are not spatially differentiated like Eq. (4) SDM regression.

Table 7 Regressions of LTAV proxy at firm-level

Variables	*AbSSCs*					
	1			2		
	Coef.	*t* stat	*p* val.	Coef.	*t* stat	*p* val.
Variable of interest						
IMMIGR	−0.002	−6.03	0.000	−0.002	−4.67	0.000
Control variables						
SIZE	−0.067	−13.36	0.000	−0.040	−18.81	0.000
AGE	−0.041	−8.03	0.000	0.008	30.87	0.000
LEVER	0.038	2.44	0.015	−0.140	−12.84	0.000
CAPINT	0.025	1.21	0.225	−0.234	−19.78	0.000
ROA	0.047	1.99	0.047	0.063	2.21	0.027
LOSS	0.002	0.30	0.767	−0.062	−8.37	0.000
GROW	0.002	4.10	0.000	−0.003	−5.51	0.000
DAC	−0.034	−4.56	0.000	−0.079	−7.81	0.000
AbMATL	0.000	−1.69	0.091	−0.001	−13.22	0.000
AbSERV	0.001	7.39	0.000	0.000	−4.91	0.000
CASHTA	−0.035	−1.68	0.092	−0.095	−4.77	0.000
ETR	−0.045	−7.46	0.000	−0.098	−13.42	0.000
SD_ROA	0.006	1.17	0.241	−0.027	−5.86	0.000
INVENTA	0.101	6.03	0.000	−0.001	−0.11	0.910
DENSITY	−0.001	−9.86	0.000	−0.002	−9.46	0.000
CRIME	0.240	6.56	0.000	0.205	3.79	0.000
UNEMPL	−0.001	−0.64	0.522	−0.002	−1.19	0.233
HGRSAL	0.133	14.32	0.000	0.148	11.04	0.000
ΔGDP	−0.007	−5.47	0.000	−0.009	−4.28	0.000
FIRM Fixed-effects	Yes			No		
YEAR (dummies)	Yes			Yes		
PROVINCE (dummies)	No			Yes		
INDUSTRY (dummies)	No			Yes		
Intercept	−2.285	−9.57	0.000	−2.904	−7.96	0.000
Number of obs.	993,606			993,606		
R^2	0.0015			0.0616		
F	30.57 ($p<0.001$)			310.52 ($p<0.001$)		

The sample period is from 2008 to 2016. The *t* statistics are based on standard errors clustered by province. The *p* values are two-tailed. Variables are defined in the "Appendix"

Nonetheless, the decision to estimate our Eq. (4) regression model using a spatial econometric approach (SDM) is supported not only by theoretical arguments, but also by the results of several statistical tests. Specifically, we first employ the Moran's I test (Kelejian and Prucha 2001), based on the residuals of the OLS model, to determine whether a spatial autocorrelation is present in our data and then a spatial model, rather than a non-spatial model, is appropriate. In addition, we follow the specific-to-general approach, suggested by Elhorst (2010), consisting in estimating first a non-spatial linear regression model (OLS model) and then testing whether the spatial autoregressive model (SAR) or the spatial error model (SEM) is more appropriate to describe the data. For this purpose, we use Lagrange multiplier (LM) tests for a spatially lagged dependent variable (LM Spatial Lag) and/or for spatial error autocorrelation (LM Spatial Error), as well as the robust LM tests which test for a spatially lagged dependent variable in the local presence of spatial error autocorrelation and for spatial error autocorrelation in the local presence of a spatially lagged dependent variable (Elhorst 2014a). These tests are based on the residuals of the OLS model and follow a Chi squared distribution with one degree of freedom (Anselin et al. 1996; Burridge 1981). Table 8 shows the results of all these tests.

Importantly, the tests reject the null hypotheses of no spatial autocorrelation in the error (Moran's I and LM Spatial Error tests) and no spatial autocorrelation in the spatial lagged dependent variable (LM Spatial Lag tests) below the 1% level, suggesting that a spatial model, rather than the OLS model, is the appropriate model to use. In this scenario,

Table 8 Tests for spatial autocorrelation

Tests	Stat.	Stat. value	p val.
Ho: error has no spatial autocorrelation			
Moran's I	$\chi^2(1)$	19.75	0.000
LM spatial error	$\chi^2(1)$	48.242	0.000
LM spatial error (Robust)	$\chi^2(1)$	34,800,000	0.000
Ho: spatial lagged dependent variable has no spatial autocorrelation			
LM Spatial Lag	$\chi^2(1)$	72.052	0.000
LM Spatial Lag (Robust)	$\chi^2(1)$	34,800,000	0.000
Ho: no general spatial autocorrelation			
LM SAC (LM Error + LM Lag (Robust))	$\chi^2(2)$	34,800,000	0.000
LM SAC (LM Lag + LM Error (Robust))	$\chi^2(2)$	34,800,000	0.000
Ho: $\theta = 0$			
Wald test: SDM versus SAR	$\chi^2(6)$	18.150	0.006
Ho: $\theta + \rho\beta = 0$			
Wald test: SDM versus SEM	$\chi^2(6)$	43.260	0.000

All tests are performed using the inverse distance spatial weight matrix (W) with a threshold distance of 57.14 km and spectrally normalized so that its largest eigenvalue is 1. Moran's I test is computed for the final year of the analysis (2016)

Table 9 SDM fixed-effect regression of LTAV at province level

Explanatory variables	Dependent variable: *LTAV_PROV*								
	Direct effect			Indirect effect			Total effect		
	Coef.	z stat	p val.	Coef.	z stat	p val.	Coef.	z stat	p val.
*W*LTAV_PROV* (ρ)	0.5630	8.44	0.000						
Variable of interest									
IMMIGR	−0.0012	−2.60	0.009	−0.0019	−3.36	0.001	−0.0031	−4.32	0.000
Control variables									
DENSITY	−0.0024	−12.67	0.000	−0.0004	−1.55	0.122	−0.0027	−12.74	0.000
CRIME	0.2256	4.07	0.000	0.2522	2.92	0.004	0.4778	4.71	0.000
UNEMPL	0.0038	2.03	0.042	0.0112	2.64	0.008	0.0150	3.28	0.001
HGRSAL	0.0838	4.43	0.000	0.0280	1.62	0.105	0.1118	6.58	0.000
ΔGDP	−0.0090	−3.42	0.001	0.0004	0.18	0.860	−0.0085	−3.05	0.002
PROVINCE FE	Yes								
Number of obs.	972								
Number of groups	108								
Obs. per group	9								
Log-likelihood	860.693								
R^2 (within)	0.567								
Wald $\chi^2(13)$	1322.69 ($p < 0.001$)								

The sample period is from 2008 to 2016. The p values are two-tailed. Variables are defined in the "Appendix"

J. LeSage and Pace (2009) recommend to first consider the SDM. Therefore, we estimate the fixed-effects[27] panel data SDM of Eq. (4) and, following the general-to-specific approach (Elhorst 2014b), we determine whether SDM is actually a better choice than the SAR and SEM models by testing the hypotheses: H_0: $\theta = 0$ and H_0: $\theta + \rho\beta = 0$. Specifically, the first hypothesis examines whether the SDM can be simplified to the SAR, and the second hypothesis whether it can be simplified to the SEM (Burridge 1981). If both hypotheses are rejected, then the SDM best describes the data (Elhorst 2014b). Table 8 shows the results of the Wald tests used to corroborate these hypotheses. As both

[27] The Hausman test [$\chi^2(12) = 109.65$; $p < 0.01$] suggests that the fixed-effect specification is more adequate than the random effect.

Table 10 SDM 2SLS fixed-effect regression of LTAV at province level

Explanatory variables	*IMMIGR* (1st stage eq.)			*LTAV_PROV* (2nd stage eq.)								
				Direct effect			Indirect effect			Total effect		
	Coef.	*t* stat	*p* val.	Coef.	*z* stat	*p* val.	Coef.	*z* stat	*p* val.	Coef.	*z* stat	*p* val.
*W*LTAV_PROV* (ρ)				0.506	4.43	0.000						
Variable of interest												
Pred_IMMIGR				−0.003	−3.96	0.000	−0.003	−2.26	0.024	−0.007	−3.71	0.000
Control variables												
DENSITY	0.003	1.22	0.225	−0.010	−3.63	0.000	0.006	1.25	0.210	−0.004	−0.76	0.448
CRIME	10.227	4.93	0.000	0.630	5.89	0.000	0.047	0.28	0.781	0.677	3.50	0.000
UNEMPL	0.221	0.76	0.449	0.010	2.79	0.005	0.018	2.26	0.024	0.027	2.88	0.004
HGRSAL	−0.256	−0.17	0.866	−0.138	−1.89	0.059	0.243	1.81	0.071	0.105	0.83	0.405
ΔGDP	0.280	0.97	0.335	−0.012	−2.51	0.012	−0.013	−1.64	0.101	−0.025	−3.52	0.000
Lag6_IMMIGR	1.146	11.38	0.000									
PROVINCE FE	No			Yes								
Number of obs.	324			324								
Number of groups				108								
Obs. per group				3								
Log-likelihood				417.093								
R^2 (within)	0.9424			0.500								
Wald $\chi^2(13)$				45.54 ($p<0.001$)								
F	550.24 ($p<0.001$)											

The sample period is from 2014 to 2016. The t statistics are based on standard errors clustered by province. The *p* values are two-tailed. *Lag6_IMMIGR* is 6-year lag of variable *IMMIGR*; *Pred_IMMIGR* is predicted *IMMIGR* from 1st stage equation. The rest of variables are defined in the "Appendix"

hypotheses are rejected below the 1% level, we can conclude that the SDM best describes the data (Elhorst 2014b). Finally, Table 9 shows the estimation results of the Eq. (4) fixed-effect panel data SDM.

First, it should be noted that the estimated regression is significant at the 0.01 level according to the Wald χ^2 test. In addition, the spatial coefficient, ρ, on (*W*LTAV_PROV*), displayed in the first row of Table 9, is positive and highly significant ($p<0.001$), suggesting that our estimation strategy is appropriate. Specifically, LTAV intensity in a province is positively associated with LTAV intensity in the neighbouring provinces because of spatially clustered determinants of LTAV, including social, cultural and economic factors, that may lead labour-intensive neighbouring firms to compete through LTAV practices aiming to reduce their labour costs. Turning to the explanatory variables, the SDM methodology allows the estimation of their direct effect (feedback), indirect effect (spillover), and total effect as the sum of the previous two effects (LeSage and Pace 2009). In our specific context, the direct effect records the impact of an explanatory variable in a specific province on LTAV in the same province, whereas the indirect effect measures the impact of the explanatory variable on LTAV in surrounding provinces. As regards our variable of interest *IMMIGR*, both its direct effect and indirect effect are negative and significant ($p<0.01$), suggesting that the non-UE immigrant concentration in a province is positively associated with the level of LTAV in that province and in the neighbouring provinces, consistent with the main hypothesis of our study.

Turning to the other control variables, all their coefficients (direct and total effects) are significant at conventional levels although only variables *CRIME* and *UNEMPL* show significant indirect effects. Specifically, it is noteworthy that positive macroeconomic trends suggested by lower unemployment *(UNEMPL)* and higher GDP growth *(ΔGDP)* are associated with higher LTAV. Indeed, in these scenarios, greater opportunities for higher quality jobs for nationals may increase the availability of more precarious and low-qualified jobs, mostly unattractive to nationals, for immigrants, especially in labour-intensive agriculture and construction industries that mostly include those types of work. To the extent that our LTAV proxy reflects the employment of UDW, our results contradict some prior studies finding a positive association between UDW and unemployment even in Mediterranean countries such as France, Spain and Greece (Buehn 2012; Dell'Anno et al. 2007; Haigner et al. 2013). The fact that our study is specifically focused on agriculture and construction industries and that our LTAV definition is wider than UDW, by including more legitimate practices, may partially explain our different results. Finally,

Table 11 SDM fixed-effect regression of LTAV excluding provinces of *Mezzogiorno*

Explanatory variables	Dependent variable: *LTAV_PROV*								
	Direct effect			Indirect effect			Total effect		
	Coef.	*z* stat	*p* val.	Coef.	*z* stat	*p* val.	Coef.	*z* stat	*p* val.
*W*LTAV_PROV* (ρ)	0.4298	5.82	0.000						
Variable of interest									
IMMIGR	−0.0016	−3.84	0.000	−0.0020	−4.08	0.000	−0.0036	−5.71	0.000
Control variables									
DENSITY	−0.0020	−9.95	0.000	−0.0008	−3.53	0.000	−0.0028	−13.30	0.000
CRIME	0.0977	1.89	0.059	0.3105	3.55	0.000	0.4082	4.07	0.000
UNEMPL	0.0040	1.39	0.164	0.0103	2.05	0.040	0.0143	2.43	0.015
HGRSAL	0.1048	5.33	0.000	0.0087	0.50	0.615	0.1135	7.08	0.000
ΔGDP	−0.0033	−1.20	0.229	0.0006	0.23	0.814	−0.0027	−1.05	0.294
PROVINCE FE	Yes								
Number of obs.	621								
Number of groups	69								
Obs. per group	9								
Log-likelihood	699.982								
R^2 (within)	0.690								
Wald $\chi^2(13)$	78.12 ($p<0.001$)								

The sample period is from 2008 to 2016. The *p* values are two-tailed. Variables are defined in the "Appendix"

within their socioeconomic context, the outcomes of our study provide empirical support for structuralist and marginalization theories predicting that marginalized and more disadvantaged populations such as the immigrants are more likely to be involved in the informal labour market and being victims of labour exploitation practices.

Additional Analyses and Robustness Checks

If immigrants are attracted to provinces where they have more opportunities to work informally, and then LTAV is higher, an endogeneity problem, in the form of reverse causality between our LTAV proxy and non-EU immigrant concentration variable (*IMMIGR*), may arise and bias our estimations. Specifically, this selective settlement would lead to an upwardly biased estimate of the effects of immigrants' concentration on province-level LTAV (Okkerse 2008). To address this concern, an instrumental variable (IV), highly correlated with endogenous *IMMIGR* but uncorrelated with LTAV (exogenous instrument), is needed. Previous studies mostly use as an instrument the immigrant concentration at some time in the past, under the assumption that immigrants tend to settle where they can find support from previously established clusters and networks of immigrants with the same cultural and linguistic background as themselves (Dustmann et al. 2017). On the other hand, pre-existing immigrant concentrations are unlikely to be correlated with current economic shocks (pull factors), if measured with a sufficient time lag, unless local economic shocks are strongly persistent (Okkerse 2008). Our endogeneity concern is confirmed by the results of the Durbin–Wu–Hausman test for endogeneity ($F(1,209)=8.58$), which lead to the rejection of the null hypothesis of exogeneity of variable *IMMIGR* with a *p* value <0.01, thus confirming the need to account for endogeneity in our model. Therefore, we use the 6-year lag non-EU immigrant concentration (*Lag6_IMMIGR*) as an instrument for contemporary variable *IMMIGR*. This new variable has a very high correlation (0.968) with *IMMIGR* as it is needed of an instrument to be valid. We then estimate a two-stage least squares (2SLS) SDM panel regression (Anselin and Lozano-Gracia 2008) by including in the second-stage SDM panel regression the predicted value of *IMMIGR* (*Pred_IMMIGR*) based on a first-stage regression of *IMMIGR* on the instrumental variable *Lag6_IMMIGR* and the other control variables of the SDM. Table 10 shows the results of our estimations.

The results of the first-stage regression show that the instrumental variable *Lag6_IMMIGR* is relevant, namely it is a strong and significant determinant of the endogenous variable *IMMIGR*. Indeed, the coefficient on *Lag6_IMMIGR* is positive, as expected, statistically significant ($p<0.01$), and the *F* test on the significance of the instrument is equal to 129.59, far above the value of 10, the minimum relevance threshold typically used in the academia (Cameron and Trivedi 2010; Staiger and Stock 1997). Regarding the second-stage SDM panel regression, both direct and indirect effects on variable *Pred_IMMIGR* are negative and significant ($p<0.01$ and $p<0.05$, respectively), confirming the

Table 12 SDM fixed-effect regression of LTAV without threshold distance for spatial weight matrix of dependent variable

Explanatory variables	Dependent variable: *LTAV_PROV*								
	Direct effect			Indirect effect			Total effect		
	Coef.	*z* stat	*p* val.	Coef.	*z* stat	*p* val.	Coef.	*z* stat	*p* val.
*W*LTAV_PROV* (ρ)	0.5369	8.99	0.000						
Variable of interest									
IMMIGR	−0.0010	−2.04	0.042	−0.0031	−4.07	0.000	−0.0041	−4.35	0.000
Control variables									
DENSITY	−0.0016	−7.92	0.000	−0.0012	−4.03	0.000	−0.0028	−9.98	0.000
CRIME	0.1732	3.07	0.002	0.2477	2.23	0.026	0.4209	3.31	0.001
UNEMPL	0.0046	2.40	0.016	0.0181	3.03	0.002	0.0227	3.65	0.000
HGRSAL	0.0467	2.34	0.019	0.0377	1.96	0.050	0.0844	3.69	0.000
ΔGDP	−0.0076	−2.82	0.005	−0.0031	−1.01	0.311	−0.0107	−2.95	0.003
PROVINCE FE	Yes								
Number of obs.	972								
Number of groups	108								
Obs. per group	9								
Log-likelihood	866.912								
R^2 (within)	0.557								
Wald $\chi^2(13)$	42.35 ($p<0.001$)								

The sample period is from 2008 to 2016. The *p* values are two-tailed. Variables are defined in the "Appendix"

results of our previous non-instrumented estimations that fully support the hypothesis of our study.

Finally, we carry out additional robustness analyses that provide results qualitatively analogous to our main estimations. Specifically, Table 11 displays estimations by excluding provinces of *Mezzogiorno* that may distort our results because of its historical economic underdevelopment, compared to Northern Italy, and the strong dominance of Mafia organizations that foster the illegality in the socioeconomic fabric (Ravenda et al. 2018). Furthermore, Table 12 presents estimations by adopting a spatial weight matrix without any threshold distance for the computation of the spatially lagged dependent variable (*W*LTAV_PROV*) to consider that LTAV practices, unlike immigrants, may potentially spill over not only into surrounding provinces, but also into higher-order neighbouring provinces (neighbours to the neighbours) without any defined distance limit.

Conclusions and Discussion

In this study, we investigate whether the geographic concentration of non-EU immigrants in the various Italian provinces is positively associated with LTAV practices adopted by firms located in the same provinces of residence of immigrants, as well as in the surrounding provinces, and operating in construction and agriculture industries that mostly employ immigrants in Italy. For this purpose, we develop a LTAV proxy, based on the financial accounting information of the employing firms, and specifically consisting in the abnormal values of the ratio of SSCs paid to lagged total assets, computed with a sample of 993,606 firm-years spread over the 108 Italian provinces over the period 2008–2016. Our results provide empirical support for the hypothesis that a higher non-EU immigrant concentration in a specific province enhances the opportunities for LTAV in that province as well as in the neighbouring provinces.

The presence of spatial spillover in LTAV intensity and determinants may suggest that targeted public interventions, also through awareness-raising campaigns and assistance to vulnerable groups in sectors and areas that are particularly sensitive and at risk of LTAV, could have beneficial effects also in spatially contiguous regions. On the other hand, the mobility of non-EU immigrants across neighbouring provinces may support the need for a close coordination between regional administrations in the definition of labour inspection plans and other social interventions, addressing practices of illegal exploitation of immigrant workers, that should be based on a "supra-regional" perspective. This close coordination across local governments, that could be fostered by the central government through common guidelines and constraints, should also underlie the development and implementation of public policies aiming to influence on other determinants of LTAV that may have spillover effects on neighbouring regions. Among them, industrial policies, such as labour and income tax relief policies and low-interest or subsidized loans for investments in technology and professional training in local labour-intensive industries like agriculture and construction, may increase the competitiveness of their firms and make LTAV less necessary for their subsistence.

The partisans of an open policy towards immigration argue that immigrants, by mostly undertaking jobs which natives refuse and would otherwise be unfilled, may support the solvency of European social security systems that suffer from significant reductions of SSCs because of population ageing, changes in labour market structure, and financial globalization (French and Jones 2012; Okkerse 2008). However, this positive effect may be undermined by LTAV practices associated with the employment of immigrants. Furthermore, although immigrants may not be perfect substitutes for native workers, they may partially compete with low-paid/skilled native workers whose working conditions may also deteriorate because of the increased immigration, especially in periods of high unemployment and in the poorest regions. Therefore, keeping immigrants in the country without recognizing their legal status may create counterproductive unfair competition in the labour market and threaten the socioeconomic rights of the nationals as well as of the regular immigrants (Triandafyllidou and Maroukis 2012). In this regard, a greater social integration and recognition of rights of immigrants even through the regularization of those undocumented and informally employed may enhance the efficacy of public policies in several areas (e.g. employment, health and education), and simultaneously safeguard the social rights of nationals and regular immigrants (Triandafyllidou and Maroukis 2012). The potential contribution of the immigrant regularization to the funding of national social security systems, by reducing LTAV, may depend on the skill level of the illegal migrant workers (Casarico et al. 2018) and may arise from the commonly large labour force participation rate of undocumented immigrants (Borjas 2017). Indeed, several studies show the positive effects of the legalization of immigrants through amnesties on the labour tax revenues of the social security. Specifically, Monras et al. (2018) document that the legalization of around 600,000 immigrants by the Spanish government in 2004 contributes positively to the social security by increasing its payroll-tax revenues by 4189 euros, on average, for each newly legalized immigrant. In addition, Di Porto et al. (2018) find that 73.5% of undeclared migrant workers, regularized through a large amnesty implemented in Italy in 2002, remain within the formal labour market and then contribute to the social security, although in different jobs, with no effects on legalized immigrant coworkers. Nonetheless, the authors also show that in the medium and long run, the regularization has no impact on the level of formal employment in the firms subscribing to the amnesty. In this regard, we interpret that, after performing the regularization, the firms may continue to engage in LTAV by replacing the regularized employees with other undeclared migrant workers. This vicious circle might persist until unprotected and exploitable immigrants are available for employers willing to breach labour regulations. In this respect, our study, providing empirical evidence of the perverse effects of non-EU immigration on LTAV and the spillover effects on neighbouring regions, supports the legalization of the status of the immigrants and their employment relationships as measures that may benefit both the interests of immigrants and the interests of nationals.

Our findings, however, are subject to some limitations. Specifically, the validity of our results depends on the ability of our proxy to properly measure LTAV variability. Furthermore, we do not account for the heterogeneity within non-EU immigrants in terms of origin, education, skills, culture, motivation and socioeconomic status that may moderate their impact on the labour market and LTAV. Finally, as suggestions for future research, our results may be corroborated by using other estimation methodology of UDW, our study may be replicated for other industries and national contexts, and additional spatial models may be tested by incorporating additional sources of territorial heterogeneity that may affect LTAV.

Compliance with Ethical Standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical Approval This article does not contain any studies with human participants or animals performed by any of the authors.

Appendix

Definition of Variables

Variable definition of Eq. (3)

$$AbSSCs_{i,t} = \beta_0 + \sum_r \beta_r PROVINCE^r_{i,t} + \sum_k \beta_k CONTROLS^k_{i,t} + \sum_s \beta_s INDUSTRY^s_{i,t} + \varepsilon_{i,t}$$

AbSSCs abnormal SSCs equal to residuals from Eq. (2) simultaneously estimated with Eq. (1), *PROVINCE* dummy variable for each of 107 Italian provinces, *CONTROLS* firm-level control variables of Eq. (3) regression model: *SIZE* natural logarithm of total assets in thousands of euros, *AGE* age of the firm in years, *LEVER* total debt divided by total assets, *CAPINT* net fixed assets and net intangible assets divided by total assets, *ROA* net income divided by total assets, *LOSS* dummy variable that takes a value of 1 if the firm had two or more consecutive years of negative income including the current and 0 otherwise, *GROW* percentage change in net sales relative to previous year, *DAC* discretionary accruals estimated based on the performance-adjusted modified Jones model (Ravenda et al. 2018), *AbMATL* abnormal material costs equal to residuals from the following Eq. (5) with material costs (*MAT*), including both raw materials and merchandise, as dependent variable, estimated cross-sectionally for each two-digit NACE industry-year

$$\frac{MAT_{i,t}(SERV_{i,t})}{\ln(TA_{i,t-1})} = \beta_0 + \beta_1 \frac{1}{\ln(TA_{i,t-1})} + \beta_2 \frac{S_{i,t}}{\ln(TA_{i,t-1})} + \beta_3 \frac{\Delta S_{i,t}}{\ln(TA_{i,t-1})} + \varepsilon_{i,t} \tag{5}$$

AbSERV abnormal service costs equal to residuals from Eq. (5) with service costs (*SERV*) as dependent variable, estimated cross-sectionally for each two-digit NACE industry-year, *CASHTA* cash and cash equivalents divided by total assets, *ETR* abnormal effective tax rate equals to industry- and size-matched GAAP ETR minus firm's GAAP ETR, where GAAP ETR is the total tax expense divided by pre-tax income. Industry- and size-matched GAAP ETR is the average GAAP ETR for the portfolio of firms in the same quintile of total assets and the same two-digit NACE industry-year, *SD_ROA* standard deviation of *ROA* over the past four years, *INVENTA* inventory divided by total assets, *INDUSTRY* dummy variable for each three-digit NACE industry.

Variable definition of Eq. (4)

$$LTAV\_PROV_{i,t} = \rho WLTAV\_PROV_{i,t} + \beta_1 IMMIGR_{i,t} + \sum_k \beta_k CONTROLS^k_{i,t} + \theta_1 WIMMIGR_{i,t} + \sum_k \theta_k WCONTROLS^k_{i,t} + u_i + v_{i,t}$$

LTAV_PROV LTAV measure at province level equal to the estimated coefficients on *PROVINCE* in Eq. (3), *W* inverse distance spatial weight matrix with a threshold distance of 57.14 km, spectrally normalized, *IMMIGR* non-EU immigrant concentration, computed as the fraction of non-EU residents per 1000 residents in each province and year, restricted to the population between 18 and 59 years of age, and spatially differentiated from the province of Rome (source: ISTAT), *CONTROLS* province-level control variables of Eq. (4) regression model: *DENSITY* province population per km^2, spatially differentiated from the province of Rome (source: ISTAT), *CRIME* natural logarithm of crimes reported by police forces to judicial authorities per 1000 residents, spatially differentiated from the province of Rome (source: ISTAT), *UNEMPL* annual unemployment rate, spatially differentiated from the province of Rome (source: ISTAT), *HGRSAL* employee hourly gross salary (CPI deflated, 2016 equivalents), spatially differentiated from the province of Rome (source: ISTAT), Δ*GDP* gross domestic product growth rate, spatially differentiated from the province of Rome (source: ISTAT), *u* province fixed-effect (*PROVINCE FE*).

References

Adom, K. (2014). Beyond the marginalization thesis: An examination of the motivations of informal entrepreneurs in sub-saharan Africa. *International Journal of Entrepreneurship and Innovation, 15*(2), 113–125.

Annisette, M., & Trivedi, V. U. (2013). Globalization, paradox and the (un)making of identities: Immigrant Chartered Accountants of India in Canada. *Accounting, Organizations and Society, 38*(1), 1–29.

Anselin, L. (2010). Thirty years of spatial econometrics. *Papers in Regional Science, 89*(1), 3–25.

Anselin, L., Bera, A. K., Florax, R., & Yoon, M. J. (1996). Simple diagnostic tests for spatial dependence. *Regional Science and Urban Economics, 26*(1), 77–104.

Anselin, L., & Lozano-Gracia, N. (2008). Errors in variables and spatial effects in hedonic house price models of ambient air quality. *Empirical Economics, 34*(1), 5–34.

Badertscher, B. A., Katz, S. P., Rego, S. O., & Wilson, R. J. (2017). Conforming Tax Avoidance and Capital Market Pressur. Kelley School of Business Research Paper.

Bastida, F., Guillamón, M.-D., & Benito, B. (2013). Municipal spending in Spain: Spatial approach. *Journal of Urban Planning and Development, 139*(2), 79–93.

Beręsewicz, M., & Nikulin, D. (2018). Informal employment in Poland: An empirical spatial analysis. *Spatial Economic Analysis, 13*(3), 338–355.

Biles, J. J. (2009). Informal work in Latin America: Competing perspectives and recent debates. *Geography Compass, 3*(1), 214–236.

Bird, R., & Davis-Nozemack, K. (2018). Tax avoidance as a sustainability problem. *Journal of Business Ethics, 151*(4), 1009–1025.

Bohn, S. (2010). The quantity and quality of new immigrants to the US. *Review of Economics of the Household, 8*(1), 29–51.

Bohn, S., & Owens, E. G. (2012). Immigration and informal labor. *Industrial Relations, 51*(4), 845–873.

Borjas, G. J. (2017). The labor supply of undocumented immigrants. *Labour Economics, 46,* 1–13.

Buehn, A. (2012). The shadow economy in German Regions: An empirical assessment. *German Economic Review, 13*(3), 275–290.

Burridge, P. (1981). Testing for a common factor in a spatial autoregression model. *Environment and Planning A, 13*(7), 795–800.

Cai, G., Li, W., & Tang, Z. (2018). Religion and the method of earnings management: Evidence from China. *Journal of Business Ethics, 25,* 1–20. https://doi.org/10.1007/s10551-018-3971-6.

Cameron, A. C., & Trivedi, P. K. (2010). *Microeconometrics using stata*. College Station: Stata Press.

Cappelen, C., & Muriaas, R. L. (2018). Polish labour migrants and undeclared work in Norway. *Scandinavian Political Studies, 41*(2), 167–187.

Casarico, A., Facchini, G., & Frattini, T. (2018). What drives the legalization of immigrants? Evidence from IRCA. *Regional Science and Urban Economics, 70,* 258–273.
Commissione Nazionale per il Diritto di Asilo. (2018). I numeri dell'asilo| Dipartimento Libertà Civili e Immigrazione. Retrieved October 23, 2018, from http://www.libertaciviliimmigrazione.dlci.interno.gov.it/it/documentazione/statistica/i-numeri-dellasilo.
Confcommercio Studies Office. (2017). *Le determinanti dell'evasione fiscale: un'analisi regionale*. http://www.confcommercio.it/documents/10180/13595649/Le+determinanti+dell%27evasione+fiscale.pdf/8fd55216-d0fc-4802-a2ef-c2254e0a56f6.
Constant, A. F., & Zimmermann, K. F. (2016). Towards a new European refugee policy that works. *CESifo DICE Report, 14*(4), 3–8.
Coppens, L., & Peek, E. (2005). An analysis of earnings management by European private firms. *Journal of International Accounting, Auditing and Taxation, 14*(1), 1–17.
Corrado, A. (2011). Clandestini in the Orange Towns: Migrations and Racisms in Calabria's Agriculture. *Race/Ethnicity: Multidisciplinary Global Contexts, 4*(2), 191–201.
D'Amuri, F., Ottaviano, G. I. P., & Peri, G. (2010). The labor market impact of immigration in Western Germany in the 1990s. *European Economic Review, 54*(4), 550–570.
Dell'Anno, R., Gómez-Antonio, M., & Pardo, A. (2007). The shadow economy in three Mediterranean countries: France, Spain and Greece. A MIMIC approach. *Empirical Economics, 33*(1), 51–84.
Desai, M. A., & Dharmapala, D. (2009). Corporate tax avoidance and firm value. *The Review of Economics and Statistics, 91*(3), 537–546.
Di Porto, E., Martino, E. M., & Naticchioni, P. (2018). Back to Black? The Impact of Regularizing Migrant Workers (No. 517). Naples. https://ideas.repec.org/p/sef/csefwp/517.html.
Directorate General of Immigration and Integration Policies. (2018). Eighth Annual Report 2018 "Foreigners in the Italian Labour market." http://www.lavoro.gov.it/priorita/Pagine/Ottavo-Rapporto-annuale-Stranieri-nel-mercato-del-lavoro-in-Italia.aspx.
Donohoe, M. P. (2015). The economic effects of financial derivatives on corporate tax avoidance. *Journal of Accounting and Economics, 59*(1), 1–24.
Dustmann, C., Fasani, F., Frattini, T., Minale, L., & Schönberg, U. (2017). On the economics and politics of refugee migration. *Economic Policy, 32*(91), 497–550.
Dustmann, C., Frattini, T., & Preston, I. P. (2013). The effect of immigration along the distribution of wages. *Review of Economic Studies, 80*(1), 145–173.
Easton, T. (2001). Immigration and natives' wages: Understanding their correlation in the 1980s. *Review of Regional Studies, 31*(3), 219–235.
EC. (2014). Special Eurobarometer 402 "Undeclared Work in the EU." European Commission (Vol. 13). European Commission Directorate-General for Employment and Social Affairs.
Elhorst, J. P. (2010). Applied spatial econometrics: Raising the bar. *Spatial Economic Analysis, 5*(1), 9–28.
Elhorst, J. P. (2014a). *Spatial econometrics*. Cambridge: Academic Press.
Elhorst, J. P. (2014b). Matlab software for spatial panels. *International Regional Science Review, 37*(3), 389–405.
Fairlie, R. W., & Meyer, B. D. (2003). The effect of immigration on native self-employment. *Journal of Labor Economics, 21*(3), 619–650.
Fama, E. F., & MacBeth, J. D. (1973). Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy, 81*(3), 607–636.
Feld, L. P., & Schneider, F. (2010). Survey on the shadow economy and undeclared earnings in OECD countries. *German Economic Review, 11*(2), 109–149.
Flai-Cgil. (2014). *Agromafias and Caporalato, Second Report. Osservatorio Placido Rizzotto Flai-Cgil.*
French, E., & Jones, J. (2012). Public pensions and labor supply over the life cycle. *International Tax and Public Finance, 19*(2), 268–287.
Fullin, G., & Reyneri, E. (2011). Low unemployment and bad jobs for new immigrants in Italy. *International Migration, 49*(1), 118–147.
Gavana, G., Guggiola, G., & Marenzi, A. (2013). Evolving connections between tax and financial reporting in Italy. *Accounting in Europe, 10*(2014), 43–70.
Gavosto, A., Venturini, A., & Villosio, C. (1999). Do immigrants compete with natives? *Labour, 13*(3), 603–622.
Gurtoo, A., & Williams, C. C. (2009). Entrepreneurship and the informal sector. *The International Journal of Entrepreneurship and Innovation, 10*(1), 55–62.
Haigner, S. D., Jenewein, S., Schneider, F., & Wakolbinger, F. (2013). Driving forces of informal labour supply and demand in Germany. *International Labour Review, 152*(3–4), 507–524.
Harney, N. D. M. (2011). Accounting for African migrants in Naples, Italy. *Critical Perspectives on Accounting, 22*(7), 644–653.
Hatton, T. J., De Haas, R., & Egger, P. (2017). Refugees and asylum seekers, the crisis in Europe and the future of policy. *Economic Policy, 32*(91), 447–496.
Hong, Y., & Andersen, M. L. (2011). The relationship between corporate social responsibility and earnings management: An exploratory study. *Journal of Business Ethics, 104*(4), 461–471.
International Labour Organization (ILO). (2013). *Measuring informality: A statistical manual on the informal sector and informal employment*. Retrieved July 10, 2018, from http://www.ilo.org/global/publications/ilo-bookstore/order-online/books/WCMS_222979/lang–en/index.htm.
Italian Institute of Statistics (ISTAT). (2018). Immigrati.Stat. Retrieved October 16, 2018, from http://stra-dati.istat.it/Index.aspx.
Jayet, H., Ukrayinchuk, N., & De Arcangelis, G. (2010). The location of immigrants in Italy: Disentangling networks and local effects. *Annals of Economics and Statistics, 97/98,* 329–350.
Kelejian, H. H., & Prucha, I. R. (2001). On the asymptotic distribution of the Moran I test statistic with applications. *Journal of Econometrics, 104*(2), 219–257.
Kim, C., & Zhang, L. (2016). Corporate political connections and tax aggressiveness. *Contemporary Accounting Research, 33*(1), 78–114.
Lanis, R., & Richardson, G. (2012). Corporate social responsibility and tax aggressiveness: A test of legitimacy theory. *Accounting, Auditing & Accountability Journal, 26*(1), 75–100.
Lanis, R., & Richardson, G. (2015). Is corporate social responsibility performance associated with tax avoidance? *Journal of Business Ethics, 127*(2), 439–457.
Lanis, R., Richardson, G., Liu, C., & McClure, R. (2018). The impact of corporate tax avoidance on board of directors and CEO reputation. *Journal of Business Ethics, 160*(2), 463–498.
LeSage, J. P. (2014). What regional scientists need to know about spatial econometrics. *SSRN Electronic Journal., 44*(1), 13–32.
LeSage, J., & Pace, R. K. (2009). *Introduction to spatial econometrics. Systems Engineering*. Boca Raton: CRC Press.
Lewis, H., Dwyer, P., Hodkinson, S., & Waite, L. (2015). Hyper-precarious lives: Migrants, work and forced labour in the Global North. *Progress in Human Geography, 39*(5), 580–600.
Longhi, S., Nijkamp, P., & Poot, J. (2010a). Meta-analyses of labour-market impacts of immigration: Key conclusions and policy implications. *Environment and Planning C: Government and Policy, 28*(5), 819–833.

Longhi, S., Nijkamp, P., & Poot, J. (2010b). Joint impacts of immigration on wages and employment: Review and meta-analysis. *Journal of Geographical Systems, 12*(4), 355–387.
Maroukis, T., Iglicka, K., & Gmaj, K. (2011). Irregular migration and informal economy in southern and Central-Eastern Europe: Breaking the vicious cycle? *International Migration, 49*(5), 129–156.
Marques, M., Rodrigues, L. L., & Craig, R. (2011). Earnings management induced by tax planning: The case of Portuguese private firms. *Journal of International Accounting, Auditing and Taxation, 20*(2), 83–96.
Monras, J., Vázquez-Grenno, J., & Elias, F. (2018). Understanding the Effects of Legalizing Undocumented Immigrants. *CEPR Discussion Paper Series*, (DP 12726).
Moulton, B. R. (1990). An illustration of a pitfall in estimating the effects of aggregate variables on micro units. *The Review of Economics and Statistics, 72*(2), 334.
Okkerse, L. (2008). How to measure labour market effects of immigration: A review. *Journal of Economic Surveys, 25,* 256. https://doi.org/10.1111/j.1467-6419.2007.00533.x.
Pajnik, M. (2016). 'Wasted precariat': Migrant work in European societies. *Progress in Development Studies, 16*(2), 159–172.
Pfau-Effinger, B. (2009). Varieties of undeclared work in European societies. *British Journal of Industrial Relations, 47*(1), 79–99.
Platikanova, P. (2015). Debt maturity and tax avoidance. *European Accounting Review., 26*(1), 97–124.
Prosser, T. (2016). Dualization or liberalization? Investigating precarious work in eight European countries. *Work, Employment & Society, 30*(6), 949–965.
Ravenda, D., Argilés-Bosch, J. M., & Valencia-Silva, M. M. (2015). Labor tax avoidance and its determinants: The case of mafia firms in Italy. *Journal of Business Ethics, 132*(1), 41–62.
Ravenda, D., Valencia-Silva, M. M., Argiles-Bosch, J. M., & García-Blandón, J. (2019). Money laundering through the strategic management of accounting transactions. *Critical Perspectives on Accounting, 60,* 65–85.
Ravenda, D., Valencia-Silva, M. M., Argilés-Bosch, J. M., & García-Blandón, J. (2018). Accrual management as an indication of money laundering through legally registered Mafia firms in Italy. *Accounting, Auditing & Accountability Journal, 31*(1), 286–317.
Seifert, S., & Valente, M. (2018). *An Offer that You Can't Refuse? Agrimafias and Migrant Labor on Vineyards in Southern Italy*. *Ssrn.*
Skillings, J. H., & Mack, G. A. (1981). On the use of a friedman-type statistic in balanced and unbalanced block designs. *Technometrics, 23*(2), 171–177.
Staiger, D., & Stock, J. H. (1997). Instrumental variables regression with weak instruments. *Econometrica, 65*(3), 557.
Strauss, K., & McGrath, S. (2017). Temporary migration, precarious employment and unfree labour relations: Exploring the 'continuum of exploitation' in Canada's Temporary Foreign Worker Program. *Geoforum, 78,* 199–208.
The World Bank. (2018). World Bank national accounts data, and OECD National Accounts data files. Accessed November 16, 2018, from https://data.worldbank.org/indicator/NY.GDP.MKTP.KD.ZG?locations=IT.
Theodore, N., Pretorius, A., Blaauw, D., & Schenck, C. (2018). Informality and the context of reception in South Africa's new immigrant destinations. *Population, Space and Place, 24*(3), 1–10.
Triandafyllidou, A., & Maroukis, T. (2012). Migration policy in southern Europe: Challenges, constraints and prospects. *A Strategy for Southern Europe - London School of Economics and Political Science, 6,* 54–63.
Trinci, S. (2006). The contribution of networks to immigrant insertion into the informal economy. *International Journal of Sociology and Social Policy, 26*(9/10), 385–396.
United Nations Department of Economic and Social Affairs (UNDESA). (2017). *Trends in International Migrant Stock: The 2017 Revision. United Nations database*.
Venkatesh, S. A., & Fiola, J. A. (2006). Off the books: the underground economy of the urban poor. *Choice Current: Reviews for Academic Libraries, 45,* 56–57.
Williams, C. C., & Horodnic, I. A. (2015). Rethinking the marginalisation thesis: An evaluation of the socio-spatial variations in undeclared work in the European Union. *Employee Relations, 37*(1), 48–65.
Williams, C. C., & Nadin, S. (2012). Tackling undeclared work in the European Union. *CESifo Forum, 13*(2), 20–25.
Williams, C. C., & Nadin, S. (2014). Evaluating the participation of the unemployed in undeclared work. *European Societies, 16*(1), 68–89.
Williams, C. C., & Round, J. (2010). Explaining participation in undeclared work: A result of exit or exclusion? *European Societies, 12*(3), 391–418.
Yea, S. (2017). The art of not being caught: Temporal strategies for disciplining unfree labour in Singapore's contract migration. *Geoforum, 78,* 179–188.
Ylönen, M., & Laine, M. (2015). For logistical reasons only? A case study of tax planning and corporate social responsibility reporting. *Critical Perspectives on Accounting, 33,* 5–23.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.