Argomenti di discussione

(Discussion topics)

SPATIAL DYNAMICS IN TAX GAP DETERMINANTS

A. Carfora, R.V. Pansini, S. Pisani

N. 01/2016

Agenzia Entrate

Argomenti di discussione è una pubblicazione che intende divulgare contributi ed analisi su argomenti di economia, statistica, econometria e scienza delle finanze, che abbiano rilevanza per la missione dell'Agenzia delle Entrate, al fine di alimentare il dibattito scientifico sui temi di interesse strategico dell'Agenzia e favorire lo scambio di opinioni. La pubblicazione ospita contributi in lingua italiana o inglese proposti da autori sia interni sia esterni all'Agenzia. Le opinioni espresse negli articoli sono lasciate alla discrezionalità degli autori e non impegnano la responsabilità dell'Agenzia delle Entrate.

Discussion topics *is a publication aiming at disclosing contributions and analysis in economics, statistics, econometrics and public finance, which are relevant to the mission of the Italian Revenue Agency (Agenzia delle Entrate) in order to feed the scientific debate on topics of strategic interest for the Agency and to facilitate the exchange of opinions. This publication admits contributions in Italian or in English proposed by authors internal or external to the Agency. The views expressed in the articles are those of the authors and do not involve the responsibility of the Revenue Agency.*

ISSN: 2420-773X

Tel. +39-06-5054-3874/5071 Fax +39-06-5054-5038
E-mail: ae.argomentididiscussione@agenziaentrate.it
Web: www.agenziaentrate.gov.it/wps/content/Nsilib/Nsi/Documentazione/Rivista/Argomenti+di+discussione

Spatial Dynamics in Tax Gap Determinants

Alfonso Carfora, Rosaria Vega Pansini, Stefano Pisani

Agenzia delle Entrate, rosariavega.pansini@agenziaentrate.it

Abstract

This paper tries to shed light on the determinants of regional level of tax evasion in Italy by analyzing the relationship between structural and socio-economic characteristics of local economies and the magnitude of tax avoidance. In fact, similarities in business opportunities and aspects related to tax morale can influence the attitude toward tax evasion. At the same time, proximity issues may lead taxpayers to mimic their neighbors in tax compliance behavior. We explicitly test these hypotheses by estimating the effects of different determinants of tax gap and the size of spatial correlation among Italian regional levels of tax gap in the tax years 2001-2011. Econometric analysis uses spatial panel models in order to control for determinants of tax gap and evaluates the existence and scope of a mimicking phenomenon in tax evading business conduct. Results show a high spatial correlation in tax gap intensities among Italian regions. Local level of tax evasion appears to react to neighbor attitudes toward tax compliance. Moreover, the analysis of significant estimated coefficients from SARSAR model specification enable us to draw a taxonomy of determinants of tax gap: environmental factors linked to the economic and institutional features (agriculture value added, diffusion of electronic money, bank deposits, incidence of self-employment, amount paid after tax amnesties and GDP growth); operational factors linked to the relative efficacy of tax evasion contrasting policies (IRA enforcement and renew of auditing schemes) and spillovers factors from neighbor regions (level of crime and tax gap of near regions). This outcome suggests few considerations and issues for further research. First, there may be spillovers in tax evading behavior between different regions. Second, auditing and contrasting policies should take into account also geographical patterns. Finally, such evidence may require measures containing the risk of collusion among taxpayers in neighbor regions.

Sommario

L'obiettivo di questo lavoro è individuare i fattori che determinano i tassi di evasione regionali analizzando la relazione tra le caratteristiche strutturali e socio-economiche delle realtà locali e la corrispondente propensione all'evasione. Si vuole infatti verificare se le caratteristiche dello sviluppo socio-economico e gli aspetti relativi alla *tax morale* possono influenzare una preferenza a evadere le tasse. Allo stesso modo, la vicinanza geografica può spingere i contribuenti ad imitare il comportamento fiscale dei contribuenti delle aree a loro più vicine. Per verificare tali ipotesi, gli autori dapprima studiano l'effetto di diverse determinanti del tax gap; successivamente, calcolano indicatori di correlazione spaziale relativi ai livelli di evasione regionale per gli anni 2001-2011 e stimano un modello econometrico di tipo "spatial-panel" per individuare sia i fattori che determinano il fenomeno evasivo che l'esistenza di tale atteggiamento imitativo da parte dei contribuenti. I risultati mostrano che esiste una forte correlazione spaziale tra le regioni italiane in termini di propensione all'evasione. I tassi regionali di

evasione fiscale, infatti, sono influenzati anche dall'attitudine alla *tax compliance* delle regioni più vicine. Inoltre, la significatività statistica dei coefficienti del modello econometrico utilizzato consente di individuare una tassonomia dei fattori che influenzano il tax gap: quelli ambientali, connessi alle caratteristiche economiche ed istituzionali (diffusione della moneta elettronica, ammontare dei depositi bancari, incidenza del settore agricolo sull'economia regionale e dei lavoratori autonomi sugli occupati, andamento del PIL e ammontare delle somme condonate); quelli che attengono all'operatività dell'Amministrazione fiscale in termini di prevenzione e contrasto all'evasione (indici regionali di presidio del territorio e di efficacia degli studi di settore); quelli di *spillover* legati alla vicinanza con le altre regioni (livelli di criminalità e tax gap delle regioni vicine). Questi risultati suggeriscono alcune considerazioni in termini di politiche di contrasto all'evasione fiscale e forniscono interessanti spunti per ricerche future. In primo luogo, che possono esistere effetti di "contaminazione" nel comportamento evasivo tra contribuenti residenti in regioni vicine. In secondo luogo, che il design delle politiche di prevenzione e contrasto del fenomeno evasivo dovrebbero tenere in considerazione anche gli aspetti legati alla localizzazione geografica. Infine, che l'attività di contrasto deve essere portata avanti anche attraverso l'adozione di misure atte a ridurre il rischio di collusione tra i contribuenti in realtà locali geograficamente vicine.

JEL classifications: C21, C23, E26, H26

Key words: Tax gap estimation, spatial econometrics, panel estimation

Index

1. Introduction

Recent statistics place Italy among countries with the highest tax burden in the EU (Eurostat, 2014). During the last decade, the weight of tax revenues on central and local governments' budget has increased due also to international obligations to maintain public debt and deficit below the requirements of the European Fiscal Compact agreement. Among the issues connected with the maximization of tax revenues, tax collection efficiency and the size of tax evasion are the most important ones. Moreover, in order to limit distortions and negative effects on economic growth due to an excessive taxation, it is necessary to minimize the space for tax evasion and its distortions for economic activity. Such issues gain importance for either national and local fiscal systems.

The relative success of any auditing or tax payment enforcement policy by tax collecting agencies lays on the capacity to identify the determinants of taxpayers' behavior. Many factors have been analyzed in the literature to explain the tax evading propensity (Allingham and Sadmo, 1972, Alm, 2012; Yitzaky, 1974, Clotfelter, 1983, Frey ad Feld, 2002, Richardson, 2006, Buehn and Schneider, 2012 among others). Some common elements, such as tax morale, the quality of public expenditure, tax rates, tax burden, tax penalties, gender (McGee, 2014; Marino and Zizza, 2012), education and the quality of institutions have been identified in these studies. Nevertheless, to the best of our knowledge there are no other analyses that explicitly address the issue of proximity as an additional determinant of tax evasion. In fact, aspects related to geography and spillover effects have been recently analyzed with respect only to tax collection efficiency (Arvate and Mattos, 2008 for Brazilian municipalities) or tax settings (Depalo and Messina, 2011 for Italian municipalities) but any correlation with tax evasion has been neglected.

Our analysis aims at filling this gap by focusing on those spatial aspects that contribute to tax evasion behavior. In particular, this paper tries to shed light on the determinants of regional level of tax evasion in Italy by analyzing the relationship between structural and socio-economic characteristics of local economies and the magnitude of tax evasion. In addition, our attention is concentrated on neighbor taxpayers' behavior as specific institutional and economic factor influencing regional attitude toward tax compliance. In fact, similarities in business opportunities, aspects related to tax morale and geographical contiguity issues may lead taxpayers to mimic their neighbors in fiscal (non) compliant behavior. Our paper contributes to the literature explicitly identifying spatial interactions on tax evasion among Italian regions with similar economic and institutional characteristics.

Previous literature on the determinants of tax evasion considers geography as an additional variable in the estimation strategy, i.e. using dummy or categorical variables in order to control for differential effect of origin (Dell'Anno and Schneider, 2006; Richardson, 2006; Yalama and Gumus, 2013; Schneider and Enste, 2000; Zizza and Marino, 2012). A slightly different example developed by Carbone and Spingola (2015) uses geography together with other socio-demographic and economic variables at the provincial level in a factor analysis to derive clusters of provinces with similar characteristics.

Nevertheless, results on the effects of spatial interactions should be cautiously interpreted. From one point of view, spatial correlations can capture strategic behavior of taxpayers in fiscal compliance. From another, we can measure spatial dependence among the residuals and thus determine inefficient estimated coefficients. In order to overcome this second issue and to isolate strategic behavior among neighbor regions, as in Depalo and Messina (2011), a spatial panel approach has been adopted in which we control for possible spatial dependence in the error terms due to mis-pecification of the model.

As for the measure of non-compliance, we calculate an index of tax intensity as the ratio of regional tax gap on spontaneously paid tax returns. We then estimate the size of spatial correlation among Italian regional percentage of tax gap in years 2001-2011. Econometric analysis uses a dynamic spatial panel estimation model in order to control for determinants of tax gap and evaluates the existence and scope of a mimicking phenomenon in tax evading business conduct. In order to justify the use of spatial panel models, we use three specifications starting from a pooled OLS to a dynamic panel and relative diagnostics to show how percentage tax gap is characterized by time and spatial persistence.

The rest of the paper is organized as follows. Section 2 sketches the main evidence on geographical distribution of tax gap among Italian regions underling some first insights on the role of mimicking in tax evading behavior. Section 3 describes the empirical strategy used in the analysis and some econometric issues related to the estimation procedure Section 4 presents the panel dataset used in the econometric analysis. Section 5 describes the results. Finally, Section 6 concludes providing few suggestions for further research.

2. Geography and the tax gap: some evidence from Italian regions

Our empirical analysis is based on a panel of 20 Italian regions observed from 2001 to 2011. Tax evasion rates are measured by the Italian Revenue Agency (IRA) as percentage tax gap (PTG). Specifically tax gap is defined as the difference between the amount of taxes the tax administration should levy and collect (the potential tax yield) and the actual tax revenues (cash due and paid in period). The potential amount is "that which could be collected if no taxpayers would voluntary breach the law and involuntary errors would amount to zero"[1]. PTG is the ratio of monetary level of total tax gap[2] and the value of the total voluntary tax returns (VTR):

$$PTG_{it} = \frac{\text{Tax Gap}_{it}}{VTR_{it}} \qquad [2.1]$$

[1] Das-Gupta, Mookherjee (2000). It is derived from national accounts data on value added and total consumption.

[2] By total tax evasion we refer to the value of the evasion estimated on taxes under the duty of the IRA, i.e. the sum of VAT, personal income tax (namely, IRPEF), corporate income tax (namely, IRES) and tax on production activities (namely, IRAP). For details on the calculation of tax gap as well as potential collectable taxes and PTG, see Braiotta et al., (2015) and D'Agosto et al. (2014).

Where *i* indicates the region and *t* the tax year. Designed in this way, PTG can be seen as an indicator of the intensity of non-compliance as it measures the amount of each monetary unit of tax gap per each unit of tax return. PTG can assume values greater or equal to 0 and can be interpreted as the regional propensity to compliance. In particular:

$$PTG_{it} = \begin{cases} = 0 \text{ in absence of evasion} \\ \leq 1 \text{ if Tax Gap } \leq \text{VTR} \\ > 1 \text{ if Tax Gap} > \text{VTR} \end{cases} \qquad [2.2]$$

Figure 1 shows the 2001-2011 distribution of PTG index for each region[3]. The heterogeneity of economic and social structure of Italian regions is reflected in the geographic distribution of tax gap and can be visualized using the box plots for each region. Median values of percentage tax gap have a large range of variation, from 0.15 of Lazio[4] to 0.88 of Basilicata.Moreover, in some regions, like for example, Emilia Romagna, Lazio, Liguria and Tuscany, the percentage tax gap remained quite stable along the ten years analyzed in our panel data, while others, like Calabria, Sardinia and Sicily experienced a large variation in PTG values. Considering the definition of PTG, its time variation can be due either to large yearly differences of voluntary tax returns or of tax gap. Due to the focus of our analysis, it is interesting to note that southern and neighbor regions experience the same large dispersion of tax gap index trough time. Summary statistics of regional 2001-2011 distributions of PTG are reported in Table 1.

Evidence of regional differences in tax evasion rates appears also by calculating the 2001-2011 mean values of the PTG[5] shown in Figure 2. The map helps visualizing how tax evasion has been distributed between Italian regions during the last ten years. Southern regions have the highest level of percentage tax evasion while, as noticed, Lazio, Emilia Romagna and Friuli Venetia Giulia are those with the lowest level of evasion compared to voluntary tax paid. The picture appears quite different and with opposite results if we consider the total value of tax evasion instead of the ratio between tax gap and as percentage of voluntary tax payments. In this case, northern Italian regions are placed first with respect to southern regions.

Descriptive analysis provides a first evidence of the relevance of proximity as an additional factor explaining the attitude toward evading tax payment. Our empirical analysis is meant to analyze more deeply the geographical dimension of tax evasion. In fact, such geographical distributions suggest the use of an empirical model, in which also a 'region-specific' component and spatial interactions are considered among explanatory variables as a possible additional determinant of tax gap.

---

[3] Both tax gap and voluntary tax returns are calculated by the Italian Revenue Agency depending on firm's registered address. The geographical distribution of these measures can be different if calculated using the region or the province where firms' plants are located.

[4] It should be noted that a large share of voluntary tax returns of Lazio are represented by public administration and central government instead of private enterprises. As a consequence, PTG in Lazio region is relatively low not because the component of VTR is high and not because tax gap is low.

[5] For each *i*-region we calculate the average PTG as the ratio between average tax gap (period 2001-2011) and average VTR (period 2001-2011) as $\overline{\mathrm{PTG_I}} = \frac{\mathrm{E(TaxGap_i)_{01-11}}}{\mathrm{E(VTR_i)_{01-11}}}$

3. Empirical issues and econometric strategy

The aim of this paper is to test the existence of a spatial interaction in Italian regions in determining tax gap rates. Studying spatial interactions raises several econometric issues. First, if there are proximity effects in tax evasion and firms react to evading decisions of neighbors, then the choice to evade taxes is endogenous and correlated with residual term (u). In fact, there can be unobserved characteristics like institutional environment, tax morale and tastes that can be spatially correlated among bordering regions. Second, if neighbor taxpayers are subject to correlated random shocks, there could be a correlation between regional levels of tax evasion. If we omit the spatial dimension of the covariates, the result could be the presence of spatial dependence in the residuals.

Therefore, we introduce and analyze spatial dependence in a spatial panel framework due to the availability of tax gap data over a ten-year time span. Moreover, in order to test our hypothesis of spatially correlated tax gap rates and their determinants, we estimates different non-spatial and spatial model specifications.

3.1 Specification 1: Pooled OLS

We perform a preliminary pooled linear OLS analysis following the specification:

$$y_{i,t} = \alpha + \beta' x_{i,t} + u_{i,t} \qquad [3.1]$$

where i = 1,…,n is the individual (region) index, t = 1, … , T is the time index and u_{it} is a random disturbance term of mean 0. This specification ignores the spatial error dependence and considers tax evasion as depending only on own regional features. As already underlined, model [3.1] may provide a misleading evidence on determinants of tax gap. Moreover, from an econometric point of view, it can be mis-specified and leads to inconsistent estimated coefficients.

3.2 Specification 2: Static and Dynamic panel model

The presence of heterosckedasticity in the residual component of the pooled OLS model and issues related to the features of the dependent variable emerged in the previous session suggest to modify [3.1] with a specification that introduces an individual (region-specific) unobserved element in the residual component:

$$y_{i,t} = \alpha + \beta' x_{i,t} + u_i + \epsilon_{i,t} \qquad [3.2]$$

The idiosyncratic error $\epsilon_{i,t}$ is supposed to be independent with regressors $x_{i,t}$ while the

individual (time-invariant regional) error component u_i may be independent or correlated with regressors. In the first case equation [3.2] becomes:

$$y_{i,t} = \alpha + \beta' x_{i,t} + \epsilon_{i,t} \qquad [3.3]$$

and consistent estimates can be obtain with the estimation of a fixed effects model. In case the error component is correlated with the explanatory variables, in order to estimate the [3.2] a random effects model is required. This is estimated using the family of generalized least squares (GLS) estimators in order to avoid correlation across the composite error terms (within individuals).

In order to get rid of serial correlation of disturbances and when the time-lagged dependent variable is included in the regressors, a dynamic specification of model [3.2] using an IV approach such as the GMM estimator by Arellano and Bond (1991) is appropriate and has been estimated.

$$y_{i,t} = \alpha + \lambda y_{i,t-1} + \beta' x_{i,t} + u_i + \epsilon_{i,t} \qquad [3.4]$$

3.3 Specification 3: Spatial panel model

Even though model [3.4] considers the time correlation, it still neglects the spatial dimension in the residual component. In order to examine simultaneously the effect of persistence and proximity, i.e. time and spatial correlation, a spatial dynamic panel model should be specified. In the literature on spatial statistics (Anselin, 1988 and Elhorst, 2010), the extent of cross-section dependence is measured with respect to a given "spatial matrix" W that is a $N \times N$ binary weighting matrix (where N is the number of regions) characterizing the pattern of spatial dependence according to a pre-specified set of rules. Matrix W can be specified using a variety of weighting schemes that allow different designs of spatial interaction. For example, the (i, j) elements of the connection matrix, w_{ij}, could be set equal to 1 if the *i*-th and *j*-th regions are joined, and zero otherwise.

The w_{ij} weights of the Italian regional spatial matrix (W) are obtained calculating contiguities using the *k*-nearest neighbors algorithm that sets that two regions r_i and r_j are said to be neighbors if their distance $d_{ij} \leq \min(d_{ik})$ for every k. This criterion ensures that each observation has exactly the same number (k) of neighbors and consequently that every row of matrix w_{ij} has exactly *k*-rows not equal to 0. In our analysis, bilateral distances d_{ij} among two different regions are calculated using geo-spatial coordinates[6] of each region while *k* is fixed equal to 5.

The w_{ij} elements of spatial weighting matrix W can be also employed to measure the extent of spatial dependence. Among the several methods used to evaluate it, one of the simplest and frequently used is to calculate the Global Moran *I*-index (Moran, 1950):

[6] Shape files with georeferenced data of Italian data are available at this link: http://www.gadm.org/

$$I = \left(\frac{N}{\sum_i \sum_j w_{ij}}\right)\left(\frac{N \sum_i \sum_j w_{ij}(y_i - \bar{y})(y_j - \bar{y})}{\sum_i (y_i - \bar{y})^2}\right) \qquad [3.5]$$

The *I*-index measures the extent to which high values of a characteristic are generally located near other high values and vice-versa. It varies from $-1/(N-1)$, if there is no spatial autocorrelation, to 1, when there is maximum autocorrelation. The nearer *I*-index to the value of 1, the stronger and positive spatial autocorrelation is, while high negative values signal a strong negative spatial autocorrelation.

Effects of the presence of spatial dependence of PTG affect residuals of both random and dynamic random models making GLS and GMM estimators inefficient. In order to overcome this risk, we use the SARSAR specification [3.6] originally suggested by Kelejian and Prucha (1999) to capture spatial interactions across spatial units and over time. It can be written in structural form as :

$$y_{i,t} = \lambda W'_{.i} y_{i,t} + \beta' x_{i,t} + \rho W'_{.i} u_i + \epsilon_{i,t} \qquad [3.6]$$

The seminal contribution of Kapoor et al., (2007) follows an approach that does not include the spatial lag of the dependent variable. In this paper, we chose to extend such specification including both spatial lagged dependent variable and spatial error components. We then estimate model [3.6] (Mutl and Pfaermayr, 2011) using a generalization of the estimation procedure suggested by Kelejian and Prucha (1999). This method employs a two-step procedure. In the first step, two initial estimators are computed: a within and a between two stage least squares coefficients. The two sets of corresponding residuals are then used in the spatial generalized moments estimator (GM) where the moments conditions of Kapoor et al. (2007) are again modified accordingly.

4. Data

Explanatory variables have been selected from a large range of about 40 indicators of different nature and sources. We use regional database and the 'Health for All' Italian module available from the National Institute of Statistics (ISTAT) and IRA DbGeo internal database[7].

Due to the high number of covariates, we proceed with the selection of variables applying a classification into homogeneous clusters from which one or at least two variables were drawn. Following this method, all the covariates were previously

[7] The first two databases are public and accessible at the following links:
1. http://www.istat.it/it/archivio/14562
2. http://sitis.istat.it/sitis/html/
the third is an internal IRA Database.

distinguished into two main groups related to operational and environmental variables. In the first group, we include all variables directly connected to the activity of IRA to increase tax compliance. In the second group we include other exogenous variables related to tax evasion but not directly managed by IRA. Nevertheless, as being the second group of variables too large, in order to select variables that contain the most of information discarding redundant ones, a principal component analysis has been performed on the database containing all (*N*) exogenous variables belonging to the second group.

Following the method proposed by King and Jackson (1999, for each of the *p*-PCA's component axes extracted, the *k* variables with the highest (positive and negative) component loadings have been selected to be included as regressors in the analysis. All *N-k* remaining variables were discarded[8]. Variables selected with this method have been divided into sub-groups related to seven specific thematic areas, as reported in Table 2.

Two variables were selected from the group including operational characteristics: *Ren_ss* and *IRA_enforcement*. The first can be considered an indicator of innovation due to changes introduced by 'studi di settore'. 'Studi di settore' are an audit scheme based on a specific interaction between the tax agency and taxpayers in which the agency unveils only part of the information used to develop its audit rule[9]. This scheme provides criteria to determine the amount of taxes that are to be paid by some categories of taxpayers (self-employed and small companies). The adherence to this scheme, which in Italy is voluntary and not mandatory, implies the determination of the number of taxpayers that are congruous with its criteria[10]. Variable *Ren_ss* is the ratio of the number of not congruous taxpayers on total taxpayers who joined the scheme[11]. Not congruous taxpayers have higher probability to be audited by the IRA. We expect a negative coefficient for this variables as a higher number of not congruous taxpayers directly relates to the strength of the audit scheme and its capacity to detect tax evasion. Moreover, the one-year lag of *Ren_ss* is included in the estimation because the effects on compliance of the application of this auditing scheme are perceived by taxpayers during the following fiscal years.

The second operational variable *IRA_enforcement* is a governance indicator and it is considered as a proxy of the probability of a generic taxpayer to be audited. It is calculated as the ratio between the number of audited taxpayers and total taxpayers' population and it can be also used as a measure of IRA enforcement. The variable is inserted as simultaneous to the dependent variable. We expect a negative coefficient for this variable too as we suppose that a higher probability to be audited should lower tax evasion intensity.

As for environmental and context variables included in the PCA analysis, *Q_Agriculture* and *Q_Industry* measure the incidence of agriculture and manufacture sectors' value added on total regional GDP. *Q_Self-employedSelf-employed* is the weight of self-employed on total employees. In the thematic area of Employment and

[8] The list of all the variables discarded by the PCA and relative results are available from the authors upon request.
[9] For details on the effects of 'studi di settore' audit scheme on tax compliance, see Santoro and Fiorio (2011).
[10] Since their institution in 1993 by law n.427, 'studi di settore' require that taxpayers subject to this audit scheme must attach to their tax return file a form containing information required in order to estimate their revenues. While they are obliged to fill in the form, they are not obliged to respect the criteria of congruity and coherence.
[11] Endogeneity issues are eluded as this variable measures the ratio between not congruous taxpayers and total taxpayers subjected to 'studi di settore', it does not refer to any amount of collectable and evaded taxes

Education (as shown in Table 2), using the principal components analysis, we discarded all variables related to the level of education because redundant and retain only the quota of self-employed individuals. *Pos_pc*, measures the number of points of sale (per resident) in the region. It can be considered as a proxy of the diffusion of the electronic money in the area and we expect a negative coefficient, as a more spread use of electronic money should reduce evasion.

Deposits_pc is the per capita amount of bank deposits. The use of *Deposits_pc* as a potential factor influencing tax evasion is justified by the fact that the amount of bank deposits can be connected with the use of cash in undeclared business operations. In fact, we can expect that the higher tax evasion especially in sectors connected to the use of cash (like for example, retail), the higher the average amount of cash deposits. Moreover, if we think of per capita amount of bank deposits as an indicator of financial wealth, we can expect higher tax evasion the higher financial wealth. For these reasons, we can question about possible endogeneity issue connected to the use of the variable D*eposits_pc* that could include tax evasion. Nevertheless, the use of a panel approach allows us to avoid endogeneity problems. Finally, the amount of bank deposits is often used by revenue agencies as one possible indicator to select taxpayers to be audited.

Both *Pos_pc*, and *Deposits_pc* variables are inserted as logs.

Size is the average number of employees as it can represent a proxy of average firm size. From the thematic area "Crime and Inequality", the selected index *Crime* measures the presence of organized crime in the region. The index is calculated as the number of crimes (weighted for their seriousness conferred by the Italian penal law) committed by organized crime associations every 1000 inhabitants. The inclusion of this variable leads to the exclusion of other variables like income poverty and inequality (regional Gini index) indices because redundant as resulting from the principal component analysis.

D_Gdp is an indicator of the time-variation of regional GDP. It is a binary variables that measures value 1 if in the *t*-year an increase of regional GDP occurred, 0 otherwise. This variable is included as a proxy of the business cycle.

Tax_amnesties indicates the amount of resources received after fiscal amnesties inserted in logs due to large disparities in regional values.

5. Results

5.1 Specification 1: Possible causes of tax gap

The first column of Table 4 contains results for model [3.1] estimated trough an OLS with robust standard errors. Coefficients are mostly significant with the expected sign and provide a first evidence of the choice of the covariates to explain the dependent variable, i.e. a first insight into the determinants of percentage of tax gap.

Results from least square estimation show that firms' size, electronic money, the revision of the auditing scheme (*studi di settore*) and the time variation of regional GDP have a negative and significant impact on the regional percentage tax gap. On the

contrary, positive and significant determinants are agriculture value added quota, the level of crime, per capita level of deposits and, finally, the relative weight of self-employed on total number of employees.

Even though the analysis of OLS coefficients are in line with expected hypotheses on the determinants of percentage tax gap, they ignore the longitudinal features of our dataset. Thus, pooling estimation may suffer from mis-specification. Moreover, the analysis of residuals (Table 5) and results of Breusch-Pagan and Wooldridge tests suggest that OLS estimators are inconsistent, given the correlation between the residual component (u_{it}) and the covariates. The variability captured by the individual error term is very large with respect to total variability (θ =0.8284) confirming the validity of the choice of an alternative estimator. Such evidence confirms the impression derived from descriptive analysis on the opportunity to use a panel model-specification as the [3.2].

5.2 Specification 2: Dynamic determinants of tax evasion

Table 4 (columns 2 and 3) show estimated coefficients of a random effects model specification. Results of the Hausman test are reported in Table 5 together with Wooldridge's test (Wooldridge, 2002). Both tests confirm the presence of unobserved individual (regional) effects. The Hausman test does not allow to reject the null hypothesis of equal coefficients between FE and RE models. As in the OLS model, RE model coefficients have the expected sign and a stronger significance level.

All the considerations made in the previous section on the determinants of percentage regional tax gap remain valid in the case of the RE model. In addition, using a panel specification, three more variables turn out to be significant. The first is the proportion of value added of the Industry sector, which has a negative effect. As expected and in line with estimates of the National Statistics Institute, firms belonging to the industrial sector have a lower propensity to evade. In fact, given the Italian fiscal law and tax procedural obligations that bigger sized firms must meet, opportunities to evade taxes are much lower than for less organized small firms and self-employed. The second is the variable indicating the probability to be audited that can be considered as a proxy of the IRA enforcement power. Its coefficient turns significant keeping the negative sign and indicating a positive effect of the IRA activity on tax compliance (see, among others, Alm, 1999 and Yitzhaki, 1974). Finally, also coefficient of tax amnesties turns significant indicating that, as expected, such legal provisions are perceived from taxpayers as opportunities to legalize their fiscal position and they can from time to time represent an incentive to evade.

In order to check if a static panel model is a right specification, we test assumptions on serial and cross-sectional independence of disturbances, i.e. if u_{it} is independent with $u_{i,t+k}$, for k=1,…T, and if $u_{i,t}$ is independent with $u_{i,j}$ every $i \neq j$. We use Baltagi-Li and Breusch-Goedfrey tests (Baltagi and Li, 1995; Godfrey, 1978) for serial correlation and Pesaran CD test (Pesaran, 2004) for cross-sectional correlation. Specifically, Pesaran CD test is implemented on the residuals of model [3.2] pre-treated as an AR structure to avoid serial correlation. Results (Table 5) lead us to reject both hypotheses of independence. Therefore, we use a dynamic panel specification [3.4] and a GMM estimator à la Arellano and Bond (1991).

Estimation results are presented in column 3 of Table 4. They show the presence of a

significant (as expected) relationship between the propensity to tax evasion at time *t* and *t+1* and, consequently, that percentage tax gap is characterized by a time-persistence pattern. The result of the Sargan test supports our choice of the instruments included in the model, i.e. the lagged values of the dependent variable.

Nevertheless, even with expected results and a better specification given the longitudinal dimension of our dataset and the persistence of tax evasion, specification [3.4] does not allow to correct serial and cross sectional dependence in the disturbances, given the results of AR(1), AR(2), Breusch-Goedfrey and Pesaran CD tests reported in Table 5.

5.3 Specification 3: Proximity as an explaining factor of tax compliance

Considerations contained in the previous section on the presence of both serial and cross-sectional correlation and results of tests contained in column 2 and 3 of Table 5 lead us to switch to a model specification able to handle the features of our data. In fact, we observe a time persistency in regional tax gap intensities as well as a spatial dependence in tax evasion pattern (see figure 2). For instance, the fact that taxpayers in one region are more virtuous given all the features considered in our model, might influence the behavior of neighbor taxpayers in order to improve tax compliance and vice versa[12]. Thus, we want to show that there can be (negative) positive spatial spillovers in attitude toward fiscal (non)compliance.

Before using spatial regression models, we check if spatial correlation exists. Values of Moran's I-statistics of PTG between Italian regions calculated for every year of analysis (2001-2011) are reported in Table 3 together with *p*-values. We can reject the hypothesis of absence of spatial covariance between PTG regional levels.

Values of I-indexes and results of the Moran's test for the null hypothesis of spatial independence (Cliff and Ord, 1981), obtained using W as spatial weights matrix, lead to try to modeling the strong spatial dependence of regional percentage tax gap. Moreover, the local variant of the CD test, CD(p) test (Pesaran et al., 2011) that takes into account spatial weight matrix W to check the null of no cross-sectional dependence against the alternative of local cross-sectional dependence, i.e. dependence between neighbors only is applied on the residuals of the models [3.4] and [3.6] (Millo and Piras, 2012 and Millo, 2014). The results are reported in Table 5 (column 3).

Results of the CD(p) test indicate that cross sectional correlation of the disturbances of the random and dynamic random models can be treated as spatial correlation. Thus PTG is influenced by time and proximity persistence as already derived from the analysis of Figure 1 and Figure 2, which clearly show that regions with high values of average PTG are located near other regions with high values and vice versa. Our third specification is then a SARSAR model that overcomes the problems of obtaining inefficient coefficients in presence of spatial correlation (Elhorst, 2003). Results are reported in Table 4 column 4.

The coefficient of the spatial lagged dependent variable, strongly significant and

[12] A similar issue has been analyzed by Arvate and Mattos (2008) with respect to efficiency in tax collection among Brazilian municipalities.

positive, confirms our a priori idea that tax evasion phenomenon is influenced by time and spatial persistence: in particular, the spatial proximity to a region with high (low) levels of PTG is a significant determinant of the high (low) PTG of a neighbor region. Thus, the introduction of the spatial lag variable as additional covariate suggests that spillovers effects maybe in action among neighbor regions in influencing taxpayers' attitude toward compliance.

All other covariates, with exception of the value added of industry, remain significant and with expected signs. Therefore, the SARSAR model appears as a better specification in order to test determinants of tax gap.

The firms' size indicated by the number of employees has a negative impact on the percentage tax gap. As stated by the Italian fiscal law, large firms are subjects to several additional duties and obligations in order to complete their tax return forms than smaller firms. Moreover, due to specific characteristics of business conduct[13] and a higher number of controls from IRA[14], as firm's dimension increases there is less room for tax evasion[15]

The same strong, significant and negative effect is connected with the relative diffusion of electronic money, i.e. the number of per capita points of sale. There is a large debate on the positive relation between the amount of transactions in cash and the level of tax evasion. This debate relates to a large evidence, here confirmed, on the link between the use of cash and increasing opportunities to evade taxes (see among the recent Immordino et al., 2014). Moreover, the currency demand is one of the many approaches employed to estimate the extent of tax evasion and underground economy (Ardizzi et al., 2014). On the contrary, even from the debate on possible legal interventions on the limitation of cash in transactions, the use of electronic money is often indicated as a deterrence for tax evasion. It should also be noted that the negative coefficient of per capita points of sale variable is related to the contribution of tax evasion mainly from sector involved in retail sales and provision of household services.

The effect of Agriculture value added quota is also significant and shows a great incidence of this sector on percentage tax gap. This result confirms other descriptive analyses conducted by the Italian National Statistics Institute on the sectorial composition of underground economy[16]. Similarly, it is not surprising that the quota of value added generated by the Industry sector does not contribute to tax gap.

The variable for IRA enforcement is negative indicating that an increase in the probability to be audited has a significant effect on the reduction of percentage of tax gap. This is a somehow expected result of our analysis that turns significant even though the probability to be audited for Italian firms considered does not have a great variability over time. Moreover, this result is also in line with those presented in

[13] As states by the Italian law, large firms are subject to a specific 'tutoring' activity by the Italian Revenue Agency consisting in a deep and long-lasting fiscal assistance. Moreover, large firms are subject to several additional rules involving the spread use of electronic payment and invoicing in business conduct that limit the opportunities of evasion.

[14] The numbers of audits from IRA on relatively large firms is higher than for smaller firms and self-employed.

[15] It should be noted that estimates tax gap used in this analysis refers strictly to the evasion connected to the exercise of activity of production of goods and services. Thus it does not consider the phenomenon of tax avoidance or specific tax frauds that can increase the incidence of tax evasion of larger firms.

[16] For latest estimates see ISTAT(2010) available at http://www.istat.it/it/archivio/4384

(Marigliani and Pisani, 2014 and Pisani, 2014)[17] .

The effect of crime is positive and significant even though it is not as strong as expected. This can be due to measurement errors in the crime variable or to the fact that our variable considers different types of illegal activities. We expect that a more specific indicator of illegal activities, for example related just in the economic sector, could have a stronger effect on tax evasion. Nonetheless, the inclusion of crime among explanatory variables confirms the results obtained also by Dell'Anno and Schneider (2006). In fact, if we attach to the crime variable the double meaning of rate of illegality and efficacy of the police force and of a perception of police protection, the variable crime can also have a negative effect on tax gap percentage, as it is perceived a cost to participate in illegal activities (Eilat and Zinnes, 2000).

As also in (Marigliani and Pisani, 2014 and Pisani, 2014), the amount of resources received after fiscal amnesties (*Tax_amnesties*) significantly determines the percentage of tax gap. This result confirms our a priori idea of a positive effect as tax amnesties can implicitly represent an incentive for taxpayers to evade tax payment with the perspective to legalize their fiscal obligations in a future moment.

As the crime variable, a double meaning can be also attached to the coefficient of per capita amount of bank deposits, which makes it difficult to have an a priori hypothesis on its sign. On the one side, it can be considered as a proxy of the regional level of wealth that itself can have a positive (cyclical) or negative (anticyclical) effect on tax gap. In our case, considering the amount of deposits as a measure of relative richness of the region, it has positive effects on tax evasion indicating that an increase in economic activity proportionally increases also underground economy[18]. On the other hand, considering the increasing development of the banking system, it is easier for people to put more money into deposits, even if they are derived from tax evading activities, instead of keeping this money as currency (Giles and Tedds, 2002).

As expected, the incidence of self-employed on total labor force has a strong and significant positive effect on percentage tax gap. As in Bordignon and Zanardi (1997) and Dell'Anno and Schneider (2006), the positive sign of the coefficient reflects one of the peculiarities of the Italian productive system characterized by a large proportion of small firms, professionals and self-employed with respect to the total workforce. Ceteris paribus, tax evading chances are clearly larger for self-employed than for employees (Braiotta et al., 2015).

Such inequality is also recorded for: the UK, where the tax gap from individual in self-assessment is 17% of the tax liabilities and the same percentage for the whole taxpayers is 7%[19], and Denmark, where the evasion rate for individual with self-reported income is equal to 37%[20].

[17] The indicator for the IRA enforcement power used in their analysis is slightly different and is calculated as the ratio of the amounts collected through the work of preventing and tackling evasion and total tax gap used as average of the ratio in the two years prior to those for which it is estimated the propensity to gap.

[18] For these reasons and considering that deposits are used as variable to select taxpayers to be audited, deposits the amount of deposits may itself contain the amount of evasion, so that being endogenous. Nevertheless, it should be noted that we control for endogeneity in the specification of panel model and that our dependent variable is not the level of tax evasion but percentage tax gap

[19] See HM Revenue & Customs (2013), pp. 7 and 36.

[20] See Klever et al. (2010).

As also expected, the sign of the lagged variable related to the number of not congruous on total taxpayers who joined the *studi di settore* audit scheme is negative and significant. In fact, not congruous taxpayers have a higher probability to be audited by IRA and their number negatively influences the power of the *studi di settore* scheme in inducing tax compliance.

Finally, time variation of regional GDP has a negative effect on percentage tax gap. The debate on the relationship between business cycle and tax evasion is large and unresolved (Giles, 1999; Marigliani and Pisani, 2014, Chiarini and Marzano, 2008, Caballe and Panades, 2000 among others). Our results from the pooled OLS model indicate that the effect on tax gap of an expansion of GDP is negative and can be linked to two drivers: the first is the reduction of demand for underground products as GDP increases; the second is the fact that a higher economic growth can induce more job opportunities into regular economy driving a reduction of opportunities to evade taxes[21].

Together with the analysis of coefficients and in order to assess the right of model specification, we perform several checks. Results of the tests of cross sectional and local cross sectional dependence, like Pesaran and Baltagi (Baltagi et al., 2003) and tests of serial correlation on disturbances of model [3.6] lead us to reject the alternative of dependence and serial correlation (Table 5). These results confirm the goodness of the choice of the SARSAR specification and that estimated coefficients are not biased by the presence of unobservable (time and spatially varying) effects

6. Concluding remarks

The analysis of the determinants of tax gap has always attracted large theoretical and, more recently, empirical research. Despite the growing number of studies about factors that influence taxpayers' attitude toward compliance, the effect of proximity on the level of tax gap has not been yet recognized. Moreover, from an empirical point of view, the use of spatial econometrics has not been exploited in the analysis of tax evasion.

This aim of the paper is then to provide an original contribution to the debate on the determining factors of tax evasion at regional level using different econometric specifications, with a focus on spatial econometric models. As far as we are concern, ours is the first example of such analysis applied to the Italian case.

Results support the hypothesis that spatial interactions have a significant and positive role in determining the regional levels of percentage tax gap. Tax evasion is then characterized by time and spatial persistence. Moreover, our other a priori ideas on additional influencing factors find a significant support in the empirical section of the paper.

[21] Our results on the relationship between GDP growth and tax evasion confirm those obtained by Dell'Anno and Schneider (2006) between GDP increase and the size of underground economy.

The econometric analysis is based upon the estimation of three different models: pooled OLS, static and dynamic panel and spatial dynamic panel models. As our focus is to test the presence of spatial correlation in regional tax gap and how this influences the level of tax compliance, we also run several tests to show how the first two specifications, pooled OLS and panel data analysis, provide inconsistent and inefficient coefficients due to the specificity of our longitudinal database. Therefore, spatial dynamic panel models are the most suitable specification in order to correct for spatial correlation in the residual component.

Results show that the percentage of tax gap with respect to total voluntary tax returns is affected by serial and spatial correlation, and in particular, a high spatial correlation in tax gap intensities among Italian regions. Local level of tax evasion appears to react to neighbor attitudes toward tax compliance. Once corrected, our estimates indicate that proximity can be considered as an additional significant and positive determinant of tax gap: a regional (low) high of tax gap intensity is determined by (low) high tax gap in the neighbor areas. Spatial correlation indicates a mimicking effect in tax evading behavior among Italian regions. Regarding the other covariates considered in our models, results indicate that percentage of tax gap is positively influenced by the proportion of value added generated by the agricultural sector, the rate of criminality, the amount of tax amnesties, the level of per capita deposits and the incidence of self-employed on total workforce. Otherwise, the number of points of sale, the revision of the *studi di settore* auditing scheme and the business cycle negatively determine tax evasion. Therefore, the analysis of significant estimated coefficients from SARSAR model specification enable us to draw a taxonomy of determinants of tax gap: environmental factors linked to the economic and institutional features (agriculture value added, diffusion of electronic money, bank deposits, incidence of self-employment, amount paid after tax amnesties and GDP growth); operational factors linked to the relative efficacy of tax evasion contrasting policies (IRA enforcement and renewal of auditing schemes) and spillovers factors from neighbor regions (level of crime and tax gap of near regions).

Our analysis represents a first systematic attempt to investigate different determinants of tax gap and the role of geography and proximity in explaining tax evasion pattern. Even though our results are informative, they need additional research. Future and further analysis can be done in the direction of checking if spatial correlation in tax gaps exists also at the provincial level as well as if further economic and institutional factors can help explaining tax compliance attitude and orient future auditing activity of tax administration.

Nevertheless, our results suggest how policy interventions should consider also spillovers (negative and positive) effects among contiguous geographical areas in tax evasion contrasting strategies. In fact, from one point of view, we show how proximity can be harmful for tax compliance as high tax gap regions can influence their neighbors. From another, an effective auditing scheme designed for a geographical area can have positive effects also in the nearby areas. Similarly, as one analyzes all the components included in the tax gap and, in general, in the measurement of hidden economy, the relevance of our results apply also to production district, supply chains and employees' internal migration. A policy intervention by fiscal agencies targeted on sectors and firms geographically correlated can have a positive and multiplier effect on tax compliance improving strategies.

References

Allingham M. G., Sandmo A., (1972), "Income tax evasion: a theoretical analysis", Journal of public Economics, 1, 323 338.

Alm, J. (1999) Tax compliance and administration, Public Administration and Public Policy 72, 741-768

Alm J., (2012) 'Measuring, explaining, and controlling tax evasion: lessons from theory, experiments, and field studies,' International Tax and Public Finance, Springer, vol. 19(1), pages 54-77, February.

Anselin, L. (1988), Spatial Econometrics: Methods and Models, Dorddrecht: Kluwer Academic Publishers.

Ardizzi, G., C. Petraglia, M. Piacenza and G. Turati (2014) 'Measuring the Underground Economy with the Currency Demand Approach: A Reinterpretation of the Methodology, with an Application to Italy'Review of Income and Wealth, Vol. 60, Issue 4, pp. 747-772, 2014

Arellano M. and Bond S., (1991), Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. Review of Economic Studies, 58, 277-297.

Arvate, P. and Mattos, E. (2008) 'Efficiency in tax collection: evidence from Brazilian municipalities', CEPESP and EESP -Getulio Vargas Foundation, working paper N. 01/2008.

Braiotta A., Carfora A., Pansini R., Pisani S. (2015). Tax Gap and Redistributive Aspects across Italy, Argomenti di Discussione, N. 02/2015, IRA.

Baltagi B. and Li Q., (1995), Testing AR(1) Against MA(1) Disturbances in an Error Component Model, Journal of Econometrics, 68, 133-151.

Baltagi, B.H., Song, S.H. and Koh, W. (2003) Testing panel data regression models with spatial error correlation. Journal of Econometrics, 117, 123–150

Bordignon, M. and A. Zanardi, (1997) 'Tax Evasion in Italy,' Giornale degli Economisti e Annali di Economia, vol. 56(3-4), pages 169-210, December.

Buehn, A. and F. Schneider, (2013) 'Shadow Economies in highly developed OECD countries: What are the driving forces?,' Economics working papers 2013-17, Department of Economics, Johannes Kepler University Linz, Austria.

Caballe, J., Panades, J. (2000) Tax evasion and Economic Growth', Public Finance / Finances Publiques 52, 318-340

Carbone, E. and A. Spingola (2015) Analisi delle Direzioni Provinciali attraverso il DB Geo, Argomenti di Discussione, N. 01/2015, IRA.

Chiarini, B., E. Marzano and F. Schneider, (2008) 'Tax Rates and Tax Evasion: An Empirical Analysis of the Structural Aspects and Long-Run Characteristics in Italy', IZA Discussion Papers N0. 3447.

Cliff A.D. and Ord, J.K., (1981), Spatial processes: models and applications, Oion LTD, London.

Clotfelter, C T. (1983) "Tax Evasion and Tax Rates: An Analysis of Individual Returns."Review of Economics and Statistics, Vol. 65, No. 3 (August), pp. 363-373

D'Agosto E., Marigliani M., Pisani S., (2014), Asymmetries in the territorial VAT gap, Argomenti di Discussione, N. 02/2014, IRA.

Das-Gupta A., Mookherjee D. (2000) Incentives and institutional reform in tax enforcement an analysis of developing country experience. Lavoisier.

Depalo, G. and G. Messina (2011), Follow the Herd. Spatial Interactions in tax setting behavior of Italian municipalities, Banca d'Italia.

Dell'Anno, R. and F. G. Schneider, (2006) 'Estimating the underground economy by using MIMIC models: A response to T. Breusch´s critique,' Economics working papers 2006-07, Department of Economics, Johannes Kepler University Linz, Austria.

Eilat and C. Zinnes, (2000) The Evolution of the Shadow Economy in Transition Countries: Consequences for Economic Growth and Donor Assistance', CAER II Discussion Paper No. 83. Harvard Institute for International Development

Elhorst, J.P. (2003), 'Specification and Estimation of Spatial Panel Data Models'. International Regional Science Review, 26: 244-268.

Elhorst, J.P. (2010), 'Spatial Panel Data Models'. In Fischer MM, Getis A (Eds.) Handbook of Applied Spatial Analysis, pp. 377-407. Springer: Berlin Heidelberg New York.

Eurostat, (2014), "Taxation Trends in the European Union, Taxation and Custom Union.

Feld, L. P. and B. S. Frey (2002). Trust Breeds Trust: How Taxpayers are Treated, Economics of Governance. 3: 87-99.

Giles, D. E. A., (1999), 'Measuring the hidden Economy: Implications for Econometric Modeling', The Economic Journal, 109, No 46, pp.370-380.

Giles, D. E. A. and Tedds, L. M. (2002) 'Taxes and the Canadian Underground Economy', Canadian Tax paper n.106, Canadian Tax Foundation. Toronto.

Godfrey L., (1978), Testing against general autoregressive and moving average error models when the regressors include lagged dependent variables, Econometrica, 46, 1293–1302.

Hausman J.A., (1978), Specification tests in econometrics, Econometrica, 46, 1251–1271.

HM Revenue & Customs (2013) "Measuring tax gaps", London, UK.

ISTAT, (2010), "La misura dell'economia sommersa secondo le statistiche ufficiali. Anni 2000-2008", Statistiche in breve, 13 luglio.

Kapoor M., Kelejian H.H. and Prucha I.R., (2007), Panel Data Model with Spatially Correlated Error Components. Journal of Econometrics, 140(1), 97-130.

Kelejian H.H. and Prucha I.R., (1999), A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Model, International Economic Review, 40(2), 509-533.

King J.R. and Jackson D.A. (1999). Variable selection in large environmental data sets using principal components analysis, Environmetrics, 10, 67-77.

Kleven H. J., Knudsen M.B., Kreiner C. T., Pedersen S., Saez E. (2010) "Unwilling or unable to cheat? Evidence from a randomized tax audit experiment in Denmak" NBER working paper series, n.15769, February, Cambridge, MA.

Immordino, G and F. F. Russo (2014) 'Taxing Cash to Fight Collaborative Tax Evasion?, Working paper No. 351, Department of Economics, University of Naples.

Marigliani, M. and S. Pisani (2014) Una stima dell'effetto deterrenza esercitato dall'Agenzia delle Entrate, IRA, mimeo.

Marino, M. R and R. Zizza, (2012) 'Personal Income Tax Evasion in Italy: An Estimate by Taxpayer Type,' in: Tax Evasion and the Shadow Economy, chapter 3 Edward Elgar.

McGee R. W., (2014), 'Gender and the Ethics of Tax Evasion: An Empirical Study of 82 Countries,' Journal of Accounting, Ethics and Public Policy, April 14, 2014

Millo G. and Piras G., (2012), splm: Spatial Panel Data Models in R, Journal of Statistical Software, 47, 1.

Millo G., (2014), "Narrow Replication of 'A Spatio-temporal Model of House Prices in the USA' Using R", Journal of Applied Econometrics. doi: 10.1002/jae.2424

Moran P. A. P., (1950), Notes on Continuous Stochastic Phenomena, Biometrika, 1/2, 17-23.

Mutl J. and Pfaffermayr M., (2011), The Hausman Test in a Cliff and Ord Panel Model, Econometrics Journal, 14, 48-76.

Pesaran M., (2004), General Diagnostic Tests for Cross Section Dependence in Panels, CESifo Working Paper Series, 1229.

Pesaran H.M. and Tosetti E., (2011), Large Panels with Common Factors and Spatial Correlations, Journal of Econometrics, 161(2), 182-202.

Pisani, S., (2014), An Approach to Access How the Activity of the Italian Revenue Agency Affects Compliance, Argomenti di Discussione dell'Agenzia delle Entrate, 01/2014

Richardson, G. (2006). Determinants of tax evasion: A cross-country investigation. Journal of International Accounting, Auditing and Taxation,15(2), 150 – 169.

Santoro A. and Fiorio C.V., (2011). Taxpayer Behavior When Audit Rules Are Known: Evidence from Italy, Public Finance Review, vol. 39(1), 103-123.

Schneider F., Enste D.H., (2000), "Shadow Economies: Size, causes, and consequences", Journal of Economic Literature, 38, pp. 77 -114.

Wooldridge J.M., (2002), Econometric Analysis of Cross-Section and Panel Data, MIT Press, Sec. 10.4.4., 264.

Yalama, G. Oz and E. Gumus (2013) 'Determinants of Tax Evasion Behavior: Empirical Evidence from Survey Data', International Business and Management, Vol. 6, No. 2, 2013, pp. 15-23

Yitzhaki, S. (1974), ‘A note on 'income tax evasion: a theoretical analysis’, Journal of Public Economics 3(2):201-202.

7. Tables and Figures

Table 1 Summary statistics of the regional 2001-2011-distributions of PTG

Region	min	I Quartile	Median	III Quartile	max	IQ range
Abruzzi	0.24	0.28	0.29	0.35	0.50	0.07
Basilicata	0.76	0.80	0.88	0.94	0.99	0.14
Calabria	0.69	0.75	0.87	0.99	1.12	0.24
Campania	0.45	0.46	0.53	0.59	0.70	0.13
Emilia Romagna	0.17	0.20	0.22	0.24	0.24	0.03
Friuli V.G.	0.19	0.22	0.25	0.26	0.29	0.04
Lazio	0.12	0.14	0.15	0.17	0.17	0.03
Liguria	0.21	0.23	0.26	0.26	0.27	0.02
Lombardy	0.16	0.20	0.22	0.28	0.30	0.08
Marche	0.33	0.39	0.40	0.43	0.44	0.04
Molise	0.58	0.67	0.69	0.76	0.81	0.09
Piedmont	0.22	0.24	0.27	0.28	0.29	0.05
Puglia	0.48	0.51	0.61	0.69	0.81	0.18
Sardinia	0.35	0.36	0.40	0.56	0.68	0.20
Sicily	0.40	0.43	0.51	0.55	0.89	0.12
Tuscany	0.24	0.30	0.32	0.34	0.35	0.04
Trentino A.A.	0.18	0.19	0.20	0.24	0.30	0.05
Umbria	0.31	0.39	0.43	0.44	0.49	0.04
Val d'Aosta	0.23	0.26	0.32	0.33	0.35	0.06
Veneto	0.23	0.27	0.29	0.31	0.35	0.04

Figure 1 2001-2011 distribution of percentage Tax Gap (PTG) for Italian regions

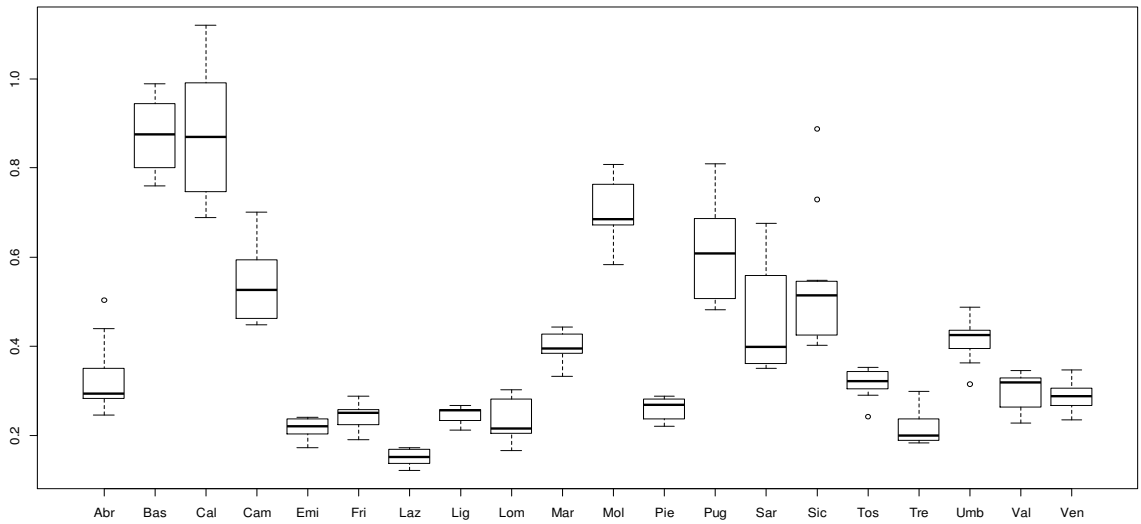

Table 2 Groups and thematic areas of covariates

Group	*Thematic areas*	*Variable*	*label*
Operational	Revenue Agency policies	Strength of sectorial studies	***Ren_ss***
		Index of audit coverage	***IRA_enforcement***
Context	Production area	Agriculture GDP quota	***Q_Agriculture***
		Industry GDP quota	***Q_Industry***
	Employment and Education	Self- employed	***Q_self-employed***
	Electronic money	Number of Pos per resident	***Pos_pc***
	Income and Private Saving	Bank Deposits	***Deposits_pc***
	Firm size	Employees	***Size***
	Crime and Inequality	Crime index	***Crime***
	Economic cycle	GDP variation rate	***D_gdp***
	Government policies	Amount of tax amnesties	***Tax_amnesties***

Figure 2 Average PTG: regional distributions over the period 2001-2011.

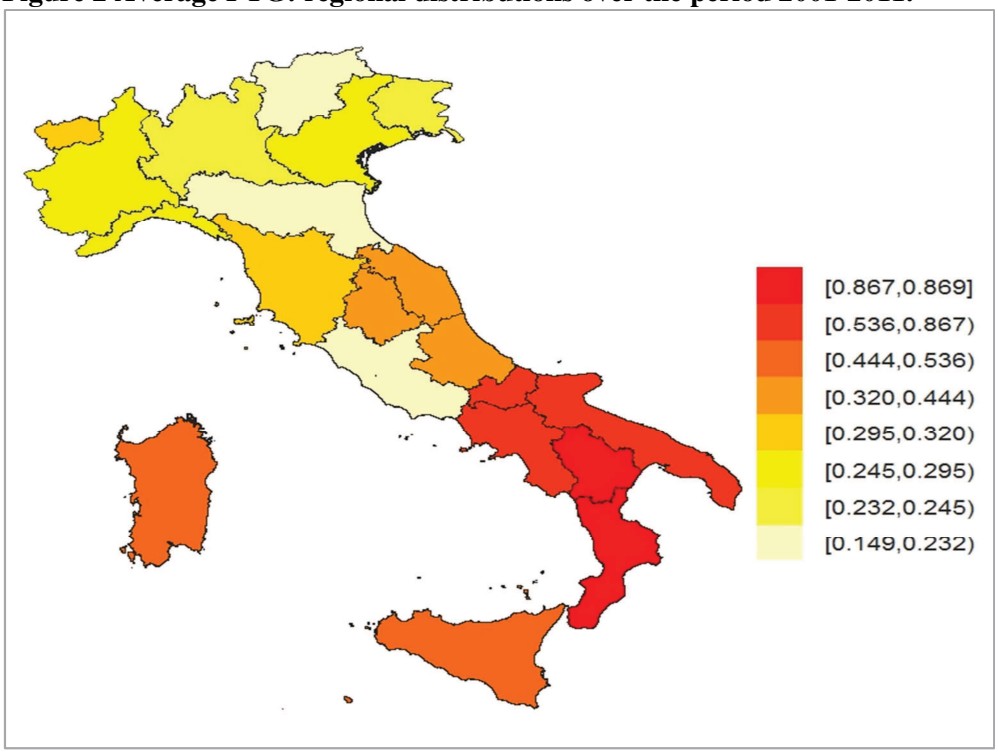

Table 3 Spatial autocorrelation of regional PTG

Year	I-statistics	Moran's test (p-value)
2001	0.6553	0.0000
2002	0.7016	0.0000
2003	0.5697	0.0000
2004	0.5230	0.0000
2005	0.5307	0.0000
2006	0.5870	0.0000
2007	0.5489	0.0000
2008	0.4840	0.0000
2009	0.4841	0.0000
2010	0.4448	0.0000
2011	0.4785	0.0000

Table 4 Dynamics of tax gap determinants

	OLS	Random	Dynamic	SARSAR
	(1)	(2)	(3)	(4)
Intercept	-0.0294	0.2012	-	0.2338***
	(0.8733)	(0.1358)		(0.0017)
Spatial lag(y)	-	-	0.1936*	0.5852***
			(0.0570)	(0.0000)
Size	-0.1277***	-0.1437***	-0.0893**	-0.1025***
	(0.0000)	(0.0000)	(0.0151)	(0.0000)
POS_pc	-0.2607***	-0.2475***	-0.1862***	-0.1140***
	(0.0000)	(0.0000)	(0.0040)	(0.0000)
Q_Agriculture	5.2179***	2.7467***	2.1646	2.3512***
	(0.0000)	(0.0084)	(0.2704)	(0.0006)
Q_Industry	0.1818	-0.5710**	-0.5885	-0.0438
	(0.2006)	(0.0262)	(0.1121)	(0.7770)
IRA_enforcement	-0.1526	-0.7221**	-0.5964	-0.5332**
	(0.6862)	(0.0168)	(0.1173)	(0.0104)
Crime	0.0013**	0.0016***	0.0011***	0.0016***
	(0.0316)	(0.0025)	(0.1879)	(0.0001)
Tax_Amnesties	0.0012	0.0086***	0.0075***	0.0066***
	(0.7164)	(0.0003)	(0.0009)	(0.0000)
Deposits_pc(log)	0.1098**	0.1050**	0.1489**	0.1035***
	(0.0181)	(0.0132)	(0.0394)	(0.0000)
Q_Self-employed	1.0611***	1.5352***	1.4859**	1.3254***
	(0.0001)	(0.0000)	(0.0289)	(0.0000)
lag(Ren_ss)	-0.2064**	-0.2079***	-0.2204***	-0.0991***
	(0.0310)	(0.0005)	(0.0000)	(0.0027)
D_gdp	-0.0317**	-0.0312***	-0.0329***	-0.0227***
	(0.0177)	(0.0001)	(0.0000)	(0.0000)

Notes: P-value are in parenthesis; Robust standard error; significance level: *** 1%, ** 5%, * 10%.

Table 5 Diagnostics of the models

	OLS	Random	Dynamic	SARSAR
	(1)	(2)	(3)	(4)
				Diagnostics
Adj.R2	0.8131	0.6182		
θ		0.8227		0.7191
ρ				-0.6832
Sargan (p-value)			0.9993	
Unobserved effects				
Breusch-Pagan	355.2111			
	(0.0000)			
Wooldridge	3.224			
	(0.0013)			
Hausman		9.764		
		(0.5517)		
Serial Correlation				
Baltagi and Li		14.6556		
		(0.0001)		
Breusch–Godfrey		39.7776	17.5579	13.2487
		(0.0004)	(0.0141)	(0.1517)
AR (1)			-2.7618	1.1525
			(0.0029)	(0.2830)
AR (2)			1.7453	
			(0.0405)	
Cross sectional correlation				
Pesaran CD		3.3166	2.5998	-0.2675
		(0.0009)	(0.0093)	(0.7891)
Local cross sectional correlation				
Pesaran CD (p)		7.6383	2.6839	0.2324
		(0.0000)	(0.0073)	(0.8162)
Baltagi Song, Koh				1.3698
				(0.1708)

Notes: P-value are in parenthesis.