

Il numero totale di cluster viene determinato utilizzando il *BIC (Bayesian Information Criterion)*, un indicatore della bontà di adattamento del modello, e sulla base di scelte che rendano possibile l'interpretazione dei risultati.

Al fine di analizzare in maniera più appropriata le caratteristiche del fenomeno delle frodi IVA, sono addestrati modelli distinti per dichiarazione IVA attiva (cessioni) e passiva (acquisti), e per i versamenti.

Esempio: modello per l'IVA sulle operazioni attive

A titolo d'esempio, si consideri il caso in cui la variabile target Y^* è l'imposta che dovrebbe essere dichiarata dal contribuente relativamente alle operazioni attive. Per quanto riguarda il modello per non dichiaranti, sono state selezionate le seguenti variabili di profilazione:

- totale IVA esposta in fattura e corrispettivi, espresso in scala logaritmica;
- rapporto tra l'IVA relativa agli acquisti e quella relativa alle cessioni;
- variabile dicotomica relativa alla data di attribuzione della partita IVA (tale variabile assume il valore 1 se la partita IVA è stata aperta negli ultimi due anni rispetto all'anno in esame e il valore 0 altrimenti);
- variabile dicotomica relativa all'iscrizione alla banca dati VIES (tale variabile assume il valore 1 se la partita IVA è presente in VIES e 0 altrimenti);
- differenza tra l'importo dell'IVA periodica dovuta e versata.

Per quanto riguarda invece il modello per coloro che hanno presentato la dichiarazione IVA, sono state considerate le seguenti variabili di profilazione Z :

- rapporto tra l'IVA relativa agli acquisti e quella relativa alle cessioni;
- variabile dicotomica relativa alla data di attribuzione della partita IVA;
- variabile dicotomica relativa all'iscrizione alla banca dati VIES.

Inoltre, come variabile X è stata utilizzata l'IVA sulle operazioni attive ricostruita sulla base delle informazioni derivanti dalla fatturazione elettronica e dai corrispettivi; invece come variabile Y è stato utilizzato il totale dell'imposta dichiarata sulle operazioni attive (quadro VE26 del modello dichiarativo).

III.4.3 Gli output

La clusterizzazione di tipo *model-based* dei soggetti analizzati consente di attribuire al soggetto *i-esimo*, la probabilità a posteriori di appartenere al k -esimo cluster G_k (classificazione Bayesiana) ovvero al gruppo associato al k -esimo livello di rischio:

$$\tau_k(y_i|z_i, x_i; \theta) = \frac{\pi_k(z_i)N_k(y_i)}{\sum_{h=0}^{K-1} \pi_h(z_i)N_h(y_i)}, \quad k = 0, \dots, K-1$$

dove le variabili Y assumono il significato appropriato a seconda che siano presenti i valori dichiarati oppure no, e, in quest'ultimo caso, le variabili Z sono omesse.

Partendo dalla specificazione del modello e utilizzando le regole algebriche della teoria della distribuzione Normale multivariata, è possibile ricavare i valori attesi condizionati della variabile target Y^* , cioè il valore corretto che il contribuente *i-esimo* dovrebbe dichiarare o versare:

$$E(Y_i^*|y_i, z_i, x_i; \theta) = \sum_{k=0}^K \tau_k(y_i, z_i, x_i; \theta) E(Y_i^*|y_i, z_i, x_i, i \in G_k; \theta).$$

Sulla base del valore atteso della variabile target Y^* è possibile ottenere una stima (previsione) del *gap totale individuale*

$$gap_i \equiv \pm(E(Y_i^*|y_i, z_i, x_i; \theta) - Y_i)$$

e delle componenti del gap associate ai vari cluster

$$gap_i^k \equiv \pm(\tau_k(y_i|z_i, x_i; \theta)E(Y_i^*|y_i, z_i, x_i, i \in G_k; \theta) - Y_i).$$

Nelle precedenti formule, il segno \pm dipende dalla natura della variabile Y_i che stiamo considerando: segno positivo (+) nel caso dell'IVA sulle operazioni attive (cessioni), segno negativo (-) nel caso dell'IVA sulle operazioni passive (acquisti).

La predetta profilazione può essere effettuata relativamente all'imposta dichiarata sulle operazioni attive o passive, ovvero all'imposta versata. I singoli profili possono essere unificati in modo da ottenere una profilazione globale del contribuente.

III.5 DERIVAZIONE DEL MODELLO ECONOMETRICO PER L'ETEROGENEITÀ DEGLI EFFETTI DELLA FATTURAZIONE ELETTRONICA

In questa appendice si descrive il modello utilizzato per l'analisi riportata al paragrafo V.I.3. In particolare, riprendendo il modello utilizzato nell'edizione 2020 della Relazione per valutare l'impatto di fatturazione elettronica (FE) sull'IVA esigibile, se ne illustrano gli sviluppi realizzati. Il modello, nella formulazione originaria, è rappresentato formula che segue:

$$(Z_2 - Z_1) = \alpha + \gamma q + \lambda'U + \varepsilon \quad (1)$$

dove Z_1 e Z_2 rappresentano (in scala logaritmica) l'IVA esigibile dichiarata nelle LIPE del primo semestre del 2018 e 2019 rispettivamente, U rappresenta un insieme di variabili di controllo⁵⁵, q identifica il rapporto tra il totale dell'IVA associata alle cessioni di beni e/o prestazioni di servizi di tipo B2B e il totale dell'IVA esigibile risultante dalle comunicazioni IVA relative allo stesso semestre.

Per quanto concerne la variabile q , in particolare, essa può assumere valori compresi tra 0 e 1, dove per valori pari, o prossimi, all'unità si è in presenza di un comportamento dichiarativo orientato verso la fatturazione elettronica mentre, per valori pari, o prossimi, allo zero si è in presenza di un gettito IVA del contribuente derivante da operazioni certificate tramite corrispettivi. Il termine ε , infine, rappresenta il residuo casuale che soddisfa le usuali ipotesi di incorrelazione e omoschedasticità, e $(\alpha, \gamma, \lambda)$ l'insieme di parametri da stimare.

Rispetto al modello di cui sopra, sono state apportate alcune modifiche al fine di verificare se il comportamento delle diverse tipologie di contribuenti fosse omogeneo. I risultati di tali estensioni sono sintetizzati nel paragrafo V.I.3 e dettagliati nel seguito.

La prima modifica apportata ha riguardato l'introduzione di una specifica variabile K per poter individuare i soggetti IRPEF, ossia persone fisiche e società di persone. Ciò ha, quindi, permesso di poter distinguere due platee di soggetti (IRPEF, noIRPEF) e di analizzarne, separatamente, il comportamento a seguito dell'introduzione della FE. Più precisamente, la variabile K è una variabile dicotomica, o *dummy*, che assume valore 1 in presenza di soggetti IRPEF e 0 altrimenti. Considerata la seguente riformulazione del modello:

$$(Z_2 - Z_1) = \alpha + \gamma q + \lambda'U + \delta K + \eta (K * q) \varepsilon \quad (2)$$

il coefficiente δ rappresenta la differenza di intercetta dei soggetti IRPEF rispetto ai soggetti noIRPEF, mentre il coefficiente η associato all'interazione tra K e q rappresenta la differenza di

⁵⁵ Le variabili di controllo sono inserite per sterilizzare l'effetto indotto da fattori esogeni alla FE quali, ad esempio, il settore di attività economica di appartenenza o alla localizzazione geografica.

pendenza. In altri termini quest'ultimo parametro identifica il diverso comportamento dei soggetti IRPEF in relazione alla fatturazione elettronica.

I risultati così ottenuti sono riportati nella Tabella III.5.1 che riporta i valori delle stime dei parametri, con i relativi errori standard, per le due sotto-platee di soggetti (IRPEF, noIRPEF).

TABELLA III.5.1: STIME DEI PARAMETRI DEL MODELLO 2 PER LA SEZIONE G		
Parametri	Coefficiente	Std.Error
q	0,0299***	0,0033
q*K	-0,0226***	0,0032
Costante	-0,0329***	0,0058
K	0,0037	0,0023
dummy settoriale	si	Si
dummy regionale	si	Si
N	393.147	
<i>Variabile dipendente:</i> IVA esigibile		
<i>K:</i> flag che vale 1 per i soggetti Irpef e 0 altrimenti		

Dai risultati emersi si desume un differente effetto della FE per le due sotto platee. Nello specifico, poiché il parametro associato al flag *K* risulta non significativo, i due gruppi condividono la stessa intercetta. D'altro canto, il parametro relativo all'interazione tra la variabile *q* e la variabile *K*, che, come richiamato in precedenza rappresenta la differenza tra il valore stimato del coefficiente della variabile *q* (parametro γ , positivo e circa pari a 0,03) e quello dei soggetti IRPEF, risulta negativo e pari -0,023. Ciò si traduce in un extraggettito pari all' 1,04% e al 0,6% per i soggetti, rispettivamente, non IRPEF e IRPEF. I risultati ottenuti permettono, dunque, di concludere che **l'extraggettito IVA riscontrato nella relazione evasione 2020, pari al 2%, sia da ascrivere principalmente al comportamento dei soggetti "noIRPEF"**.

Successivamente, entrambi i modelli sopra definiti nelle equazioni (1) e (2) sono stati utilizzati anche per valutare l'impatto della FE sul reddito da impresa e lavoro autonomo dichiarato, al fine di comprendere se il comportamento dell'intera popolazione considerata, e delle due sotto-platee individuate (IRPEF, noIRPEF), fosse coerente tra le due imposte considerate (IVA-IRPEF). La formulazione dei due modelli rimane, quindi, invariata, con l'unica eccezione della variabile di output *Z* che in questo caso non rappresenta l'IVA esigibile bensì il reddito da impresa e lavoro autonomo dichiarato⁵⁶.

Le Tabelle III.5.2 e III.5.3 riportano le risultanze del modello applicato sul reddito da impresa e lavoro autonomo. Nel dettaglio, i risultati in Tabella III.5.2 definiscono il comportamento dell'intera popolazione considerata, mentre nella Tabella III.5.3 si riportano i parametri per le singole sotto platee.

TABELLA III.5.2: STIME DEI PARAMETRI DEL MODELLO 1 PER LA SEZIONE G		
Parametri	Coefficiente	Std.Error
Q	-0,1096***	0,0174
Costante	-0,2023***	0,0459
dummy settoriale	si	Si
dummy regionale	si	Si
N	393.147	
<i>Variabile dipendente:</i> Reddito da impresa e lavoro autonomo		

⁵⁶ Si tenga presente come l'utilizzo del reddito da impresa e lavoro autonomo, e non dell'imposta, come variabile di output nel modello sull'IRPEF non distorce il confronto con l'IVA. In quest'ultimo caso, infatti, poiché l'aliquota media e mediana IVA per i due anni considerati (2018-2019) rimane pressoché costante e pari al 22%, replicare il modello sulle basi produrrebbe risultati equivalenti. D'altra parte, invece, l'utilizzo dell'imposta nel modello sui redditi rischierebbe di essere influenzato da altri fattori (deduzioni e detrazioni) non riguardanti l'attività d'impresa.

TABELLA III.5.3: STIME DEI PARAMETRI DEL MODELLO 2 PER LA SEZIONE G

Parametri	Coefficiente	Std.Error
q	-0,0081	0,0279
q*K	-0,0724 **	0,0266
Costante	-0,4249***	0,0492
K	0,2349***	0,0193
dummy settoriale	si	si
dummy regionale	si	si
N	393.147	

Variabile dipendente: Reddito da impresa e lavoro autonomo
K: flag che vale 1 per i soggetti Irpef e 0 altrimenti

I risultati ottenuti per l'intera platea mostrano un effetto FE sul reddito da impresa e lavoro autonomo contrario rispetto a quello sull'IVA esigibile (Tabella III.5.2). Il coefficiente della variabile q assume valore negativo e pari a -0,11, al quale corrisponde una contrazione del reddito da impresa e lavoro autonomo pari al 6,58%.

I risultati riportati nella Tabella III.5.3 indicano che l'effetto FE negativo riscontrato sull'intera platea è in realtà riconducibile al solo comportamento dei soggetti IRPEF. Il coefficiente relativo alla variabile q risulta essere non significativo, mentre quello dell'interazione con variabile *dummy* assume valore negativo determinando un effetto FE pari a -0,0805 che si traduce in una contrazione del reddito da lavoro autonomo e di impresa pari al 3,31%.

IV. RIFERIMENTI BIBLIOGRAFICI

Braiotta A., Carfora A., Pansini R.V., (2015), "Tax Gap and Redistributive Aspects across Italy", *Argomenti di Discussione dell'Agenzia delle Entrate*, n.2.

Braiotta A., Pisani S., Pisano E. (2013), "Evasione fiscale e distribuzione primaria del reddito", *Rivista di diritto finanziario e scienza delle finanze*, n. 2.

D'Agosto E., Marigliani M, Pisani S., (2016) "A general framework for measuring Vat compliance in Italy". *Argomenti di discussione Agenzia delle Entrate*, N.02/2016.

D'Agosto E., Marigliani M, Pisani S., (2014) "Asymmetries in the territorial VAT gap". *Argomenti di discussione Agenzia delle Entrate*, N.02/2014.

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

European Commission (2020), "Study and reports on the VAT gap in the EU-28 Member States 2020 final report".

Eurostat (2010) *European System of Account - ESA 2010*, Luxembourg.

Fiscalis Tax Gap Project Group (2018), *The concept of tax gaps - Report II: Corporate Income Tax Gap Estimation Methodologies*, Directorate General Taxation and Customs Union, European Commission.

Fiscalis Tax Gap Project Group (2018), "The concept of tax gaps - Report III: MTIC Fraud Gap estimation methodologies", Directorate General Taxation and Customs Union, European Commission

Gallucci M., Pansini R.V., Pisani S. (2020), *Direct taxes gap estimates: methodology and preliminary results*, *Argomenti di Discussione dell'Agenzia delle entrate*, n. 02/2020.

Heckman, James (1976). "The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models". *Annals of Economic and Social Measurement*, 5 (4): 475-492.

HM Revenue & Customs (2021), "Measuring tax gaps 2021 edition - tax gap estimates for 2019 to 2020", March, London

Imperioli L., Manzo M., Monteduro M. T. (2015), *Il tax gap sulle imposte immobiliari e la perequazione fiscale*, XXXVI Conferenza Italiana di Scienze Regionali.

Liberati C.; Pisani S.; Serra A. (2002) *Nota metodologica per effettuare un confronto tra i dati IRAP e di Contabilità Nazionale nel settore privato Documento di lavoro dell'Ufficio Studi*, n. 2, Agenzia delle Entrate.

Little, Roderick J. A. (1988). "Missing-Data Adjustments in Large Surveys". *Journal of Business & Economic Statistics*. 6 (3): 287-296.

Mavrokonstantis P.; Seibold A. (2022), "Bunching and adjustment costs: Evidence from cyriot tax reforms", *Journal of Public Economics* 214: 104727

McParland D., Gormley I. C. (2015). "Model based clustering for mixed data: clustMD".

Murphy K., Murphy T. B. (2020). "Gaussian parsimonious clustering models with covariates and a noise component". *Advances in Data Analysis and Classification*. 14 (2): 293-325.

Pisani S., (2014), "An approach to assess how the activity of the Italian Revenue Agency affects compliance", *Argomenti di Discussione dell'Agenzia delle Entrate*, n.1.

Pisani S., Polito C (2006), "Metodologia di integrazione tra i dati IRAP e quelli di Contabilità Nazionale", Agenzia delle Entrate, Ufficio Studi, documenti di lavoro

Reckon, (2009), "Study to quantify and analyze the VAT gap in the EU-25 Member States", Working paper Reckon, 21 September, London, UK.

Ueda J. (2018), *Estimating the Corporate Income Tax Gap: the RA Gap Methodology*, IMF Technical Notes and Manual, August, Washington.