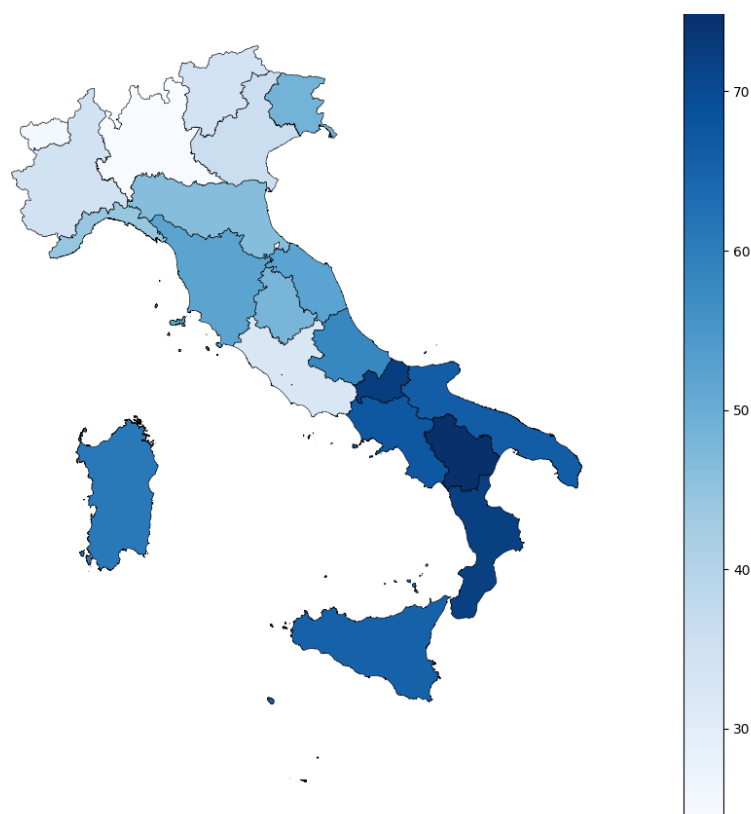


Il campione così identificato, subordinatamente ad opportune ipotesi di stabilità delle relazioni tra predittori e variabili risposta, è potenzialmente utilizzabile per stimare il *tax gap* riguardante gli anni dal 2014 in poi. Nella sperimentazione condotta, come sopra anticipato, l'anno di interesse è il 2021. Per l'illustrazione dettagliata della metodologia impiegata, si rimanda alla relazione 2022.

Come già sottolineato, lo scopo principale della metodologia sviluppata è quello di permettere la disaggregazione e la scomposizione del *gap* IVA sulla base dei diversi domini di interesse, tra cui quello geografico (ad esempio, regionale, provinciale, comunale), settoriale e dimensionale. In linea di principio, il metodo *bottom up* consente di ottenere previsioni del *gap* IVA e della propensione ad evadere a livello di singolo contribuente, e pertanto derivare stime su qualsiasi dominio di interesse per aggregazione delle previsioni micro. Occorre tenere presente, tuttavia, che all'aumentare del dettaglio richiesto, si riduce l'accuratezza della stima. Inoltre, il risultato a livello micro è importante non tanto per una valutazione del livello di evasione del singolo contribuente, quanto per il valore informativo che tale disaggregazione implica.

A titolo di esempio, si riportano i risultati del *gap* e della propensione a livello regionale (Figura III.1.1.1). I risultati evidenziano una propensione minore al nord, e maggiore al sud e nelle isole. Inoltre, tramite il modello *bottom-up* è possibile incrociare i diversi domini di interesse per effettuare analisi dettagliate sul fenomeno evasivo.

FIGURA III.1.1.1: PROPENSIONE REGIONALE AL GAP IVA



III.1.2 La stima tematica per singole componenti

Come già menzionato, la metodologia di tipo *bottom-up* descritta nella relazione evasione 2022 è stata applicata per distinguere le diverse componenti del *gap* IVA, in funzione delle diverse operazioni effettuate dai contribuenti (ad esempio cessioni o acquisti). Analizzando i singoli campi

di dichiarazione, infatti, si può verificare l'origine della maggiore imposta accertata, e ricostruire quella totale.

A differenza di quanto fatto per la metodologia *bottom-up* totale, poiché si intende studiare contemporaneamente le singole componenti della maggiore imposta accertata, la variabile di *output* (MIA totale) è ora da considerare come un vettore, le cui componenti, che andranno a comporre il totale della maggiore imposta accertata, sono state stimate simultaneamente.

In particolare, i due aspetti principali che contribuiscono alla formazione del *gap* IVA sono la sotto-dichiarazione delle vendite e la sovra-dichiarazione dei costi, che rappresentano, dal lato dell'imposta, la MIA dal lato attivo e la MIA da lato passivo, rispettivamente. La prima è stata ricavata dalle informazioni presenti in dichiarazioni e accertamenti relativi al quadro VE, rigo 26 mentre la seconda proviene da quelle del quadro VF, rigo 71. Inoltre, come approfondimento sono stati considerati due ulteriori fenomeni: l'evasione da aliquota, sia sul lato attivo che su quello passivo, e la scomposizione della sotto-dichiarazione dei ricavi nelle componenti B2B e B2C mediante il quadro VT.

Riguardo la formulazione dell'evasione da aliquota, in questa fase di analisi ne è stato definito un limite inferiore. Infatti, la maggiore imposta da aliquota è stata stimata unicamente nei casi in cui, in fase di accertamento, è stata attribuita solamente una maggiore imposta e non un maggior imponibile, poiché si tratta dei soli casi in cui l'evasione deriva sicuramente da un'aliquota errata. Tale calcolo è stato effettuato sia per gli acquisti che per le cessioni, dunque è stata ottenuta una maggiore imposta accertata da aliquota sia attiva che passiva, che rappresentano una parte del totale della maggiore imposta attiva e passiva, rispettivamente.

Oltre alla stima dell'evasione da aliquota, è stata calcolata la ripartizione della MIA totale attiva nella sua quota B2B, cioè derivante dalla cessione di beni e servizi ad altre imprese, e B2C, relativa alle operazioni destinate al consumo finale. Infatti, la quota della maggiore imposta B2C, oltre ad essere una stima di per sé informativa, potrebbe fornire una stima più realistica del *gap* IVA aggregato. L'ipotesi di base è che, una volta non dichiarata l'IVA in un'operazione B2B, essa non verrà né dichiarata e detratta nelle operazioni successive, né pagata dal consumatore finale. In questo caso, il mancato gettito per lo Stato sarebbe rappresentato dall'omesso pagamento dell'IVA ascrivibile al consumatore finale, più le componenti non detraibili come i costi d'impresa. Dunque, considerare la somma delle maggiori imposte IVA individuali porterebbe ad una sovra-stima del *gap* IVA totale in quanto potrebbe essere contato più volte lo stesso ammontare di imposta evasa. Rimane tuttavia importante la stima del *bottom-up* globale, che come già evidenziato risponde ad esigenze differenti.

Da un punto di vista metodologico, la stima del *gap* IVA tematica è stata realizzata estendendo il modello predittivo già applicato per la stima *bottom-up* del *gap* IVA. Nel dettaglio, il primo dei due stadi di analisi, ossia la stima della probabilità di selezione per accertamento, è stato riproposto esattamente con le stesse modalità descritte nella Relazione evasione 2022.

Il secondo stadio, relativo alla predizione della maggiore imposta accertata IVA, è stato sottoposto a delle modifiche. In particolare, il *bagging* di alberi di regressione è stato utilizzato per stimare separatamente le previsioni della MIA lato attivo e lato passivo⁴³. Infine, per associare ad ogni contribuente non accertato il contribuente accertato con valori previsti più prossimi (*nearest neighbour*) è stata utilizzata una funzione di distanza multivariata (distanza di Mahalanobis), che permette di tenere conto contemporaneamente sia di entrambe le variabili di output principali che di eventuali correlazioni tra di esse. Una volta individuato, attraverso la previsione della MIA lato attivo e della MIA lato passivo, il contribuente accertato più prossimo, a ciascun soggetto non accertato vengono assegnate anche la maggiore imposta da aliquota e quella attribuibile alla componente B2C, oltre alle già citate MIA attiva e passiva.

Di seguito si riportano i risultati ottenuti dall'applicazione della metodologia descritta per la sola platea delle ditte individuali per l'anno 2019.

⁴³ Nel lato attivo consideriamo anche il quadro VJ, relativo alle operazioni in *reverse charge*, anche se in questa trattazione lo escludiamo per motivi di parsimonia espositiva.

TABELLA III.1.2.1: QUOTE DI MAGGIORE IMPOSTA ATTIVA E PASSIVA SU BASE REGIONALE

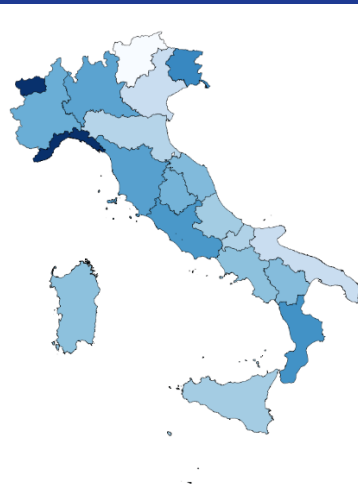
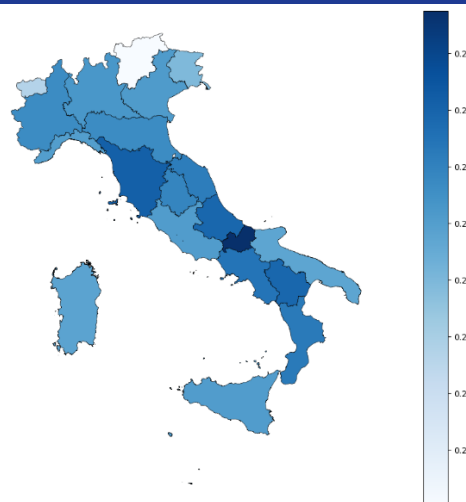
	MIA attiva	MIA passiva
PIEMONTE	57,8%	42,2%
VALLE D'AOSTA	54,4%	45,6%
LOMBARDIA	59,3%	40,7%
TRENTINO-ALTO ADIGE	53,5%	46,5%
VENETO	59,9%	40,1%
FRIULI VENEZIA GIULIA	58,6%	41,4%
LIGURIA	59,8%	40,2%
EMILIA ROMAGNA	59,9%	40,1%
TOSCANA	62,1%	37,9%
UMBRIA	60,6%	39,4%
MARCHE	59,1%	40,9%
LAZIO	62,2%	37,8%
ABRUZZO	58,3%	41,7%
MOLISE	54,8%	45,2%
CAMPANIA	59,2%	40,8%
PUGLIA	53,4%	46,6%
BASILICATA	51,6%	48,4%
CALABRIA	58,6%	41,4%
SICILIA	56,6%	43,4%
SARDEGNA	54,4%	45,6%
TOTALE	58,5%	41,5%

In Tabella III.1.2.1 è riportato la quota delle componenti della MIA: quella da sotto-dichiarazione dei ricavi (MIA attiva) e quella da sovra-dichiarazione dei costi (MIA passiva). Nella platea selezionata, il 58,5% della maggiore imposta totale predetta è ascrivibile al lato attivo, mentre il 41,5% al lato passivo. La distribuzione tra lato attivo e passivo è abbastanza omogenea tra le regioni. In tutti i casi la MIA attiva costituisce più del 50% della totale, con punte di oltre il 60% per Lazio, Toscana e Umbria.

Per quanto riguarda l'evasione da aliquota, questa può esser considerata unicamente come limite inferiore, nel senso che riguarda i casi in cui il mancato gettito deriva esclusivamente da una aliquota errata. Questa riguarda almeno l'1% dell'evasione da sotto-dichiarazione dei ricavi (lato attivo) e l'1,8% dell'evasione lato passivo.

Nella Figura III.1.2.1 è riportata la distribuzione regionale della maggiore imposta pro capite derivante dalla sotto-dichiarazione dei ricavi riferita alle transazioni B2C. Poiché si considerano le sole operazioni B2C, il *gap* è imputato alla regione in cui vi è la sede legale delle ditte che effettuano l'operazione che coinvolge il consumatore finale. Le regioni che presentano i valori più elevati sono la Valle d'Aosta e la Liguria; il Trentino-Alto Adige si conferma tra le regioni a *gap* IVA pro capite più basso. I risultati ottenuti per il Trentino-Alto Adige risultano essere coerenti con quelli relativi alla propensione al *gap* IVA (Figura III.1.2.1b): questa, infatti, risulta essere la regione con la propensione al *gap* IVA inferiore. Al contrario, la regione in cui si evidenzia la più alta propensione al *gap* è il Molise, con un valore eccedente lo 0,27. Infine, per quanto riguarda il resto delle regioni sembra esserci una distribuzione regionale piuttosto uniforme.

Qualche considerazione sulle differenze visibili fra la Figura III.1.2.1a, rappresentativa dell'approccio globale, e la III.1.2.1b, che illustra i risultati dell'approccio tematico: nel primo caso, la mappa può essere considerata un'indicazione del comportamento evasivo individuale dei contribuenti di quella regione; nel secondo caso, invece, l'obiettivo è valutare il mancato gettito effettivo, focalizzandosi però su un particolare aspetto del fenomeno. Questo confronto mette bene in luce la differenza nello scopo dei due approcci considerati.

FIGURA III.1.2.1a: GAP IVA B2C REGIONALE PROCAPITE**FIGURA III.1.2.1b: PROPENSIONE GAP IVA B2C REGIONALE**

III.2 LA STIMA BOTTOM UP DEL GAP IRAP

L'approccio *bottom-up* adottato per stimare il *tax gap* IRAP delle persone fisiche e società, si basa sull'applicazione di modelli econometrici che utilizzano sia le informazioni dei dati dichiarati al Fisco, sia quelle dedotte dalle fonti di accertamento fiscale, per il sottoinsieme di contribuenti sottoposti a tale processo. Per quanto riguarda la popolazione di stima, il modello prende a riferimento solo i contribuenti che hanno ricevuto un accertamento unificato sull'anno di imposta oggetto di studio, in quanto tale tipologia di controllo prevede uno specifico processo di selezione finalizzato alla ricostruzione della situazione patrimoniale e contabile e quindi dell'eventuale *gap* tra quanto dichiarato e quanto dovuto in termini di base imponibile e di imposta.

L'applicazione del modello riguarda la stima della Base Imponibile Potenziale (BIT) IRAP⁴⁴ sia per i soggetti accertati che per coloro che non hanno subito un accertamento, attraverso la proiezione su tutta la popolazione dei contribuenti dei risultati del modello statistico. Questa stima è importante poiché se confrontata con la base imponibile dichiarata (BID), consente una quantificazione della mancata base imponibile IRAP: il *gap* nella base (BIND).

La platea di studio è composta da Persone Fisiche, Società di Persone e Società di Capitali, che hanno effettivamente dichiarato reddito da impresa e/o da lavoro autonomo ovvero quei contribuenti che hanno dei dati IVA significativi. L'anno di imposta di riferimento per la stima dei modelli è il 2016.

Al fine di indentificare la platea di stima per il modello, sono stati identificate ed escluse le casistiche fuori contesto e non utili ai fini dello studio. Dalla popolazione iniziale sono stati esclusi i soggetti non passivi dell'imposta regionale sulle attività produttive⁴⁵ e i soggetti per i quali non si dispone di nessun dato reddituale. Per la definizione della popolazione di stima, state individuate ed escluse ulteriori casistiche specifiche dell'accertamento e sono stati applicati criteri per verificare e garantire l'affidabilità del dato. In seguito alla fase di *preprocessing*, la platea di riferimento è stata partizionata in due sotto-gruppi, all'interno dei quali si proceduto, in modo separato alla stima della variabile obiettivo: un gruppo per Persone Fisiche e Società di Persone studiate congiuntamente e uno per Società di Capitali.

⁴⁴ L'aggregato di riferimento nella stima del gap IRAP è il valore della Produzione Netta.

⁴⁵ I criteri specifici per la definizione della reale popolazione IRAP si basano sulle informazioni che derivano dai quadri del modello IRAP e per le sole Persone Fisiche anche sul concetto di autonoma organizzazione.

III.2.1 Il modello a due stadi di Heckman

L'approccio di tipo inferenziale alla base della metodologia è incentrato sul processo di stima a due stadi introdotto da Heckman negli anni Settanta, espresso da due equazioni differenti ma tra loro relazionate, ideate nell'ottica del superamento del problema del *selection bias* dovuto in questo caso a eventuali specificità dei Soggetti sottoposti ad accertamento fiscale, spesso frutto di un processo di selezione mirato da parte degli Enti preposti.

Tale metodologia si applica, infatti, nei casi in cui si debba stimare un modello econometrico a partire da un campione le cui osservazioni sulla variabile endogena sono disponibili solo se si verifica una data condizione che determina l'appartenenza del soggetto a un gruppo e non al suo complemento.

Nel caso in esame la variabile obiettivo, la Base Imponibile IRAP Potenziale, viene stimata tenendo conto della possibile distorsione dovuta al fatto che i contribuenti sottoposti ad accertamento fiscale, proprio per tale eventualità, potrebbero differire in maniera sistematica da coloro che non hanno subito tale processo.

Primo stadio: modello di selezione

Il primo modello, riconducibile ad un modello probit, ha come risposta una variabile dicotomica dove il valore 1 indica che il contribuente è stato sottoposto ad accertamento unificato e 0 in caso contrario; si evidenzia che il modello stimato non è utilizzato per fini previsivi, ma solo per determinare la probabilità di essere selezionati per un accertamento fiscale sull'anno d'imposta oggetto di studio⁴⁶.

Le probabilità stimate dal modello di selezione sono utilizzate per calcolare un fattore λ , denominato Lambda di Mills, coincidente con l'inverso del rapporto di Mills, da inserire come ulteriore covariata nella successiva fase di stima del Modello al secondo stadio con l'intento di catturare l'effetto sulla variabile obiettivo anche dell'eventualità di essere accertato e tener conto nelle stime quindi della distorsione legata alla selezione non casuale dei contribuenti.

Secondo stadio: modello sostanziale

Il modello di Heckman al secondo stadio prevede un modello di regressione per la stima della Base Imponibile Potenziale IRAP, eseguita solamente sul gruppo dei contribuenti sottoposti a controllo unificato sull'anno d'imposta 2016 e corretta per il coefficiente di distorsione (Lambda di Mills) ottenuto al primo stadio con il modello di selezione.

Dal punto di vista metodologico l'approccio di Heckman è incentrato sull'esistenza di correlazione tra i termini d'errore delle due equazioni supponendo l'influenza di elementi non osservabili direttamente ma che agiscono sia sull'aspetto della selezione che su quello strutturale.

Se i termini d'errore delle due equazioni sono incorrelati, il meccanismo di selezione può essere considerato ignorabile e il metodo di stima dei minimi quadrati ordinari (OLS) fornisce stimatori corretti: la selezione dipende perciò solo dalle variabili ausiliarie che sono poi impiegate per trattare la distorsione.

Se invece i termini d'errore sono correlati la probabilità di selezione è una funzione monotona della variabile obiettivo e quindi la selezione dipende dalla variabile oggetto dell'indagine. Pertanto, lo stimatore OLS è affetto da una distorsione dovuta alla selezione campionaria.

⁴⁶ È importante sottolineare che questa selezione riguarda diversi periodi temporali sui quali hanno insistito diversi momenti economici e diverse direttive dell'Agenzia: i controlli sull'anno d'imposta 2016 sono frutto di selezioni effettuate negli anni successivi alla presentazione della dichiarazione.

Assunzione del modello

L'impostazione fin qui illustrata fa riferimento a un contesto di tipo statistico-inferenziale che presuppone un insieme di assunzioni, non ultimo il fatto stesso di considerare l'insieme degli accertati un campione dell'universo dei dati di analisi e in particolare, un campione distorto.

Il modello di Heckman fornisce una rappresentazione parametrica sia della funzione strutturale che di selezione, assumendo indipendenza tra gli errori e i regressori e la normalità bivariata della distribuzione dei termini d'errore.

Risultati del modello

Le variabili utilizzate nella stima, sia nel primo che nel secondo stadio, derivano dalle informazioni dichiarate fiscalmente dal contribuente in materia IRPEF, IVA e IRAP e da informazioni presenti negli archivi anagrafici che ne caratterizzano il profilo, come la tipologia di soggetto fiscale, la regione del domicilio fiscale, il settore economico.

Una descrizione dettagliata delle variabili incluse nei modelli è riportata nella seguente tabella (Tabella III.2.1.1)⁴⁷:

TABELLA III.2.1.1. VARIABILI UTILIZZATE	
VARIABILI	SIGNIFICATO
Ricavi delle vendite e delle prestazioni	Ricavi delle vendite e delle prestazioni (dichiarazione Unico)
Attività totali	Totale compensi e componenti positive (dichiarazione Unico)
Redditi da lavoro 770	Totale redditi per i quali è possibile fruire della detrazione
Reddito lordo	Reddito lordo (dichiarazione Unico)
Presenza di ulteriori deduzioni	Dummy pari a 1 se il contribuente presenta ulteriori deduzioni; 0 in assenza di deduzioni.
Costi funzionamento	Costi funzionamento (dichiarazione IRAP)
Costi materie prime	Costi materie prime (dichiarazione IRAP)
Costi per servizi	Costi per servizi (dichiarazione IRAP)
Dimensione aziendale	Dummy dimensionali calcolate in base al numero di addetti.
IVA sulle operazioni effettuate nell'anno	Ammontare complessivo dell'IVA sulle operazioni effettuate nell'anno
IVA a credito	Totale IVA a credito (VL39)
Altri acquisti ed importazioni IVA	Altri acquisti ed importazioni IVA
Valore aggiunto fiscale	Differenza tra il volume d'affari dichiarato e l'ammontare complessivo degli acquisti effettuati nell'anno.
Tipologia Soggetto	Dummy pari a 1 se il contribuente è una Società di Persone; 0 altrimenti.
Regime fiscale	Dummy pari a 1 se è presente il reddito d'impresa semplificata; 0 altrimenti
Settore economico	Dummy macrosezione attività prevalente (Industria, Costruzioni, Commercio, Servizi Pubblici e Privati)
Ripartizione Geografica	Dummy Geografiche (Nord Ovest, Nord Est, Centro, Sud e Isole)

I risultati ottenuti dai modelli di Heckman (modello congiunto Persone Fisiche-Società di Persone e modello Società di Capitali), sono molto soddisfacenti, in quanto le correlazioni tra gli errori delle equazioni al primo e secondo stadio sono alte e significative, confermando il vantaggio dell'utilizzo della metodologia di Heckman per correggere la distorsione da selezione dei soggetti sottoposti ad accertamento.

I modelli stimati al primo e secondo stadio, sono caratterizzati da una buona efficacia interpretativa, in particolare: per quanto riguarda l'equazione al primo stadio, influenzano significativamente e risultano discriminanti nel processo di selezione le caratteristiche strutturali del contribuente, quali la ripartizione geografica di appartenenza, il settore di attività economica e la dimensione aziendale in base al numero di addetti. Da un punto di vista dichiarativo, i predittori con un peso maggiore nella stima della probabilità di ricevere un accertamento unificato, sono le attività totali da Modello Unico, il "Costo delle materie prime" dichiarate ai fini IRAP e la variabile

⁴⁷ Nella specificazione dei modelli analizzati, la trasformazione considerata per le variabili continue è quella della Radice Quadrata.

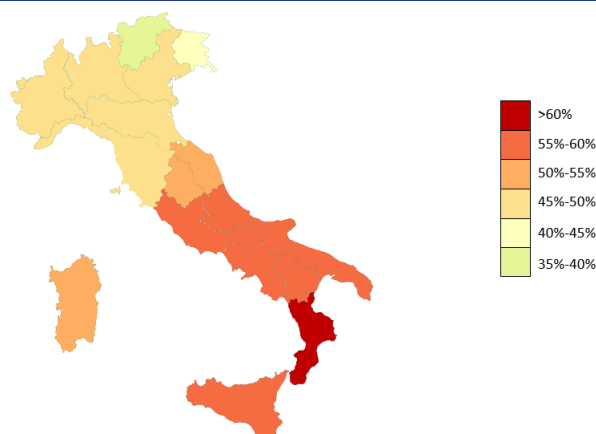
“Altri acquisti e importazioni⁴⁸”, informazione presente nella dichiarazione IVA dove possono essere trascritte anche tutte le operazioni di compensazione a seguito di fatturazioni stornate.

Per quanto riguarda l’equazione di regressione al secondo stadio (modello sostanziale), le variabili con maggior peso e più importanti nella stima della Base Imponibile Potenziale sono riconducibili a grandezze economiche inerenti il volume delle attività dell’impresa, siano esse desunte dalla dichiarazione IVA sia dalla dichiarazione dei redditi Unico.

I modelli di Heckman stimati sono stati utilizzati per ottenere una stima della Base Imponibile Totale e del Gap IRAP delle Società e Ditte Individuali per gli anni che vanno dal 2016 in poi. Naturalmente, partendo dalle previsioni micro relative ai singoli contribuenti, è possibile ottenere una stima aggregata della Base Imponibile Potenziale e del Gap IRAP per qualsiasi dominio di interesse.

Di seguito, si riportano i risultati della propensione al Gap⁴⁹ IRAP a livello regionale⁵⁰ (Figura III.2.1.1). Dalla figura è possibile desumere una scarsa variabilità regionale all’interno della stessa macroarea ma anche una discreta contrapposizione delle regioni settentrionali rispetto all’area meridionale. In sintesi, nelle regioni meridionali, si evince una forte propensione all’evasione rispetto al resto del territorio, con la Calabria con un valore superiore al 60%.

FIGURA III.2.1.1. PROPENSIONE REGIONALE AL GAP IRAP



Questi risultati sono abbastanza allineati con quelli ottenuti tramite la metodologia top down, e questo ci conforta sulla possibilità di utilizzare tali studi al fine di ottenere una maggiore granularità dei dati ufficiali.

⁴⁸ La variabile “Altri acquisti e importazioni imponibile” dal punto di vista dichiarativo riporta tutti gli acquisti di beni e servizi che non siano beni ammortizzabili, beni strumentali non ammortizzabili o beni e/o servizi acquistati e funzionali alla attività economica (rigo VF 27 col 4)

⁴⁹ La propensione al Gap IRAP si ottiene come rapporto dell’ammontare della Base Imponibile Non Dichiarata (gap) su quello della Base Imponibile Potenziale.

⁵⁰ La figura riporta il valore medio della propensione al Gap IRAP calcolato sull’arco temporale 2016-2020 per ripartizione territoriale in livelli.

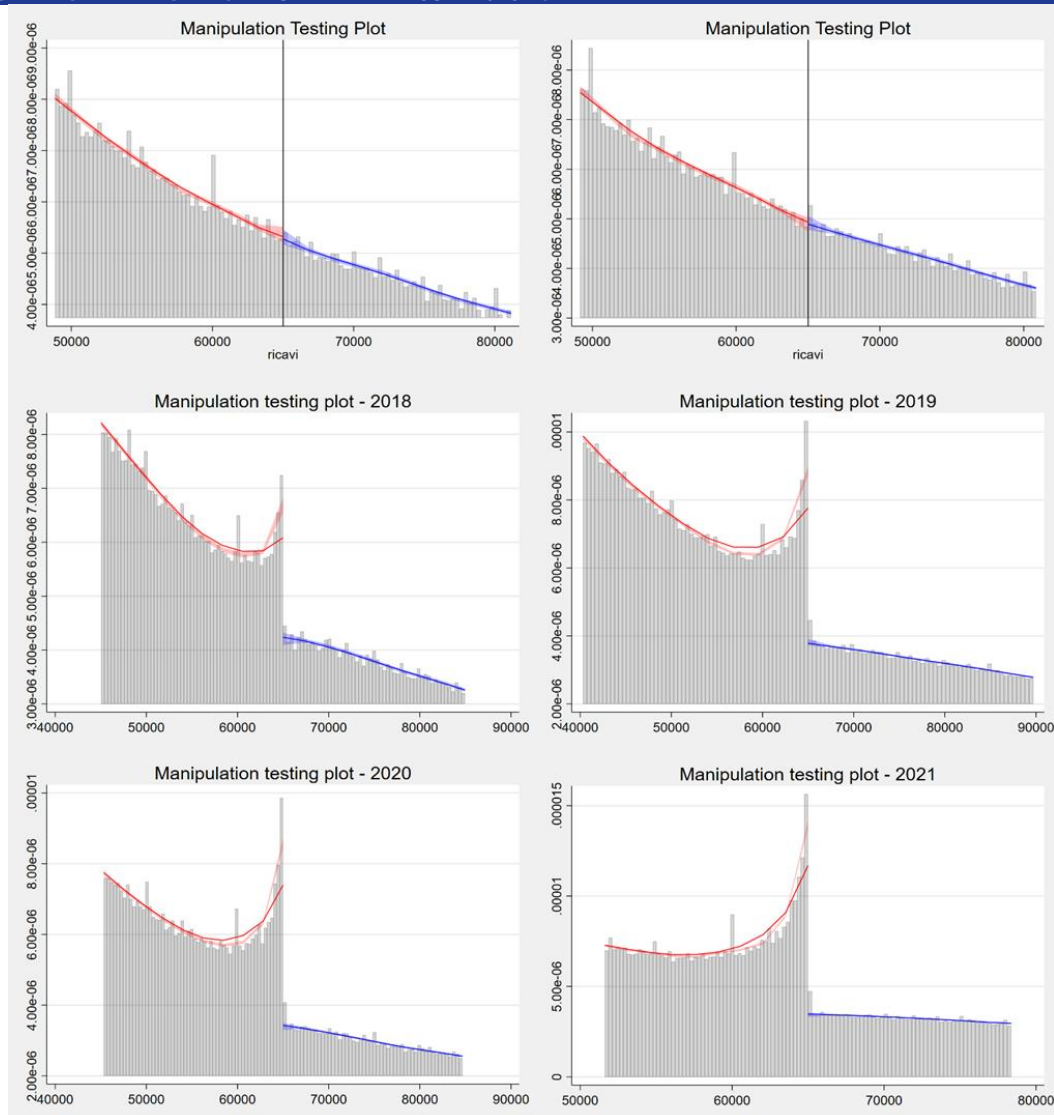
III.3 VALUTAZIONE EX-POST DELL'ESTENSIONE DELLA SOGLIA DI RICAVI NEL NUOVO REGIME FORFETARIO

Analisi della discontinuità nella distribuzione dei ricavi e stima dell'elasticità rispetto alla variazione dell'aliquota marginale

L'analisi si basa sui dati relativi a tutti i contribuenti che compilano i quadri RE (i redditi derivanti dall'esercizio di arti e professioni rientranti nel regime analitico), RG (redditi derivanti dall'esercizio di attività commerciali in contabilità semplificata) e LM (contribuenti che aderiscono al regime di vantaggio e al regime forfetario) negli anni di imposta dal 2016 al 2021.

La Figura III.3.1 mostra la distribuzione dei ricavi stimata con il *manipulation test*, attraverso lo stimatore basato su *kernel density* e regressioni polinomiali locali. L'analisi conferma e rafforza i risultati già emersi sulla base della semplice ispezione grafica e identifica una forte discontinuità nella distribuzione dei ricavi intorno alla soglia dei 65 mila euro per gli anni di imposta 2018-2021. Tale fenomeno, come atteso, non si verifica nel 2016 e nel 2017, anni precedenti all'introduzione della norma.

FIGURA III.3.1 MANIPULATION TEST ANNI D'IMPOSTA 2016-2021



L'analisi di *bunching*, attraverso una stima non parametrica, fornisce una evidenza empirica nelle risposte comportamentali agli incentivi di tipo fiscale. Nella fattispecie, questa tecnica è in grado di analizzare fenomeni in cui la distribuzione di una variabile è soggetta a discontinuità a causa dell'implementazione di una *policy* che genera un incentivo/disincentivo fiscale, alterando i comportamenti degli individui. Il *bunching* è, quindi, particolarmente indicato per analizzare dati amministrativi (perché questi presentano un'ampia numerosità campionaria e bassi errori di misurazione) ed è stata spesso utilizzata per analizzare fenomeni relativi all'offerta di lavoro e all'evasione fiscale.

Si suddividono, quindi, i contribuenti per gruppi di ricavi ("bins"); successivamente, si analizza la distribuzione della variabile di interesse e si stima un controfattuale artificiale dato dalla distribuzione che si sarebbe osservata intorno alla soglia in assenza di *policy*, sotto l'ipotesi di assenza di discontinuità (Chetty et al. 2011). Una volta stimata la distribuzione controfattuale artificiale, questa si confronta con la distribuzione osservata nei dati e si calcolano: i) la differenza di densità rispetto al controfattuale artificiale (*excess mass*), cioè il numero di soggetti in eccesso, che si localizzano al di sotto della soglia, rispetto a quello che si osserverebbe in assenza di *policy*; ii) l'elasticità dei ricavi, cioè la variazione percentuale dei ricavi dovuta alla variazione di aliquota, che viene stimata localmente intorno al punto di discontinuità (*kink*).

L'analisi attorno alla soglia dei 65 mila euro è stata eseguita considerando i soggetti con ricavi compresi fra 30 mila euro e 100 mila euro, che sono stati divisi in 140 gruppi corrispondenti a classi di ricavi ordinate in misura crescente e di ampiezza pari a 500 euro. Per il calcolo dell'elasticità, sono state considerate l'aliquota del 15% per i soggetti al di sotto della soglia dei 65 mila euro e l'aliquota IRPEF del 38% per i soggetti con ricavi superiori a 65 mila euro⁵¹.

Nella Figura III.3.2 si riportano i risultati dell'analisi principale⁵². La curva *bordeaux* rappresenta la distribuzione teorica controfattuale che si sarebbe osservata in assenza di *policy* e di discontinuità, mentre la linea nera rappresenta la distribuzione empirica dei ricavi. L'analisi grafica conferma i risultati del *Manipulation test* e mostra assenza di autoselezione intorno alla soglia dei 65 mila euro (indicata dalla linea verticale) negli anni di imposta 2016 e 2017. Al contrario, si evince una forte autoselezione dei soggetti al di sotto dei 65 mila euro negli anni in cui la normativa era in vigore (dal 2018 al 2021). Nella Figura III.3.2 vengono inoltre riportati i valori della "excess mass", *b*, dell'elasticità, *e*, ed i corrispondenti *bootstrapped standard errors*.

È importante considerare che il valore dell'elasticità che osserviamo approssima solo in parte quella strutturale del fenomeno perché: i) è calcolata localmente nell'intorno della soglia; ii) ci possono essere delle "optimization frictions", cioè dei costi di aggiustamento o inerzia che alterano la reazione degli individui all'incentivo determinando una elasticità osservata che è minore di quella strutturale del fenomeno; iii) la scelta dei parametri (per esempio l'ampiezza delle classi di ricavi, l'area di *bunching*, etc) può influire sulla stima. Al fine di tenere conto di quest'ultimo aspetto, nella Tabella III.3.1 si mostrano le stime di *excess mass* e dell'elasticità al variare dei parametri.

Osservando i valori delle *excess mass*, nelle diverse specificazioni, si nota come nel 2019 in corrispondenza della soglia ci siano fra il 52 e il 69% di contribuenti in più rispetto a quelli che si sarebbero osservati nello scenario controfattuale (cioè, in assenza di *policy* e di discontinuità nella distribuzione). Questo valore cresce nel 2020 e si attesta a valori compresi fra l'85 e il 102% nel 2021. Al contrario, nel 2016 e nel 2017 (quando la soglia dei 65 mila euro non era in vigore) la distribuzione empirica non presenta punti di discontinuità e si sovrappone a quella teorica non presentando significative *bunching mass* in corrispondenza della soglia.

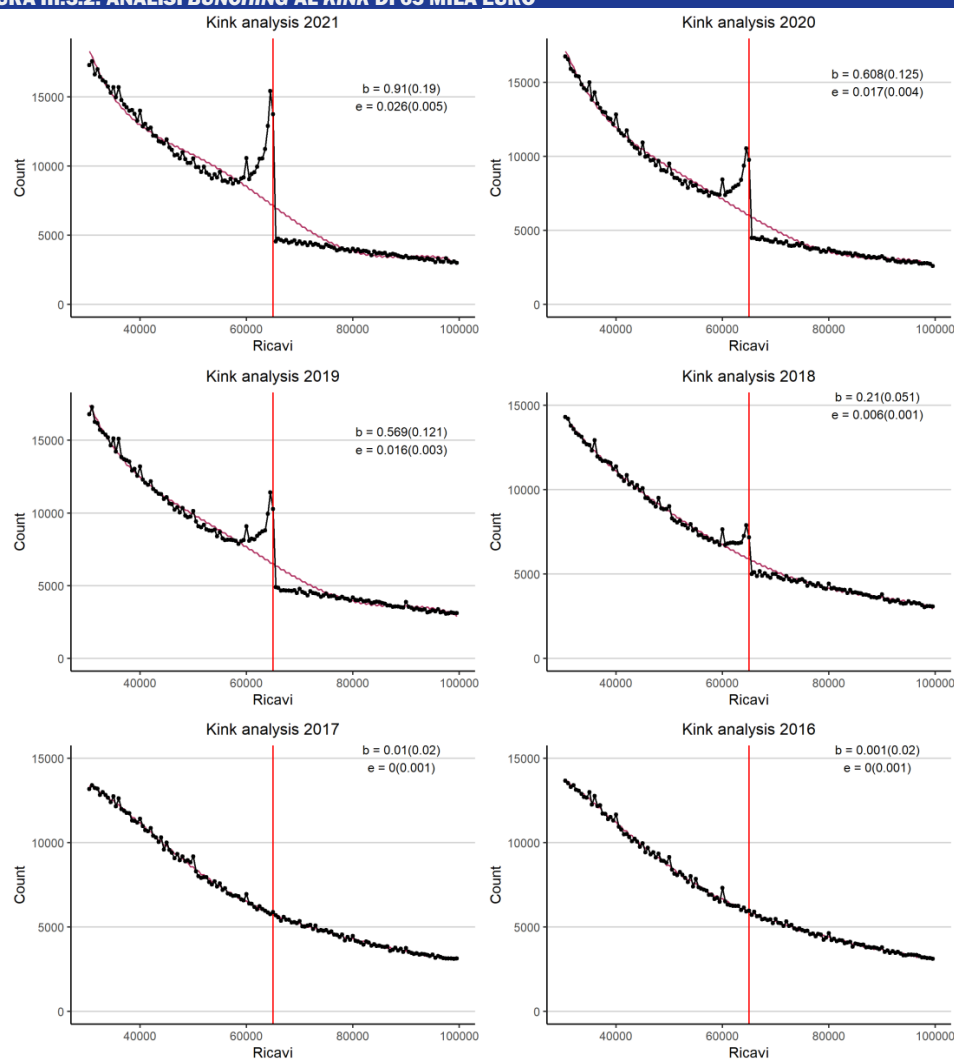
Anche l'elasticità dei ricavi alla soglia cresce dopo l'introduzione della norma attestandosi ad un valore compreso fra il 2.6% e il 3.4% nel 2021.

Nella Figura III.3.3 si mostra, per alcuni settori economici, l'analisi *bunching* relativa alle soglie del vecchio regime forfetario, in vigore fino al 2018. L'analisi è implementata considerando l'aliquota del 15% per i soggetti con ricavi al di sotto della vecchia soglia e quella del 27% per i soggetti con ricavi superiori rispetto alle soglie previgenti⁵³. L'analisi conferma i risultati ottenuti con il *manipulation test*, indicando una significativa autoselezione degli individui al di sotto delle soglie dei ricavi massimi, in particolare per i settori ATECO del "Commercio all'ingrosso e al dettaglio", delle "Attività Professionali" e delle "Altre attività economiche", per cui osserviamo rispettivamente il 15%, il 68% e il 57% di contribuenti in più rispetto al controfattuale artificiale che si osserverebbe in caso di assenza di *policy*. Al contrario, per il settore delle "Attività dei servizi di alloggio e di ristorazione" si osserva una *excess mass* del 5% non significativa.

⁵¹ Per ragioni di ordine pratico, non è stata apportata una distinzione fra soggetti con partita IVA in essere da più di 5 anni (aliquota del 15%) e nuove partite IVA (aliquota del 5%), quindi l'elasticità osservata è da considerarsi leggermente sottostimata rispetto a quella strutturale del fenomeno. La scelta dell'aliquota del 38% dipende dal fatto che applicando il coefficiente di redditività a ricavi pari a 65 mila euro, la maggior parte dei contribuenti sarebbe soggetta ad un'aliquota IRPEF del 38%.

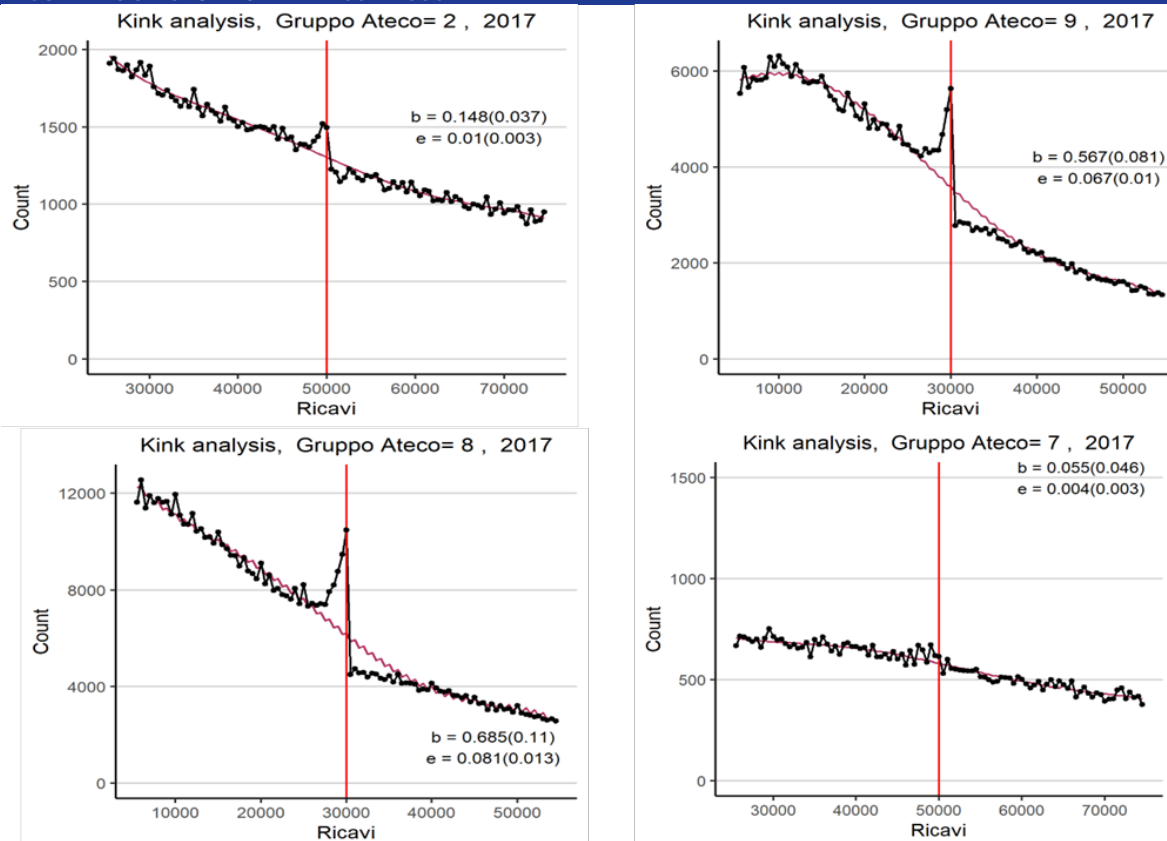
⁵² L'analisi del *bunching* viene sviluppata in R (si veda Mavrokonstantis, P. 2022, Introduction to the bunching Package). La stima controlla per i "round numbers" pari a 1000, cioè per valori di ricavi in corrispondenza dei quali spesso si creano dei punti di discontinuità legati all'arrotondamento dei ricavi dichiarati ma che non costituiscono un punto di discontinuità determinato dalle *policy*.

⁵³ La scelta dell'aliquota del 27% dipende dal fatto che applicando il coefficiente di redditività a ricavi pari alle soglie del precedente Regime forfetario, tutti i contribuenti sarebbero soggetti ad un'aliquota IRPEF del 27%.

FIGURA III.3.2: ANALISI BUNCHING AL KINK DI 65 MILA EURO**TABELLA III.3.1: ANALISI BUNCHING AL KINK DI 65 MILA EURO – ANALISI DI SENSITIVITÀ**

N. bins	Ampiezza bins		2016	2017	2018	2019	2020	2021
200	350 euro	Excess mass	0.031 (0.022)	0.014 (0.023)	0.249 (0.08)	0.687 (0.181)	0.731 (0.185)	1.021 (0.304)
		Elasticità	0.001 (0)	0 (0)	0.005 (0.002)	0.014 (0.004)	0.015 (0.004)	0.02 (0.006)
		Excess mass	0.01 (0.02)	0.001 (0.02)	0.21 (0.051)	0.569 (0.121)	0.608 (0.125)	0.91 (0.19)
140	500 euro	Elasticità	0 (0.001)	0 (0.001)	0.006 (0.001)	0.016 (0.003)	0.017 (0.004)	0.026 (0.005)
		Excess mass	0.008 (0.023)	0.018 (0.022)	0.211 (0.055)	0.523 (0.131)	0.569 (0.134)	0.85 (0.208)
		Elasticità	0 (0.001)	0.001 (0.001)	0.008 (0.002)	0.021 (0.005)	0.023 (0.005)	0.034 (0.008)

Nota: In parentesi vengono riportati i *bootstrapped standard errors*.

FIGURA III.3.3: BUNCHING ALLE VECCHIE SOGLIE

Nota: Settore Ateco 2: "Commercio all'ingrosso e al dettaglio"; Settore Ateco 7: "Attività dei servizi di alloggio e di ristorazione"; Settore Ateco 8: "Attività Professionali"; Settore Ateco 9: "Altre attività economiche".

Stima econometrica dell'impatto del nuovo regime forfetario sulla variazione dei ricavi dichiarati

Ai fini della stima econometrica si è costruito un panel bilanciato per gli anni d'imposta 2016-2021 e si è circoscritta l'analisi ai contribuenti con ricavi inferiori a 100 mila euro e maggiori di 0. Il dataset si riduce a circa 1,2 milioni di contribuenti per ciascuna annualità. L'approccio metodologico utilizzato si basa su una stima panel ad effetti fissi.

Nella stima panel ad effetti fissi la variabile dipendente è rappresentata dal logaritmo dei ricavi dichiarati e la variabile esplicativa di interesse è costituita da una variabile dummy "Forfetari Nuovo Regime" (FNR) che assume valore uno per i contribuenti che aderiscono al nuovo regime forfetario negli anni d'imposta 2019-2021 e zero altrimenti⁵⁴. I risultati delle stime sono riportati nella Tabella III.3.2.

Nella Colonna (1) della Tabella III.3.2 risulta un incremento positivo e significativo, pari al 5,9%, dei ricavi dichiarati da parte dei contribuenti che aderiscono al nuovo regime forfetario rispetto agli altri contribuenti. Nell'ipotesi in cui siano validi gli effetti di variazione della base imponibile dei ricavi dichiarati su entrambe le soglie, ovvero un effetto di incremento associato al venir meno delle soglie precedenti e un effetto di riduzione corrispondente all'istituzione della nuova soglia, tale risultato sembra suggerire la prevalenza del primo effetto rispetto al secondo. In altri termini, la riduzione dei ricavi dichiarati attorno alla soglia dei 65 mila euro risulterebbe

⁵⁴ A tal proposito, occorre sottolineare che l'analisi ad effetti fissi fornisce una stima della variazione dei ricavi tenendo conto delle caratteristiche "time invariant" proprie di ogni contribuente.

più che compensata dall'aumento dei ricavi dichiarati attorno alle precedenti soglie di ricavi, con una prevalenza degli effetti di emersione di base imponibile o di stimolo all'attività economica innescata dall'estensione del regime forfetario.

Tuttavia, nella stima riportata nella Colonna (2) emerge una differenziazione degli effetti rispetto ai vari settori di attività economica. L'impatto risulta, infatti, positivo e significativo soprattutto per le attività professionali (pari al 14,5%), gli intermediari del commercio e il settore delle costruzioni e delle altre attività immobiliari. Viceversa, si ottiene un risultato opposto soprattutto nel caso del commercio ambulante (in media del 28,5%) e del settore dell'alloggio e della ristorazione (pari al 18,5%).

Molto interessanti ai fini dell'analisi differenziata per le soglie di ricavi sono i risultati della stima riportata nella Colonna (3), che confermano gli effetti ipotizzati sul piano teorico e desumibili dall'analisi della discontinuità nella distribuzione dei ricavi. Le classi di ricavi sono state costruite nell'anno d'imposta 2017, ovvero prima che il nuovo regime forfetario producesse i suoi effetti. In particolare, risulta un effetto positivo e significativo di incremento dei ricavi per le classi di ricavi da zero a 20 mila euro (con un aumento del 12,7%) e da 20 mila a 40 mila euro (con un aumento del 4,3%). Viceversa, per le classi di ricavi più elevate si osserva un effetto opposto, ovvero una riduzione di ricavi dichiarati del 6,6% per i contribuenti con ricavi compresi tra 40 mila e 65 mila euro, del 25,5% per i contribuenti con ricavi dichiarati compresi tra 65 mila e 85 mila euro e addirittura del 47,3% per i contribuenti con ricavi dichiarati da 85 mila fino a 100 mila euro. Vala la pena, a tale proposito, precisare che l'impatto sulle classi di ricavi più basse, pur essendo inferiore, risulta nel complesso più rilevante alla luce della maggiore numerosità dei contribuenti che aderiscono al nuovo regime forfetario all'interno di ciascuna classe.

Tuttavia, la stima riportata nella Colonna (4) risulta più esaustiva e affidabile in quanto consente un confronto immediato tra i contribuenti che aderiscono o meno al nuovo regime forfetario. Nella classe di ricavi più bassa si osserva una variazione positiva maggiore per i contribuenti che aderiscono al nuovo regime forfetario; viceversa, nella classe di ricavi compresi tra 20 mila e 40 mila euro la diminuzione dei ricavi dichiarati è inferiore nel caso dei contribuenti che aderiscono al nuovo regime forfetario rispetto agli altri contribuenti. Al contrario, nelle classi di ricavi più elevate la riduzione dei ricavi è inferiore nel caso dei contribuenti che non aderiscono al nuovo regime forfetario.

Infine, dalla stima riportata nella Colonna (5) si evince che anche nel caso delle attività professionali l'effetto di incremento dei ricavi si concentra nelle classi di ricavi inferiori, mentre per le classi di ricavi attorno alla nuova soglia del regime forfetario si osserva sempre e comunque una riduzione dei ricavi dichiarati.

Nel complesso, comunque, i risultati, sebbene molto differenziati per classi di ricavi e settore di attività economica, non sembrano suggerire un effetto di riduzione significativa della base imponibile, in quanto l'impatto positivo che si registra attorno alle precedenti soglie, caratterizzate da una maggiore numerosità di contribuenti, sembra in grado di compensare l'effetto negativo che si osserva generalmente attorno alla nuova soglia di ricavi prestabilita. Ciò nonostante, occorre interpretare questi risultati con molta cautela, in quanto la variabile esplicativa considerata, nonché la strategia econometrica impiegata, sono in grado di catturare soltanto parzialmente l'impatto della riforma, essendo difficile isolare tutti gli altri fattori confondenti.

TABELLA III.3.2: EFFETTO DEL NUOVO REGIME FORFETARIO SUI RICAVI DICHIARATI – STIMA PANEL EFFETTI FISSI -

	(1)	(2)	(3)	(4)	(5)
Forfetari Nuovo Regime (FNR)	0.05883*** 0.001061	-0.01854*** 0.001846			FNR - AMB_ALI - CL_0-20 -0.2362*** 0.032452
FNR - Industrie Alimentari (ALIM)		-0.07083*** 0.017723			FNR - AMB_ALI - CL_20-40 -0.2705*** 0.026913
FNR - Commercio Ingrosso (COMM)		-0.03402*** 0.003776			FNR - AMB_ALI - CL_40-65 -0.4611*** 0.042339
FNR - Comm. Ambulante Alim. (AMB_ALI)		-0.28155*** 0.019805			FNR - AMB_ALI - CL_65-85 -1.0845*** 0.242921
FNR - Comm. Ambulante Altro (AMB_ALT)		-0.29298*** 0.006846			FNR - AMB_ALI - CL_85-100 -0.7333*** 0.130075
FNR - Costruzioni Immobiliare (COSTR)		0.11113*** 0.00324			FNR - AMB_ALT - CL_0-20 -0.2969*** 0.009297
FNR - Intermediari del Commercio (INT)		0.08759*** 0.003536			FNR - AMB_ALT - CL_20-40 -0.3008*** 0.009693
FNR - Alloggio e Ristorazione (RIST)		-0.185*** 0.008405			FNR - AMB_ALT - CL_40-65 -0.4214*** 0.022962
FNR - Attività Professionali (PROF)		0.14509*** 0.002123			FNR - AMB_ALT - CL_65-85 -0.6028*** 0.05249
FNR - Altre Attività Economiche (ALTR)					FNR - AMB_ALT - CL_85-100 -0.8724*** 0.16794
FNR - Classe Ricavi 0-20k (CL_0-20)			0.127478*** 0.001592	0.084039*** 0.001817	FNR - RIST - CL_0-20 -0.1168*** 0.017053
FNR - Classe Ricavi 20k-40k (CL_20-40)			0.04291*** 0.001344	-0.00307*** 0.001502	FNR - RIST - CL_20-40 -0.1971*** 0.010899
FNR - Classe Ricavi 40k-65k (CL_40-65)			-0.06573*** 0.001877	-0.1142*** 0.001987	FNR - RIST - CL_40-65 -0.2864*** 0.015606
FNR - Classe Ricavi 65k-85k (CL_65-85)			-0.25549*** 0.005727	-0.31548*** 0.005778	FNR - RIST - CL_65-85 -0.522*** 0.065676
FNR - Classe Ricavi 85k-100k (CL_85-100)			-0.47262*** 0.01856	-0.54886*** 0.018579	FNR - RIST - CL_85-100 -0.9349*** 0.242693
Non Forfetari - Classe Ricavi 0-20k (CL_0-20)				0.069079*** 0.002061	FNR - PROF - CL_0-20 0.2075*** 0.002174
Non Forfetari - Classe Ricavi 20k-40k (CL_20-40)				-0.03475*** 0.001281	FNR - PROF - CL_20-40 0.1023*** 0.001776
Non Forfetari - Classe Ricavi 40k-65k (CL_40-65)				-0.0788*** 0.001211	FNR - PROF - CL_40-65 -0.0102*** 0.002294
Non Forfetari - Classe Ricavi 65k-85k (CL_65-85)				-0.15012*** 0.001553	FNR - PROF - CL_65-85 -0.1817*** 0.006902
Non Forfetari - Classe Ricavi 85k-100k (CL_85-100)				-0.23256*** 0.002837	FNR - PROF - CL_85-100 -0.3474*** 0.020674
Costante	10.03064*** 0.000503	10.0311*** 0.006301	10.03064*** 0.000502	10.03064*** 0.000503	10.037*** 0.006239
Dummy Anno	√	√	√	√	√
Dummy Ateco		√	√	√	√
Osservazioni	1,216,058	1,216,058	1,216,058	1,216,058	1,216,058
R-quadro	0.000	0.0203	0.0161	0.0571	0.0051

Nota: La variabile dipendente è espressa in logaritmi. La stima nella Colonna (5) riporta soltanto i coefficienti associati alle principali variabili esplicative di maggior interesse. Standard errors robusti in corsivo. *** p<0.01, ** p<0.05, *p<0.1.

III.4 UN MODELLO A CLASSI LATENTI PER L'ANALISI DELLE FRODI IVA

III.4.1 Premessa

Nella presente appendice viene illustrata la metodologia che l'Agenzia delle Entrate sta sviluppando per l'analisi del fenomeno delle frodi IVA, in particolare da *missing trader*. Si tratta di un approccio di tipo non supervisionato che prevede l'utilizzo di modelli a classi latenti volti a profilare i contribuenti in termini di propensione all'evasione IVA ad "alto rischio", e si affianca all'approccio utilizzato finora, di tipo supervisionato.

La metodologia supervisionata, i cui risultati sono presentati nella Relazione Evasione, si basa sull'utilizzo di dati di controllo derivanti da verifiche e accertamenti unificati relativi ai cinque anni precedenti l'anno di analisi, e risulta costituita da una platea di riferimento che include tutti i soggetti che non presentano dichiarazione IVA. Al contrario, il metodo non supervisionato non utilizza le informazioni relative ai dati di controllo, consentendo così da un lato di aumentare la tempestività della stima e, dall'altro, di considerare anche i meccanismi fraudolenti di coloro che presentano dichiarazione IVA ma non versano l'IVA all'Erario. L'approccio non supervisionato consente quindi una rappresentazione dei fenomeni ad elevato rischio di frode sull'intera platea delle partite IVA, profilandosi come un efficace strumento di analisi del rischio.

Per meglio catturare e analizzare la pluralità di comportamenti evasivi, sono stati formalizzati due diversi modelli a seconda della presenza della dichiarazione (versamento) IVA; in particolare:

- *Modelli di mistura per variabili miste*, per analizzare i soggetti che non hanno presentato una dichiarazione IVA o che non hanno effettuato versamenti nell'anno di imposta oggetto dello studio;
- *Modelli mixtures of experts*, per analizzare i soggetti che hanno presentato la dichiarazione IVA o che hanno effettuato versamenti.

I modelli sono stati implementati in R utilizzando i pacchetti *clustMD* (McParland and Gormley, 2015) e *MoEClust* (Murphy and Murphy, 2020).

In ogni caso, la metodologia impiegata si avvale di tecniche di cluster *analysis* di tipo *model-based*: le diverse componenti del modello mistura sono associate ad altrettanti cluster (classi di contribuenti caratterizzate da diversi livelli di rischio) corrispondenti a distinte distribuzioni (eventualmente condizionate) di probabilità nell'ambito di una stessa famiglia parametrica (in questo caso Gaussiana).

III.4.2 Il modello

La metodologia proposta ha l'obiettivo di stimare il valore dell'IVA che dovrebbe essere dichiarata e/o versata in caso di perfetta *compliance* sulla base delle informazioni desumibili principalmente dai dati della fatturazione elettronica, dei corrispettivi inviati telematicamente e delle bollette doganali. A causa della complessità della normativa fiscale e dell'eventuale incompletezza dei dati disponibili, non è tuttavia sempre possibile effettuare una ricostruzione esatta dell'IVA dovuta o a credito; viene pertanto adottato un approccio statistico a più stadi.

Nel primo stadio, l'imposta teorica (l'IVA che dovrebbe essere dichiarata e/o versata in caso di perfetta *compliance*) è espressa, mediante un modello di regressione, in funzione dell'imposta ricostruita da fatture elettroniche e corrispettivi telematici.

In termini formali, il modello di primo stadio può essere specificato come segue:

$$Y^* = \alpha + \beta X + U,$$

dove Y^* è l'imposta teorica che dovrebbe essere dichiarata e/o versata in caso di *compliance*, X è l'IVA ricostruita da fatture e corrispettivi, U è il termine di errore che si assume seguire una distribuzione Gaussiana a media zero e varianza σ^2 , e α e β sono i parametri da stimare.

Nel secondo stadio, si ipotizza che l'insieme dei contribuenti sia suddiviso in due sottoinsiemi: un sottoinsieme composto da contribuenti *compliant*; un altro sottoinsieme in cui i contribuenti adottano comportamenti evasivi che si concretizzano in omesse o infedeli dichiarazioni e/o mancati o insufficienti versamenti. Si assume, inoltre, che quest'ultimo insieme si suddivida a sua volta in gruppi G_k ($k = 1, \dots, K - 1$), ciascuno corrispondente ad un determinato livello di *intensità* di rischio fiscale.

Per ogni contribuente, la probabilità π_k ($k = 0, \dots, K - 1$) di appartenere al k -esimo cluster G_k (probabilità *a priori*) può dipendere, a seconda che il contribuente dichiari (versi) oppure no, da un insieme di variabili di profilazione Z attraverso una funzione logistica secondo la relazione: $\log(\pi_k/\pi_0) = b'Z$, dove b' è un vettore di parametri e $k = 0$ identifica l'assenza di evasione; in tal modo, il cluster G_0 rappresenta il gruppo di riferimento.

Il modello risultante è un modello di tipo *logit-multinomiale*, in cui le variabili I_k , indicatrici di appartenenza ai diversi cluster ($I_k = 1$ in caso di appartenenza a G_k e $I_k = 0$ altrimenti), non sono osservate (variabili latenti).

Il terzo e ultimo stadio della formulazione del modello consiste nella specificazione della distribuzione dell'imposta effettivamente dichiarata (versata) condizionatamente al corrispondente valore dell'importo corretto. Tale distribuzione condizionata ha carattere semi-continuo a causa dell'inflazione di zeri determinata dai casi di omessa dichiarazione o omesso versamento; per questo motivo, vengono implementati due diversi modelli a seconda della presenza di valori dichiarati/versati.

Specificamente, i dati relativi ai soggetti non dichiaranti (o che non versano) vengono modellati mediante una opportuna distribuzione di probabilità per variabili miste (cioè sia quantitative che categoriche) che, previa ricodifica delle variabili stesse, può essere ricondotta ad una mistura finita di Gaussiane.

Per i contribuenti per cui sono presenti i valori dichiarati o versati, si specifica un "modello di evasione" in cui l'IVA dichiarata/versata viene espressa in funzione dell'imposta che dovrebbe essere dichiarata/versata in caso di perfetta *compliance*:

$$Y = \delta_k + \gamma_k Y^* + \epsilon_k,$$

dove ϵ_k è una variabile aleatoria distribuita secondo una gaussiana a media zero e varianza ξ_k^2 , indipendente da U , e δ_k e γ_k sono parametri da stimare.

Dalle assunzioni fatte, discende che la distribuzione delle variabili "manifeste" Y può essere espressa come mistura di K distribuzioni gaussiane:

$$f(Y|x, z; \theta) = \sum_{k=0}^{K-1} \pi_k(z; w) N(y; \mu_k, \sigma_k^2) \quad \text{con} \quad \sum \pi_k = 1$$

dove θ è il vettore dei parametri del modello ($\alpha_k, \beta_k, \sigma_k^2$) e $\pi_k(z; w)$, detti *mixing weights*, sono le probabilità a priori determinate dal modello *logit-multinomiale* specificato sopra.

Nel caso di omessa dichiarazione (o versamento), le variabili manifeste Y sono le stesse variabili che caratterizzano i diversi profili di rischio, mentre i pesi π_k sono costanti da stimare (non ci sono variabili Z).

Nel caso di importo dichiarato (versato) diverso da zero, sarà proprio quest'ultimo a giocare il ruolo di variabile manifesta Y e i *mixing weights*, $\pi_k(z; w)$, saranno funzioni delle variabili Z . Inoltre, in questo caso, le medie μ_k della variabile Y nei cluster saranno funzioni lineari del valore dell'imposta x ricostruito da fatture e corrispettivi (*mistura di regressioni*): $\mu_k = \alpha_k + \beta_k x$.

In entrambi i casi citati, i parametri del modello sono stimati con il metodo della massima verosimiglianza mediante l'algoritmo *EM* (*Expectation-Maximization*). Quest'ultimo si compone di due passi: nel primo passo viene calcolato il valore atteso del logaritmo della verosimiglianza utilizzando i valori correnti dei parametri (*Expectation*), nel secondo passo viene massimizzato tale valore atteso e dunque viene aggiornata la stima dei parametri (*Maximization*).