

UNIVERSITÀ DEGLI STUDI DI PADOVA

Dipartimento di Scienze Statistiche

Corso di Laurea in Statistica per l'Economia e l'Impresa

Un indicatore continuo di aliquota fiscale effettiva: analisi
dei dati suddivisi per modalità di estrazione del reddito

Relatrice:

Ch.mo Prof.ssa

Maria Michela Dickson

Laureando:

Alessandro Gobbo

Matricola: 2042159

Indice

1	Introduzione	3
2	Prelievo fiscale e relativi indicatori	5
2.1	Brevi richiami	5
2.2	Indicatori macro	6
2.3	Indicatori su lavoro e famiglie	8
2.4	Indicatori su imprese e investimenti	9
2.5	Indicatori di progressività e redistribuzione . . .	12
3	Aliquota fiscale effettiva: indicatore continuo	15
3.1	Scopo	15
3.2	Formalizzazione dell'indicatore	17
3.3	Misure di dispersione	19
3.4	Misure di pendenza	19
4	Analisi dei dati	21
4.1	Raccolta dati, fonti e problematiche	21
4.2	Metodologia	22
4.3	Reddito da lavoro dipendente	24
4.4	Reddito da lavoro autonomo in regime ordinario	30
4.5	Reddito da microimpresa	35
4.6	Reddito societario	39
4.7	Analisi comparativa	44
5	Conclusioni	49
5.1	Sintesi dei risultati	49
5.2	Limiti del lavoro	50

5.3	Sviluppi futuri	51
Appendice		53
A.1	Kernel Density Estimation	53
A.2	Simulazione di microdati da aggregati per classe	54
A.3	Lo stimatore di Nadaraya-Watson	56
A.4	LOESS: locally estimated scatterplot smoothing	56
A.5	Intervalli di ampiezza minima	58
A.6	Indici di concentrazione	58

Capitolo 1

Introduzione

Il gettito fiscale non si forma in modo uniforme lungo la distribuzione dei redditi. È lecito aspettarsi che in qualche modo venga a formarsi, per la maggior parte, in prossimità della fascia di reddito medio-basso, ovvero dove si concentrano i contribuenti. D'altra parte, alcune zone della distribuzione contribuiscono in misura sproporzionata rispetto alla loro numerosità. Valutare questo aspetto passa necessariamente attraverso un'analisi della distribuzione del gettito.

Da qui, le strade per approfondire il tema sono due. La prima opzione, quella su cui si sviluppa l'elaborato, si basa sull'analisi grafica della distribuzione del gettito o meglio sull'idea che possa essere informativa sul tema. Ipotizzando di ricavare la distribuzione del gettito come prodotto della densità di probabilità del reddito per una qualche funzione che descriva il prelievo, si può costruire in parallelo uno studio dell'aliquota da collegare a indicatori fiscali noti.

La sezione di analisi dati e commento dei risultati che ne segue va interpretata come costruzione e test dell'indicatore individuato e non come analisi economica: sono leggibili la forma e l'ordine di grandezza dei risultati, non il dettaglio fine, soprattutto considerando la mancanza di una quantificazione dell'incertezza e vari problemi sorti nella reperibilità di dati adatti e in alcune decisioni in fase di stima.

Resta aperta la seconda direzione, introdotta come sviluppo futuro della tesi. Accennata brevemente alla fine del capitolo 4, consiste nell'esaminare quantitativamente la concentrazione del gettito non solo come indice sintetico, ma nella sua forma lungo la distribuzione.

Capitolo 2

Prelievo fiscale e relativi indicatori

2.1 Brevi richiami

Per assolvere le sue funzioni pubbliche lo Stato necessita di ingenti risorse finanziarie. Una quota significativa la ottiene tramite emissione di debito, ma la parte più rilevante è costituita dai tributi, ovvero pagamenti obbligatori richiesti dallo Stato o dagli enti locali per finanziare la spesa pubblica. I tributi si suddividono in tre categorie:

- Imposte: pagamenti obbligatori che servono a finanziare i servizi generali e non sono legati a una richiesta specifica. Si parla di imposte dirette se colpiscono la ricchezza che già si possiede o che si guadagna, e di imposte indirette se colpiscono la ricchezza quando viene spesa.
- Tasse: pagamenti obbligatori dovuti in caso di utilizzo di un servizio pubblico specifico.
- Contributi: versamenti obbligatori che finanziano un'opera, un servizio pubblico specifico o una prestazione futura.

La percentuale che, applicata alla base imponibile (un reddito o il valore di un bene), determina l'importo di un tributo da versare è chiamata aliquota. Si parla di aliquota legale per indicare l'aliquota fissata dalla norma, di aliquota effettiva in

riferimento al rapporto fra imposta effettivamente pagata e base imponibile, e di aliquota marginale, definita come l'aliquota che si applica sull'ultimo euro di reddito guadagnato.

L'ammontare complessivo delle entrate tributarie incassate dallo Stato costituisce il gettito fiscale.

Tributi, aliquote e gettito fiscale sono nozioni elementari del linguaggio fiscale, ma da sole dicono poco su quanto pesi in concreto il prelievo fiscale su un dato soggetto o su una data categoria di contribuenti: un'aliquota legale non coincide quasi mai con quella effettivamente sopportata e il gettito complessivo nulla dice su come il carico si distribuisca fra chi guadagna poco e chi guadagna molto. Per rendere questo carico misurabile, confrontabile e dunque informativo, serve un ulteriore passo di sintesi che condensi basi imponibili e aliquote eterogenee in misure costruite appositamente per rispondere a domande più specifiche: è il compito degli indicatori fiscali.

Nei paragrafi che seguono se ne propone una rassegna organizzata per famiglie: indicatori macro, costruiti sui conti nazionali per fotografare il prelievo complessivo di un paese; indicatori su lavoro e famiglie, che si spostano alla scala del singolo lavoratore o nucleo familiare; indicatori su imprese e investimenti, pensati per valutare l'impatto fiscale su una decisione di investimento o sulla competitività di un'impresa; indicatori di progressività e redistribuzione, che misurano come il sistema fiscale ridisegna la distribuzione del reddito.

2.2 Indicatori macro

La prima famiglia riunisce gli indicatori pensati per una descrizione aggregata del sistema fiscale nel suo complesso. Non sono costruiti su dati individuali, ma sui conti nazionali e descrivono quanto lo Stato preleva nel suo complesso.

Il primo indicatore da presentare è la **Pressione Fiscale**, che misura quantitativamente il prelievo fiscale attuato dallo Stato, ovvero le entrate tributarie e contributive complessive, rapportato al PIL. Si distingue la pressione fiscale “legale” da quella “effettiva”, perché quest’ultima tiene conto anche dell’economia sommersa.

$$\text{Pressione fiscale (\%)} = \frac{\text{Imposte} + \text{Contributi sociali}}{\text{Prodotto Interno Lordo (PIL)}} \times 100$$

La versione OCSE della pressione fiscale è detta **Tax-to-GDP Ratio**. Concettualmente è identica, ma può differire lievemente a causa di diverse metodologie di trattamento dei dati.

Il **Fiscal Drag** misura l’aumento della pressione fiscale che si verifica quando i redditi nominali crescono per via dell’inflazione, ma l’aliquota legale non viene ricalibrata, spingendo i contribuenti in scaglioni più alti senza un corrispondente aumento di reddito reale.

Un approccio complementare a queste misure consiste nel ricostruire, sempre a partire dai conti nazionali, la base imponibile effettivamente colpita da ciascun tipo di prelievo (lavoro, capitale, consumi), così da calcolare un’aliquota implicita su ciascuna di esse.

Implicit Tax Rate (ITR)

Mendoza, Razin e Tesar, autori del paper *Effective Tax Rates in Macroeconomics (1994)*, propongono la costruzione di aliquote fiscali effettive aggregate su lavoro, capitale e consumo a partire dai conti nazionali, applicando ai grandi aggregati (monte salari, redditi da capitale, consumi) le imposte effettivamente

riscosse su quelle basi.

Eurostat adotta una metodologia simile e pubblica l'Implicit Tax Rate (ITR), calcolato come rapporto fra il gettito effettivamente riscosso su una base economica (lavoro, capitale, consumi) e la base imponibile potenziale corrispondente, stimata dai conti nazionali. Si scompone in ITR on labour, ITR on capital e ITR on consumption, ed è l'indicatore standard per confrontare il carico fiscale fra Stati membri UE.

L'ITR sul lavoro, ad esempio, assume la forma:

$$ITR_{\text{lavoro}} = \frac{\text{imposte personali sul reddito} + \text{contributi}}{\text{costo del lavoro totale}}$$

2.3 Indicatori su lavoro e famiglie

Questa seconda famiglia guarda al prelievo dal punto di vista di chi lavora, per capire come lo stesso prelievo si traduca in un incentivo o un disincentivo per le scelte del singolo. Non si elabora più il gettito aggregato, ma la differenza fra quanto costa un lavoratore all'impresa, quanto gliene resta in tasca, e quanto gli converrebbe, in termini fiscali, lavorare piuttosto che restare inattivo.

Nel contesto del lavoro dipendente e della competitività imprenditoriale va nominato il **cuneo fiscale** (tax wedge), che misura la differenza fra il costo del lavoro e la retribuzione netta che arriva al lavoratore, espressa come percentuale del costo del lavoro:

$$\text{Cuneo fiscale (\%)} = \frac{\text{Imposte sul reddito} + \text{Contributi}}{\text{Costo totale sostenuto dall'impresa}} \times 100$$

Nello standard OCSE viene calcolato su un lavoratore rappresentativo, a livelli di reddito fissati e per diverse tipologie di composizione del nucleo familiare, per rendere i confronti comparabili.

Sempre nell'ambito delle imposte dirette, ma limitato alla sola componente reddituale, si definisce il **Net Personal Average Tax Rate** (NPATR), l'aliquota media effettivamente sopportata su un dato livello di reddito:

$$NPATR(y) = \frac{\text{imposta netta dovuta}(y)}{\text{reddito lordo}(y)}$$

Il **Participation Tax Rate** (PTR) è l'indicatore OCSE/UE che misura la perdita netta di reddito disponibile quando una persona passa dalla disoccupazione/inattività a un lavoro, includendo sia le imposte pagate sia i sussidi persi: mostra quanto “conviene” davvero tornare al lavoro. Misura quanto del guadagno da un passaggio dalla non-attività al lavoro viene sottratto da imposte e perdita di trasferimenti:

$$PTR = 1 - \frac{y_{\text{occupato}}^{\text{netto}} - y_{\text{non occupato}}^{\text{netto}}}{y_{\text{occupato}}^{\text{lordo}}}$$

2.4 Indicatori su imprese e investimenti

La terza famiglia di indicatori si sposta al punto di vista di chi decide di investire capitale. Qui l'unità di analisi non è più il reddito percepito, ma il progetto di investimento: si chiede di quanto il fisco riduca il rendimento che un'impresa, o chi la

finanzia, può attendersi dapprima in scenari ipotetici standardizzati, poi su dati contabili effettivamente osservati.

Marginal Effective Tax Rate su investimento (METR)

King e Fullerton (1984), *The Taxation of Income from Capital*, NBER

Calcola l'aliquota effettiva marginale su un investimento marginale ipotetico, a partire dallo scarto fra rendimento lordo richiesto dall'investitore e costo del capitale al netto di imposte e ammortamenti fiscali:

$$METR = \frac{p - s}{p}$$

dove p è il tasso di rendimento minimo che un investimento deve generare prima delle tasse per poter essere realizzato, e s è il tasso di rendimento che l'investitore riceve in tasca, dopo che sono state pagate tutte le imposte (sia a livello societario sia personale).

Si tratta di una stima predittiva del carico fiscale di un investimento marginale standardizzato, che permette di valutare l'opportunità di un investimento dal punto di vista del rendimento atteso.

Effective Average Tax Rate (EATR)

Devereux e Griffith (1999, 2003)

La metodologia di King-Fullerton viene estesa a investimenti non marginali e a elevata redditività, particolarmente rilevante per le scelte di localizzazione dell'impresa:

$$EATR = \frac{R^* - R}{p(1 + r)}$$

dove R^* è il valore attuale netto del progetto di investimento senza imposte, R è il VAN del progetto con le imposte applicate. La differenza $R^* - R$ rappresenta quindi il valore sottratto al progetto sotto forma di imposte, mentre il denominatore rappresenta il valore attuale del rendimento finanziario totale del progetto (p è il tasso di profitto atteso, r il tasso di interesse reale). Anche questo indicatore si basa su una stima di uno scenario rappresentativo non osservato.

Effective Tax Rate contabile (GAAP ETR) e Cash ETR Fullerton (1984), *Which Effective Tax Rate?*, NBER/National Tax Journal

Oltre agli indicatori di tipo predittivo appena visti, esistono anche misure basate su un approccio *backward-looking*, cioè osservate su dati contabili. Le più utilizzate nella letteratura contabile/finanziaria sono:

$$GAAP\ ETR = \frac{\text{imposte a conto economico}}{\text{utile ante imposte}}$$

$$Cash\ ETR = \frac{\text{imposte effettivamente pagate in cassa}}{\text{utile ante imposte}}$$

La loro differenza è utile per studiare il fenomeno della *tax avoidance* nelle grandi multinazionali.

Il **Total Tax and Contribution Rate** è l'indicatore usato da World Bank/PwC nel report *Paying Taxes*: misura il carico fiscale complessivo su un'impresa tipo come percentuale del profitto commerciale, utile per confronti di competitività fiscale fra paesi.

2.5 Indicatori di progressività e redistribuzione

Le famiglie di indicatori viste finora descrivono il livello del prelievo, ma non dicono nulla su come tale livello si distribuisca fra chi ha redditi diversi. Gli indicatori di progressività servono proprio a misurare questo fenomeno, basandosi sulla stessa metodologia applicata per quantificare la disuguaglianza a livello nazionale.

Indice di Kakwani

Kakwani (1977), *Measurement of Tax Progressivity: An International Comparison*, Economic Journal

Misura la progressività di un'imposta come differenza fra il coefficiente di concentrazione delle imposte e il coefficiente di Gini del reddito lordo:

$$K = C_T - G_X$$

Range teorico $[-1, 1]$: $K > 0$ progressivo, $K = 0$ proporzionale, $K < 0$ regressivo.

Indice di Suits

Suits (1977), *Measurement of Tax Progressivity*, American Economic Review

Analogo di Kakwani, ma applicato al carico fiscale rispetto al reddito, cioè basato su una curva di concentrazione con il reddito cumulato sull'asse x:

$$S = 1 - 2 \int_0^1 L_T(p) dp$$

dove $L_T(p)$ è la quota cumulata di imposta pagata dal p -esimo percentile cumulato di reddito. Range $[-1, 1]$, con la stessa in-

interpretazione dell'indice di Kakwani.

Indice di Reynolds-Smolensky

Reynolds e Smolensky (1977)

Misura l'effetto redistributivo complessivo del sistema fiscale, confrontando il Gini ante imposte e il Gini post imposte/trasferimenti:

$$RS = G_X - G_N$$

dove G_N è il Gini del reddito netto.

Indice di Musgrave-Thin

Musgrave e Thin (1948), *Income Tax Progression, 1929-48*, Journal of Political Economy

L'approccio di Musgrave-Thin è composto da tre misure di progressività locale:

- **Average Rate Progression (ARP):** $\frac{d\tau(x)}{dx}$

la pendenza locale dell'aliquota media, valuta se esiste progressività per un reddito fissato.

- **Liability Progression (LP):** $\frac{\tau'(x) \cdot x}{\tau(x)}$

elasticità dell'imposta pagata rispetto al reddito. Se $LP > 1$, l'aliquota marginale è maggiore dell'aliquota media, cioè il sistema è localmente progressivo.

- **Residual Income Progression (RIP):** $\frac{d\log(x - T(x))}{d\log x}$

elasticità del reddito netto rispetto al reddito lordo, cioè come reagisce il reddito netto al variare del reddito lordo. $RIP < 1$ indica un sistema progressivo (l'aliquota marginale m eccede quella media t); $RIP = 1$ proporzionale ($m = t$); $RIP > 1$ regressivo ($m < t$).

Parametro di progressività di Heathcote-Storesletten-Violante (HSV)

Heathcote, Storesletten e Violante (2017), *Optimal Tax Progressivity: An Analytical Framework*

Forma funzionale parametrica per il reddito netto:

$$y^{\text{netto}} = \lambda \cdot (y^{\text{lordo}})^{1-p}$$

equivalente a un'aliquota media $\tau(y) = 1 - \lambda y^{-p}$, dove p è il singolo parametro di progressività e λ il livello medio dell'aliquota.

$p = 0$: aliquota flat; $p > 0$: progressivo; $p < 0$: regressivo.

È il riferimento più citato in macro per sintetizzare in un solo parametro, oltre al livello, l'intera progressività di un sistema fiscale.

Capitolo 3

Aliquota fiscale effettiva: indicatore continuo

3.1 Scopo

Gli indicatori macro, come l'ITR o la pressione fiscale, restituiscono un solo numero ottenuto rapportando due totali. Sono estremamente sintetici, ma per costruzione non dicono nulla su come il prelievo si distribuisca. Gli indicatori sul lavoro sono funzioni del reddito, ma vengono calcolati solo in pochi punti fissati e rappresentativi, non lungo l'intera distribuzione empirica dei redditi. Gli indicatori di progressività guardano all'intera distribuzione, ma la sintetizzano in un unico numero che non permette di localizzare a quale livello di reddito la progressività sia più o meno marcata.

Come anticipato nel capitolo introduttivo, la finalità principale del progetto è trovare un indicatore sintetico in grado di fornire un supporto numerico e grafico utile a valutare quale fascia di reddito sia più rilevante per la costituzione del gettito fiscale. La rilevanza di un dato reddito consiste in due aspetti distinti: una sorta di “efficienza” del singolo prelievo fiscale su quel reddito, cioè qualcosa che abbia a che fare con l'aliquota e con l'ammontare stesso del prelievo individuale, e la capacità di quel reddito di portare risorse in virtù della sua frequenza nella distribuzione complessiva.

Nessuno degli approcci evidenziati nel capitolo precedente si adatta adeguatamente a questo scopo: da qui l'ipotesi di costruire un indicatore che sia al tempo stesso coerente con le misure già esistenti in letteratura e capace di rispondere alle esigenze appena descritte.

Onorare i requisiti stabiliti richiede di tenere insieme due informazioni che, prese singolarmente, sono insufficienti: quanto è frequente un dato livello di reddito nella popolazione, e quanto viene prelevato nominalmente a quel livello. Per avere un quadro completo, inoltre, non basta guardare un singolo reddito esemplificativo o pochi punti fissati, ma serve l'intera distribuzione dei redditi. Occorrono, dunque, due ingredienti: una densità di probabilità del reddito $f(x)$ e una funzione di prelievo $\tau(x)$ definita punto per punto lungo tutto il suo supporto. Il loro prodotto, soprattutto se rappresentato graficamente, soddisfa già le aspettative: mostra dove, lungo la distribuzione, si concentra la formazione del gettito. Affinché si tenga conto anche dell'efficienza del prelievo appena descritta, è necessario non perdere le informazioni sulla coda alta della distribuzione, dove i redditi sono rari ma il loro contributo al gettito totale può essere sproporzionato. Per questo motivo $f(x)$ viene pesata per il reddito stesso, ottenendo $f_{IW}(x)$. Integrando infine il prodotto $f_{IW}(x) \cdot \tau(x)$ su tutto il supporto si ottiene un'unica aliquota media effettiva aggregata che sintetizza in un solo numero l'informazione contenuta nelle due funzioni di partenza, una misura coerente nella sua struttura con quanto già visto per l'ITR e per gli ETR del capitolo 2. Per brevità, nel seguito ci si riferirà a questo indicatore con la sigla CETR (Continuous Effective Tax Rate).

3.2 Formalizzazione dell'indicatore

Sia X la variabile che rappresenta il reddito, con densità di probabilità $f(x)$ e media $E[X]$. Si definisce la densità del reddito pesata per il reddito stesso:

$$f_{IW}(x) = \frac{x \cdot f(x)}{E[X]} \quad (3.1)$$

$f_{IW}(x)$ ha la stessa interpretazione di una funzione di densità (è non negativa e integra a 1), ma a differenza di $f(x)$ non descrive i contribuenti, bensì l'unità di reddito.

Sia poi $\tau(x) = T(x)/x$ la funzione di prelievo, dove $T(x)$ è il tributo versato al reddito x : $\tau(x)$ restituisce, per ciascun livello di reddito, la corrispondente aliquota media effettiva.

L'indicatore centrale del lavoro è:

$$\text{CETR} = \int_0^\infty f_{IW}(x) \tau(x) dx \quad (3.2)$$

Sostituendo $f_{IW}(x)$ con la sua definizione:

$$\begin{aligned} \text{CETR} &= \frac{\int_0^\infty x \tau(x) f(x) dx}{E[X]} = \frac{E[X \cdot \tau(X)]}{E[X]} \\ &= \frac{\text{Gettito totale atteso}}{\text{Reddito totale atteso}} \end{aligned} \quad (3.3)$$

Oltre a spostare il focus dal soggetto giuridico all'unità di reddito, l'aver pesato $f(x)$ per il reddito stesso permette di ottenere, come misura sintetica, una struttura di rapporto aggregato concettualmente molto vicina a indicatori noti come la pressione

fiscale, l'ITR o un ETR generalizzato ad altre categorie a seconda di quali versamenti $T(x)$ comprende. Il CETR, dunque, coincide con l'aliquota fiscale effettiva aggregata, qui ricavata da un'operazione di integrazione di due funzioni stimate empiricamente lungo l'intera distribuzione.

La questione sulla rilevanza di un dato reddito per la costituzione del gettito è invece affidata all'integrando $g(x) = f_{IW}(x) \cdot \tau(x)$.

Riscrivendo $g(x)$ e sostituendo $f_{IW}(x)$ con la sua definizione:

$$g(x) = \frac{x \cdot \tau(x) \cdot f(x)}{E[X]} = \frac{T(x) \cdot f(x)}{E[X]} \quad (3.4)$$

dove $T(x) = x \cdot \tau(x)$ è il versamento corrispondente al reddito x . $g(x)$ è quindi, a meno della costante $1/E[X]$, la densità di gettito lungo la distribuzione: descrive quanto gettito si genera per unità di reddito attorno al livello x , pesando sia quanti contribuenti si trovano a quel livello ($f(x)$) sia quanto versano ($T(x)$).

L'area sotto $g(x)$ fra due redditi x_1 e x_2 è:

$$\int_{x_1}^{x_2} g(x) dx = \frac{\text{gettito dalla fascia } [x_1, x_2]}{\text{reddito totale}} \quad (3.5)$$

Dividendo quest'area per l'area totale su tutto il supporto $[0, \infty[$ si ricava:

$$\frac{\text{area}([x_1, x_2])}{\text{CETR}} = \frac{\text{gettito dalla fascia}}{\text{gettito totale}} \quad (3.6)$$

cioè la quota di gettito complessivo proveniente da quella fascia di reddito, esattamente l'informazione richiesta da cui ha origine il progetto.

3.3 Misure di dispersione

Il solo valore di CETR non dice se l'aliquota effettiva sia costante lungo tutta la distribuzione pesata o se vari molto da un punto all'altro. Per quantificare questa variabilità si costruisce il secondo momento di $\tau(X)$ sotto la stessa misura f_{IW} :

$$\begin{aligned}\text{Var}_{IW}(\tau) &= E_{IW}[\tau(X)^2] - \text{CETR}^2 \\ &= \int_0^\infty f_{IW}(x) \tau(x)^2 dx - \text{CETR}^2\end{aligned}\tag{3.7}$$

con deviazione standard $\text{SD}_{IW}(\tau) = \sqrt{\text{Var}_{IW}(\tau)}$, espressa nella stessa unità di τ (punti percentuali di aliquota). Si tratta di una misura di dispersione delle aliquote effettive sotto una distribuzione che attribuisce maggior peso ai livelli di reddito che rappresentano una quota maggiore del reddito aggregato. Per confrontare categorie con un livello di CETR molto diverso fra loro, si affianca una versione normalizzata, il coefficiente di variazione:

$$\text{CV}_{IW} = \frac{\text{SD}_{IW}(\tau)}{\text{CETR}}\tag{3.8}$$

3.4 Misure di pendenza

Oltre alla dispersione, interessa sapere se l'aliquota tenda sistematicamente ad aumentare o a diminuire al crescere del reddito. Una misura naturale è la pendenza della retta di regressione di $\tau(X)$ su X : coerentemente con la costruzione del CETR, covarianza e varianza sono calcolate rispetto alla stessa misura f_{IW} :

$$\beta = \frac{\text{Cov}_{IW}(X, \tau(X))}{\text{Var}_{IW}(X)}\tag{3.9}$$

Per $\beta > 0$ l'aliquota pesata cresce con il reddito, per $\beta = 0$ l'aliquota pesata è costante, per $\beta < 0$ l'aliquota pesata diminuisce al crescere del reddito.

È lo stesso oggetto misurato puntualmente dall'Average Rate Progression di Musgrave-Thin ($d\tau(x)/dx$), qui ricostruito come pendenza aggregata sull'intera distribuzione pesata anziché come derivata locale in un singolo punto. La formulazione è simile, ma il significato in questo caso si allontana da quello standard di progressività sulla popolazione: l'oggetto di analisi è la relazione tra reddito e aliquota calcolata su una distribuzione pesata, non dice nulla sulla progressività dell'imposta dal punto di vista del contribuente.

Per confrontare categorie con range di reddito molto diversi fra loro si affianca una versione adimensionale, la correlazione:

$$\rho_{IW} = \frac{\text{Cov}_{IW}(X, \tau(X))}{\text{SD}_{IW}(X) \cdot \text{SD}_{IW}(\tau)} \in [-1, 1] \quad (3.10)$$

Capitolo 4

Analisi dei dati

4.1 Raccolta dati, fonti e problematiche

La struttura del CETR formalizzata nel capitolo 3 richiede implicitamente di avere a disposizione microdati di redditi e di tributi versati. Non solo: questi stessi microdati devono contenere informazioni riguardo la provenienza dei redditi. I diversi regimi fiscali rendono infatti necessaria una suddivisione delle casistiche, in modo da ottenere risultati interpretabili e non distorti da un'operazione di sintesi cieca.

Accedere a questo tipo di dati richiede lunghi tempi di attesa e, nella maggior parte dei casi, permessi di ricerca. In mancanza di entrambi gli elementi, l'approvvigionamento dei dati e la conseguente metodologia di analisi sono al servizio di una finalità più orientata al testing dell'indicatore proposto e a un'esplorazione di superficie, più che a una vera e propria analisi economica.

Entrando nello specifico, le modalità di estrazione del reddito individuate sono quattro: reddito da lavoro dipendente, reddito da lavoro autonomo in regime ordinario, reddito da microimpresa e reddito societario. Tali categorie sono trattate separatamente proprio perché soggette a regimi fiscali strutturalmente non comparabili: una distribuzione unica per l'intera platea di contribuenti sommerebbe basi imponibili e aliquote eterogenee, appiattendolo la variazione che l'indicatore è costruito per cogliere.

Sebbene il progetto prevedesse inizialmente di includere anche i contributi previdenziali, $T(x)$ comprende solo le imposte dirette (IRPEF netta e addizionali, oppure imposte correnti, IRES e IRAP, nel caso delle società) a causa delle difficoltà riscontrate nel reperire dati adatti e comparabili fra le categorie. Questo allontana il CETR concettualmente da indicatori come il tax wedge o la pressione fiscale, riducendolo ad un effective tax rate limitato alla sola componente costituita dalle imposte dirette.

Per microimprese e società si utilizza un campione di bilanci individuali provenienti dalla banca dati AIDA (Bureau van Dijk): qui i dati sono già molto vicini alle caratteristiche ricercate. Per le altre due categorie, la fonte è invece il Ministero dell'Economia e delle Finanze (MEF), che pubblica statistiche aggregate per classi di reddito: per ciascuna classe si conosce il numero di contribuenti e l'ammontare complessivo di reddito, di imposta IRPEF e di addizionali. Il limite più rilevante della fonte MEF è che $f(x)$ e $\tau(x)$ non sono osservabili direttamente a partire da microdati: la fonte non contiene alcun reddito individuale, solo frequenza e ammontare per classe. Per ottenere le due funzioni continue richieste dal CETR è quindi necessario simulare microdati coerenti con le frequenze di classe osservate, e stimare $f(x)$ e $\tau(x)$ su questo campione simulato anziché su osservazioni reali. I dati sono tutti relativi all'anno d'imposta 2023.

4.2 Metodologia

La stima di $f(x)$ si basa in tutti i casi su un kernel density estimate (KDE) gaussiano condotto su scala logaritmica del reddito, con larghezza di banda fissata dalla regola di Silverman. Il metodo per ottenere $\tau(x)$ dipende invece dalla natura della fonte.

Categorie MEF (dipendenti, autonomi)

Poiché il dato di partenza è aggregato per classi, si simula un

campione di redditi individuali coerente con la frequenza di ciascuna classe, assegnando a ogni reddito simulato l'aliquota media della classe di appartenenza. Da questo campione si stimano poi due KDE sulla stessa griglia di punti: una pesata per il solo numero di contribuenti, che restituisce $f(x)$; una pesata per la “massa fiscale” (contribuenti $\times$ aliquota di classe), che restituisce la densità del gettito atteso. Il rapporto punto per punto fra le due KDE,

$$\tau(x) = \frac{\text{KDE pesata su massa fiscale}(x)}{f(x)}$$

costituisce una stima continua di $\tau(x)$: vicino ai confini di classe, il procedimento mescola gradualmente le osservazioni evitando il salto netto che si avrebbe lasciando τ_{classe} come funzione a gradini. Questo stimatore è, formalmente, uno stimatore di Nadaraya-Watson per la regressione non parametrica di τ su X , con pesi campionari aggiuntivi che tengono conto di quanti contribuenti reali rappresenta ciascuna osservazione simulata. La procedura è approfondita in appendice.

Categorie AIDA (microimprese, società)

Qui il dato è già individuale ma rumoroso: il rapporto puntuale $\tau_i = T_i/EBT_i$ oscilla molto da un'impresa all'altra. Si usa quindi un regressore loess di τ su $\log(\text{reddito})$, che stima una $\tau(x)$ liscia mediando localmente su molte osservazioni vicine: la funzione di prelievo emerge direttamente dai dati osservati. Analogamente al caso precedente si rimanda in appendice per approfondimenti.

Per ciascuna categoria si confronta il CETR ottenuto per integrazione con CETR_{agg} , il rapporto diretto fra imposta totale e reddito totale calcolato sui dati grezzi puliti prima di qualunque stima: uno scarto contenuto fra i due valori indica che la pro-

cedura di stima non introduce distorsioni sistematiche rispetto all'aggregato di partenza.

4.3 Reddito da lavoro dipendente

Fonte: MEF, portale Analisi statistiche delle dichiarazioni fiscali — Dichiarazioni 2024, anno d'imposta 2023; Tipo di imposta *IRPEF*; Modello *Persone fisiche totali*; Tipologia contribuente *Lavoratori dipendenti*.

La tabella è classificata per classi di reddito complessivo, con numero di contribuenti, reddito complessivo, reddito da lavoro dipendente o assimilati, imposta netta *IRPEF* e addizionali regionale/comunale. La tabella è tarata sul reddito complessivo: per imposta e addizionali non è quindi possibile isolare la sola quota dovuta al lavoro dipendente. Assunto questo come limite dichiarato della fonte, si utilizza il reddito complessivo come denominatore nella stima di $\tau(x)$, in modo da mantenerne la coerenza. L'effetto dell'approssimazione è trascurabile fino a redditi di circa 100.000 euro, soglia oltre la quale diventa significativo, probabilmente per effetto di pensioni, redditi da lavoro autonomo e partecipazioni.

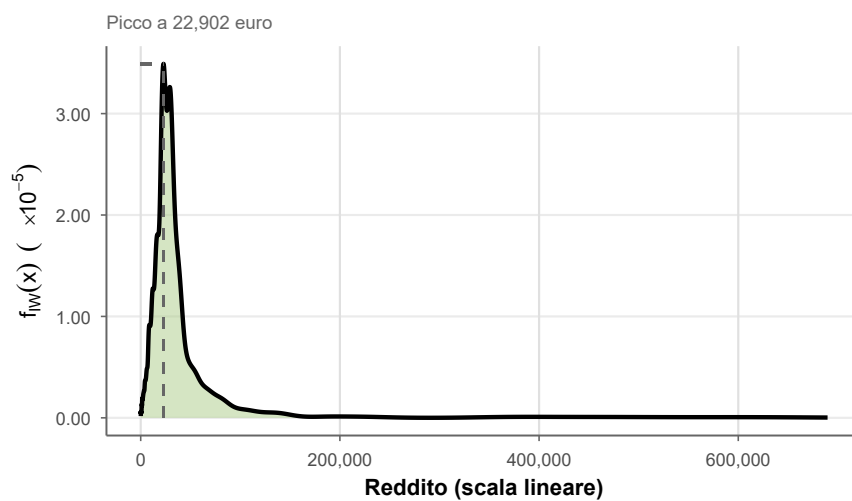


Figura 1.1: $f_{IW}(x)$ per il reddito da lavoro dipendente (scala lineare).

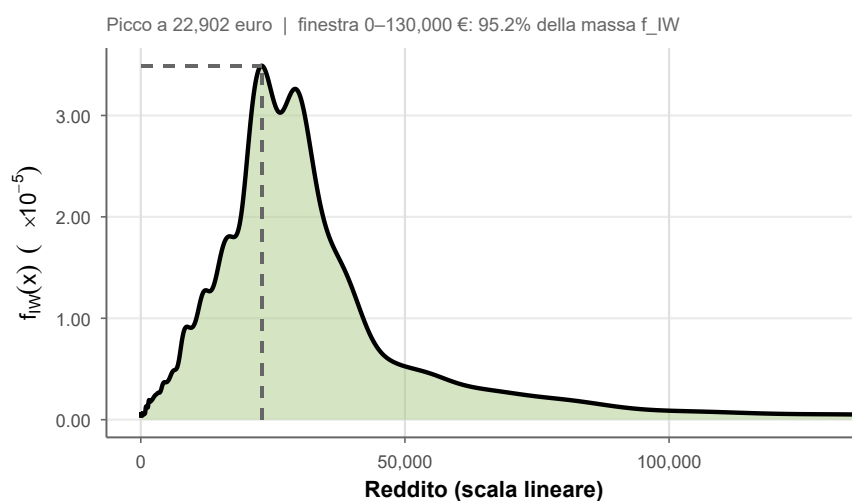


Figura 1.2: $f_{IW}(x)$ per il reddito da lavoro dipendente, zoom 0–130.000 euro (scala lineare).

Media 48.706 €, mediana 29.357 €, primo quartile 21.646 € e terzo quartile 41.846 €; skewness 5.20, curtosi in eccesso 29.49.

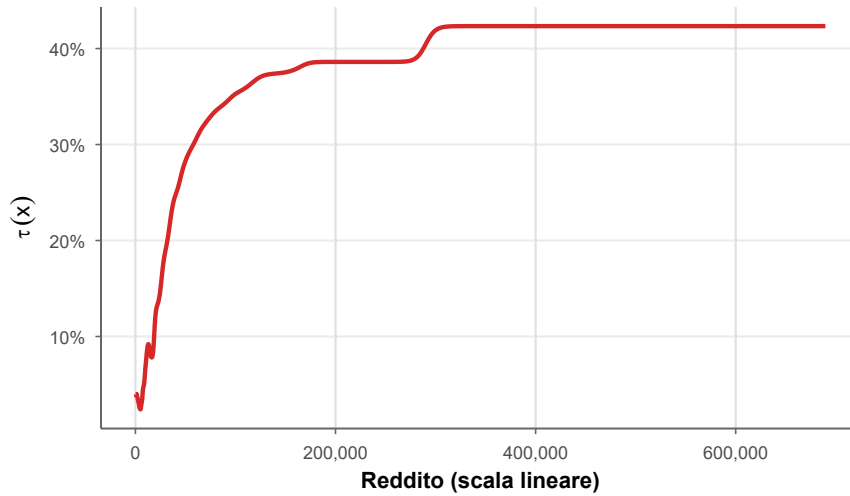


Figura 1.3: $\tau(x)$ per reddito da lavoro dipendente (scala lineare).

Si noti che $\tau(x)$ raggiunge il 43% (aliquota marginale massima IRPEF) solo oltre i 300.000 euro, non subito dopo i 50.000 euro in cui inizia lo scaglione corrispondente: l'IRPEF infatti è un'imposta progressiva e $\tau(x)$ è un'aliquota media, non marginale, dunque converge al 43% solo asintoticamente man mano che il peso dei redditi tassati agli scaglioni più bassi diventa trascurabile, effetto rinforzato dalle detrazioni da lavoro dipendente molto generose a reddito medio-basso e decrescenti verso l'alto. I due gradini visibili intorno a 150-180 mila euro e a 300-320 mila euro non sono un effetto diretto degli scaglioni IRPEF, ma del dato aggregato originario: si tratta delle ultime due classi di reddito, molto ampie e a cui è assegnata un'unica aliquota. Il comportamento non monotono per redditi bassi riflette, invece, le caratteristiche del dato aggregato di partenza.

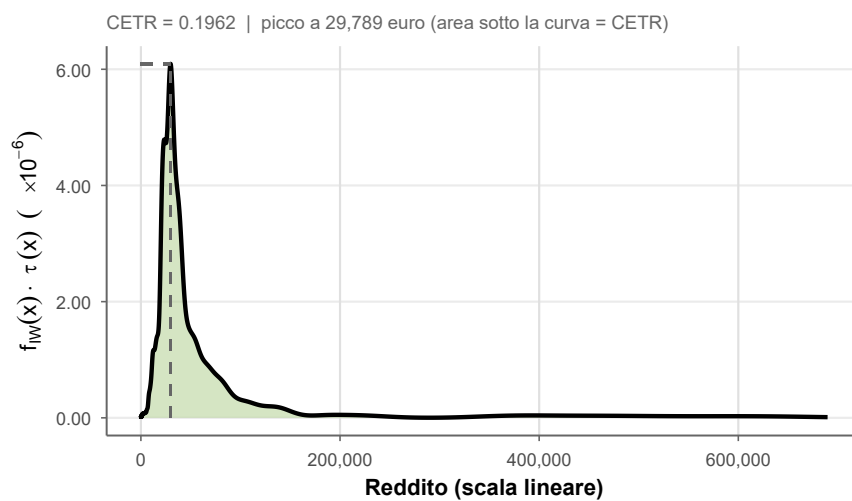


Figura 1.4: Integrando del CETR per il reddito da lavoro dipendente (scala lineare)

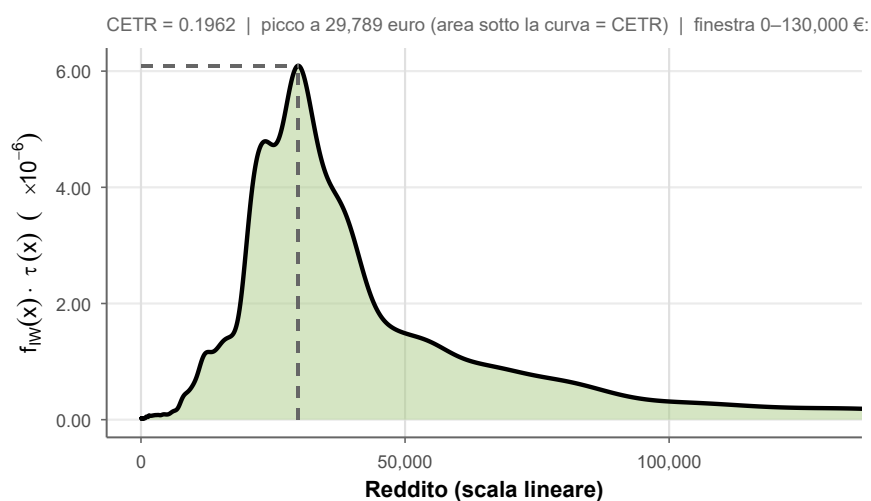


Figura 1.5: Integrando del CETR per il reddito da lavoro dipendente, zoom 0–130.000 euro (scala lineare)

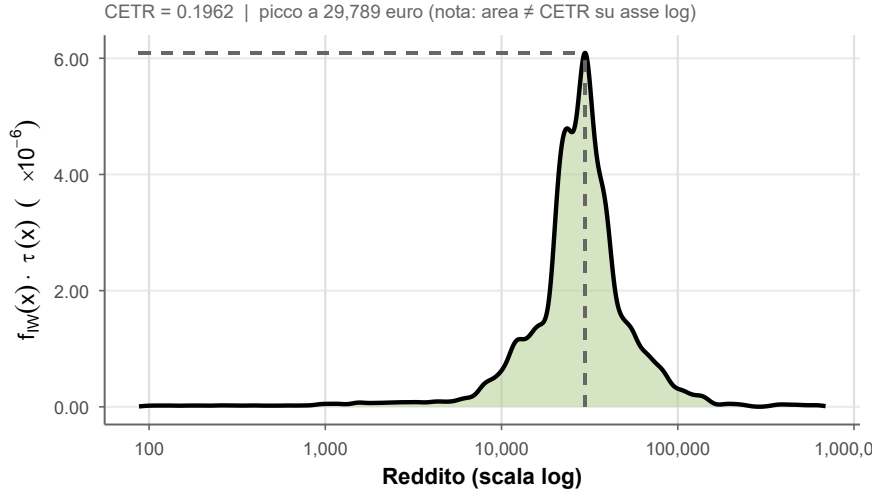


Figura 1.6: Integrando del CETR per il reddito da lavoro dipendente (scala logaritmica)

$\text{CETR} = 0.1962$ (19.6%), $\text{SD}_{IW}(\tau) = 9.3$ pp, $\text{CV}_{IW} = 0.47$. La correlazione $\rho_{IW} = 0.65$ conferma la tendenza generale dell'aliquota a crescere lungo l'intera distribuzione pesata, comportamento coerente con la struttura a scaglioni IRPEF combinata con le detrazioni decrescenti. Il confronto di robustezza con l'aggregato grezzo ($\text{CETR}_{agg} = 0.197$), calcolato come media delle aliquote di classe pesate per il reddito, mostra uno scarto di -0.1% : la catena di stima non introduce distorsioni sistematiche rispetto all'aggregato di partenza.

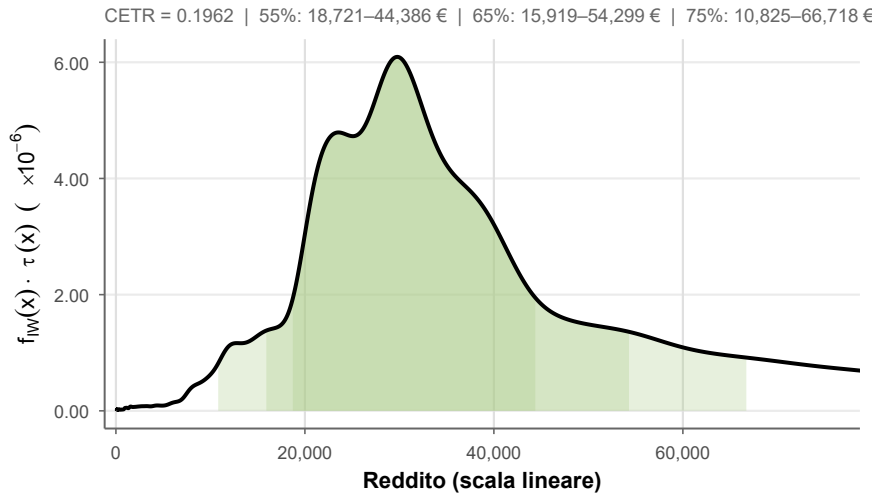


Figura 1.7: Integrando del CETR per il reddito da lavoro dipendente, con evidenziati gli intervalli di ampiezza minima che contengono il 55%, 65% e 75% del CETR.

Il grafico mostra le fasce di reddito più rilevanti per la costituzione del gettito. Scelte arbitrariamente tre quote, identificate dai seguenti intervalli (in euro):

- 55% $\rightarrow$ [18.721 , 44.386]
- 65% $\rightarrow$ [15.919 , 54.299]
- 75% $\rightarrow$ [10.825 , 66.718]

4.4 Reddito da lavoro autonomo in regime ordinario

Fonte: MEF, portale Analisi statistiche delle dichiarazioni fiscali — Dichiarazioni 2024, anno d'imposta 2023; Tipo di imposta *IRPEF*; Modello *Persone fisiche titolari di partita IVA*; Tipologia contribuente *Professionisti*.

Come per i dipendenti, le classi sono definite sul reddito complessivo, mentre la variabile di interesse è il reddito da lavoro autonomo, sistematicamente più basso. L'imposta netta è calcolata sull'intera posizione fiscale del contribuente, non isolata sul solo reddito professionale. Le addizionali regionale e comunale non sono disponibili per i soli professionisti in questa tabella e sono omesse (effetto aggiuntivo stimato di circa 1-2 punti percentuali). Anche in questo caso, il limite strutturale dei dati è accettato per le stesse ragioni già esposte nel paragrafo 4.1 riguardo alle problematiche di reperimento dei dati e al conseguente effetto sul significato del progetto.

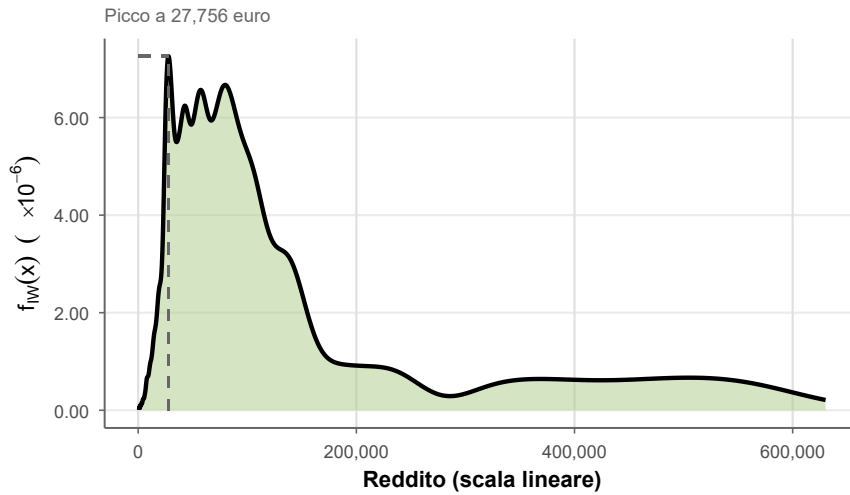


Figura 2.1: $f_{IW}(x)$, per il reddito da lavoro autonomo ordinario (scala lineare)

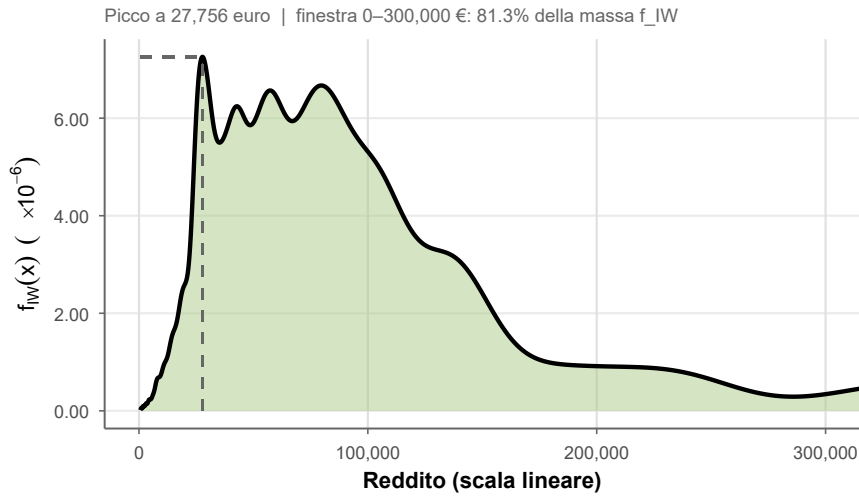


Figura 2.2: $f_{IW}(x)$ per il reddito da lavoro autonomo ordinario, zoom 0–300.000 euro (scala lineare)

Media 163.253 €, mediana 99.401 €, primo quartile 58.974 € e terzo quartile 198.525 €; skewness 1.43, curtosi in eccesso 0.86.

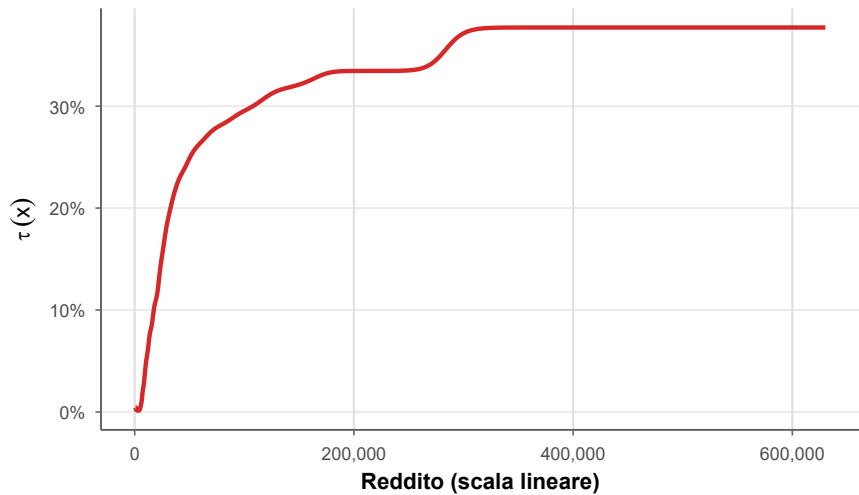


Figura 2.3: $\tau(x)$ per il reddito da lavoro autonomo ordinario (scala lineare)

A differenza dei dipendenti, la curva satura attorno al 38% senza avvicinarsi ulteriormente al 43% marginale. La spiegazione è nei dati di origine: l'ultima classe MEF, aperta (“oltre 300.000

euro” di reddito complessivo), raccoglie da sola il 3,2% dei professionisti ma il 18,6% del reddito complessivo aggregato e il 24,2% dell’imposta netta totale. Il confine superiore usato per collocarla sull’asse del reddito da lavoro autonomo è però arbitrario, dunque la curva resta piatta per costruzione, non perché l’aliquota smetta realmente di crescere per i redditi molto alti.

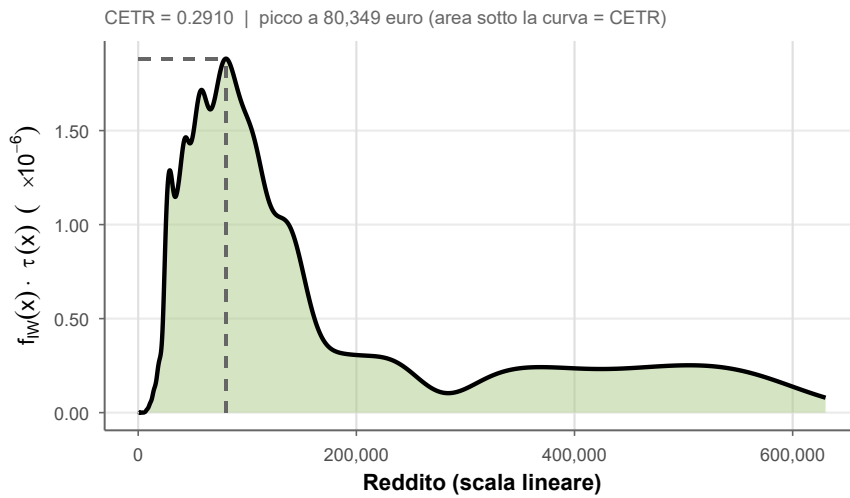


Figura 2.4: Integrando di CETR per il reddito da lavoro autonomo ordinario (scala lineare)

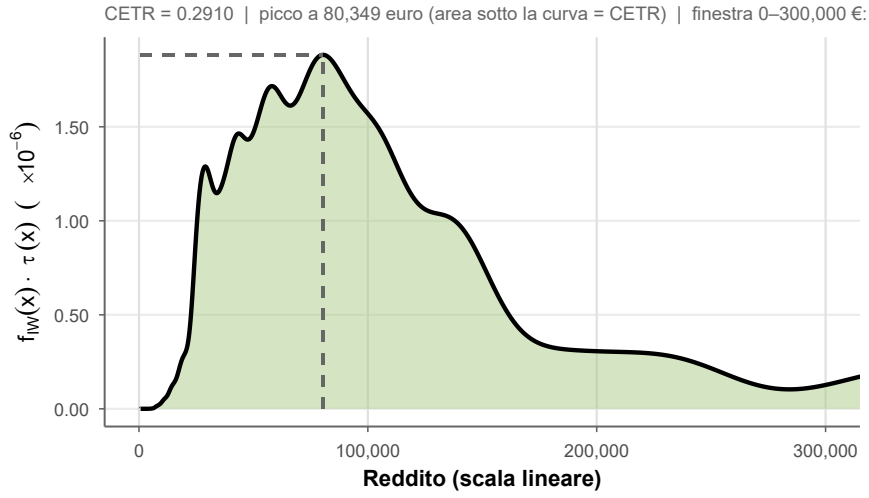


Figura 2.5: Integrando del CETR per il reddito da lavoro autonomo ordinario, zoom 0–300.000 euro (scala lineare)

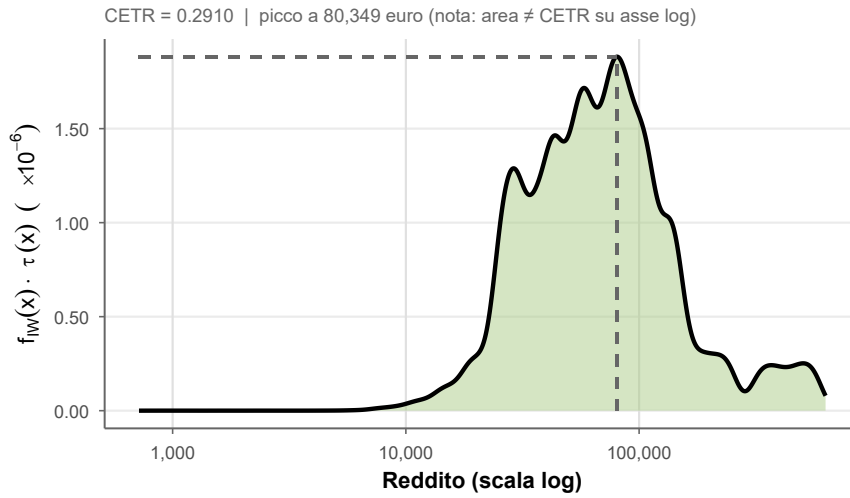


Figura 2.6: Integrando del CETR per il reddito da lavoro autonomo ordinario (scala logaritmica)

$\text{CETR} = 0.2910$ (29.1%), $\text{SD}_{IW}(\tau) = 6.5$ pp, $\text{CV}_{IW} = 0.22$. La correlazione $\rho_{IW} = 0.78$ mostra un andamento simile nella forma a quello commentato nel caso del reddito da lavoro dipendente, su un range di reddito più alto (media di $f(x) \approx 73.000$ € contro ≈ 23.000 € dei dipendenti). Il confronto con l'aggregato grezzo ($\text{CETR}_{agg} = 0.292$), costruito analogamente a prima pesando le

aliquote di classe per il reddito da lavoro autonomo, mostra uno scarto di -0.5% .

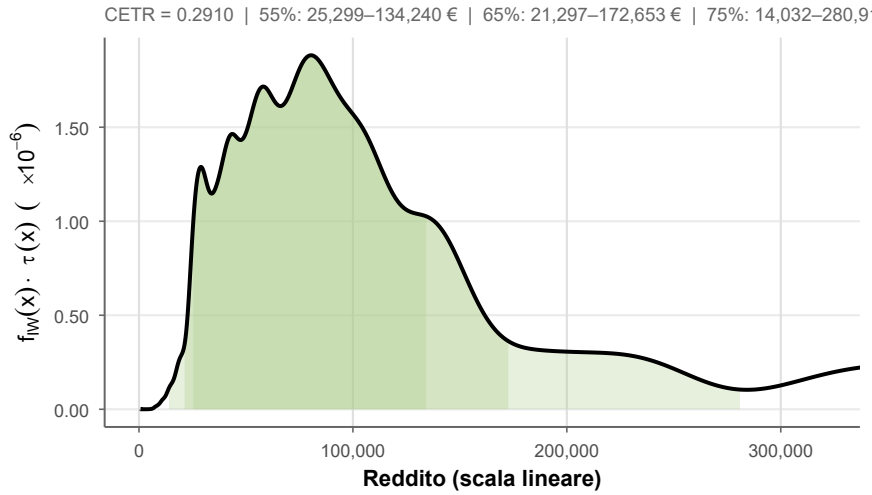


Figura 2.7: Integrando del CETR per il reddito da lavoro autonomo ordinario, con evidenziati gli intervalli di ampiezza minima che contengono il 55%, 65% e 75% del CETR.

Il grafico mostra le fasce di reddito più rilevanti per la costituzione del gettito. Scelte arbitrariamente tre quote, identificate dai seguenti intervalli (in euro):

- 55% $\rightarrow$ [25.299 , 134.240]
- 65% $\rightarrow$ [21.297 , 172.653]
- 75% $\rightarrow$ [14.032 , 280.912]

4.5 Reddito da microimpresa

Fonte: AIDA; campione di bilanci individuali, anno d'imposta 2023. Estrazione di imprese con meno di 10 addetti e ricavi delle vendite < 500.000 euro (proxy fatturato). Tabella classificata per utile netto, risultato prima delle imposte (EBT) e imposte correnti. È applicato il filtro $EBT > 0$ per escludere le imprese in perdita, non passibili di imposte dirette.

L'EBT massimo osservato nel campione grezzo (703 milioni di euro) è incompatibile con qualunque definizione di microimpresa: il filtro di estrazione basato sui soli ricavi delle vendite risulta quindi insufficiente, come evidenziato da questo dato. Si esegue perciò un taglio al 97,5° percentile ($EBT \approx 461.000$ €), da considerarsi un limite dichiarato dei criteri di estrazione del campione.

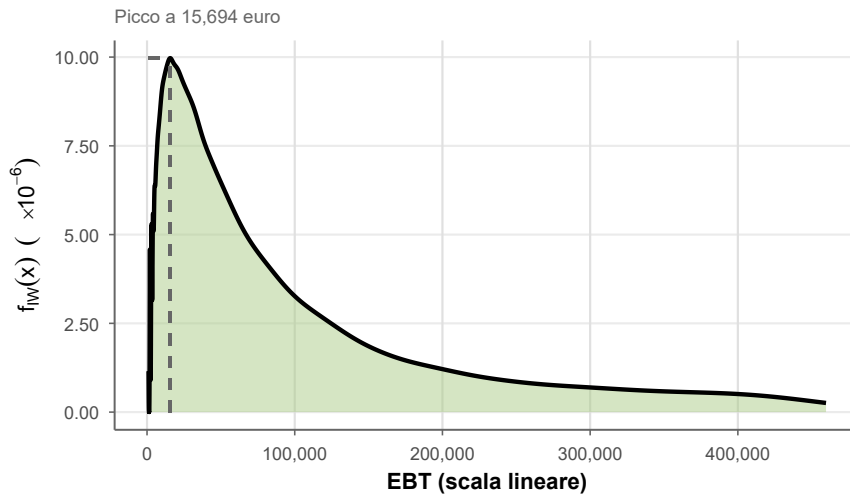


Figura 3.1: $f_{IW}(x)$ per il reddito da microimpresa (scala lineare)

Media 108.019 €, mediana 68.611 €, primo quartile 30.786 € e terzo quartile 148.528 €; skewness 1.42, curtosi in eccesso 1.26.

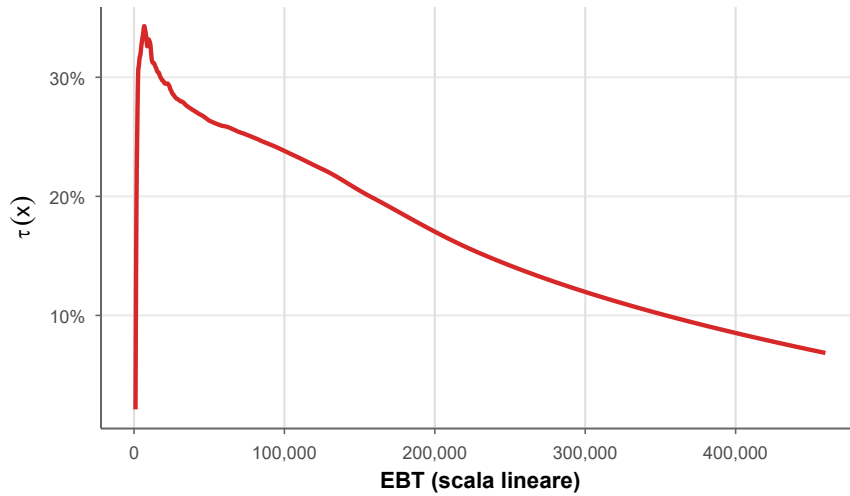


Figura 3.2: $\tau(x)$ per il reddito da microimpresa (scala lineare)

Nella parte sinistra del grafico, per EBT molto basso, l'arrotondamento alle migliaia di euro nei dati di bilancio fa sì che $\tau(x)$ possa assumere solo pochi valori razionali semplici ($1/3$, $1/4$, 0 , $\dots$), dunque il comportamento della curva in questa fascia va interpretato con cautela. Il calo nella parte destra è invece un fenomeno diverso: oltre metà delle microimprese con EBT fra 300.000 e 460.000 euro ha un'aliquota effettiva inferiore al 5%. La causa specifica non è isolabile con i soli dati di bilancio disponibili.

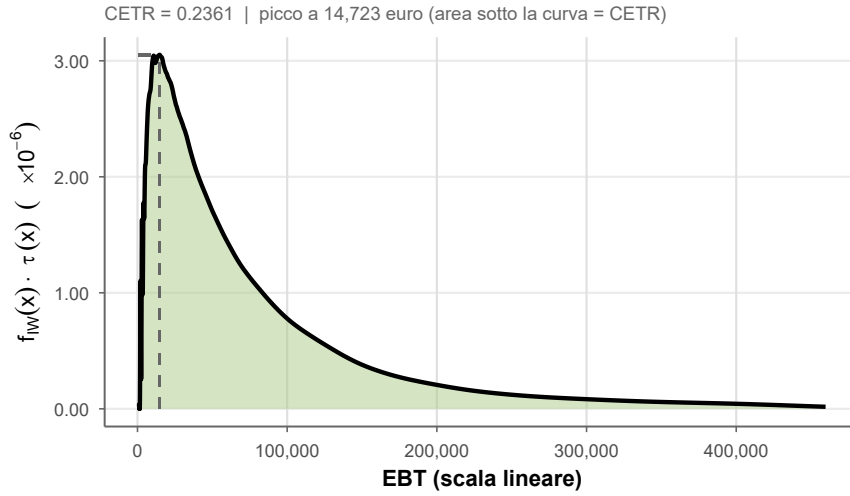


Figura 3.3: Integrando del CETR per il reddito da microimpresa (scala lineare)

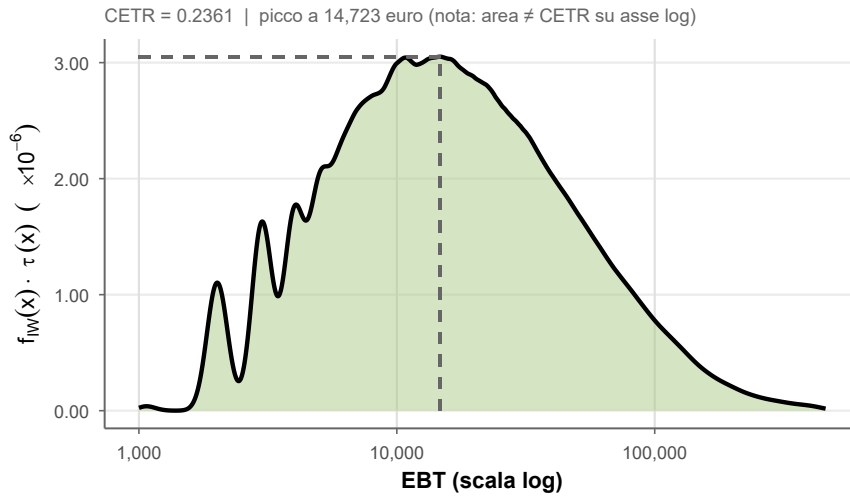


Figura 3.4: Integrando del CETR per il reddito da microimpresa (scala logaritmica)

$CETR = 0.2361$ (23.6%), $SD_{IW}(\tau) = 6.4$ pp, $CV_{IW} = 0.27$. La correlazione aliquota pesata-reddito è negativa ($\rho_{IW} = -0.98$). Il confronto con l'aggregato grezzo ($CETR_{agg} = 0.237$) mostra uno scarto di -0.3% .

Il comportamento a EBT basso è reso particolarmente evidente

dalla scala logaritmica, ma dipende unicamente dall'arrotondamento alle migliaia dei dati AIDA.

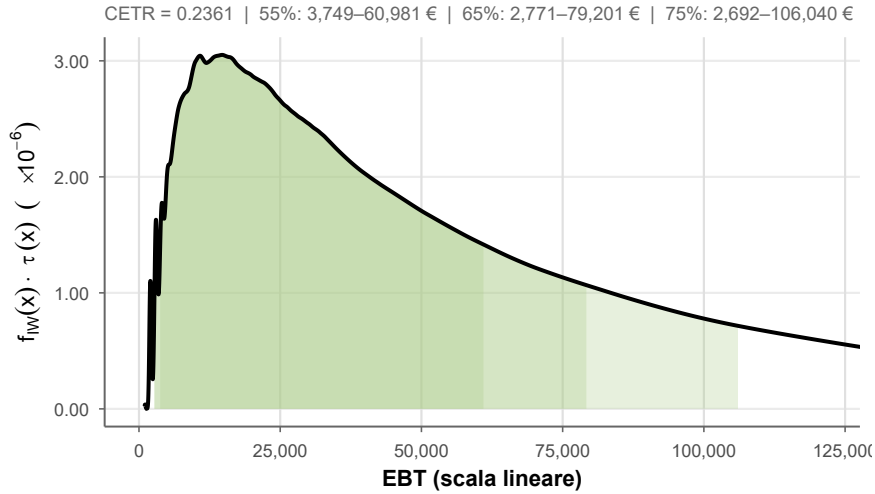


Figura 3.5: Integrando del CETR per il reddito da microimpresa, con evidenziati gli intervalli di ampiezza minima che contengono il 55%, 65% e 75% del CETR.

Il grafico mostra le fasce di EBT più rilevanti per la costituzione del gettito. Scelte arbitrariamente tre quote, identificate dai seguenti intervalli (in euro):

- 55% → [3.749 , 60.981]
- 65% → [2.771 , 79.201]
- 75% → [2.692 , 106.040]

4.6 Reddito societario

Fonte: AIDA; campione di bilanci individuali, anno d'imposta 2023. Estrazione di imprese con più di 10 addetti e ricavi delle vendite > 500.000 euro (proxy fatturato). Tabella classificata per utile netto, risultato prima delle imposte (EBT) e imposte correnti. È applicato il filtro $EBT > 0$ per escludere le imprese in perdita, non soggette ad imposte dirette.

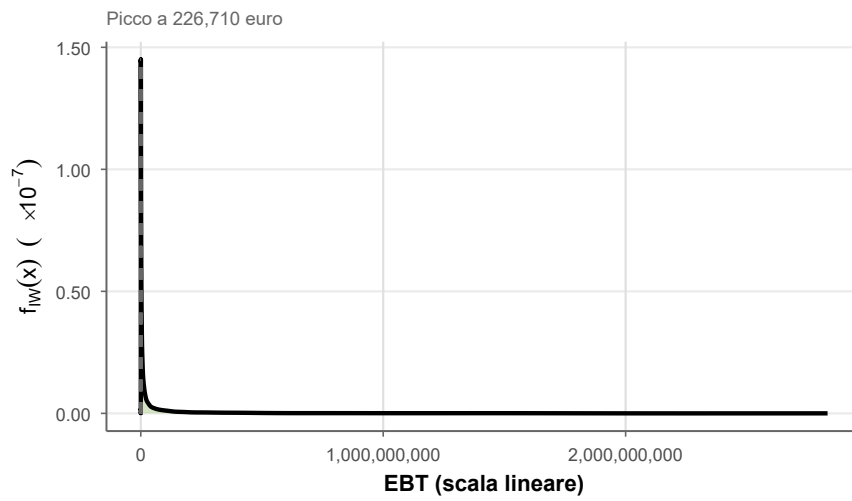


Figura 4.1: $f_{IW}(x)$ per il reddito da società di capitali (scala lineare)

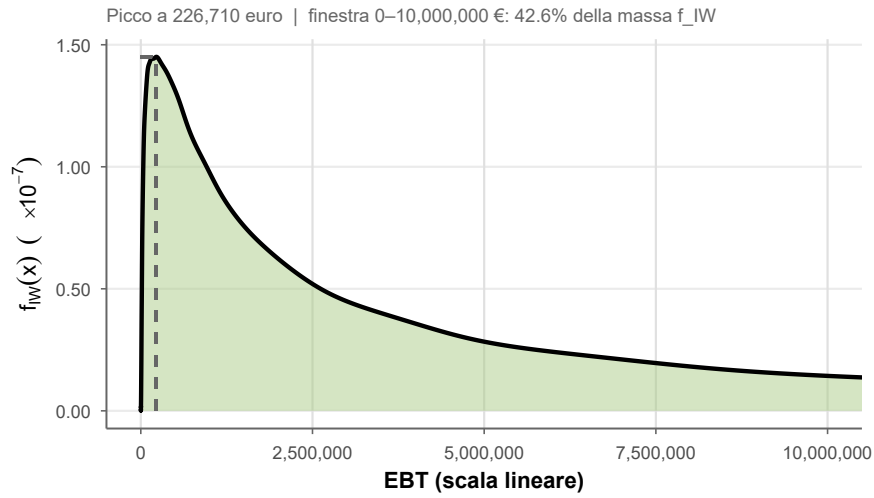


Figura 4.2: $f_{IW}(x)$ per il reddito da società di capitali, zoom 0–10 milioni di euro (scala lineare)

Media 195.987.453 €, mediana 16.719.548 €, primo quartile 2.929.589 € e terzo quartile 130.794.706 €; skewness 3.03, curtosi in eccesso 9.66.

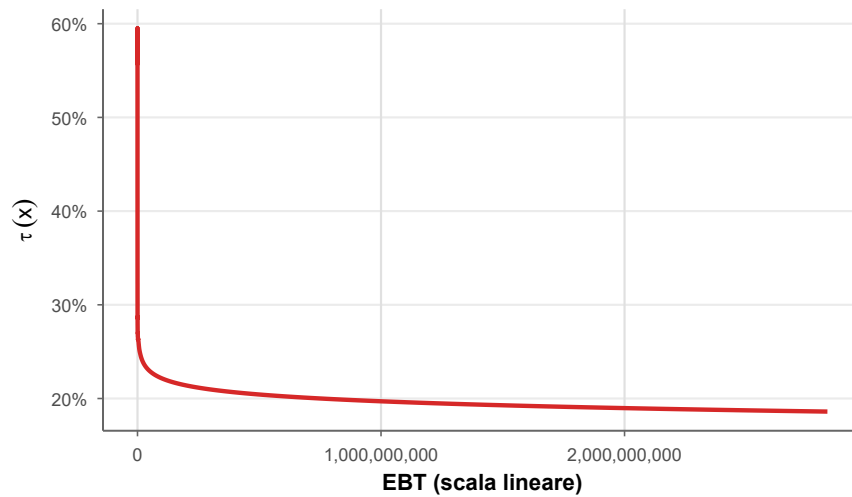


Figura 4.3: $\tau(x)$ per il reddito da società di capitali (scala lineare)

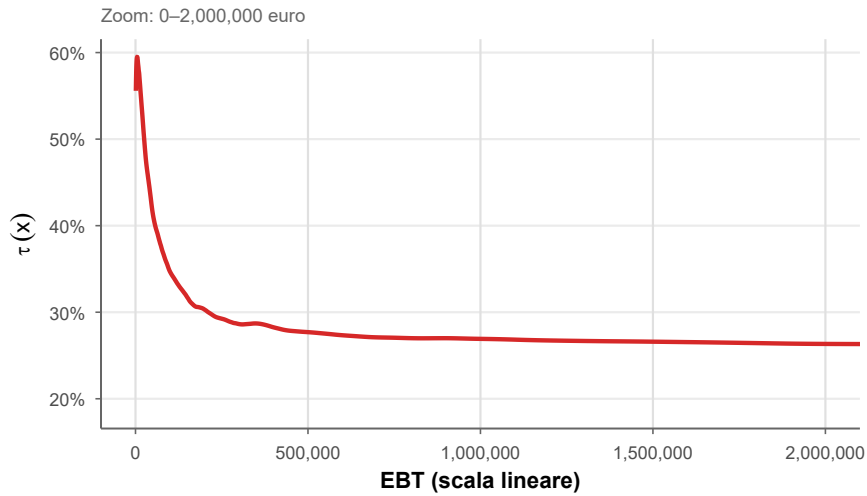


Figura 4.4: $\tau(x)$ per il reddito da società di capitali, zoom 0–2 milioni di euro (scala lineare)

La parte sinistra della curva, dove $\tau(x)$ supera il 50%, probabilmente riflette un problema di arrotondamento alle migliaia dei dati su ebt e imposte, molto marcato per via di rapporti fra numeri interi piccolissimi. Il calo nella parte restante della curva è invece liscio e continuo: da circa il 28% oltre i 300.000 euro di EBT fino a circa il 24,7% oltre i 5 milioni, un valore molto vicino all'aliquota IRES nominale (24%) da sola, coerente con società di grandi dimensioni la cui aliquota effettiva converge verso il solo tributo proporzionale con l'IRAP che pesa sempre meno.

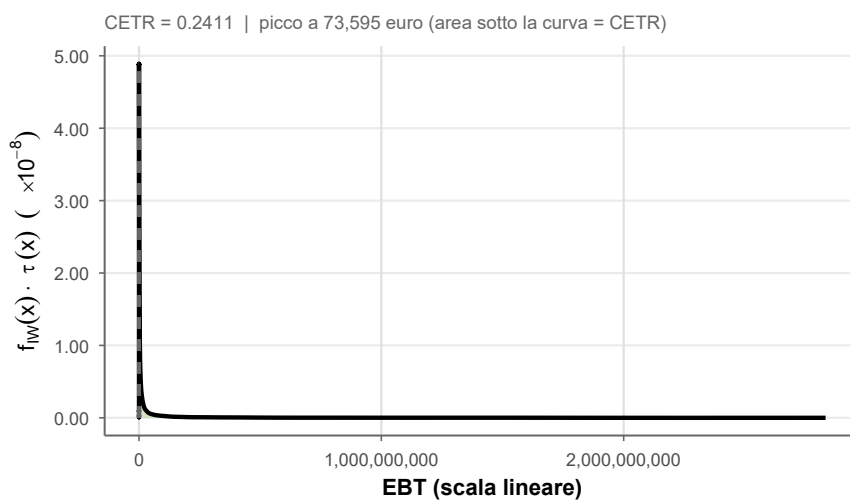


Figura 4.5: Integrando del CETR per il reddito da società di capitali (scala lineare)

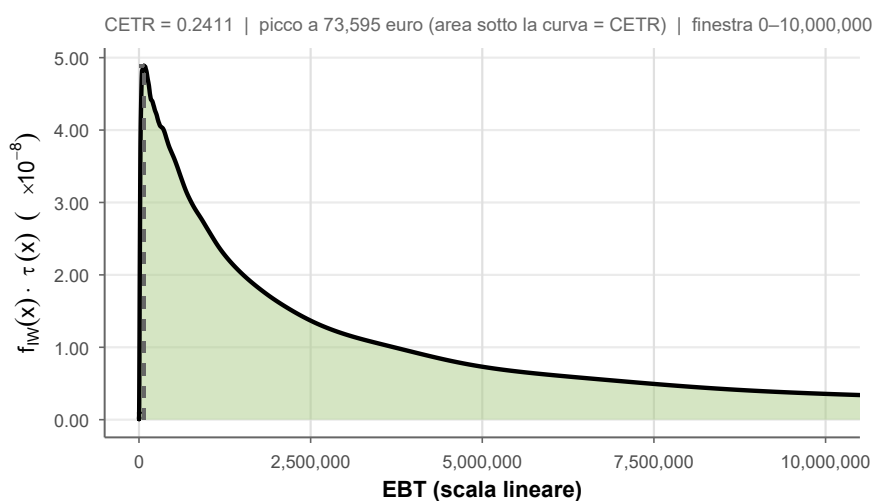


Figura 4.6: Integrando del CETR per il reddito da società di capitali, zoom 0–10 milioni di euro (scala lineare)

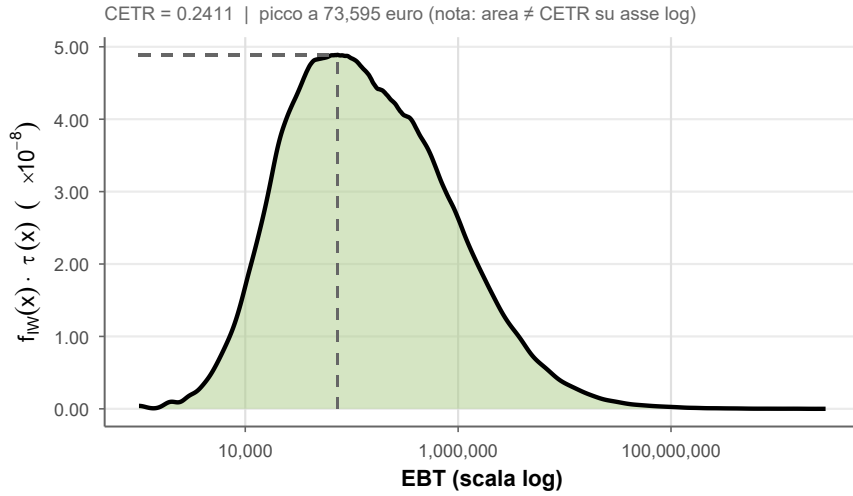


Figura 4.7: Integrando del CETR per il reddito da società di capitali (scala logaritmica)

$\text{CETR} = 0.2411$ (24.1%), $\text{SD}_{IW}(\tau) = 3.3$ pp, $\text{CV}_{IW} = 0.14$. L'aliquota pesata generalmente segue un pattern decrescente al crescere dell'EBT ($\rho_{IW} = -0.59$), andamento compatibile con molteplici spiegazioni che i dati non permettono di discriminare. Il confronto con l'aggregato grezzo sul campione ($\text{CETR}_{agg} = 0.244$) mostra uno scarto di -1.2% .

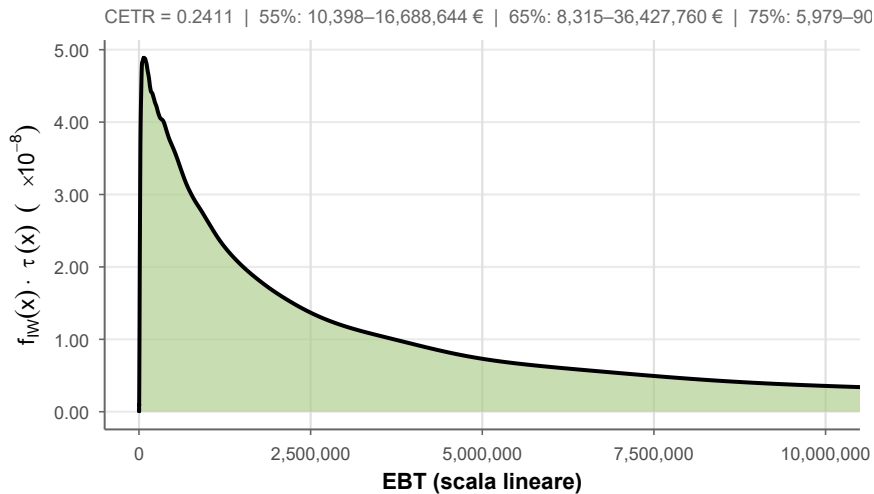


Figura 4.8: Integrando del CETR per il reddito da società di capitali, con evidenziati gli intervalli di ampiezza minima che contengono il 55%, 65% e 75% del CETR.

Il grafico mostra le fasce di EBT più rilevanti per la costituzione del gettito. Scelte arbitrariamente tre quote, identificate dai seguenti intervalli (in euro):

- 55% $\rightarrow$ [10.398 , 16.688.644]
- 65% $\rightarrow$ [8.315 , 36.427.760]
- 75% $\rightarrow$ [5.979 , 90.396.158]

4.7 Analisi comparativa

Le sezioni precedenti trattano ciascuna categoria isolatamente; qui le casistiche vengono affiancate sugli stessi assi per rendere confrontabili livello e forma.

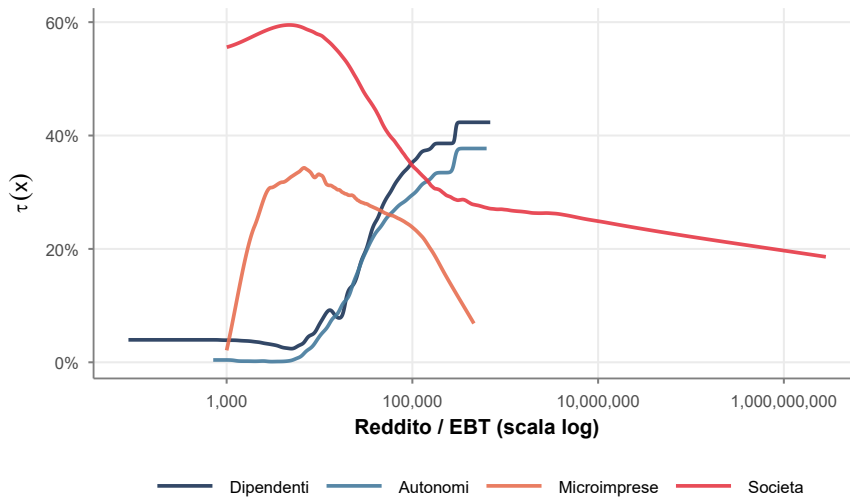


Figura 5.1: Funzione di prelievo $\tau(x)$ a confronto fra le quattro categorie (scala logaritmica).

Le quattro curve confermano visivamente la separazione già commentata categoria per categoria: dipendenti e autonomi (le due serie IRPEF) crescono verso il proprio massimo, mentre microimprese e società, dopo il picco iniziale, decrescono. La

sovrapposizione rende immediato anche il confronto tra imposta progressiva e proporzionale, ovviamente tenendo conto solo delle porzioni centrali e ignorando il comportamento alle estremità in molti casi anomalo e già discusso.

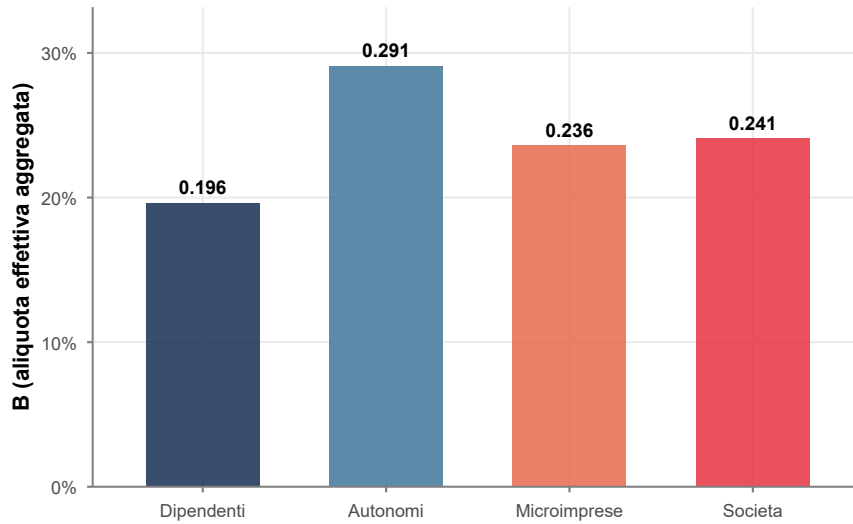


Figura 5.2: CETR per categoria: aliquota effettiva aggregata a confronto.

Categoria	$SD_{IW}(\tau)$ (pp)	CV_{IW}	ρ_{IW}
Reddito da lavoro dipendente	9.27	0.47	+0.650
Reddito da lavoro autonomo	6.51	0.22	+0.784
Reddito da microimpresa	6.42	0.27	-0.979
Reddito societario	3.27	0.14	-0.589

Tabella 1: Dispersione e pendenza dell'aliquota effettiva sotto una distribuzione pesata dal reddito; separazione per categoria.

Un'ultima chiave di lettura riprende gli intervalli di ampiezza minima già introdotti per ciascuna categoria, qui affiancati fra loro e confrontati con l'analogo intervallo calcolato sulla sola massa di reddito $f_{IW}(x)$, indipendentemente da quanto essa venga tassata.

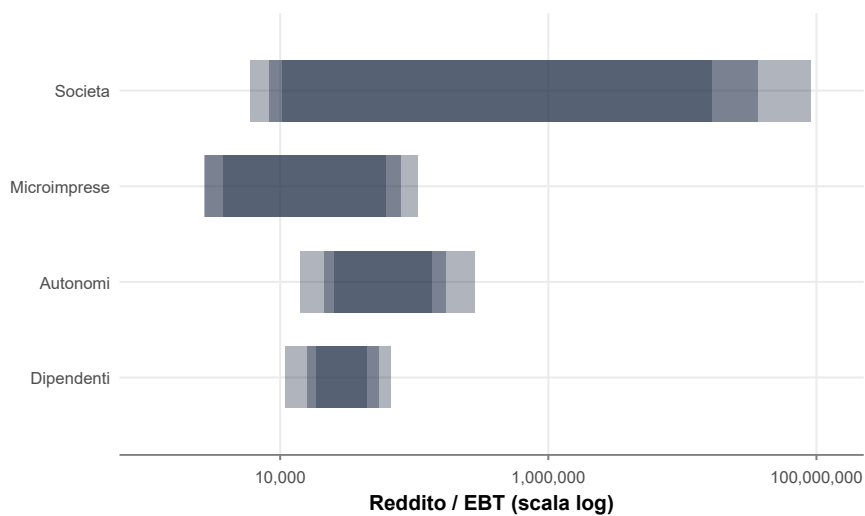


Figura 5.3: Dove si concentra il gettito: intervalli di ampiezza minima al 55/65/75% a confronto fra categorie

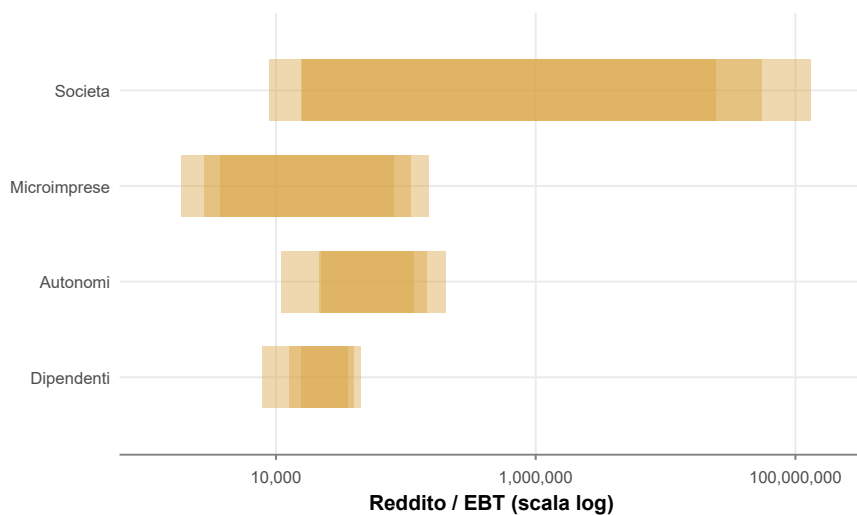


Figura 5.4: Dove si concentra il reddito (f_{IW}): stessa costruzione della figura precedente, applicata alla distribuzione f_{IW} .

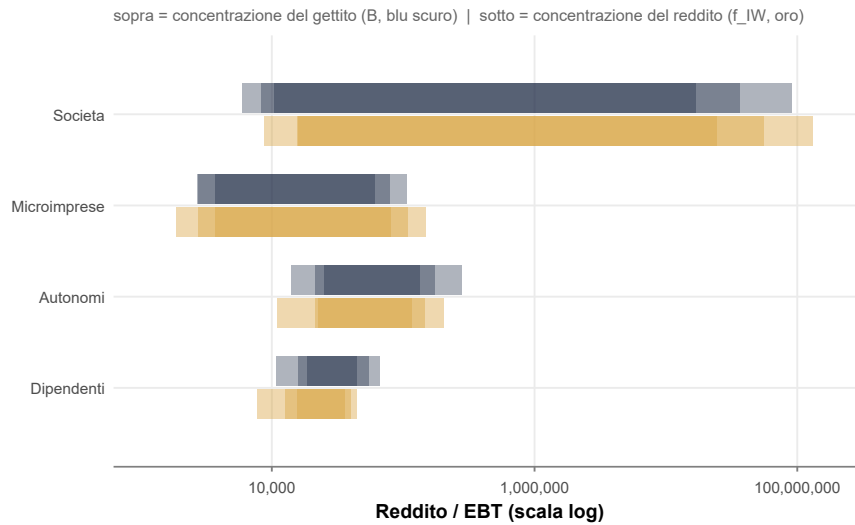


Figura 5.5: Gettito e reddito a confronto diretto per ciascuna categoria

Lettura del grafico: per costruzione le due corsie coincidono quando τ è costante lungo il supporto. Si può tenere a mente questo fatto come termine di paragone.

Il verso dello scostamento indica il segno della relazione fra aliquota pesata e reddito. Una corsia del gettito spostata a destra segnala un'aliquota media crescente nella fascia dove si concentra la massa di reddito, una corsia spostata a sinistra il contrario.

La stessa lettura trova una versione numerica negli indici di concentrazione. L'indice di Gini G_Y misura quanto è concentrato il reddito fra i contribuenti ed è calcolato su $f(x)$; l'indice di concentrazione C_T misura quanto è concentrato il gettito (si veda appendice). La loro differenza $\Pi = C_T - G_Y$ è l'indice di Kakwani introdotto nella sezione 2.5

Categoria	G_Y	C_T	Π
Reddito da lavoro dipendente	0.394	0.611	+0.217
Reddito da lavoro autonomo	0.462	0.570	+0.109
Reddito da microimpresa	0.589	0.497	-0.091
Reddito societario	0.858	0.828	-0.030

Tabella 2: Concentrazione del reddito e del gettito per categoria.

Nonostante Π coincida con l'indice di Kakwani, l'interpretazione dei risultati richiede alcune cautele. L'indice è di norma calcolato sulla distribuzione del reddito dell'intera popolazione, mentre qui G_Y è il Gini del reddito della singola categoria, dunque non è confrontabile con valori noti.

Come già osservato, l'imposta netta pubblicata dal MEF non è scomponibile per fonte di reddito: con dati aggregati per classe non è possibile stabilire in che direzione questo influisca su Π , che va perciò letto come stima approssimata. Inoltre, le addizionali sono incluse per i dipendenti e assenti per gli autonomi, e i contributi sociali sono esclusi in tutte le categorie. Non si tratta di imprecisioni della stima ma di una diversa definizione del prelievo considerato.

Capitolo 5

Conclusioni

5.1 Sintesi dei risultati

Il lavoro ha costruito, per quattro modalità di estrazione del reddito un indicatore continuo di aliquota fiscale effettiva aggregata, $CETR = \int f_{IW}(x) \tau(x) dx$, stimando $f(x)$ e $\tau(x)$ come funzioni empiriche anziché fissarle su un contribuente rappresentativo o ricavarle dalla normativa. È questo il tratto che distingue l'indicatore dalla letteratura sull'aliquota implicita (un unico rapporto fra due totali, senza funzioni sottostanti): qui il CETR nella sua sintesi finale coincide concettualmente con un'aliquota implicita, ma è ottenuto integrando l'intera distribuzione empirica osservata.

I quattro valori di CETR non si allineano su un unico livello. Il reddito da lavoro autonomo in regime ordinario mostra il prelievo aggregato più alto, $CETR = 0.2910$; il reddito da lavoro dipendente il più basso, $CETR = 0.1962$, ma è la categoria in cui l'aliquota pesata varia di più lungo la distribuzione (9.3 pp). Quella in cui varia meno (3.3 pp), invece, è il reddito societario, dunque nemmeno la dispersione $SD_{IW}(\tau)$ è omogenea.

Il segno di ρ_{IW} separa le categorie in due gruppi. Per le categorie soggette ad imposta IRPEF la correlazione tra aliquota pesata e reddito è positiva. Per il reddito da microimprese e società la correlazione è invece negativa, pattern che nelle rispettive se-

zioni è stato segnalato come risultato di difficile interpretazione.

Gli indici di concentrazione restituiscono una lettura da un'angolazione diversa. Il gettito è più concentrato per il reddito societario ($C_T = 0.828$), meno per il reddito da microimpresa ($C_T = 0.497$). L'indice di Kakwani è massimo per il reddito da lavoro dipendente ($\Pi = 0.217$) e minimo per il reddito da microimpresa ($\Pi = -0.091$), con segno positivo per lavoro dipendente e autonomo e negativo per microimprese e società. L'ordinamento per segno coincide con quello di ρ_{IW} ; i valori vanno letti come descrittivi del campione e delle scelte di stima.

5.2 Limiti del lavoro

I limiti del lavoro nascono dalla disponibilità dei dati, da imprecisioni nella selezione dei campioni e dalle procedure di stima. Si citano i più evidenti.

Il primo limite riguarda la definizione dell'indicatore. Il progetto prevedeva inizialmente di includere anche i contributi previdenziali in $T(x)$, per costruire un indicatore più vicino alla pressione fiscale. L'idea partiva dal ragionare sullo scarto tra reddito netto/utile e reddito lordo/EBT. L'impossibilità di reperire dati sui contributi e formule comparabili fra le quattro categorie ha costretto a rivedere l'ambito stesso del CETR già in fase di impostazione: l'indicatore misura quindi la sola componente delle imposte dirette, non il prelievo fiscale e contributivo complessivo. Non è un errore di stima, ma un ridimensionamento dei compiti che l'indicatore può svolgere.

Il secondo limite nasce dalla natura aggregata delle fonti MEF (dipendenti, autonomi). Non disponendo di microdati individuali, è stato necessario simulare un campione di redditi coerente con le frequenze di classe osservate. Questo passaggio

introduce per costruzione delle imperfezioni evidenti nelle funzioni stimate: i gradini nella curva $\tau(x)$ intorno alle classi più ampie, l'appiattimento della coda alta dovuto al confine arbitrario assegnato all'ultima classe aperta e la forma multimodale assunta dalle distribuzioni sono solo alcuni esempi. Sono tutti sintomi della stessa causa: $f(x)$ e $\tau(x)$ per queste due categorie non sono osservate, ma ricostruite a partire da un'informazione sintetizzata che risulta inadeguata per l'indicatore, nella sua formulazione teorica.

Il terzo limite riguarda invece le categorie AIDA (microimprese, società). In questo caso la principale responsabilità ricade nell'imprecisione durante la selezione delle voci di bilancio più adatte a fungere da filtri campionari o da variabili su cui costruire l'analisi. Come precisato nelle relative sezioni, i risultati vanno trattati con cautela, specialmente per quanto riguarda la forma delle funzioni di prelievo con particolare attenzione alle estremità.

$f(x)$ e $\tau(x)$, per tutte e quattro le categorie, non sono pienamente rappresentative delle rispettive popolazioni di riferimento. Questo si ripercuote inevitabilmente sui risultati e su qualsiasi lettura più fine dei numeri. Un'interpretazione approfondita di questo tipo di dettagli rischia quindi di essere fuorviante o addirittura errata, soprattutto considerando che manca una vera e propria misura dell'incertezza per tutte le quantità presentate. Il lavoro va letto, come dichiarato fin dall'introduzione, come un esercizio di costruzione e test dell'indicatore su dati reali ma imperfetti, più che come un'analisi economica in grado di sostenere conclusioni interpretative.

5.3 Sviluppi futuri

Due direzioni sembrano le più naturali per estendere il lavoro.

La prima è orientata verso una ripetizione della parte di analisi dei dati, questa volta cercando dati più adatti (ad esempio tentando di includere anche i contributi) e affinando la metodologia di stima. Perfezionando questi punti, si potrebbe ampliare ulteriormente l'approfondimento ipotizzando nuove forme della funzione di prelievo e applicandole alla distribuzione dei redditi oppure applicando forme note o rivisitate.

La seconda, più interessante, sposta il focus dall'aliquota effettiva e relativi indicatori. Individuare la quota di reddito più rilevante per la costituzione del gettito fiscale può tradursi nello studio della concentrazione del gettito, non solo come misura sintetica (un indice di concentrazione), ma anche nella sua forma. Una buona base di partenza viene già introdotta nel capitolo 4.7 con il calcolo dell'indice di Kakwani.

Appendice

A.1 Kernel Density Estimation

Sia $X_1, \dots, X_n$ un campione di redditi. Lo stimatore kernel della densità è

$$\hat{f}_h(x) = \frac{1}{n} \sum_{i=1}^n K_h(x - X_i), \quad K_h(u) = \frac{1}{h} K\left(\frac{u}{h}\right) \quad (\text{A.1})$$

dove K è una gaussiana e $h > 0$ è la larghezza di banda. La stima consiste nel mediare piccole distribuzioni centrate sulle osservazioni ed è per costruzione una densità, non negativa e a integrale unitario.

Nel caso dei dati simulati da tabelle per classe ogni osservazione rappresenta un numero diverso di contribuenti reali, e si usa quindi la versione pesata

$$\hat{f}_h(x) = \sum_{i=1}^n w_i K_h(x - X_i), \quad w_i = \frac{\omega_i}{\sum_j \omega_j}, \quad (\text{A.2})$$

con ω_i pari al numero di contribuenti rappresentati dall'osservazione i .

L'ampiezza costante della banda costituisce un problema. Siccome tutte le variabili in esame sono positive, una h fissata in euro è necessariamente troppo larga alla base della distribuzione e troppo stretta nella coda, dove i livelli di reddito sono di

ordini di grandezza superiori. Per questo motivo la stima è condotta sulla scala logaritmica. Posto $U = \log X$ e detta $\hat{g}$ la stima kernel della densità di U , il cambiamento di variabile restituisce

$$\hat{f}_X(x) = \frac{\hat{g}(\log x)}{x}, \quad x > 0, \quad (\text{A.3})$$

La banda è fissata dalla regola di Silverman applicata alla scala logaritmica, cioè dal valore predefinito in R (`bw.nrd0`):

$$h = 0,9 \cdot \min\left(\hat{\sigma}, \frac{\text{IQR}}{1,349}\right) \cdot n^{-1/5}, \quad (\text{A.4})$$

dove $\hat{\sigma}$ e IQR sono calcolati su $\log x$.

	Dipendenti	Autonomi	Microimprese	Società
n	47.596	49.068	178.736	132.706
h (logaritmi)	0,073	0,085	0,093	0,147
N (griglia)	2.048	2.048	4.096	4.096

Tabella 1: Larghezze di banda ottenute dalla (A.4). Essendo espresse in logaritmi naturali, si leggono come variazione relativa del reddito.

A.2 Simulazione di microdati da aggregati per classe

Le tabelle MEF riportano, per ciascuna classe $j = 1, \dots, J$, il numero di contribuenti n_j e gli ammontari complessivi di reddito e di imposta, da cui si ricavano la media di classe μ_j e l'aliquota media di classe $\tau_j = T_j/Y_j$. La procedura estrae m_j osservazioni con media μ_j , assegnando a ciascuna il peso $\omega_j = n_j/m_j$ e l'aliquota τ_j della classe di appartenenza.

Si tratta di una scelta minimale, ma si assume che all'interno della classe il logaritmo del reddito sia uniformemente distribuito:

$$\log X \sim U(\log a_j, \log b_j),$$

$$\text{cioè } f_j(x) = \frac{1}{x \log(b_j/a_j)}, \quad a_j \leq x < b_j.$$

Si applica quindi un fattore di calibrazione in modo che la media della classe coincida con quella del dato aggregato:

$$c_j = \mu_j \cdot \frac{\log(b_j/a_j)}{b_j - a_j}, \quad X^* = c_j \exp(U), \quad U \sim U(\log a_j, \log b_j),$$

(A.5)

così che $\mathbb{E}[X^*] = \mu_j$.

(l'estrazione intraclassa è generata con `set.seed(123)`).

Le tabelle MEF classificano i contribuenti per reddito complessivo, mentre la variabile di interesse è sistematicamente più bassa. I confini sono quindi ricostruiti a partire dalle medie osservate, che sono espresse nella variabile giusta:

$$a_j = \frac{\mu_{j-1} + \mu_j}{2}, \quad b_j = a_{j+1} = \frac{\mu_j + \mu_{j+1}}{2}$$

Restano da trattare i due estremi. La prima classe non ha una media precedente da cui prendere il punto medio e viene chiusa per simmetria rispetto alla propria media, $a_1 = 2\mu_1 - b_1$. L'ultima classe è aperta a destra e viene chiusa a $b_J = 2a_J$.

Quest'ultima è una scelta arbitraria. L'indicatore pesa i redditi per il loro valore, quindi la coda alta conta molto. Chiuderla a un multiplo fissato ne comprime l'estensione e appiattisce il contributo dei redditi più elevati: la coda alta delle categorie MEF è assunta, non stimata, e va letta come limite dichiarato del lavoro.

A.3 Lo stimatore di Nadaraya-Watson

Stimare $\tau(x)$ è un problema di regressione non parametrica. Lo stimatore di Nadaraya-Watson lo risolve con una media locale pesata:

$$\hat{m}_h(x) = \frac{\sum_{i=1}^n w_i \tau_i K_h(x - X_i)}{\sum_{i=1}^n w_i K_h(x - X_i)}, \quad (\text{A.6})$$

Il denominatore è esattamente $\hat{f}_h(x)$. Il numeratore, posto $k = \sum_i w_i \tau_i$ e $\tilde{w}_i = w_i \tau_i / k$ si riscrive come

$$\sum_i w_i \tau_i K_h(x - X_i) = k \sum_i \tilde{w}_i K_h(x - X_i) = k \hat{f}_h^\tau(x),$$

dove $\hat{f}_h^\tau$ è la kde calcolata con pesi proporzionali alla “massa fiscale” $w_i \tau_i$ anziché alla sola massa di contribuenti. Quindi la forma implementata è:

$$\hat{\tau}(x) = k \frac{\hat{f}_h^\tau(x)}{\hat{f}_h(x)} \quad (\text{A.7})$$

È il rapporto fra la densità del gettito atteso e la densità del reddito, calcolate con la stessa larghezza di banda e sulla stessa griglia di punti.

La stima non può produrre valori che i dati non contengono. Ne segue anche un limite importante: dove le classi sono molto ampie, tipicamente nella coda alta, la curva restituisce un andamento piatto. Ciò non va interpretato come caratteristica del sistema fiscale, ma solo come effetto del dato originale discreto.

A.4 LOESS: locally estimated scatterplot smoothing

Nei campioni AIDA l'aliquota è osservata a livello di impresa e il grafico di τ_i contro EBT_i è una nuvola.

La regressione locale generalizza la media locale sostituendo alla costante un polinomio. Fissato un punto x_0 , si risolve il problema dei minimi quadrati pesati

$$\min_{\beta_0, \beta_1} \sum_{i=1}^n W_i(x_0) [\tau_i - \beta_0 - \beta_1(x_i - x_0)]^2, \quad (\text{A.8})$$

e si pone $\hat{\tau}(x_0) = \hat{\beta}_0$, cioè il valore in x_0 della retta adattata localmente; ripetendo per ogni punto della griglia si ottiene la curva.

La regressione è condotta su $\log(EBT)$. La curva è poi valutata nei punti corrispondenti alla griglia su cui è definita $\hat{f}_{IW}$, così che le due funzioni possano essere moltiplicate punto per punto.

R non dispone di un selettore automatico per lo span del loess: il valore predefinito di `loess` è 0.75, ma è una costante convenzionale e non una quantità stimata dai dati. Lo span è quindi fissato minimizzando il criterio AICc corretto:

$$\begin{aligned} \text{AICc}(\alpha) &= \log \hat{\sigma}^2 + 1 + \frac{2(\text{tr}(L_\alpha) + 1)}{n - \text{tr}(L_\alpha) - 2}, \\ \hat{\sigma}^2 &= \frac{1}{n} \sum_i (\tau_i - \hat{\tau}(x_i))^2 \end{aligned} \quad (\text{A.9})$$

dove L_α è la matrice di lisciamento associata allo span α . Il primo termine premia l'aderenza ai dati, il secondo penalizza i gradi di libertà spesi. La minimizzazione è condotta da 0,05 a 0,75; il grado è fissato a 1.

	Microimprese	Società
Span selezionato per AICc	[0.08]	[0.08]

Tabella 2: Span del *loess* selezionato.

A.5 Intervalli di ampiezza minima

L'equazione (3.6) dà la quota di gettito di una fascia $[x_1, x_2]$ assegnata; qui la fascia è invece l'incognita. Per una quota $q \in \{0,55; 0,65; 0,75\}$ si cerca l'intervallo di reddito più stretto che contenga almeno quella quota del gettito:

$$\min_{x_1 < x_2} (x_2 - x_1) \quad \text{sotto il vincolo} \quad \int_{x_1}^{x_2} g(x) dx \geq q \cdot \text{CETR}.$$

La stessa costruzione, applicata a f_{IW} anziché a g , individua la fascia più compatta che contiene la quota q della massa di reddito.

Il calcolo è esatto sulla griglia, non approssimato. Detta $G(x_k)$ la cumulata trapezoidale di g sui nodi $x_1 < \dots < x_n$ della griglia di integrazione, per ogni indice i si individua il minimo indice $j > i$ tale che $G(x_j) - G(x_i) \geq q \cdot \text{CETR}$ e si conserva la coppia che minimizza $x_j - x_i$. Poiché $g \geq 0$ implica G non decrescente, l'indice j è a sua volta non decrescente in i : una sola scansione esamina tutte le coppie ammissibili.

A.6 Indici di concentrazione

La concentrazione del reddito è misurata con l'indice di Gini. Posto

$$L(F(x)) = \frac{1}{\mathbb{E}[X]} \int_0^x t f(t) dt, \quad G_Y = 1 - 2 \int_0^1 L(p) dp,$$

entrambe le cumulate e l'integrale finale sono calcolati con la regola dei trapezi sulla stessa griglia usata per il CETR.

La concentrazione del gettito è misurata con l'indice di concentrazione dell'imposta, che si distingue dal Gini per il fatto che il gettito è cumulato rispetto all'ordinamento per reddito e non rispetto al proprio:

$$C_T(F(x)) = \frac{1}{\text{CETR}} \int_0^x g(t) dt, \quad C_T = 1 - 2 \int_0^1 C_T(p) dp.$$

La differenza fra i due indici è l'indice di Kakwani $\Pi = C_T - G_Y$.