

UNIVERSITÀ DEGLI STUDI DI PADOVA
DIPARTIMENTO DI SCIENZE STATISTICHE
CORSO DI LAUREA TRIENNALE IN
STATISTICA PER L'ECONOMIA E L'IMPRESA



RELAZIONE FINALE

**ANALISI DELL'INFLUENZA DELLA GENERAZIONE
SOLARE SUI CONSUMI ELETTRICI: LA DUCK
CURVE**

Relatore Prof. Luigi Grossi
Dipartimento di Scienze Statistiche

Laureando Omar Sbrissa
Matricola 2001865

Anno Accademico 2022/2023

Indice

Introduzione	1
1 Analisi preliminari sui consumi di energia elettrica in Italia	5
1.1 Il sistema elettrico nazionale e la generazione solare	5
1.1.1 Il sistema elettrico nazionale	5
1.1.2 <i>Terna</i>	6
1.1.3 Dati per le analisi	7
1.1.4 La generazione solare	8
Il solare termico	8
Il solare fotovoltaico	8
1.2 Andamento del fabbisogno totale di energia in Italia	10
1.2.1 Valutazioni generali	10
1.2.2 Analisi esplorativa della serie dei <i>total load</i>	12
1.2.3 Analisi con modello <i>SARIMA</i> della serie dei <i>total load</i>	12
1.3 Modelli per il consumo infragiornaliero: la <i>Duck Curve</i>	16
1.3.1 Analisi esplorative sulla produzione di fotovoltaico in Italia	17
1.3.2 Modello lineare con variabili <i>dummy</i>	21
1.3.3 Modello con funzioni trigonometriche	23
1.3.4 Modello <i>SARIMA</i>	24
2 Valutazione della stagionalità mediante indice di Gini: aspetti metodologici	27
2.1 Scomposizione Indice di Gini	27
2.2 Test sul cambiamento dell'indice di concentrazione	29
3 Scomposizione della stagionalità: applicazione	33
3.1 Valutazioni preliminari sulla stagionalità dei consumi elettrici	33
3.2 Scomposizione dell'indice di Gini e test applicati ai dati sui consumi elettrici	35
3.3 Applicazione ai dati suddivisi per stagione meteorologica	37
3.4 Confronto Italia-California	40
Conclusioni	41
Manovre da attuare per mitigare il fenomeno	43

Introduzione

Con l'avanzare della tecnologia e la necessità di ridurre l'utilizzo di combustibili fossili, a causa della loro scarsità e negativo impatto ambientale, gli stati stanno promuovendo la penetrazione di fonti per la produzione di energia rinnovabile. Questo fenomeno rappresenta sicuramente un punto di svolta importante nella vita delle persone, sia a livello etico, sia a livello pratico. La consapevolezza di vivere in un mondo più “*green*”, sostenibile e meno inquinato dovrebbe però essere accompagnato da un cambio negli stili di vita delle persone e nelle loro abitudini. Se da una parte vediamo il lato dei consumatori (dell'energia elettrica), dall'altro lato i fornitori con l'inserimento delle fonti rinnovabili nel mercato, devono affrontare una grande sfida: quella del bilanciamento energetico.

I gestori della rete e del servizio di dispacciamento devono infatti assicurare al cliente la continuità del servizio e la qualità della tensione, evitando interruzioni del servizio e l'“*overgeneration*” (o sovrapproduzione) che se non può essere immagazzinata crea un grosso problema di sovraccarico della rete che può provocare danni irreversibili, rotture ed altre situazioni di pericolo, oltre che l'interruzione del servizio.

Per assicurare la fornitura di energia elettrica è necessario sviluppare algoritmi previsivi che agiscono sul controllo della fase di produzione dell'energia per garantirne poi un'opportuna trasmissione e distribuzione alle utenze finali.

L'aspetto positivo delle energie rinnovabili è che sono ricavate da risorse naturali e inesauribili e quindi sono in grado di auto-rigenerarsi in continuazione.

Le principali fonti rinnovabili sono: solare, eolica, geotermica, idroelettrica, biomassa.

In Italia a guidare la transizione ecologica sono l'energia idroelettrica, solare, nello specifico il fotovoltaico, e l'energia eolica.

“L’Italia è l’unico tra i principali Paesi dell’Unione Europea ad aver raggiunto già nel 2017 (18,3%) l’obiettivo fissato per il 2020 dalla Direttiva Europea 2009/28/CE, che prevede una quota dei consumi finali lordi complessivi coperta da fonti rinnovabili pari almeno al 17%”. GSE (2019)

Tali fonti, in particolare il fotovoltaico e l’eolico, dipendono fortemente dalla variabile tempo oltre che da altre variabili esogene quali la temperatura e altri fenomeni atmosferici. Al crescere della penetrazione del fotovoltaico il principale compito delle centrali termiche tradizionali non è più quello di soddisfare il fabbisogno energetico del paese, ma quello di “coprire” con la produzione di energia da fonti tradizionali (gas e petrolio) la domanda nei periodi della giornata (o dell’anno) che non riescono ad essere soddisfatti dal fotovoltaico. Questo cambio di focus ha introdotto una serie di problematiche da risolvere.

La presente tesi si pone l’obiettivo di studiare l’andamento storico della differenza tra il carico totale dell’energia fornita ai consumatori e la produzione da fonte fotovoltaica con frequenza oraria nel periodo che va da inizio 2018 a fine 2022, in Italia. Tale differenza considerata nell’arco di una giornata prende il nome di “*Duck Curve*”, per la forma caratteristica che assume. Questo andamento non regolare rappresenta un notevole problema per chi ha il ruolo di garantire una fornitura stabile dell’energia elettrica.

La valutazione della stagionalità della serie viene effettuata mediante la scomposizione dell’indice di Gini proposta da Grossi & Mussini (2021).

L’elaborato è suddiviso in capitoli.

Nel primo capitolo si descrive brevemente il sistema elettrico nazionale e la generazione fotovoltaica, il dataset e si propongono alcune analisi esplorative. Vengono infine presentati alcuni modelli di lungo e breve periodo sui consumi di energia in Italia.

Nel secondo capitolo sono introdotti gli aspetti metodologici per la scomposizione dell’indice di Gini al fine di valutare la stagionalità della *Duck Curve* e i relativi test statistici.

Nel terzo capitolo si applicano le metodologie presentate nel capitolo 2 ai dati dei consumi orari in Italia e si propone un confronto qualitativo con la situazione del mercato elettrico in California nel 2013.

Nella sezione conclusiva sono menzionati alcuni accorgimenti per mitigare almeno parzialmente il problema.

Capitolo 1

Analisi preliminari sui consumi di energia elettrica in Italia

1.1 Il sistema elettrico nazionale e la generazione solare

1.1.1 Il sistema elettrico nazionale

Il sistema elettrico è diviso in quattro fasi: produzione, trasmissione, distribuzione e utilizzo da parte delle utenze finali.

Il mercato dell'energia in Italia è stato liberalizzato a partire dal 1999 con il decreto legislativo 16 marzo 1999, n. 79, detto decreto Bersani, seguendo le orme della Norvegia nel 1991 e degli altri paesi scandinavi qualche anno più tardi. Tale manovra era volta a esautorare dalla posizione monopolistica la società *leader* del mercato energetico fin dal 1962 : Enel. A tale società venne posto il vincolo di non superare la produzione del 50% dell'energia in Italia così da creare un'apertura nel mercato e le condizioni di concorrenza a beneficio dell'acquirente. Storicamente i grandi consumatori di energia (grandi industrie) sono stati i primi a poter scegliere di cambiare fornitore, ora è possibile anche per le utenze domestiche.

Un'altra manovra necessaria fu quella di creare società diverse per la gestione delle altre fasi interne al sistema elettrico.

La fase di trasmissione comprende l'impiego delle infrastrutture che portano l'energia prodotta dalle centrali elettriche ad alta e media tensione in prossimità ai luoghi dove

viene utilizzata. L'ente che si occupa della fase di trasmissione è Terna che opera in un regime di monopolio naturale e all'interno di un mercato regolato dall'Autorità di Regolazione per Energia, Reti e Ambiente (ARERA). Si definisce monopolio naturale una configurazione che impone un solo produttore per minimizzare i costi di produzione. La funzione di costo infatti è subadditiva ($f(x + y) \leq f(x) + f(y)$, Katz et al. (2015)), i costi fissi sono elevati mentre i costi marginali sono decrescenti. Basti pensare che la presenza di un ulteriore operatore nella fase di trasmissione implicherebbe la costruzione di un impianto di rete elettrica parallelo a quello esistente e avente il medesimo scopo.

Terna si occupa inoltre della fase di dispacciamento dell'energia, ovvero il controllo e la gestione nel tempo della domanda e dell'offerta di energia che devono essere sempre equivalenti per evitare sovraccarichi o interruzioni del servizio.

La fase di distribuzione, quindi la resa a disposizione di corrente elettrica a bassa frequenza, ha caratteristiche di monopolio locale, questo implica che internamente ad un territorio comunale sia presente un solo distributore di energia. La fase di vendita alle utenze è lasciata in regime di concorrenza.

1.1.2 *Terna*

Terna è l'azienda che in Italia si occupa della pianificazione, sviluppo, manutenzione della rete di trasmissione nazionale (*RTN*) e della gestione dei flussi di energia. Tra gli impegni (*mission*) di Terna ci sono quello di implementare un sistema in sinergia con quello del resto dell'Europa ed essere regista della transizione ecologica che permette uno sviluppo basato sulle energie rinnovabili e nel rispetto dell'ambiente.

L'azienda investe inoltre per il miglioramento, l'innovazione e la digitalizzazione del sistema elettrico ed energetico nazionale e opera nella fase di manutenzione delle linee elettriche ad alta tensione del territorio nazionale.

Si occupa anche di gestire il servizio di dispacciamento dell'energia, un servizio continuo che rende possibile l'equilibrio tra domanda e offerta. Per fare questo nasce la necessità di fare previsioni a breve e lungo termine sulla quantità di energia richiesta dal mercato, tenendo conto di variabili concomitanti quali: i fenomeni climatici (temperatura, vento, irraggiamento) che hanno un effetto diretto sulla produzione di energia rinnovabile, quindi sulla domanda, la disponibilità di generazione termica e la capacità di previsione, oltre che agli aspetti legati alla variabile tempo. Per misurare l'energia immessa (e quindi consumata) nel mercato, l'azienda si avvale un sistema di *metering* che agisce in una duplice modalità: il sistema diretto che sfrutta dei contatori posizionati all'uscita degli impianti di produzione e in maniera indiretta misurando attraverso operatori esterni le quantità per poi convalidarle ed elaborarle.

Terna fornisce pubblicamente i dati relativi alle misurazioni di carico energetico con una frequenza di quattro misurazioni orarie equidistanti nel tempo disaggregate per *bidding zone* (*Centre-North, Centre-South, North, Sardinia, Sicily, South, Calabria, Italy*), mentre con misurazioni orarie il carico disaggregato per fonte di produzione a livello statale. Fornisce inoltre le misure di *import* e *export* orari, di energia, con alcuni paesi europei. Sono resi disponibili *open data* anche le previsioni di breve periodo di carico totale e la generazione divisa per fonte (*Geothermal, Hydro, Photovoltaic, Self-consumption, Thermal, Wind*). (TERNA, 2023)

1.1.3 Dati per le analisi

Dai dati prodotti da Terna ho ricavato il *dataset* alla base della ricerca. I dati sono considerabili dati secondari esterni e provenienti da fonte ufficiale, in quanto l'Ufficio statistico di Terna fa parte del *Sistan* (Sistema Statistico Nazionale). L'azienda è anche responsabile per l'Italia delle comunicazioni statistiche ufficiali a organismi internazionali come *Eurostat, IEA, OCSE, ONU* per quanto concerne il sistema elettrico nazionale.

Il *dataset* che ho ricavato copre il periodo da inizio 2018 alla fine del 2022 e presenta le osservazioni orarie del fabbisogno totale di energia elettrica del territorio nazionale e la produzione di fotovoltaico compreso l'autoconsumo.

Le variabili in tabella 1.1 sono:

- *Date* : data e ora secondo lo standard americano
- *total_load* : fabbisogno totale di energia in Megawatt, ottenuto come media aritmetica delle osservazioni ogni quarto d'ora rilevate da Terna nell'ora di riferimento. La scelta di aggregare i dati è stata fatta per renderli confrontabili con i dati relativi alla produzione di fotovoltaico e per non appesantire troppo la trattazione e le analisi del fenomeno di riferimento
- *photovoltaic* : produzione sul territorio nazionale di energia da fonte fotovoltaica compreso l'autoconsumo e trasformata dai dati originali da Gigawatt a Megawatt (*1000)
- *tlf* : $total_load - photovoltaic$
- *perc* : $photovoltaic / total_load * 100$
- *year* : anno della rilevazione
- *mon* : mese della rilevazione

- *yday* : giorno dell'anno della rilevazione (da 1 a 365, 366 per 2020 anno bisestile)
- *wday* : giorno della settimana della rilevazione
- *hour* : ora del giorno della rilevazione.

	Date	total_load	photovoltaic	tif	perc	year	mon	yday	wday	hour
1	2018-01-01 00:00:00	24336.50	0	24336.50	0.00000000	2018	1	1	lun	0
2	2018-01-01 01:00:00	23186.00	0	23186.00	0.00000000	2018	1	1	lun	1
3	2018-01-01 02:00:00	21741.00	0	21741.00	0.00000000	2018	1	1	lun	2
4	2018-01-01 03:00:00	20629.75	0	20629.75	0.00000000	2018	1	1	lun	3
5	2018-01-01 04:00:00	20034.75	0	20034.75	0.00000000	2018	1	1	lun	4
6	2018-01-01 05:00:00	20282.75	0	20282.75	0.00000000	2018	1	1	lun	5
7	2018-01-01 06:00:00	21169.50	0	21169.50	0.00000000	2018	1	1	lun	6
8	2018-01-01 07:00:00	21810.00	30	21780.00	0.13755158	2018	1	1	lun	7
9	2018-01-01 08:00:00	22230.25	400	21830.25	1.79935003	2018	1	1	lun	8
10	2018-01-01 09:00:00	24001.50	1040	22961.50	4.33306243	2018	1	1	lun	9
11	2018-01-01 10:00:00	26088.75	1420	24668.75	5.44295908	2018	1	1	lun	10

FIGURA 1.1: Prime righe del dataset in esame.

1.1.4 La generazione solare

L'energia solare rappresenta la principale fonte di energia rinnovabile nel nostro pianeta in quanto il nostro pianeta è irradiato costantemente dai raggi solari e per questo sono stati sviluppati tre principali modi per sfruttare tale energia e renderla utilizzabile sia a livello domestico che industriale.

Il solare termico Il solare termico è un sistema che permette di immagazzinare il calore derivante dalle radiazioni solari e trasferirlo all'acqua sanitaria o agli impianti di riscaldamento. In breve il suo funzionamento è diviso in tre fasi: il collettore termico assorbe il calore delle radiazioni solari e scalda un fluido termovettore, il fluido con una pompa è portato in uno scambiatore di calore che trasferisce l'energia in uno serbatoio di stoccaggio, in seguito si scalda l'acqua. Questo sistema può essere collegato alla pompa di calore che aiuta a portare alla temperatura idonea l'acqua nei momenti (dell'anno e della giornata) in cui l'energia del sole non è sufficiente a soddisfare il bisogno di riscaldamento (solare termodinamico). Questo sistema consente di risparmiare energia elettrica ma non ne genera.

Il solare fotovoltaico Un impianto fotovoltaico è costituito da più moduli fotovoltaici (pannelli) che sono formati a loro volta da molteplici celle fotovoltaiche. Le celle

di silicio (un materiale semiconduttore) sono formate da due strati caricati il primo positivamente e il secondo negativamente e da una superficie anti-riflettente. Grazie all'effetto fotovoltaico il fotone (particella della luce, priva di massa e costituente elementare della radiazione elettromagnetica) che entra a contatto con la superficie a carica positiva, rilascia un elettrone che entra in un circuito a cui sono collegate tutte le altre celle, generando così corrente continua, che in seguito passa in un *inverter* che la trasforma in corrente alternata fruibile nelle abitazioni e nelle industrie. Il *surplus* di energia generato viene accumulato in impianti di stoccaggio (apposite batterie AGM, al litio o altro) e fornito quando la richiesta supera la produzione, o immesso nella rete nazionale e venduto.

I fattori che influenzano la resa della produzione dell'impianto fotovoltaico sono molteplici. In primo luogo abbiamo la temperatura che, se troppo elevata surriscalda il sistema e impatta negativamente sull'efficienza delle celle. Un altro fattore importante sono gli ombreggiamenti che possono essere passeggeri (solo per alcune ore del giorno) oppure no, l'impatto di questi non è proporzionale alla superficie coperta dalle zone d'ombra ma compromette il rendimento dell'intero impianto. Altri fattori sono la sporcizia sulla superficie del pannello, le dispersioni dei cablaggi, il *mismatch* (irregolarità dovute alla diversità dei moduli tra loro), l'efficienza dell'*inverter* e l'anzianità delle componenti.

La produzione di energia generata dall'impianto dipende strettamente dalla disponibilità di radiazione solare presente in quel momento e in quel luogo che dipende dalla posizione del sole, infatti minore è l'angolo che il sole forma con la superficie del pannello, maggiore è lo spessore di atmosfera che la luce deve attraversare e di conseguenza minore è la potenza della radiazione che incide sulla superficie. Il moto del sole è descritto da due angoli: quello relativo alla sua altezza che dipende dal periodo dell'anno e dalla latitudine (che è massimo nel giorno del solstizio d'estate, 21 giugno) e l'angolo orario (o *azimuth*) che cambia nel corso della giornata. La radiazione solare che arriva al modulo è scomponibile in tre componenti: la radiazione diretta che arriva con un angolo definito, quella diffusa che è riflessa dall'atmosfera e l'albedo (riflessione della luce da parte dell'ambiente circostante).

Supponendo che le variabili esogene siano fissate per massimizzare la produzione del singolo impianto, è di particolare interesse valutare come la produzione totale di energia elettrica generata dal solare (quindi fotovoltaica) in Italia, dipenda dalla variabile tempo. Si può intuire facilmente che tale produzione presenta una doppia stagionalità (essendo quest'ultima un concetto multiscale), una legata alla disponibilità delle radiazioni nel corso dell'anno solare e una interna alla giornata.

1.2 Andamento del fabbisogno totale di energia in Italia

Per capire come funziona la richiesta, e quindi l'offerta di energia elettrica in Italia è necessario scomporre tale serie in due componenti: una che chiameremo di lungo periodo e l'altra che misura le variazioni all'interno della giornata, dove la componente stagionale si esaurisce al suo interno. In questo capitolo si analizza la parte di lungo periodo, lasciando la componente interna alla giornata a trattazioni successive.

1.2.1 Valutazioni generali

Come prima cosa valutiamo l'andamento generale della serie a livello grafico, considerando tutte le osservazioni orarie del fabbisogno totale di energia in Italia dal 01/01/2018 al 31/12/2022. Come si può facilmente notare nella figura 1.2 la pandemia e le restri-

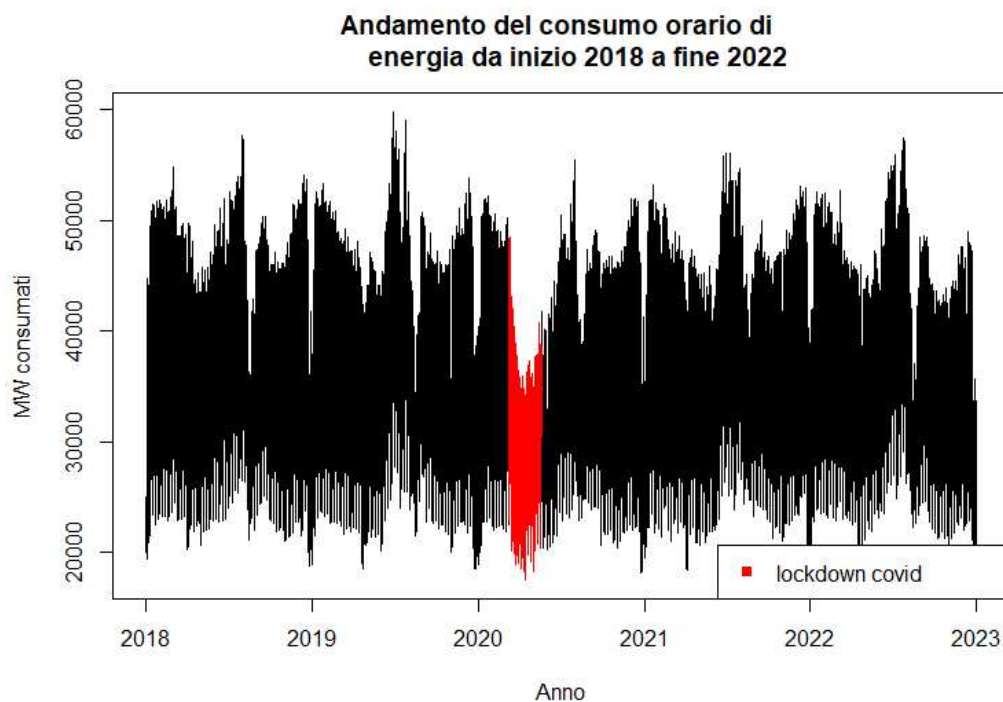


FIGURA 1.2: Serie storica complessiva del consumo orario di energia in Italia (2018-2022). In evidenza il periodo di *lockdown* per *COVID-19*.

zioni messe in atto per arginare la diffusione del virus di *COVID-19* hanno causato una drastica diminuzione del consumo di energia elettrica nel nostro Paese, arrivando ad un calo del 17,2% ad Aprile 2020 rispetto allo stesso mese dell'anno precedente (variazione relativa tendenziale), imputabile soprattutto alla temporanea chiusura delle aziende.

L'indicatore che Terna usa per misurare l'indice dei consumi mensili elettrici industriali (*IMCEI*), che valuta un campione di 1000 aziende "energivore" divise per settore (*ATECO 2007*) collegate alla rete di trasmissione elettrica nazionale, è considerabile come una *proxy* dell'andamento dell'evoluzione economica del Paese e quindi del PIL. Tra marzo e aprile 2020 tale indice era sceso infatti del 59% rispetto allo stesso mese dell'anno precedente. (TERNA (2023))

Si vuole valutare ora se la media dell'energia consumata nell'intero intervallo di tempo è significativamente maggiore della media nel periodo di *lockdown* (09/03/2020-18/05/2020).

Sia (y_1, \dots, y_n) con $n = 43824$ (totale delle osservazioni nel *dataset*), un campione casuale semplice da una variabile casuale $N(\mu, \sigma^2)$ e il sistema di ipotesi

$$H_0 : \mu = \mu_0 \text{ contro } H_1 : \mu > \mu_0.$$

L'obiettivo è rifiutare l'ipotesi nulla di uguaglianza delle due medie. Con μ vero valore della media e stimato dalla media campionaria:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n x_i. \tag{1.1}$$

Con varianza ignota e stimata dalla varianza campionaria corretta (stima non distorta e consistente della varianza):

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{X})^2. \tag{1.2}$$

La statistica test:

$$T = \frac{\mu - \mu_0}{\sqrt{\frac{S^2}{n}}} \sim t_{(n-1)} \tag{1.3}$$

Applicazione ai dati:

$$\bar{\mu} = 36036.89, \bar{\mu}_0 = 28865.11, S^2 = 71150246, t_{oss} = \frac{36036.89 - 28865.11}{\sqrt{\frac{71150246}{43823}}} = 177.9898$$

Con un livello di significatività α del 5% :

$$t_{n-1}(1 - \alpha) = 1.644888$$

essendo $t_{oss} \gg t_{n-1}(1 - \alpha)$ rifiuto l'ipotesi nulla di uguaglianza delle medie (*p-value* prossimo a 0), pertanto la media durante il periodo di *lockdown* è significativamente minore della media generale. (si veda Azzalini, 2001, paragrafo 4.4.2)

Nel caso in cui non si potesse accettare l'assunzione di normalità delle y_i , si può condurre un test approssimato per grandi campioni. Le stime di massima verosimiglianza delle medie rimangono invariate, cambia la distribuzione asintotica della statistica test

che è approssimativamente $N(0, 1)$.

Essendo $z_{1-\alpha} = 1.644854$ (livello $\alpha=0.05$) le conclusioni riportate sopra continuano a valere.

1.2.2 Analisi esplorativa della serie dei *total load*

Analizzando la distribuzione del carico totale, si possono notare delle peculiarità specifiche dell'intervallo di tempo considerato. A livello orario si vede come il consumo mediano aumenti fino alle 8 del mattino per poi stabilizzarsi nel corso della giornata, dove aumenta anche la variabilità dell'osservazione, fino alle ore 19, per poi scendere, seguendo l'orario lavorativo standard. La stessa corrispondenza si può notare attraverso l'osservazione dei *boxplot* suddivisi per giorno della settimana dove nei giorni feriali vediamo una mediana e una simiglianza molto forte tra i livelli dei quartili, mentre nel fine settimana il consumo cala drasticamente (in linea con la chiusura degli impianti lavorativi). Analizzando mensilmente il fenomeno notiamo un livello maggiore di consumi nei mesi di luglio, gennaio e febbraio imputabile all'aumento di energia utilizzata rispettivamente per raffreddamento e riscaldamento. Ad agosto il consumo cala per la chiusura estiva delle imprese (effetto vacanza). Annualmente i consumi sono abbastanza stabili, con una flessione nel 2020 dovuta al *lockdown* per la pandemia. I valori anomali identificati con i *boxplot* 1.3 (indicatori robusti della variabilità) non portano a pensare alla presenza di errori di rilevazione, ma ad alcuni valori estremi nei consumi in corrispondenza della coda destra della distribuzione.

1.2.3 Analisi con modello *SARIMA* della serie dei *total load*

Il modello *SARIMA* è un'estensione del modello *ARIMA* che permette di modellare andamenti periodici di tipo stagionale, trattata come stocastica e correlata con le componenti non stagionali. Tali modelli permettono di descrivere due tipi di correlazioni, quelle tra valori consecutivi (modellata dal modello *ARIMA*) e quelle tra valori distanti un multiplo del periodo nel tempo (modellata dalla componente stagionale).

Un modello *SARIMA*(p, d, q)(P, D, Q) $_S$ può essere definito come segue:

$$\phi(B) * \Phi(B^S) * (1 - B)^d * (1 - B^S)^D * Y_t = \phi_0 + \theta(B) * \Theta(B^S) * \epsilon_t. \quad (1.4)$$

con:

- Y_t : valore osservato della serie al tempo t
- B : operatore *Backshift* (lag o ritardo)

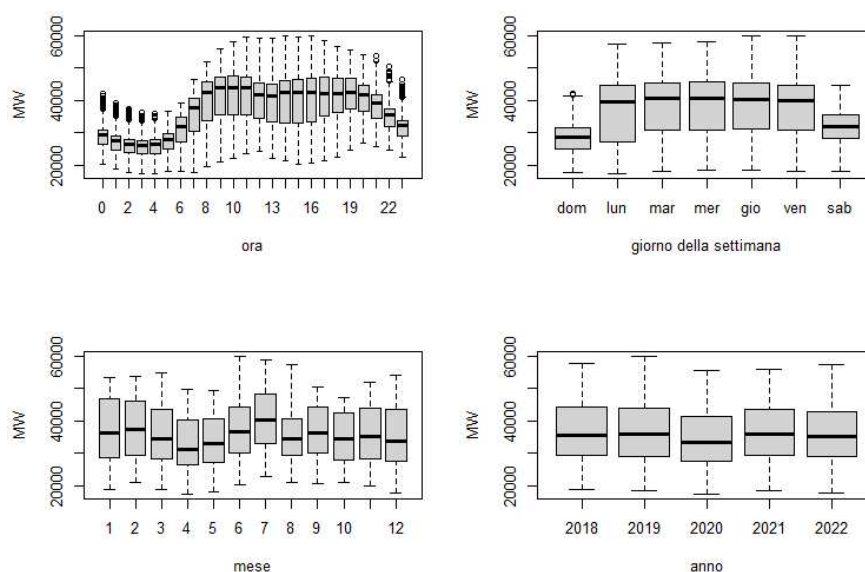


FIGURA 1.3: *Boxplot* dei consumi totali aggregati per ora, giorno della settimana, mese, anno.

- $\phi(B)$: operatore autoregressivo di ordine p del tipo $(1 - \phi_1 * B - \phi_2 * B^2 - \dots - \phi_p * B^p)$
- $\Phi(B)$: operatore autoregressivo stagionale di ordine P del tipo $(1 - \Phi_1 * B^S - \Phi_2 * B^{2*S} - \dots - \Phi_p * B^{P*S})$
- $(1 - B)^d$: differenziazione semplice di ordine d
- $(1 - B^S)^D$: differenziazione stagionale di ordine D
- ϕ_0 : primo coefficiente dell'operatore autoregressivo (legato alla media)
- $\theta(B)$: operatore a media mobile di ordine q del tipo $(1 - \theta_1 * B - \theta_2 * B^2 - \dots - \theta_p * B^q)$
- $\Theta(B)$: operatore a media mobile stagionale stagionale di ordine Q del tipo $(1 - \Theta_1 * B^S - \Theta_2 * B^{2*S} - \dots - \Theta_p * B^{Q*S})$
- ϵ_t : errore $\sim WN(0, \sigma_\epsilon^2)$

(Di Fonzo et al., 2005, paragrafo 6.5)

Analizziamo ora con il modello *SARIMA* la serie del fabbisogno nazionale, per semplicità si utilizza la serie aggregata dei consumi a livello giornaliero ottenuta come media aritmetica delle 24 osservazioni orarie nei vari giorni, rimandando la trattazione sui dati orari alla sezione 1.3. In particolare, lo studio si focalizza sullo studio della periodicità con periodo 365, per comprendere le dinamiche annuali.

Effettuando una differenziazione stagionale, dal grafico delle autocorrelazioni globali e parziali nella figura 1.4 si nota che la componente residua predominante è di periodo 7. Il processo generatore identificato è un *Random Walk* stagionale, $SARIMA(0, 0, 0)(0, 1, 0)_{365}$.

$$(1 - B^{365}) * Y_t = +\epsilon_t \quad (1.5)$$

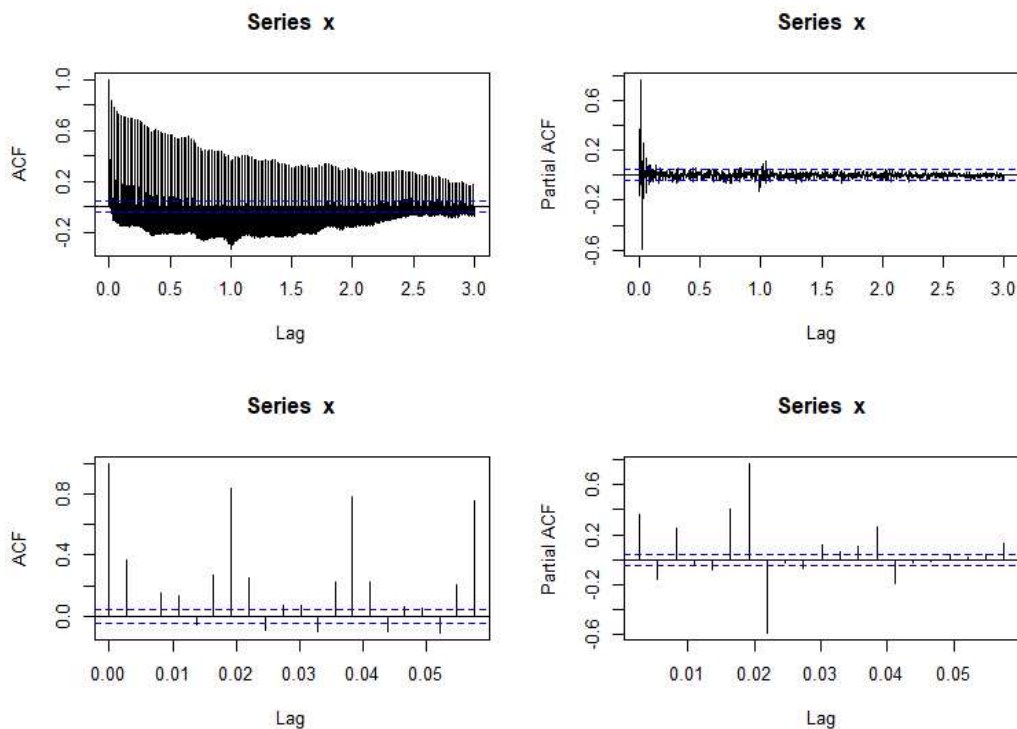


FIGURA 1.4: Autocorrelazione globale (a sinistra) e parziale (a destra), in basso misurate su un sottoinsieme dei dati per coglierne l'andamento periodico.

Alla luce di questo risultato si stima un modello *SARIMA* di periodo 7 (260 settimane in tutto), sui residui del modello precedente di periodo 365 (un anno, trascurando il fatto che il 2020 è bisestile). Il modello stimato attraverso il confronto di più modelli possibili è un $SARIMA(1, 0, 0)(0, 1, 1)_7$.

$$(1 - 0.675148 * B) * (1 - B^7) * Y_t = (1 - (-0.6776276) * B^7) * \epsilon_t \quad (1.6)$$

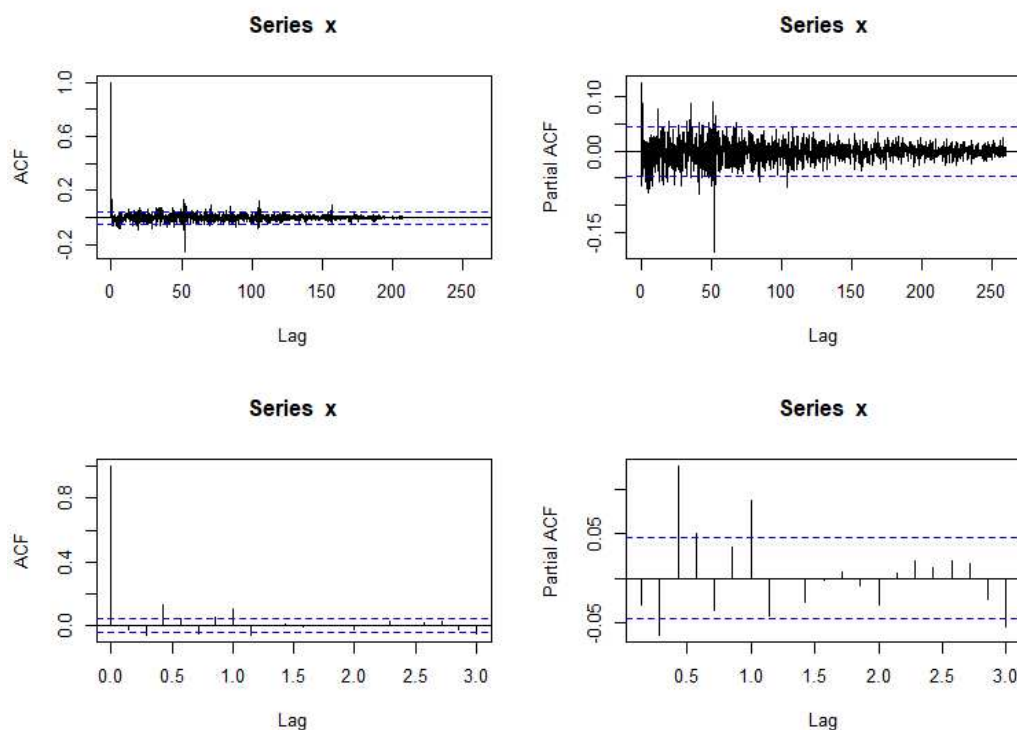


FIGURA 1.5: Autocorrelazione globale (a sinistra) e parziale (a destra), in basso misurate su un sottoinsieme dei dati.

Dai grafici delle autocorrelazioni nella figura 1.5 si vede un elemento significativo in corrispondenza dei multipli della 52-esima settimana, che corrisponde al numero di settimane in un anno, per rimediare a questo sulla serie dei residui di quest'ultimo modello stimo un modello settimanale di periodo 52. Si nota dai grafici che la componente predominante è quella di un *MA* stagionale. Il modello stimato è dunque un $SARIMA(0, 0, 0)(0, 0, 1)_{52}$.

$$Y_t = (1 - 0.1073326 * B^{52}) * \epsilon_t \quad (1.7)$$

Nel grafico delle autocorrelazioni dei residui (figura 1.6) sono presenti dei ritardi significativi, tuttavia non si riescono a riconoscere andamenti ciclici precisi e riconducibili ad una stagionalità di qualche tipo. Questi effetti sono imputabili a effetti di calendario, festività o al fatto che in anno sia composto da 52 settimane e un giorno per gli anni 2018, 2019, 2021, 2022 e 52 settimane e 2 giorni il 2020 che è bisestile. Dall'analisi dei residui infatti risulta che i residui sono correlati (con test di *Ljung Box*) e non si può accettare l'ipotesi di normalità (mediante il test di *Jarque-Bera*).

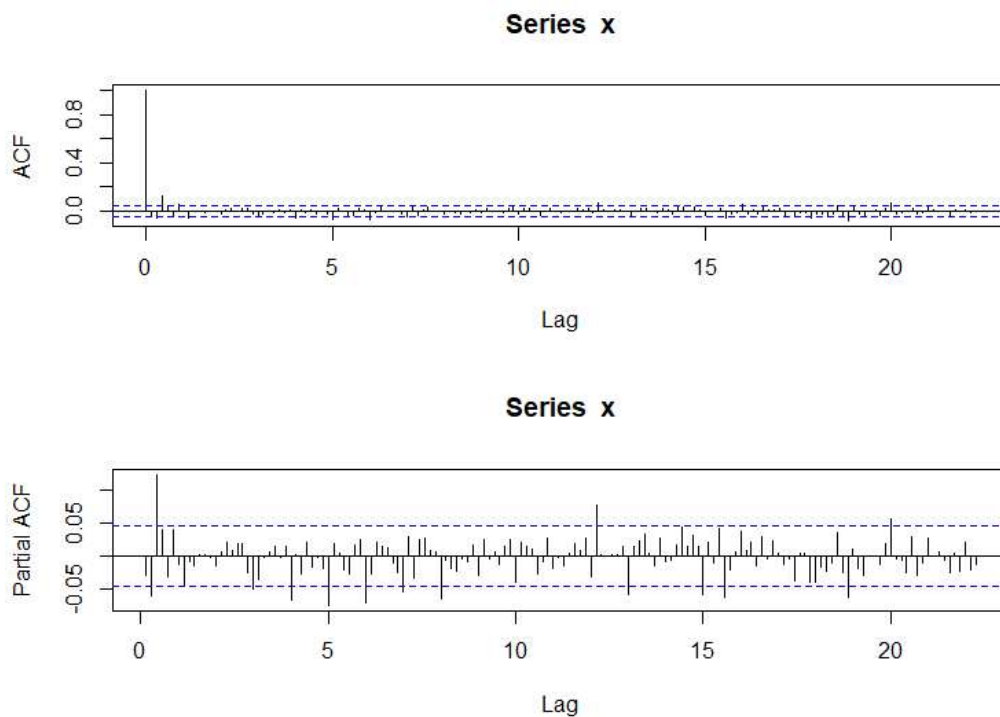


FIGURA 1.6: Autocorrelazione globale e parziale del modello sui residui.

1.3 Modelli per il consumo infragiornaliero: la *Duck Curve*

Come già accennato, la domanda di energia elettrica e quindi l'offerta che deve essere regolata a sua volta per mantenere efficiente il servizio, subiscono importanti variazioni all'interno della giornata. Questo causa notevoli problemi nella fase predittiva dell'offerta che devono essere risolti da società come Terna che fanno della gestione dell'energia immessa nel mercato il loro *business* principale. La stagionalità interna alla giornata è riconducibile principalmente a due fattori, il primo è l'usuale ciclo della giornata degli individui e delle imprese che suggerisce un maggiore consumo dell'energia durante il periodo diurno mentre l'altro fattore è quello più problematico ovvero la non linearità nel corso della giornata della produzione da parte delle fonti rinnovabili, che dipende spesso da fattori esogeni. Risulta facile pensare come la produzione di energia da impianti fotovoltaici dipenda fortemente dalla quantità di radiazioni solari che arrivano al suolo dai cieli in prossimità dell'impianto, un altro fattore influente è la portata dei fiumi in corrispondenza delle centrali per quanto riguarda la produzione idroelettrica.

Nelle ore serali si osserva un brusco innalzamento della domanda per compensare

Anno	Contributo percentuale fotovoltaico
2018	5.33
2019	5.82
2020	6.58
2021	6.08
2022	6.91

TABELLA 1.1: Contributo percentuale annuale della produzione fotovoltaica sul totale.

l'assenza di fotovoltaico e un abbassamento importante della richiesta nella parte centrale della giornata, con un conseguente rischio di sovrapproduzione e sovraccarico della rete, di difficile gestione da parte delle centrali termiche che non possono essere spente e riaccese di frequente a causa di uno scompenso economico che si verrebbe a creare.

Il grafico che si traccia mettendo in ascissa l'ora del giorno preso a riferimento e in ordinata l'energia totale richiesta dal mercato al netto della produzione di fotovoltaico relativa alla stessa ora prende il nome di *Duck Curve*, così chiamata per la caratteristica forma che assume.

Nonostante la maggior fonte di energia *green* in Italia sia l'idroelettrico, il fotovoltaico ricopre un ruolo rilevante e in crescita grazie anche agli incentivi finanziari statali messi in campo per favorire la transizione ecologica. Questo *trend* è confermato anche dai dati *GSE* (Gestore Servizi Energetici) che nel 2021 segnavano un totale di più di un milione di impianti installati, un aumento del 4% rispetto all'anno precedente, arrivando a coprire il 22% dell'energia elettrica prodotta da fonti rinnovabili.

1.3.1 Analisi esplorative sulla produzione di fotovoltaico in Italia

Di seguito sono proposte alcune analisi esplorative sulla produzione fotovoltaica in Italia. I *boxplot* orari (in figura 1.7) mettono in luce come la produzione di energia fotovoltaica (e la sua variabilità) aumenti nella parte centrale della giornata (in corrispondenza delle ore di maggior luce), indipendentemente dalla stagione in cui ci si trova o dalla temperatura registrata; è pressochè nulla nelle ore di buio.

La tabella 1.1 mostra la produzione percentuale annuale da fonte fotovoltaica sul totale dei consumi energetici.

Nel giorno domenica 12-04-2020 alle ore 14 l'Italia ha raggiunto il massimo rapporto di produzione da fonte fotovoltaica su totale della domanda raggiungendo circa il 48,27%. C'è da considerare che in quel periodo lo stato era in *lockdown* e quel giorno era la

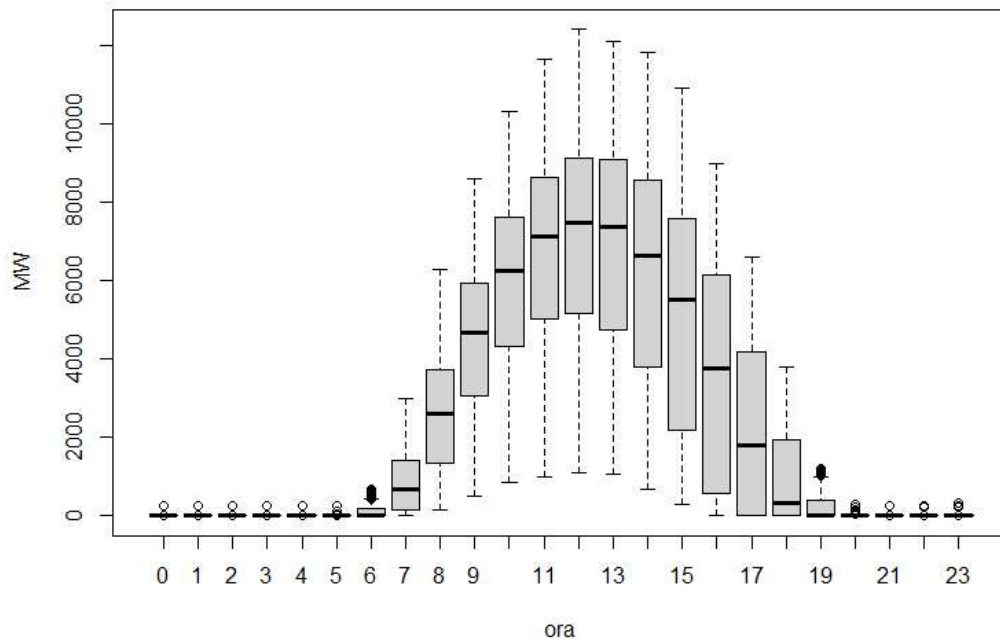


FIGURA 1.7: *Boxplot* orari della produzione di energia fotovoltaica in Italia nel periodo 2018-2022.

domenica di Pasqua quindi il consumo totale era relativamente basso (denominatore), mentre il meteo era soleggiato o poco nuvoloso in tutta Italia (numeratore del rapporto alto).

Segue il grafico 1.8 orario per anno (nel giorno 12-04) con tracciato in nero la linea che rappresenta il carico totale consumato per ora, in rosso il carico totale al netto della produzione fotovoltaica e in verde la produzione di energia da fonte fotovoltaica in *MW* (su scala che parte da 0 e riportata nella parte bassa del grafico per migliorare l'impatto visivo). Si è evidenziato il punto di massima indipendenza dalle altre fonti del 12-04-2020.

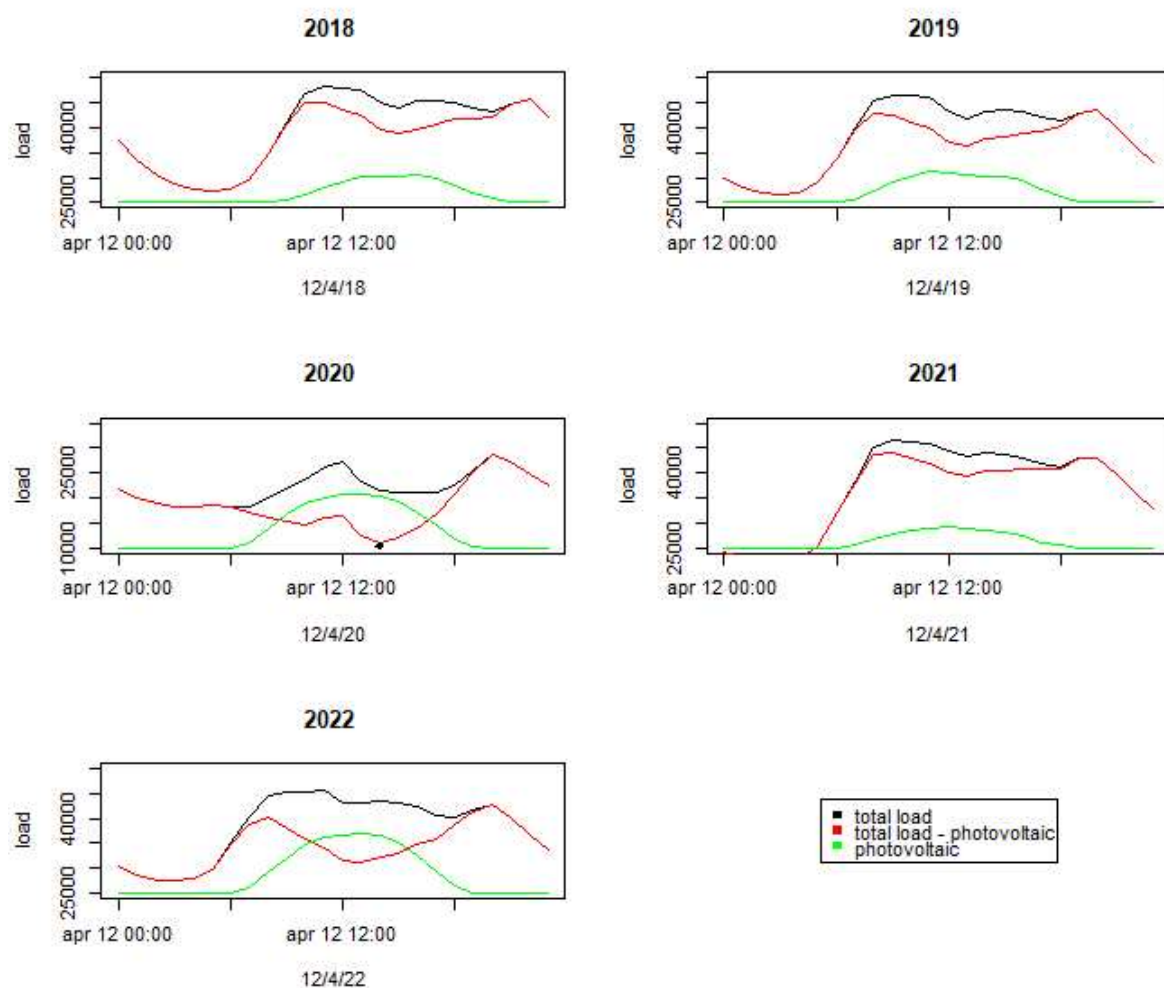


FIGURA 1.8: Apporto alla produzione totale negli anni suddivisa per fonti fotovoltaiche e non.

Si è deciso di esaminare la *Duck Curve* per l'Italia prendendo come giorno a riferimento il 5 Maggio (nella figura 1.9) per i seguenti motivi: tale giorno è a metà del periodo primaverile dunque si possono escludere i consumi dovuti al riscaldamento e al raffreddamento, la radiazione solare in primavera è buona e permette un'importante produzione della componente fotovoltaica.

Come ci si aspettava l'andamento del carico energetico richiesto dal mercato al netto del fotovoltaico tende a mostrare un divario nel corso delle ore che si amplifica sempre maggiormente con l'avanzare degli anni, andando a formare la "curva dell'anatra". Tra le 7:00 e le 8:00 del mattino con l'inversione della curva (quando gli impianti fotovoltaici iniziano a produrre) si forma la coda della papera. La differenza assume importanza nella parte centrale della giornata, arrivando al suo punto di massimo alle ore 13:00 dove, con un po' di astrazione, possiamo immaginare il ventre dell'anatra. Alle ore 20:00 si osserva, sempre compresa tra le curve annuali, quella che si può identificare

come la parte alta della testa, in corrispondenza al picco massimo di energia richiesta (al netto della produzione di fotovoltaico, che attorno a quell'orario tende ad annullarsi, con l'avvicinarsi del tramonto). In questa fase le centrali energetiche termiche hanno la maggiore difficoltà nel bilanciare le richieste di energia a causa del rapido aumento della domanda (visibile dalla forte pendenza e quindi alto coefficiente angolare in valore assoluto della tangente alla curva). Durante il corso della notte la richiesta si abbassa notevolmente andando a delineare il becco dell'animale in questione. Il grafico presenta la serie del 2020 che è da considerarsi come un *outlier*, in quanto i consumi reali sono falsati dalla situazione di *lockdown*, tuttavia se immaginiamo la serie traslata verso l'alto come sarebbe stato naturale ipotizzare a causa dei consumi aziendali, possiamo immaginare che la forma dell'andamento sia circa in linea con quello degli altri anni.

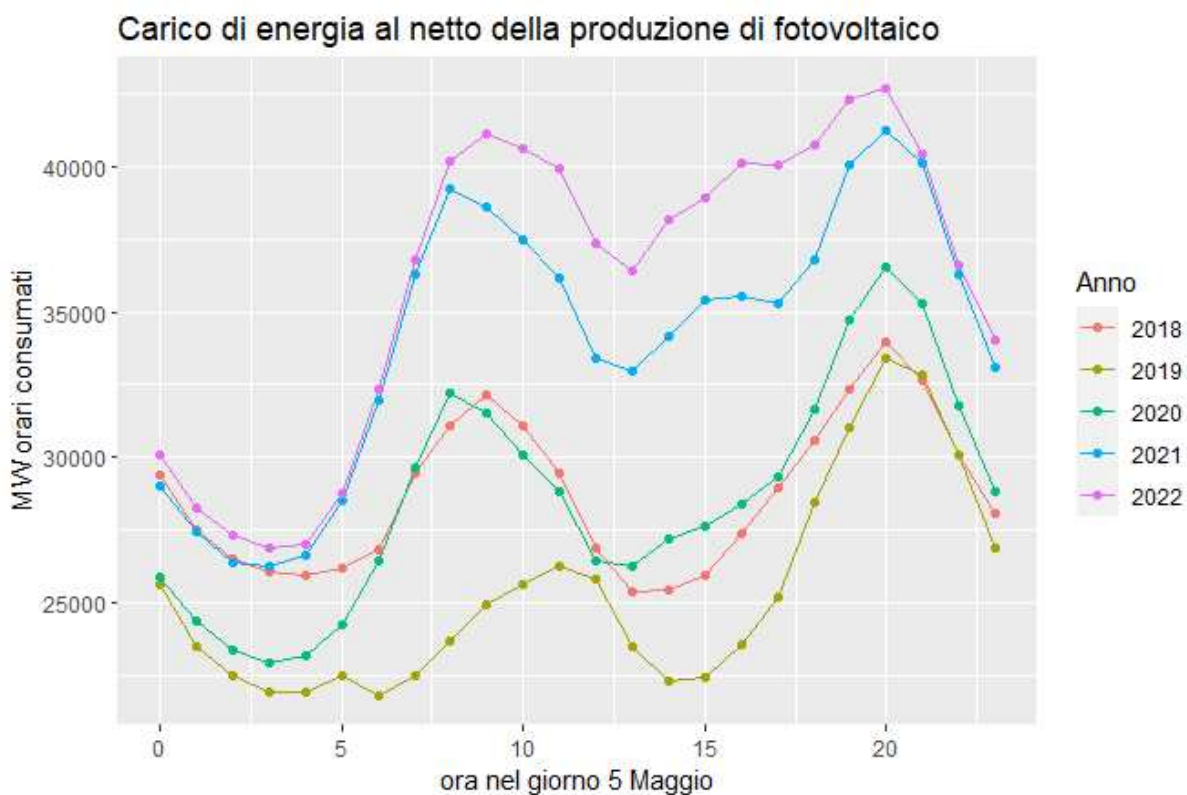


FIGURA 1.9: *Duck Curve* tracciata nel giorno 05-05 per gli anni 2018-2022.

La forma ad anatra (escludendo comunque il 2020), è ancora più riconoscibile confrontando le curve dello scostamento del consumo di energia, al netto della produzione di fotovoltaico, dalle rispettive medie giornaliere nei diversi anni (figura 1.10).

Per valutare la significatività della stagionalità oraria all'interno della giornata, quindi la stagionalità di breve periodo, si è deciso di prendere in considerazione il periodo del 4,5,6 Maggio 2022 per i motivi che seguono, in aggiunta alle motivazioni legate alle

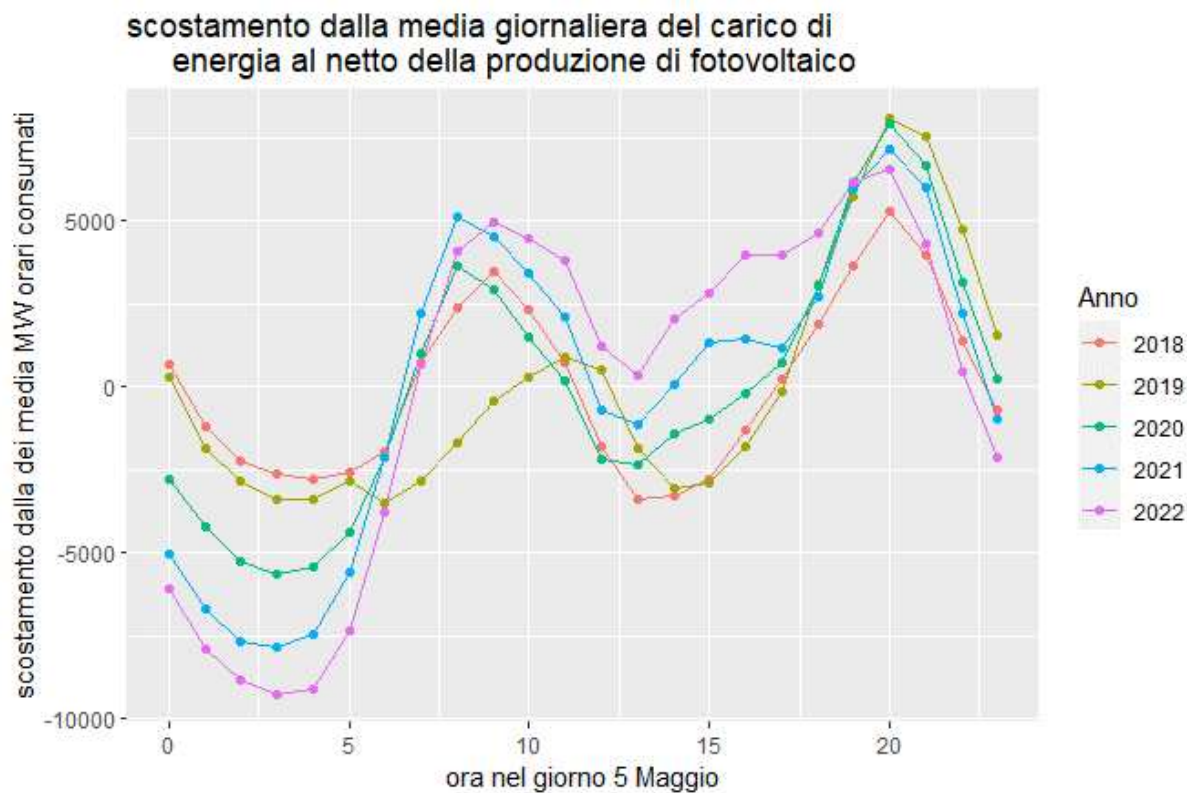


FIGURA 1.10: Scarti dalla media giornaliera dei consumi elettrici al netto della produzione di fotovoltaico nei giorni 05-05 negli anni 2018-2022.

stagioni solari spiegate in precedenza. I giorni in considerazione sono: mercoledì, giovedì e venerdì, dunque giorni feriali senza festività. Si possono inoltre escludere *bias* derivanti dall'effetto *covid*. Essendo i giorni vicini nel tempo possiamo assumere che l'effetto stagionale di lungo termine (annuale) sia trascurabile. Prendendo in considerazione tre giorni completi partendo dalle 00:00 del primo fino alle 23:00 dell'ultimo si hanno $24 * 3 = 72$ osservazioni.

1.3.2 Modello lineare con variabili *dummy*

Per stimare la stagionalità, il metodo più semplice è quello di utilizzare un modello lineare con l'ausilio delle variabili *dummy* (o indicatrici). (Di Fonzo et al., 2005, paragrafo 2.2.1)

Ipotizziamo inizialmente che sia presente solo una componente stagionale di breve periodo (S) dunque:

$$Y_t = S_t + \epsilon_t \tag{1.8}$$

La componente stagionale può essere spiegata da 24 variabili *dummy* (una per ogni ora),

tali che:

$$d_{j,t} = \begin{cases} 1 & \text{nell'ora } j - \text{esima del giorno a cui appartiene } t, j = 0, \dots, 23; t = 1, \dots, 72 \\ 0 & \text{altrimenti} \end{cases}$$

In forma matriciale il modello di regressione associato è:

$$\mathbf{y} = \mathbf{D}\boldsymbol{\beta} + \boldsymbol{\epsilon} \quad (1.10)$$

con:

$$\mathbf{D}_{72 \times 24} = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & & \\ 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & & \end{bmatrix} \quad \boldsymbol{\beta}_{24 \times 1} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_{24} \end{bmatrix} \quad \boldsymbol{\epsilon}_{72 \times 1} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_{24} \end{bmatrix}$$

$$\epsilon_t \sim N(0, \sigma_\epsilon^2).$$

I β_i con $i = 1, \dots, 24$ sono stimati con il metodo dei minimi quadrati dunque $\hat{\boldsymbol{\beta}} = (\mathbf{D}^T \mathbf{D})^{-1} \mathbf{D}^T \mathbf{y}$ e rappresentano i coefficienti grezzi di stagionalità.

Il modello stimato (senza intercetta, per evitare la “trappola delle dummy”) prendendo a riferimento le ore 23:00 (intercetta) risulta:

$$\hat{y}_t = 33435.75 - 3510.9164 * d_{0,t} - 5259.3335 * d_{1,t} - 6223.3333 * d_{2,t} - 6465.2498 * d_{3,t} - 6198.8339 * d_{4,t} - 4527.1669 * d_{5,t} - 962.9164 * d_{6,t} + 3594.7500 * d_{7,t} + 6512.8337 * d_{8,t} + 6684.7500 * d_{9,t} + 6017.7499 * d_{10,t} + 5160.9998 * d_{11,t} + 2385.0001 * d_{12,t} + 1871.9167 * d_{13,t} + 3598.1666 * d_{14,t} + 4644.2502 * d_{15,t} + 5230.7503 * d_{16,t} + 5020.0833 * d_{17,t} + 5837.3333 * d_{18,t} + 7580.4999 * d_{19,t} + 8136.2502 * d_{20,t} + 6090.9167 * d_{21,t} + 2728.3331 * d_{22,t}$$

Il modello si adatta bene ai dati, infatti l' R^2 è di 0.9206, l' $\overline{R^2}$ è 0.8826 (si abbassa notevolmente a causa dell'elevato numero di variabili inserite nel modello), l' AIC misura 1297.552.

Il criterio di informazione di *Akaike* si basa sul una penalizzazione della log-verosimiglianza, dunque è utilizzabile per il confronto di modelli correttamente specificati e non annidati.

Si definisce come:

$$AIC(F_d) = 2d - 2l(\hat{\theta}^{(d)}; y) \quad (1.11)$$

con:

- d : numero di parametri nel modello corrente

- $l(\hat{\theta}^{(d)}; y)$: log verosimiglianza massimizzata in $\theta^{(d)}$ del modello.

Il criterio predilige modelli parsimoniosi e che si adattano bene ai dati, dunque sarà migliore il modello con *AIC* minimo. Un criterio alternativo è il criterio di informazione Bayesiano (*BIC*) o Criterio di *Schwarz* (che pesa i parametri per $\log(n)$, con n numero di osservazioni). (Salvan et al., 2020, paragrafo 1.5.4)

1.3.3 Modello con funzioni trigonometriche

Un modello più parsimonioso rispetto a quello con le variabili *dummy*, è quello che modella la stagionalità attraverso le funzioni trigonometriche (Di Fonzo et al., 2005, paragrafo 2.2.2). In questo caso la componente stagionale è rappresentata da funzioni del tipo:

$$S_t = \sum_{i=1}^m A_i \cos\left(\frac{2\pi_i}{S}t - \phi_i\right) \quad (1.12)$$

con:

- m : numero di armoniche
- A_i : ampiezza dell' i -esima armonica
- $\frac{2\pi_i}{S}$: frequenza angolare dell' i -esima armonica
- ϕ_i : angolo di fase dell' i -esima armonica

Tralasciando ancora la componente di *trend* che supponiamo trascurabile nel sottoinsieme di dati considerato, il modello ipotizzato può essere reso lineare nei parametri sfruttando l'identità trigonometrica $\cos(x - y) = \cos x \cos y + \sin x \sin y$. (si veda Bertsch et al., 2011, Appendice 1.A)

Dunque

$$S_t = \sum_{i=1}^m \beta_{i,1} \cos\left(\frac{2\pi_i}{S}t\right) + \beta_{i,2} \sin\left(\frac{2\pi_i}{S}t\right) \quad (1.13)$$

con:

- $\beta_{i,1} = A_i \cos \phi_i$
- $\beta_{i,2} = A_i \sin \phi_i$.

Dalle analisi effettuate si vede che risultano significative le coppie di coefficienti delle prime tre armoniche (con cicli di periodo 24, 12 e 6 ore). Dunque il modello in forma matriciale si scrive:

$$\mathbf{y} = \mathbf{T}\boldsymbol{\beta} + \boldsymbol{\epsilon} \quad (1.14)$$

con:

$$T_{72 \times 7} = \begin{bmatrix} 1 & \cos \frac{2\pi 1}{24} & \sin \frac{2\pi 1}{24} & \cos \frac{2\pi 1}{12} & \sin \frac{2\pi 1}{12} & \cos \frac{2\pi 1}{6} & \sin \frac{2\pi 1}{6} \\ 1 & \cos \frac{2\pi 2}{24} & \sin \frac{2\pi 2}{24} & \cos \frac{2\pi 2}{12} & \sin \frac{2\pi 2}{12} & \cos \frac{2\pi 2}{6} & \sin \frac{2\pi 2}{6} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 1 & \cos \frac{2\pi 72}{24} & \sin \frac{2\pi 72}{24} & \cos \frac{2\pi 72}{12} & \sin \frac{2\pi 72}{12} & \cos \frac{2\pi 72}{6} & \sin \frac{2\pi 72}{6} \end{bmatrix} \beta_{7 \times 1} = \begin{bmatrix} \beta_0 \\ \beta_{1,1} \\ \vdots \\ \beta_{3,2} \end{bmatrix} \epsilon_{72 \times 1} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_{24} \end{bmatrix}$$

$\epsilon_t \sim N(0, \sigma_\epsilon^2)$. ?

Il modello stimato con intercetta risulta:

$$\hat{y}_t = 35433.5 - 2606.7 * \cos \frac{2\pi t}{24} - 4355.4 * \sin \frac{2\pi t}{24} + 658.4 * \cos \frac{2\pi t}{12} - 4198.7 * \sin \frac{2\pi t}{12} - 831.5 * \cos \frac{2\pi t}{6} - 519.4 * \sin \frac{2\pi t}{6}$$

Il modello ha un buon adattamento infatti ha un R^2 di 0.9062, l' $\overline{R^2}$ è 0.8976. L' AIC è di 1275.56. I criteri che pesano la parsimonia del modello portano a prediligere il modello con funzioni trigonometriche rispetto a quello con le *dummy*.

1.3.4 Modello *SARIMA*

Un ulteriore metodo per la stima del modello stagionale è mediante il modello *ARIMA* stagionale o *SARIMA* (la trattazione teorica è equivalente a quella riportata in 1.2.3). Il modello stimato (con coefficienti significativi) risulta essere appartenente alla classe *SARIMA*(2, 0, 0)(0, 1, 0)₂₄.

$$(1 - 1.3892077B - (-0.4648333B^2))(1 - B^{24})Y_t = \epsilon_t \quad (1.15)$$

Il modello ha un buon adattamento e presenta un AIC che vale 750.1123, che consente di preferirlo a tutti gli altri modelli trattati finora per parsimonia e bontà di adattamento basata sulla verosimiglianza.

Da tutti e tre i modelli stimati (il cui confronto dei *fitted values* è proposto in figura 1.11) si ha una netta significatività statistica dei coefficienti che regolano la stagionalità di breve periodo, dunque come ci si poteva aspettare, la componente oraria è fondamentale per lo studio dell'andamento dei consumi stagionali del bene trattato.

Ci si propone nel seguente capitolo 2 di comprendere la natura della stagionalità e di darne un'interpretazione per il caso in esame.

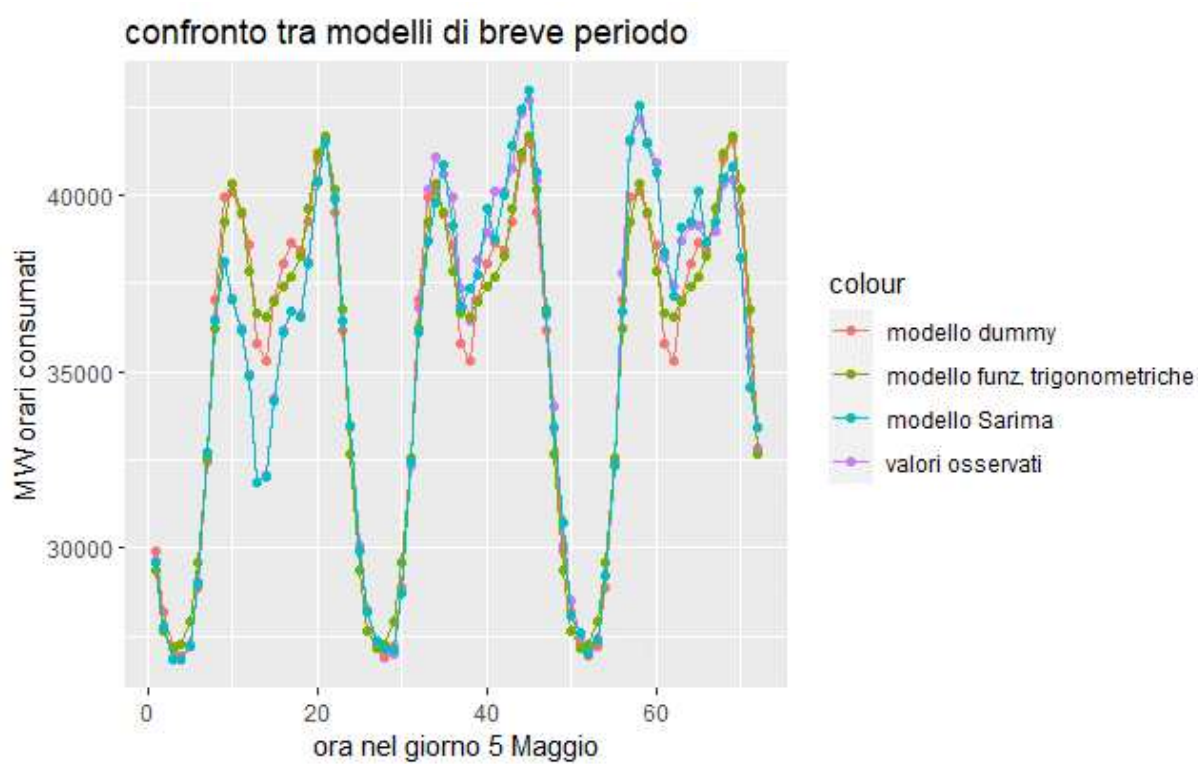


FIGURA 1.11: Confronto valori osservati e *fitted values* per i modelli trattati.

Capitolo 2

Valutazione della stagionalità mediante indice di Gini: aspetti metodologici

2.1 Scomposizione Indice di Gini

Un ulteriore modo per valutare la stagionalità avviene mediante l'applicazione di un indice di concentrazione, la cui variazione rispecchia un cambiamento nella concentrazione stagionale. In particolare in seguito si prenderà in considerazione l'Indice di Gini, introdotto dallo statistico italiano Corrado Gini nel 1914, ripercorrendo e adattando, al caso della domanda oraria di energia elettrica (al netto della produzione di fotovoltaico) la trattazione teorica e le applicazioni sviluppate da Grossi & Mussini (2021) in *“Seasonality in tourist flows: Decomposing and testing changes in seasonal concentration”*.

Il valore della variazione dell'indice di concentrazione (calcolato rispettivamente in due anni diversi) può essere scisso in due componenti, una che misura la stabilità stagionale (R), l'altra la magnitudine (entità) delle fluttuazioni stagionali (M), assumendo che l'andamento delle fluttuazioni stagionali sia stabile negli anni.

Sia $y_{0,t}, \dots, y_{n,t}$ la serie storica delle medie orarie dei consumi (e richieste) di energia al netto della produzione fotovoltaica con t da 2018 a 2022 e n il periodo della serie (da 0 a 23), dunque $y_{i,t}$ è la media del numero di MW consumati nell'anno t all'ora i -esima.

Sia $r(y_{i,t})$ il rango (livello o posizione) del periodo i nella graduatoria in ordine crescente per MW consumati dell'anno t .

L'indice di Gini può essere espresso come:

$$G_t = \frac{2cov(y_{i,t}, r_{i,t})}{n\bar{y}_t} \quad (2.1)$$

con \bar{y}_t il numero medio di MW consumati per ora nell'anno t .

La variazione di concentrazione dei consumi orari dall'anno t all'anno $t+k$ è misurato come differenza degli indici di Gini nei due periodi:

$$\Delta G = G_{t+k} - G_t = \frac{2cov(y_{i,t+k}, r_{i,t+k})}{n\bar{y}_{t+k}} - \frac{2cov(y_{i,t}, r_{i,t})}{n\bar{y}_t}. \quad (2.2)$$

Sia $C_{t+k|t}$ l'indice di concentrazione in $t+k$ utilizzando l'ordine dei periodi al tempo t

$$C_{t+k|t} = \frac{2cov(y_{i,t+k}, r_{i,t})}{n\bar{y}_{t+k}}. \quad (2.3)$$

Sottraendo e sommando $C_{t+k|t}$ alla formula di ΔG si ottiene la scissione della differenza degli indici di Gini nei due periodi come segue:

$$\Delta G = \left(\frac{2cov(y_{i,t+k}, r_{i,t+k})}{n\bar{y}_{t+k}} - C_{t+k|t} \right) + \left(C_{t+k|t} - \frac{2cov(y_{i,t}, r_{i,t})}{n\bar{y}_t} \right) \quad (2.4)$$

$$\Delta G = \left(\frac{2cov(y_{i,t+k}, r_{i,t+k})}{n\bar{y}_{t+k}} - \frac{2cov(y_{i,t+k}, r_{i,t})}{n\bar{y}_{t+k}} \right) + \left(\frac{2cov(y_{i,t+k}, r_{i,t})}{n\bar{y}_{t+k}} - \frac{2cov(y_{i,t}, r_{i,t})}{n\bar{y}_t} \right) \quad (2.5)$$

$$\Delta G = R + M. \quad (2.6)$$

La prima componente (R) tiene traccia della posizione occupata nella graduatoria delle medie orarie da parte di ciascun'ora nell'anno di riferimento. Un cambiamento nel *ranking* orario negli anni è sintomo di instabilità stagionale. Tale componente è nulla se la graduatoria nei due anni scelti a riferimento rimane invariata, è maggiore di zero se varia ed è massima se il *ranking* dei periodi si inverte completamente. Maggiore è il suo valore, maggiore è l'instabilità stagionale rilevata. R risulta quindi maggiore o uguale a 0.

La componente di magnitudine (M) è uguale a zero se non ci sono stati cambiamenti nella distribuzione dei consumi o se i cambiamenti sono stati proporzionali in tutte le ore del giorno. La componente assume valore positivo se c'è stato un aumento del divario (magnitudine) tra picco e punto di minimo dei consumi, negativo se la differenza è diminuita. M misura quanto nell'anno $t+k$, i consumi sono più concentrati nell'ora

con maggiore utilizzo di energia (prima nella graduatoria, rango massimo) dell'anno t . M è calcolato mantenendo la graduatoria dei periodi in t , supponendo dunque che non ci sia stata variazione nell'andamento stagionale dall'anno t all'anno $t + k$.

Le due componenti possono elidersi a vicenda, dunque una differenza tra indici (negli anni) uguale a zero non è da considerarsi esclusivamente come l'assenza di esistenza di cambiamenti stagionali. La valutazione numerica della differenza dell'indice di Gini non è sufficiente per un'interpretazione e può essere fuorviante, è dunque necessario analizzare gli 8 diversi casi possibili dei valori di R e M che possono portare al valore di ΔG trovato.

Caso:

1. $\Delta G < 0$ con $M < 0 \wedge R = 0$, tutta la variazione dell'indice è dovuta alla sola variazione della componente di magnitudine
2. $\Delta G = 0$ con $M = 0 \wedge R = 0$
3. $\Delta G > 0$ con $M > 0 \wedge R = 0$, tutta la variazione dell'indice è dovuta alla sola variazione della componente di magnitudine
4. $\Delta G < 0$ con $M < 0 \wedge R > 0$ e $|M| > R$, contributi opposti di R e M
5. $\Delta G = 0$ con $M < 0 \wedge R > 0$ e $|M| = R$, contributi esattamente opposti di R e M
6. $\Delta G > 0$ con $M < 0 \wedge R > 0$ e $|M| < R$, contributi opposti di R e M
7. $\Delta G > 0$ con $M = 0 \wedge R > 0$, aumento dell'indice di Gini dovuto solo all'aumento della componente di *re-ranking*
8. $\Delta G > 0$ con $M > 0 \wedge R > 0$, aumento della concentrazione stagionale calcolato come somma delle due componenti positive.

2.2 Test sul cambiamento dell'indice di concentrazione

Essendo l'indice di Gini una statistica descrittiva, non sono applicabili gli usuali test per la significatività del cambiamento rilevato. In particolare si vuole anche valutare la significatività delle due componenti considerate distintamente. Per fare questo *Grossi e Mussini* hanno proposto un adattamento del test suggerito da *Rey & Sastré-Gutiérrez* (2010) per testare la significatività del cambiamento della concentrazione in funzione del tempo.

Il test non parametrico si basa su permutazioni casuali, ed è stato applicato inizialmente all'*indice di Theil*, ma può essere utilizzato anche con altri indici di concentrazione (nel nostro caso l'indice di Gini).

Siano I_t, I_{t+k} gli indici di concentrazione al tempo t e $t+k$, $\Delta I = I_{t+k} - I_t$ è la statistica test e le ipotesi del test sono le seguenti:

$$\begin{cases} H_0 : I_{t+k} - I_t = 0 \\ H_1 : I_{t+k} - I_t \neq 0. \end{cases}$$

La distribuzione campionaria della statistica test sotto ipotesi nulla si costruisce riassegnando casualmente i pedici relativi al tempo delle osservazioni. Essendo $y_{i,t}$ e $y_{i,t+k}$ i valori osservati nei rispettivi tempi, si definiscono y_{i,t^*} e $y_{i,t+k^*}$ i valori ottenuti assegnando casualmente gli elementi in ogni coppia di osservazioni che si riferiscono al periodo i (con $i=1, \dots, n$) nel tempo t e $t+k$ (e i rispettivi vettori di valori sono le due distribuzioni miste). Permutando le osservazioni all'interno delle coppie, le due distribuzioni miste in t^* e $t+k^*$ si ottengono calcolando la correlazione tra i dati riferiti allo stesso periodo. Secondo *Rey e Sastré-Gutérrez*, t^* e $t+k^*$ sono i valori osservati in un processo in cui la concentrazione del fenomeno rimane invariata nel corso del tempo. Una volta eseguita la permutazione casuale degli indici del tempo nelle osservazioni, si possono calcolare gli indici di concentrazione e la loro differenza. Ripetendo molte volte (J) la riassegnazione degli indici si costruisce una distribuzione della statistica test sotto ipotesi nulla. Essendo $\Delta I_j = I_{j,t+k^*} - I_{j,t^*}$ la differenza di indici calcolati alla j -esima riassegnazione casuale con $j = 1, \dots, J$, si crea una variabile indicatrice d_j tale che :

$$d_{j,t} = \begin{cases} 1 & \text{se } |\Delta I_j| \geq |\Delta I| \\ 0 & \text{se } |\Delta I_j| < |\Delta I|. \end{cases}$$

Sia $D = \sum_{j=1}^J d_j$ si può calcolare uno pseudo p -value per la statistica con sistema di ipotesi specificato sopra:

$$p(\Delta i | H_0) = \frac{1 + D}{1 + J}. \quad (2.9)$$

A questo punto si rifiuta H_0 se lo pseudo p -value è piccolo (< 0.05), in questo caso si può affermare che c'è stato un cambiamento significativo della concentrazione nell'arco di tempo da t a $t+k$. A livello intuitivo il numeratore del p -value cresce (e porta ad accettare l'ipotesi di uguaglianza della concentrazione nei due tempi considerati) quando aumentano le volte in cui si verifica che la differenza tra indici randomizzati (in cui per

ipotesi si è in una situazione di costante concentrazione) è maggiore della differenza degli indici osservati nei due tempi, a parità di randomizzazioni J effettuate.

È possibile anche valutare la singola significatività di R e M per ottenere informazioni più utili per le valutazioni del caso, per farlo *Grossi e Mussini* propongono una modifica del metodo di *Rey and Sastré-Gutiérrez* proposto sopra, chiamato *scaled permutation test* (test delle permutazioni riscalato). Si formulano quindi, due diversi sistemi di ipotesi per le statistiche test R e M :

$$\begin{cases} H_0 : R = 0 \\ H_1 : R \neq 0 \end{cases} \quad \begin{cases} H_0 : M = 0 \\ H_1 : M \neq 0. \end{cases}$$

La distribuzione sotto ipotesi nulla si basa sull'applicazione di permutazioni casuali tra n coppie di dati osservati e ipotizzati in $t+k$. Sia $\lambda = \bar{y}_{t+k}/\bar{y}_t$ il rapporto tra la media dei consumi al tempo $t+k$ e t , moltiplicando gli $y_{i,t}$ per λ , si ottiene la distribuzione $t+k^\lambda$ che rappresenta la distribuzione che si sarebbe osservata in $t+k$ se la componente stagionale fosse rimasta stabile e con magnitudine pari a quella in t , il cui livello di concentrazione è per costruzione equivalente a quello in t . Dunque confrontando la distribuzione in $t+k^\lambda$ e in t l'andamento stagionale e la magnitudine sono costanti nel tempo. I consumi medi in $t+k^\lambda$ e in $t+k$ sono distribuiti egualmente per costruzione. Si ottiene poi la distribuzione $t+k^*$ attraverso una permutazione casuale tra ogni coppia di osservazioni al tempo i dalle distribuzioni di $t+k^\lambda$ e $t+k$ ($y_{i,t+k^\lambda}, y_{i,t+k}$), mentre la distribuzione in t è quella osservata. Per una permutazione casuale, le due componenti sono ricalcolate sostituendo la distribuzione osservata al tempo $t+k$ con la distribuzione $t+k^*$, che coincide con la $t+k^\lambda$ se i consumi sono cambiati in modo proporzionale nel tempo, cosa che porterebbe a valori nulli di R e M . Ripetendo molte volte (J) la permutazione casuale si costruisce una distribuzione della statistica test sotto ipotesi nulla per la componente del cambiamento di *ranking* e per quella di magnitudine.

Essendo R_j la componente di *re-ranking* calcolata alla j -esima permutazione casuale con $j = 1, \dots, J$, si crea una variabile indicatrice e_j tale che:

$$e_j = \begin{cases} 1 & \text{se } R_j \geq R \\ 0 & \text{se } R_j < R. \end{cases}$$

Sia $E = \sum_{j=1}^J e_j$ il numero di volte in cui il *re-ranking* è almeno uguale all' R osservato, si può calcolare lo pseudo p -value per la statistica:

$$p(R|H_0) = \frac{1+E}{1+J}. \quad (2.11)$$

Per un p -value piccolo (minore di 0.05) si rifiuta l'ipotesi di nullità di R quindi i dati indicano che la stagionalità ha un andamento instabile nel corso del tempo considerato (da t a $t+k$).

Essendo M_j la componente di magnitudine calcolata alla j -esima permutazione casuale con $j = 1, \dots, J$, si crea una variabile indicatrice f_j tale che :

$$f_j = \begin{cases} 1 & \text{se } |M_j| \geq |M| \\ 0 & \text{se } |M_j| < |M|. \end{cases}$$

Sia $F = \sum_{j=1}^J f_j$ il numero di volte in cui il modulo della magnitudine è almeno uguale all' M osservato, si può calcolare lo pseudo p -value:

$$p(M|H_0) = \frac{1 + F}{1 + J}. \quad (2.13)$$

Per un p -value piccolo non si può accettare l'ipotesi di nullità di M quindi i dati indicano che la magnitudine è cambiata significativamente nel tempo considerato (tra t e $t+k$).

Con lo stesso approccio si può testare la significatività di ΔG nel suo complesso che però, come detto in precedenza, non è completamente esplicativo per quanto riguarda la composizione della stagionalità in sé. In particolare definito $\Delta G_j = G_{t+k^*} - G_t$ e h_j ($j = 1, \dots, J$) con la variabile *dummy* tale che:

$$h_j = \begin{cases} 1 & \text{se } |\Delta G_j| \geq |\Delta G| \\ 0 & \text{se } |\Delta G_j| < |\Delta G|. \end{cases}$$

con $H = \sum_{j=1}^J h_j$ il numero di volte in cui il modulo di ΔG_j in J permutazioni casuali è almeno uguale al ΔG osservato, si trova lo pseudo p -value come:

$$p(\Delta G|H_0) = \frac{1 + H}{1 + J}. \quad (2.15)$$

Capitolo 3

Scomposizione della stagionalità: applicazione

3.1 Valutazioni preliminari sulla stagionalità dei consumi elettrici

Nella seguente sezione si propone un'applicazione ai dati come mostrato nella 1.1.3 della scomposizione dell'indice di Gini e dei relativi test di *Rey e Sastré-Gutérrez* e *scaled permutation* test su ΔG , R e M . Si effettua una valutazione delle variazioni tendenziali (rispetto all'anno precedente) e una valutazione che prende in considerazione l'arco dei 5 anni considerati (da 2018 a 2022). L'obiettivo di tale analisi è valutare la significatività statistica delle componenti in cui è scomposto l'indice di Gini nei vari intervalli annuali e darne un'interpretazione al fine di individuare le cause che generano il fenomeno che crea la *Duck Curve*.

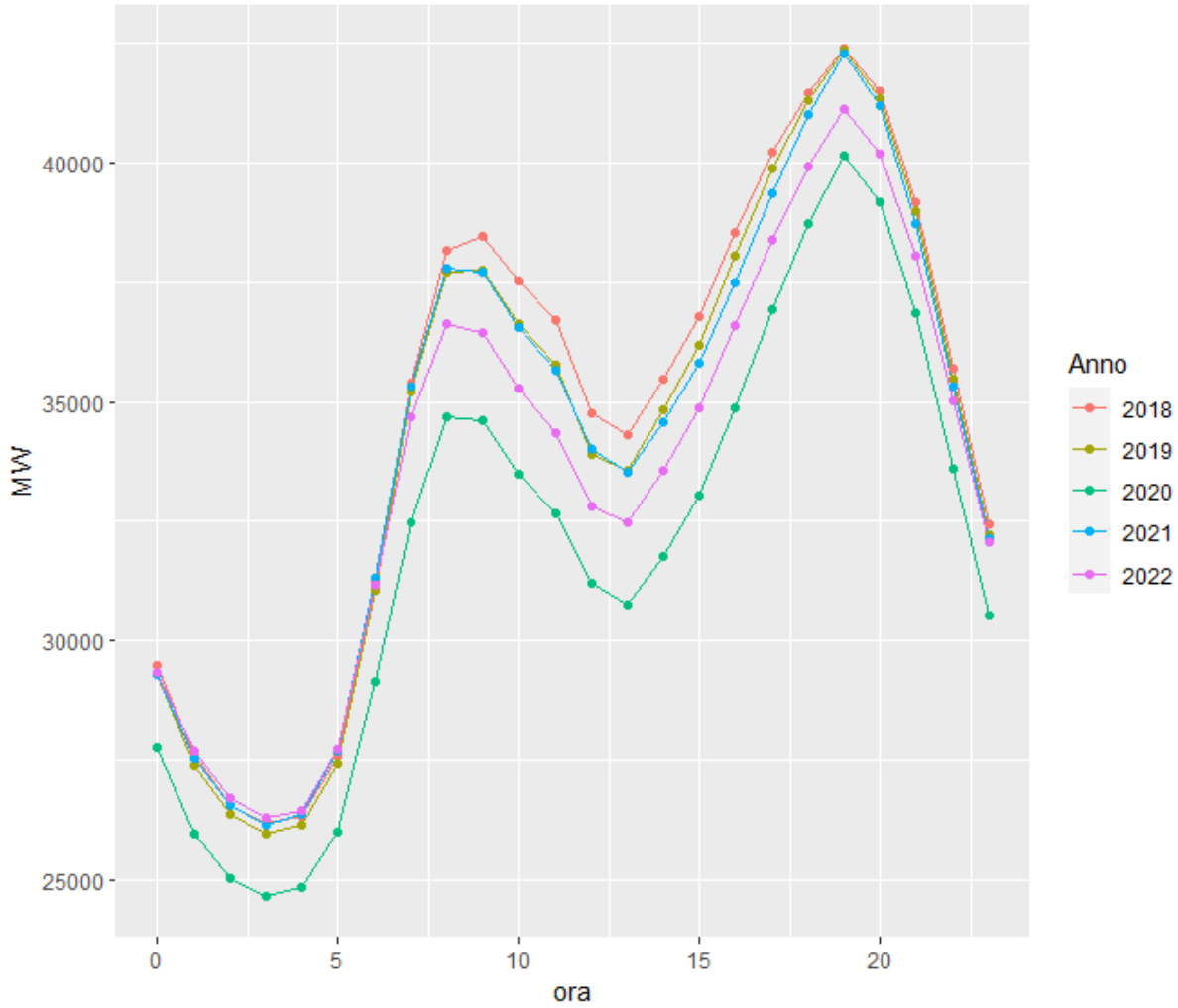


FIGURA 3.1: *Duck curve* tracciata utilizzando le medie orarie annuali tra 2018 e 2022.

Per procedere con le analisi è necessario preliminarmente definire un *dataset* di dati aggregati, per superare *bias* dovuti ai fattori esogeni: i dati utilizzati per calcolare l'indice di concentrazione sono le medie orarie dei consumi all'interno dell'anno corrente nei 5 anni, arrivando a 24 osservazioni all'anno per 5 anni, dunque un vettore di 120 valori $y_{i,t}$ con $i = 0, \dots, 23$ e $t = 2018, \dots, 2022$. Per iniziare sono stati calcolati degli indici stagionali sui consumi medi orari nell'intervallo di tempo considerato in tabella 3.1:

- l'indice di Gini G_t è definito come riportato in equazione 2.1
- il rapporto di stagionalità al tempo t , $SR_t = \frac{\max_i y_{i,t}}{\bar{y}_t}$ con \bar{y}_t media campionaria delle osservazioni nell'anno t .
- il coefficiente di variazione stagionale è $CV_t = \frac{\sigma_t}{\bar{y}_t} * 100$, con σ_t deviazione standard delle osservazioni al tempo t e \bar{y}_t media campionaria delle osservazioni nell'anno t .

t	G_t	SR_t	CV_t
2018	0.08223	1.22031	14.54451
2019	0.08255	1.23158	14.5377
2020	0.08067	1.25296	14.16488
2021	0.08017	1.23228	14.12419
2022	0.07447	1.22149	13.07997

TABELLA 3.1: Indici stagionali sui consumi orari medi annuali.

La concentrazione stagionale massima rilevata mediante l'indice di Gini si verifica nell'anno 2018 e diminuisce nel corso del tempo. Ci si aspetta che a una variazione negativa dell'indice di Gini ($\Delta G < 0$, diminuzione della concentrazione dei consumi) tra due periodi, corrisponda una diminuzione del coefficiente di variazione stagionale CV_t tra i due periodi (ΔCV), in altre parole se $G_{t+k} < G_t$ allora anche $CV_{t+k} < CV_t$. In questo caso il rapporto di stagionalità non segue l'andamento degli altri due indicatori, si nota che il massimo di SR_t è raggiunto nell'anno 2020 e questo è imputabile, più che ad un aumento del picco dei consumi orari, ad una diminuzione del denominatore in quanto in primavera il *lockdown* ha portato ad una riduzione dei consumi (come si vede nel paragrafo 1.2.1) che si riflette in una diminuzione dei consumi medi orari e quindi della loro media.

3.2 Scomposizione dell'indice di Gini e test applicati ai dati sui consumi elettrici

In questa sezione viene eseguita la scomposizione della variazione dell'indice di Gini nei vari intervalli di tempo considerati e si procede alla sua valutazione statistica e delle sue componenti mediante i test mostrati nel capitolo 2. In linea con la letteratura statistica, si considera un livello di significatività pari ad $\alpha = 0.05$, dunque si rifiuta l'ipotesi nulla per valori di α_{oss} (pseudo p -value, che da ora sarà denominato semplicemente come p -value) minori di 0.05 in tutti i test (evidenziati nella tabella 3.2). Dato che tutte le ipotesi nulle fissano la componente uguale al valore nullo, l'obiettivo è quello di rifiutare H_0 e osservare un valore significativo delle componenti in analisi. Per tutti i test si è fissato un numero di replicazioni, per trovare la distribuzione di riferimento sotto ipotesi nulla attraverso le permutazioni casuali, J pari a 1000.

Nelle colonne della tabella 3.2 sono riportati rispettivamente: gli anni di riferimento e gli indici di Gini calcolati sulle medie orarie dell'anno (prime 4 colonne), l'indice di concentrazione in $t + k$ utilizzando l'ordine dei periodi al tempo t , la differenza tra gli

$t+k$	t	G_{t+k}	G_t	$C_{t+k t}$	$\Delta G_{t+k t}$	R	M	$p(\Delta G H_0)$	$p(\Delta G_\lambda H_0)$	$p(R H_0)$	$p(M H_0)$
2019	2018	0.08255	0.08223	0.08251	0.00032	0.00004	0.00028	0.71029	0.44455	0.00299	0.49850
2020	2019	0.08067	0.08255	0.08049	-0.00188	0.00017	-0.00205	0.77223	0.10889	0.00699	0.11389
2021	2020	0.08017	0.08067	0.07992	-0.0005	0.00026	-0.00076	0.93906	0.52447	0.00099	0.36563
2022	2021	0.07447	0.08017	0.07433	-0.00571	0.00013	-0.00584	0.00999	0.00099	0.05194	0.00099
2022	2018	0.07447	0.08223	0.07421	-0.00776	0.00026	-0.00802	0.01798	0.00199	0.14386	0.00099

TABELLA 3.2: Scomposizione dell'indice di concentrazione stagionale: valori osservati dei test e p -value.

indici e la scomposizione di tale differenza in R e M (colonne dalla 6 alla 8), il p -value per il test di *Rey e Sastré-Gutérrez* (colonna 9). Le ultime tre colonne mostrano i p -value per i test ottenuto mediante le permutazioni riscalate per ΔG , R e M .

In tutti gli intervalli di tempo considerati si può affermare che il ranking della stagionalità varia nel corso del tempo dato che $R > 0$ in ogni riga.

Nell'intervallo di tempo tra 2018 e 2019 non si è rilevata una variazione significativa nell'indice di concentrazione (né in base al test di *Rey & Sastré-Gutiérrez* (2010) né con il test riscalato proposto da *Grossi & Mussini* (2021)). Il valore di ΔG è essenzialmente nullo (0.00032), come confermato dai test, e lo stesso si può dire per il valore della componente di *re-ranking* R (0.00004), il valore della componente di magnitudine M è maggiore di zero, dunque si rientra nel caso 3 tra quelli visti al paragrafo 2.1.

Nonostante il valore di R sembri molto prossimo a zero, il p -value del test per R ottenuto attraverso le permutazioni scalate ($0.00099 < 0.05$) mostra che non si può accettare l'ipotesi nulla di uguaglianza della componente R a 0. Si accetta invece la nullità di M e quindi non si sono verificati aumenti (dato che $M > 0$) statisticamente significative tra le intensità dei picchi dei consumi medi orari annuali, cosa che si può vedere anche graficamente dato che la forma dei grafici nei due anni rimane sostanzialmente invariata.

Ragionamenti simili si possono fare per le variazioni degli indici tra 2020 e 2019 e per 2021 e 2020. In questi due casi si verificano dei valori di ΔG negativi e prossimi a 0 non significativi per entrambi i test delle colonne 9 e 10, dei valori di R numericamente positivi, ma con valore basso però statisticamente significativo e valori di M negativi, ma che risultano non significativi. In questi due intervalli di tempo si verifica un comportamento di ΔG come quelli descritti nel caso 1 e nel caso 5, quindi un $\Delta G \geq 0$ con $R \geq 0$ e $M < 0$.

Il fatto che R sia statisticamente diverso da 0 indica che l'andamento stagionale si è modificato nel corso degli anni per quanto riguarda la graduatoria delle ore nella quale avvengono i picchi dei consumi. Ci si poteva aspettare questo tipo di comportamento considerando che i dati non comprendono l'energia prodotta da fonte fotovoltaica e, dunque, un aumento di questa, porta la popolazione a consumare meno energia prodotta

da altre fonti (quella utilizzata per calcolare le medie orarie annuali) nelle ore di picco di produzione fotovoltaica. Questo comportamento ha causato un cambiamento nel *ranking* orario nei diversi anni portando a valori di R diversi da 0. Un'ulteriore riflessione si può fare in particolare sulla variazione dell'indice tra 2019 e 2020: durante il *lockdown* infatti sono mutate drasticamente le abitudini dei consumi della popolazione, che hanno portato a cambiamenti significativi nella media oraria annuale, per questo si registra un significativo aumento di R .

Altre circostanze riguardano la variazione dell'indice di concentrazione tra il 2022 e il 2021, in questo caso ΔG è negativo (-0.00571) e per entrambi i test (*Rey e Sastré-Gutierrez* e *scaled permutation test*) è significativamente diverso da 0. R è maggiore di 0 (0.00013) e M è negativo (-0.00584) si rientra quindi nel caso 4 trattato in 2.1.

Considerando l'intervallo complessivo tra 2022 e 2018 si traggono le stesse conclusioni viste per l'ultimo intervallo annuale trattato, infatti $\Delta G = -0.00776$ e risulta statisticamente significativo per entrambi i test delle permutazioni casuali, $R = 0.00026$ maggiore di 0 in valore assoluto ma statisticamente non diverso da 0 e $M = 0.00802$ diverso da 0 con un p -value inferiore all'1%. Sia dai valori assoluti che dalla significatività statistica si può attribuire la quasi totalità della variazione di G alla componente di magnitudine. Un M diverso da 0 rivela una variazione non proporzionale dei consumi nelle ore e, con segno negativo ci indica una diminuzione del divario tra picco e minimo dei consumi orari medi.

3.3 Applicazione ai dati suddivisi per stagione meteorologica

Nella figura 3.2 vengono tracciate le curve delle medie dei consumi orari nell'intervallo di tempo considerato suddivisi per stagione meteorologica di appartenenza. Si rendono necessarie alcune precisazioni: essendo la stagione invernale a cavallo tra due annate diverse, si è considerato l'inizio dell'inverno 2018 in corrispondenza della prima osservazione del *dataset*, facente riferimento al giorno 01-01-2018. Si è ritenuto opportuno escludere dal grafico l'inverno che inizia a fine 2022, in quanto sarebbe stato la media delle sole osservazioni comprese tra il 21 e il 31 dicembre. Questo intervallo non può essere considerato rappresentativo della stagione in atto perché composto prevalentemente da giorni festivi e questi sottostimano ampiamente i consumi medi. Si è tenuto conto del fatto che il 2020 è un anno bisestile.

Da notare come il *lockdown* nella primavera del 2020 abbia portato ad una riduzione

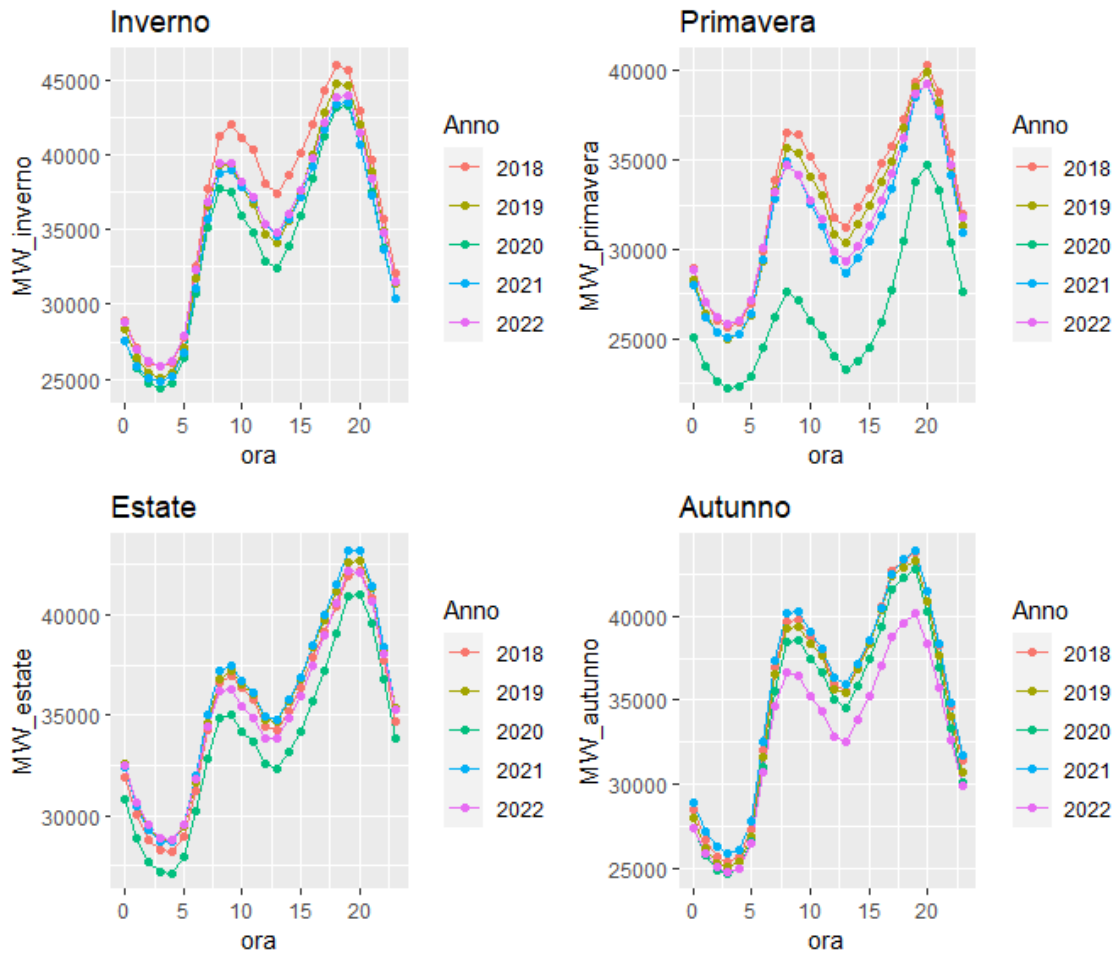


FIGURA 3.2: *Duck curve* tracciata utilizzando le medie orarie per stagione metereologica (2018-2022).

<i>stagione</i>	G_{2018}	G_{2022}	SR_{2018}	SR_{2022}	CV_{2018}	CV_{2022}
inverno	0.09966	0.09046	1.20056	1.20105	11.93425	11.25447
primavera	0.07507	0.06923	1.22586	1.23291	13.16292	12.12882
estate	0.06788	0.06398	1.20056	1.20105	11.93425	11.25447
autunno	0.09386	0.08278	1.20056	1.20105	11.93425	11.25447

TABELLA 3.3: Indici stagionali sui consumi orari medi per stagione metereologica.

drastica dei consumi medi, mantenendo comunque il *pattern* invariato nel corso della giornata.

In tabella 3.3 emerge la riduzione nel 2022 dell'indice di Gini in tutte le stagioni metereologiche (rispetto al 2018). Come mostrato con il calcolo degli indici stagionali sulle medie annuali, anche disaggregando i dati nelle varie stagioni, ad una diminuzione della concentrazione nel periodo, corrisponde un aumento del rapporto di stagionalità (SR_t) e un decremento del coefficiente di variazione stagionale (CV_t).

In tabella 3.4 sono calcolate le componenti della variazione dell'indice di Gini, le statistiche test e i rispettivi p -value considerando nel complesso l'intervallo che va dal 2018 al 2022, suddividendo i dati per stagione meteorologica. Si nota un andamento simile per tutte le stagioni delle variazioni degli indici di concentrazione e delle loro componenti, analogo a quello registrato per i consumi medi annuali nel periodo (ultima riga della tabella 3.2). In tutte le stagioni (inverno, primavera, estate e autunno) si misura una variazione dell'indice negativa, con una componente R di *re-ranking*, che rileva l'instabilità stagionale, positiva e una componente M di magnitudine negativa (situazione descritta nel caso 4 descritto alla sezione 2.1).

Per tutte le stagioni si individua variazione statisticamente significativa dell'indice di concentrazione con il test riscaldato proposto da Grossi & Mussini (2021). Si rifiuta l'ipotesi nulla di uguaglianza a 0 di M per tutte le stagioni, lo stesso vale per il test sulla componente R ad esclusione del periodo autunnale che si può considerare non nulla conducendo un test con livello di significatività al 10%. La significatività statistica di R dimostra che in tutte le stagioni meteorologiche, con il passare degli anni, si ha un cambiamento del *ranking* dei consumi orari. La significatività del test su M , con valore del test osservato a segno negativo, ci informa che su tutte le stagioni dell'anno si verifica una diminuzione della magnitudine, cioè si riduce nel tempo il divario tra picchi e minimi nella curva dei consumi orari medi. In definitiva si può dire che la concentrazione dei consumi e la scomposizione della sua variazione, nelle stagioni, mostrano un comportamento simile a quello dell'anno nel suo complesso come visto in 3.2.

Questo risultato sembra contrario rispetto a quanto atteso. Un aumento della produzione di fotovoltaico negli anni, assolutamente comprensibile dato il numero di installazioni produttive che sale costantemente, dovrebbe portare ad un valore di M negli anni statisticamente significativo e maggiore di 0. L'aspettativa è che la "pancia" dell'anatra nella curva aumenti la sua entità nel corso del tempo grazie alla sempre maggiore produzione di energia rinnovabile fotovoltaica, aumentando a sua volta i picchi in corrispondenza della sera e della mattina. Si sviluppano due interpretazioni possibili: da un lato la problematica di compensazione dell'energia non è ancora così grave e urgente come in altri paesi (si veda un esempio al paragrafo 3.4), dall'altro si può notare come non si stia ancora facendo abbastanza per promuovere e finanziare l'installazione e produzione del rinnovabile che, nonostante sia in aumento, non riesce a stare al passo della richiesta crescente di energia elettrica che nel periodo post-pandemico si è verificata da parte soprattutto delle grandi aziende energivore, per supportare la crescita del PIL nazionale (si veda il paragrafo 1.2.1).

<i>stagione</i>	G_{2022}	G_{2018}	$C_{2022 2018}$	$\Delta G_{2022 2018}$	R	M	$p(\Delta G H_0)$	$p(\Delta G_\lambda H_0)$	$p(R H_0)$	$p(M H_0)$
inverno	0.09047	0.09966	0.08998	-0.00919	0.00049	-0.00969	0.02597	0.00099	0.00099	0.00099
primavera	0.06923	0.07507	0.06852	-0.00584	0.00071	-0.00655	0.05195	0.00999	0.00999	0.00399
estate	0.06398	0.06788	0.06378	-0.00389	0.00021	-0.00410	0.01399	0.00299	0.00399	0.00199
autunno	0.08278	0.09386	0.08254	-0.01107	0.00025	-0.01132	0.15085	0.00099	0.09191	0.00099

TABELLA 3.4: Scomposizione dell'indice di concentrazione stagionale: valori osservati dei test e p -value nelle singole stagioni per l'intervallo di tempo 2018-2022.

Se si considera l'intervallo che copre i 5 anni, si potrebbe concludere che la magnitudine è l'aspetto sul quale è necessario porre attenzione e agire a livello politico ed economico per limitare l'aggravarsi del fenomeno. Tuttavia analizzando i dati annualmente si è visto che l'andamento stagionale non è stabile, dunque si è di fronte ad un fenomeno abbastanza eterogeneo che richiede un monitoraggio costante e delle azioni pratiche per mitigare la stagionalità interna alla giornata e i disagi che crea.

3.4 Confronto Italia-California

Come abbiamo visto dai dati, l'Italia, è ancora lontana dal rischio di sovrapproduzione di energia da fonte fotovoltaica e questa bassa produzione media percentuale consente agli impianti termici tradizionali di non essere soggetti a troppo bruschi innalzamenti della richiesta di energia nelle ore di picco, ma questo non deve farci sottovalutare il problema; è auspicabile infatti che la produzione di energia elettrica fotovoltaica nei prossimi anni aumenti e porti a problematiche di elevato impatto al servizio di fornitura energetica. Basti considerare che in California il problema è sorto già nel 2013 quando la “*California Independent System Operator*” (CAISO) ha pubblicato e ideato per la prima volta i risultati per una previsione sulla *Duck Curve* dello stato Americano che, grazie alle condizioni climatiche favorevoli e al progresso tecnologico in campo energetico, rappresenta un precedente importante per l'Italia.

Dal grafico in figura 3.3 tratto dall'articolo citato, ammettendo che gli sviluppi reali dell'andamento energetico siano simili a quelli proposti per gli anni fino al 2020, a livello qualitativo possiamo dire che ci si attende che il valore dell'andamento stagionale nella scomposizione dell'indice di Gini (R) sia nullo in quanto il pattern e quindi il *ranking* delle ore rimane invariato negli anni mentre dall'espansione della “pancia dell'anatra” si suppone un valore positivo e significativo della componente di magnitudine M , dato che aumenta il divario tra picchi e minimi della serie.

Attualmente il carico al netto della produzione di fotovoltaico registrato da CAISO (Aprile 2023), ha registrato valori nulli o negativi in alcune ore del giorno tanto da far rinominare la curva come “*Canyon Curve*” per sottolineare l'urgenza della problematica.

Ad oggi l'unica possibilità sembra quella di dismettere alternativamente la produzione da altre fonti che poi non possono essere riattivate agevolmente creando uno spreco di risorse e denaro notevoli.

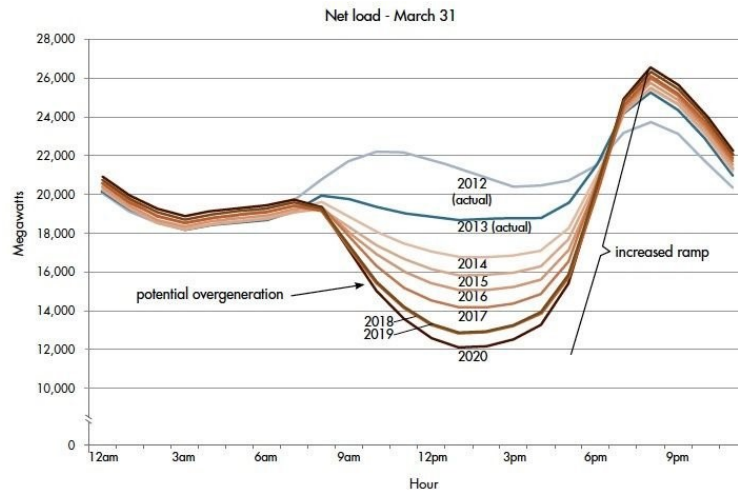


FIGURA 3.3: Grafico della *Duck Curve* stimata in California dall'articolo (Jones-Albertus, 2017).

Un altro caso preoccupante si è verificato nelle isole Hawaii in data 08-08-2013 (rappresentato nella figura 3.4) quando la sovrapproduzione (da fotovoltaico), un abbassamento della richiesta e una gestione non ottimizzata del sistema di rete si sono combinate e hanno causato una “*backfeed condition*” (condizione di ritorno) in tutta la rete creando notevoli disagi, tanto da far cambiare il nome alla curva in “*Nessie Curve*” (riferendosi ironicamente al mostro di *Loch Ness*).

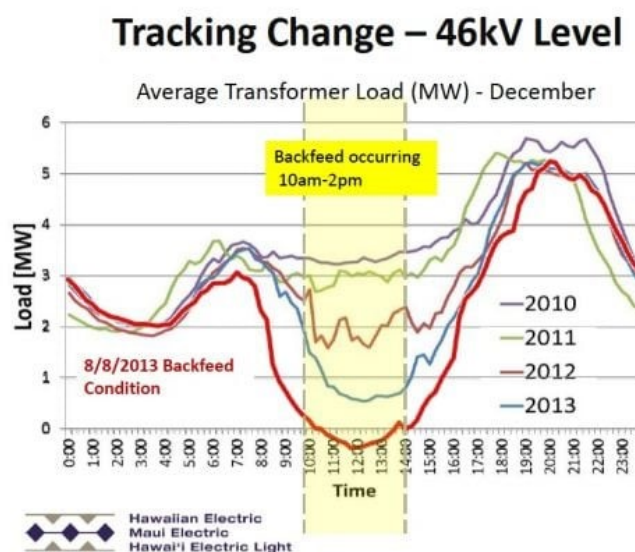


FIGURA 3.4: Grafico della *Duck Curve* isole Hawaii dall'articolo John (2014).

Conclusione

Manovre da attuare per mitigare il fenomeno Per contrastare l'acuirsi dei picchi di richiesta di energia mattutini, ma soprattutto serali, è fondamentale migliorare e aggiornare costantemente le metodologie di previsione dei consumi in tempo reale, applicare delle manovre e creare infrastrutture in grado di sopperire alla discontinuità della fonte fotovoltaica nell'arco della giornata. È necessario preventivare importanti investimenti per favorire l'immagazzinamento di energia, sia in ambito pubblico che privato, oltre al controllo costante della rete di trasmissione dell'energia.

Un modo per ritardare il picco di un intervallo di tempo (relativamente breve) è quello di installare pannelli fotovoltaici rivolti verso ovest (dove possibile) in modo da catturare maggiore energia nelle ore serali, momento della giornata in cui si hanno i maggiori problemi di compensazione, oltre a continuare a sviluppare e installare pannelli rotanti che seguono il movimento del Sole nella giornata.

La manovra più importante da attuare è favorire il cambiamento delle abitudini di consumo della popolazione; infatti se si riuscisse a programmare la maggior parte dei consumi durante la parte centrale della giornata si eviterebbero consumi esagerati nelle ore serali. Per spingere la popolazione a cambiare si potrebbe, oltre che a fare informazione, alzare i prezzi dell'energia nelle fasce orarie serali in modo da stimolare gli acquirenti a consumare durante le ore di luce. Sempre in questo ambito bisogna fare una riflessione sulla diffusione delle auto elettriche, che non devono essere ricaricate durante la sera/notte, ma durante il giorno, magari mediante l'utilizzo di colonnine installate nei posti di lavoro.

Si rende necessario lo sviluppo e la diversificazione delle fonti "*green*" che non dipendano così strettamente da fattori stagionali e uno scambio di energia (prodotta da fonti rinnovabili) da zone in cui si sta producendo a zone in cui la produzione è ferma sia per zona climatica e meteorologica differente, e su questo l'Italia può lavorare molto grazie alle diverse fasce climatiche che presenta (motivo per il quale è difficile una suddivisione geografica netta per l'applicazione dello studio in questione).

Bibliografia

- AZZALINI, A. (2001). *Inferenza Statistica. Una Presentazione Basata sul Concetto di Verosimiglianza*. Milano: Springer-Verlag Italia.
- BERTSCH, M., DAL PASSO, R., GIACOMELLI, L. et al. (2011). Analisi matematica-seconda edizione. In *Analisi matematica-Seconda edizione*. The McGraw-Hill Companies, Srl, pp. 1–611.
- DI FONZO, T., LISI, F. et al. (2005). *Serie storiche economiche. Analisi statistiche e applicazioni*. Carocci.
- GROSSI, L. & MUSSINI, M. (2021). Seasonality in tourist flows: Decomposing and testing changes in seasonal concentration. *Tourism Management* **84**, 104289.
- GSE (2019). Fonti rinnovabili in italia e in europa verso gli obiettivi al 2020 e al 2030 .
- JOHN, J. S. (2014). Hawaii’s solar-grid landscape and the ‘nessie-curve’. *Green Tech Media* **10**.
- JONES-ALBERTUS, B. (2017). Confronting the duck curve: How to address over-generation of solar energy. *US Department of Energy* .
- KATZ, M., ROSEN, H., BOLLINO, C. & MORGAN, W. (2015). Microeconomia (quinta edizione).
- REY, S. J. & SASTRÉ-GUTIÉRREZ, M. L. (2010). Interregional inequality dynamics in mexico. *Spatial Economic Analysis* **5**, 277–298.
- SALVAN, A., SARTORI, N., PACE, L., SALVAN, A., SARTORI, N. & PACE, L. (2020). *Modelli lineari generalizzati*. Springer.
- TERNA (2023). Terna driving energy. <http://https://www.terna.it/it>.

