



Università degli Studi di Padova

DIPARTIMENTO DI SCIENZE STATISTICHE

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

RELAZIONE FINALE

Modelli di diffusione per l'analisi del numero di contagiati da Covid-19 in Italia

Relatore:
Dott.ssa Mariangela Guidolin
Dipartimento di Scienze Statistiche

Laureando:
Matteo Salmaso
Matricola 1160481

Indice

Introduzione	5
1 Covid-19: caratteristiche della malattia e timeline della diffusione iniziale	7
1.1 Introduzione generale al Covid-19	7
1.1.1 Sintomatologia e tempo di incubazione	8
1.1.2 Modalità di trasmissione	8
1.1.3 Diagnostica	9
1.1.4 Ruolo degli asintomatici nella diffusione di Covid-19	9
1.2 Dai primi casi in Cina alla dichiarazione della pandemia	10
2 Diffusione del contagio: il modello di Bass e le sue estensioni	13
2.1 Teoria della diffusione dell'innovazione	13
2.2 Parentesi storica sui modelli di diffusione	14
2.2.1 Dall'equazione logistica al modello di Bass	14
2.3 Modello di Bass	15
2.4 Modello di Bass Generalizzato	17
2.5 Modello di Guseo-Guidolin	19
2.5.1 Posizionamento temporale delle fasi di contagio e di diagnosi di positività	24
2.6 Metodi di stima e identificazione statistica: regressione non lineare	25
2.7 Adattamento del modello ai dati e confronto tra modelli annidati	28
2.8 Analisi dei residui e statistica di Durbin-Watson	29
2.9 Affinamento ARMAx	32
3 Analisi statistica della diffusione di Covid-19 in Italia	35
3.1 Presentazione dei dati utilizzati nell'analisi	35
3.2 Applicazione del modello di Bass	36
3.3 Applicazione del modello di Bass Generalizzato	38

3.3.1	Modello di Bass Generalizzato con uno shock esponenziale	38
3.3.2	Modello di Bass Generalizzato con uno shock rettangolare	40
3.4	Applicazione del modello di Guseo–Guidolin	43
3.4.1	GGM: posizionamento temporale delle due fasi	45
3.4.2	GGM: analisi dei residui e affinamento SARMAx	48
Conclusioni		51
Bibliografia		55
Sitografia		57

Introduzione

Lo studio dei processi di diffusione delle innovazioni risulta un tema di centrale importanza, dal momento che prodotti e tecnologie innovative coinvolgono molteplici aspetti della vita delle persone, dei Paesi, della società e dell'economia. La diffusione di un'innovazione è un fenomeno di natura essenzialmente sociale: tuttavia esso ha sempre attratto l'attenzione di molti studiosi per il suo carattere interdisciplinare, che permette di combinare ed integrare teorie e concetti provenienti da discipline molto diverse, come le scienze naturali, la matematica, la fisica, la statistica, le scienze sociali, l'economia, il marketing.

L'interdisciplinarità, propria dei processi di diffusione, è un elemento caratterizzante di questa relazione, che nasce con l'obiettivo di analizzare un fenomeno di natura epidemiologica, quale la dinamica della diffusione del contagio da Covid-19 in Italia, utilizzando modelli di diffusione sviluppati nell'ambito del marketing quantitativo. Questi modelli descrivono il ciclo di vita di un prodotto, con lo scopo di prevedere l'andamento delle vendite del prodotto stesso durante il suo ciclo di vita. Come viene spiegato più approfonditamente nel capitolo 2, si può vedere un'analogia tra il processo di diffusione di un virus in un sistema sociale e il processo di diffusione di un prodotto innovativo in un mercato; questo giustifica l'applicazione, anche in ambito epidemiologico, di modelli pensati originariamente per il marketing.

I dati utilizzati nell'analisi sono stati messi a disposizione dalla Protezione Civile con frequenza giornaliera e si collocano in una finestra temporale che va dal 21 febbraio 2020, giorno in cui sono stati diagnosticati in Italia i primi casi positivi al virus del Covid-19, al 10 giugno 2020.

Gli strumenti utilizzati nell'analisi statistica sono i modelli di diffusione: in particolare, vengono utilizzati il modello di Bass e due sue estensioni, il modello di Bass Generalizzato (GBM) e il modello di Guseo-Guidolin (GGM). Il GBM introduce una componente di intervento che tiene conto della diversa velocità del processo di diffusione nel tempo, mentre il GGM considera il mercato potenziale

come una variabile dipendente dal tempo e modella in maniera separata le fasi di comunicazione e di adozione, che si assume co-evolvano nel tempo. Una volta identificato il miglior modello per la serie storica analizzata, si cerca di perfezionare tale modello, correggendo la problematica di autocorrelazione dei residui, di cui spesso i modelli di diffusione sono affetti, con la tecnica dell'affinamento ARMA_x (o SARMA_x).

La relazione è articolata come segue: nel primo capitolo vengono illustrate le principali caratteristiche del nuovo coronavirus, responsabile della malattia denominata Covid-19; successivamente, viene fornita una linea del tempo della diffusione iniziale del virus, dai primi casi diagnosticati in Cina fino alla dichiarazione della pandemia.

Nel secondo capitolo viene introdotta la teoria della diffusione dell'innovazione e viene aperta una parentesi storica sui modelli di diffusione; in seguito, vengono presentati nel dettaglio i modelli di diffusione utilizzati nell'analisi, interpretandoli da un punto di vista epidemiologico. Il capitolo termina con alcuni approfondimenti relativamente ad aspetti di stima, identificazione statistica ed analisi dei residui.

Nel terzo capitolo si passa all'applicazione pratica: vengono presentati i risultati ottenuti, supportati da interpretazioni e commenti non solo sul piano statistico, ma anche di contesto, con opportuno riferimento alla diffusione di Covid-19 in Italia. Seguono alcune conclusioni finali dell'analisi e considerazioni relative allo sviluppo di un vaccino anti-Covid e a possibili nuove ondate di contagi.

Capitolo 1

Covid-19: caratteristiche della malattia e timeline della diffusione iniziale

1.1 Introduzione generale al Covid-19

Covid-19 è una malattia respiratoria causata da un virus sconosciuto fino alla fine del 2019 e che è stato denominato SARS-CoV-2; questo virus appartiene alla famiglia dei coronavirus, così chiamati in quanto presentano in tutta la loro superficie delle punte che formano una specie di corona. I coronavirus possono causare malattie che vanno dal comune raffreddore a malattie più gravi, come la Sars (sindrome respiratoria acuta grave) e la Mers (sindrome respiratoria mediorientale). La comunità scientifica, nazionale e internazionale, sta ancora cercando di identificare la fonte del virus; al momento, l'ipotesi più accreditata è che questo nuovo coronavirus abbia origine zoonotica. Nonostante il potenziale *animal reservoir* (ossia l'animale nel quale il virus vive, cresce e si moltiplica) e gli eventuali *host* intermedi non siano ancora noti, studi suggeriscono che il nuovo coronavirus sia nato dalla combinazione tra il coronavirus del pipistrello e un coronavirus di origine sconosciuta; questo, tuttavia, deve essere ancora confermato.

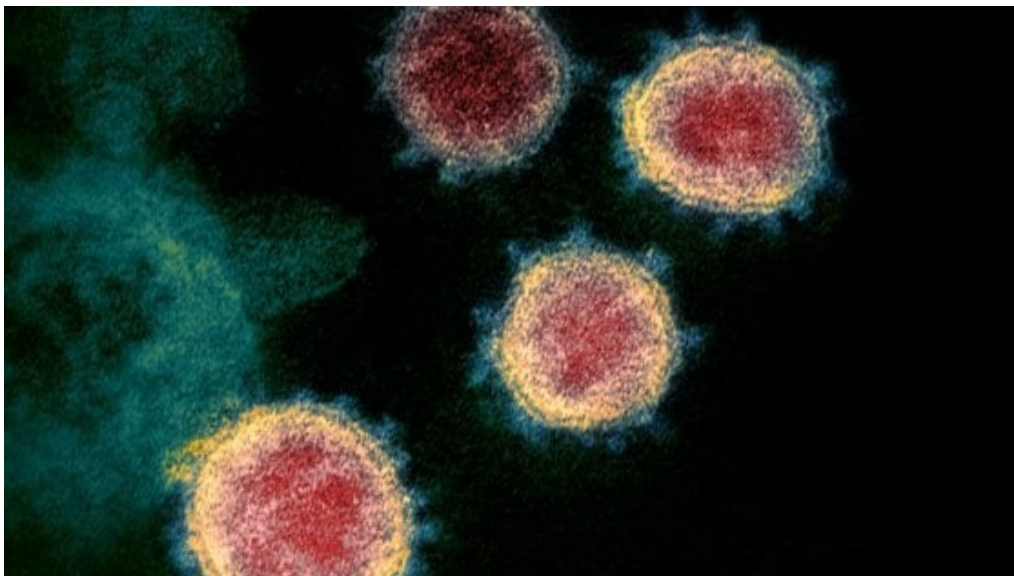


Figura 1.1: SARS-CoV-2 visto al microscopio elettronico (fonte: NIAID-RML)

1.1.1 Sintomatologia e tempo di incubazione

I sintomi più comuni di Covid-19 sono febbre, stanchezza e tosse secca. Alcuni pazienti possono presentare indolenzimento e dolori muscolari, congestione nasale, naso che cola, mal di gola, diarrea, perdita del gusto e perdita dell'olfatto. Questi sintomi iniziano gradualmente e sono generalmente lievi, motivo per cui molte persone guariscono da Covid-19 senza ricorrere al ricovero ospedaliero. Nei casi più gravi, l'infezione può causare polmonite interstiziale, sindrome respiratoria acuta grave, insufficienza renale e persino la morte. Le persone anziane e quelle con patologie preesistenti, come ipertensione arteriosa, problemi cardiaci o diabete, hanno maggiori probabilità di sviluppare forme gravi della malattia, in quanto più fragili.

La stima del tempo di incubazione (dove per "tempo di incubazione" si intende il periodo di tempo che intercorre tra il contagio e la manifestazione dei primi sintomi) va dai 2 ai 14 giorni, con una mediana dai 5 ai 6 giorni.

1.1.2 Modalità di trasmissione

Il nuovo coronavirus, responsabile della malattia respiratoria Covid-19, può essere trasmesso da persona a persona tramite un contatto stretto con un individuo positivo a SARS-CoV-2, sia esso sintomatico o asintomatico. La via primaria di trasmissione sono i cosiddetti *droplets*, ossia goccioline di diametro $\geq 5 \mu\text{m}$ prodotte respirando, starnutando, tossendo o parlando, che possono penetrare nell'organismo da occhi, naso o bocca.

Per le loro dimensioni, i *droplets* viaggiano nell'aria per brevi distanze, generalmente inferiori a un metro; oltre a raggiungere in maniera diretta soggetti suscettibili nelle immediate vicinanze, possono anche depositarsi su oggetti o superfici che diventano quindi fonte di contagio. In questo caso, le mani che sono entrate in contatto con gli oggetti così contaminati possono costituire veicolo di trasmissione per contatto indiretto quando toccano bocca, naso e occhi.

In uno studio effettuato dai National Institutes of Health (NIH), agenzia del Dipartimento della Salute e dei Servizi Umani degli Stati Uniti, i ricercatori hanno studiato la sopravvivenza di SARS-CoV-2 in diverse superfici. Quello che i

ricercatori hanno scoperto è che questo nuovo coronavirus può resistere fino a 4 ore sul rame, fino a 24 ore sul cartone e fino a 3 giorni sulle superfici in acciaio inox e in plastica. Questi risultati vanno però interpretati con cautela, tenendo in considerazione il fatto che la presenza di RNA virale non indica necessariamente che il virus sia vitale e la sua carica potenzialmente infettiva. Il contatto con una superficie contaminata, pur restando un rischio potenziale di contagio da Covid-19, non è ritenuto il principale mezzo di contagio.

1.1.3 Diagnostica

Lo strumento attualmente più attendibile ed efficace che viene utilizzato per una corretta diagnosi di Covid-19 è il tampone faringeo, mentre il test diagnostico di riferimento è basato sulla Real Time PCR (RT-PCR), che consente l'amplificazione dei microrganismi virali e l'individuazione di casi positivi al virus del Covid-19.

I soggetti a cui viene effettuato il tampone sono coloro che presentano i sintomi riconducibili a Covid-19 o che sono stati a contatto con soggetti positivi al virus; anche gli operatori sanitari che prestano servizio nelle unità di terapia intensiva e di malattie infettive vengono sottoposti al tampone con una certa regolarità.

Al soggetto sottoposto al tampone viene prelevato materiale biologico (muco) presente nelle prime vie respiratorie con un bastoncino cotonato; dopo averlo accuratamente sigillato, il campione viene inviato direttamente ad un laboratorio di microbiologia, dove viene schedato e successivamente analizzato.

1.1.4 Ruolo degli asintomatici nella diffusione di Covid-19

Uno degli aspetti maggiormente discussi durante la pandemia di Covid-19 riguarda gli asintomatici, ossia coloro che sono positivi al virus del Covid-19 ma non presentano alcun sintomo. Dato che il tampone generalmente non viene effettuato a chi non presenta sintomi, molti asintomatici, che quindi non vengono sottoposti al tampone, potrebbero inconsapevolmente essere dei portatori sani.

L'OMS sostiene, anche sulla base di quanto già osservato su altri coronavirus come MERS-CoV, che il contatto con i casi sintomatici sia il motore principale della trasmissione del nuovo coronavirus, mentre il contagio attraverso gli asintomatici

sia possibile, ma non frequente. Questa dichiarazione è stata contestata da molti scienziati, tra cui il professore Andrea Crisanti, microbiologo e direttore del Dipartimento di Medicina molecolare dell'Università di Padova. Crisanti è anche il promotore del noto studio epidemiologico effettuato a Vo' Euganeo, primo focolaio del Veneto, per studiare la diffusione e la trasmissione del nuovo coronavirus. Da questo studio è risultato che ben 63 soggetti, positivi al test sierologico, sono sempre risultati negativi al tampone. Ciò significa che queste 63 persone si sono ammalate e sono guarite in diversi periodi dalla fine di gennaio fino al 21 febbraio (il primo screening sulla popolazione di Vo' Euganeo è stato effettuato dal 21 al 29 febbraio). Stando alle risposte dei 63 soggetti sopracitati, questi non presentavano alcun sintomo: ciò indica che il virus si trasmette sottotraccia, senza essere notato, e buona parte della popolazione non presenta sintomi. Calcolando le persone positive al tampone e quelle che erano già guarite e che avevano sviluppato anticorpi, risulta che a Vo' il 22 febbraio c'erano già 150 individui entrati in contatto col virus; per Crisanti, questo studio conferma il ruolo importante determinato dagli asintomatici nella trasmissione del nuovo coronavirus.

1.2 Dai primi casi in Cina alla dichiarazione della pandemia

Con questo paragrafo si intende ricostruire la storia del nuovo coronavirus, partendo dai primi casi diagnosticati in Cina e percorrendo le tappe che hanno portato alla dichiarazione della pandemia di Covid-19.

Il 31 dicembre 2019 la Cina riporta all'OMS un focolaio di 41 casi di polmonite a Wuhan, nella provincia di Hubei, la cui causa era sconosciuta. Gran parte dei pazienti contagiati nel focolaio del 31 dicembre dichiara di aver frequentato il mercato del pesce di Wuhan, un *wet market*: per questo motivo, il 1° gennaio 2020 questo mercato viene chiuso.

Il 7 gennaio le autorità cinesi identificano il virus responsabile dei casi di polmonite a Wuhan come un nuovo tipo di coronavirus, che è stato provvisoriamente chiamato 2019-nCoV (dove la "n" sta per *novel*, ossia nuovo).

L'11 gennaio la Cina riporta la prima morte nel proprio Paese a causa del nuovo coronavirus.

Nella settimana dal 13 al 19 gennaio vengono riportati casi isolati di contagio da Covid-19 in alcune nazioni, tra cui Thailandia, Stati Uniti, Corea del Sud e Giappone.

Il 23 gennaio, la città di Wuhan è messa in quarantena, misura di contenimento estesa al resto della provincia di Hubei qualche giorno dopo.

Il 30 gennaio, l'OMS dichiara il nuovo coronavirus un'emergenza sanitaria globale. In Italia, due turisti cinesi provenienti dalla provincia di Hubei vengono diagnosticati positivi al nuovo coronavirus il 31 gennaio, mentre il 6 febbraio viene confermato come infetto uno degli italiani residenti a Wuhan e rimpatriati dalla Cina con un volo speciale dell'Aeronautica Militare italiana.

L'11 febbraio l'OMS annuncia che la malattia provocata dal nuovo coronavirus è ufficialmente denominata COVID-19 (dove "CO" sta per corona, "VI" per virus, "D" per *disease* e "19" indica l'anno in cui si è manifestata). Qualche giorno prima, invece, il coronavirus responsabile del Covid-19 viene rinominato SARS-CoV-2.

Nel mese di febbraio, diversi Paesi del mondo registrano le prime morti a causa del nuovo coronavirus, tra cui l'Italia, in cui il primo decesso è avvenuto il 21 febbraio. In questo mese, i primi focolai cominciano a formarsi anche al di fuori della Cina: Corea del Sud, Iran e Italia sono i primi paesi in cui il numero di contagi aumenta significativamente. A questi paesi, si aggiunge la Spagna, dove, dal 3 marzo, il numero di nuovi contagiati dal nuovo coronavirus è cominciato ad aumentare esponenzialmente.

Prendo una parentesi al contesto italiano, il giorno dopo la scoperta dei primi focolai nel Lodigiano e a Vo' Euganeo, queste zone vengono messe in quarantena. Progressivamente vengono attuate misure di contenimento in varie regioni d'Italia, fino ad arrivare al DPCM dell'8 marzo, che impone una quarantena a livello nazionale, limitando gli spostamenti, al di fuori dal proprio Comune di residenza, solo per motivi di lavoro o di salute.

Per poter dichiarare la pandemia di una malattia, questa malattia deve rispondere a tre criteri: si diffonde tra le persone, la sua diffusione avviene su scala mondiale ed è mortale. Poiché all'11 marzo sono stati registrati più di 118000 casi di Covid-19

in 114 paesi e 4291 persone hanno perso la vita per via dell'infezione da Covid-19, l'OMS ha dichiarato il passaggio da epidemia a pandemia.

Capitolo 2

Diffusione del contagio: il modello di Bass e le sue estensioni

2.1 Teoria della diffusione dell'innovazione

I modelli di diffusione utilizzati nell'analisi e presentati da un punto di vista teorico in questo capitolo affondano le loro radici nella teoria presentata dal sociologo E.M. Rogers nel libro "Diffusion of Innovations" (2003). In questo libro, Rogers definisce la diffusione come un processo con cui un'innovazione viene trasmessa attraverso determinati canali, nel tempo, tra i membri di un sistema sociale. L'adozione di tale innovazione non avviene simultaneamente in un sistema sociale (alcuni esempi di innovazioni sono un'idea, un prodotto o un virus, come SARS-CoV-2). A tal proposito, Rogers ipotizza 5 categorie di adottanti, sulla base della tempestività degli individui nell'adottare l'innovazione: gli innovatori (*innovators*), gli adottanti precoci (*early adopters*), la maggioranza precoce (*early majority*), la maggioranza ritardataria (*late majority*) e infine i ritardatari (*laggards*).

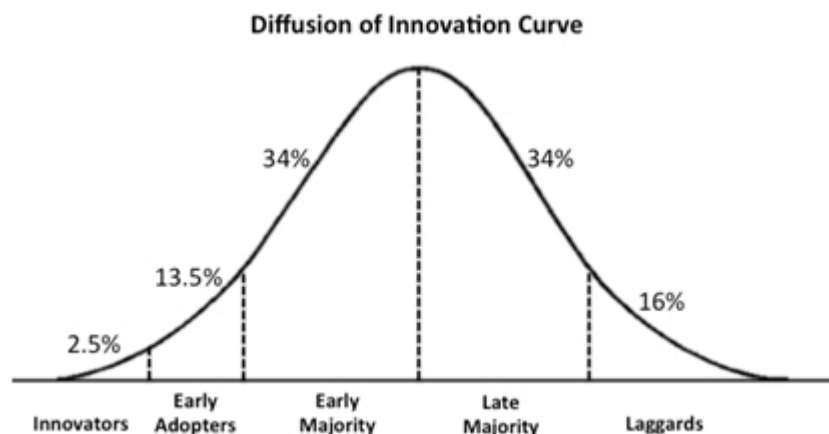


Figura 2.1: Categorie di adottanti ipotizzate da Rogers (fonte: Rogers, 2003)

2.2 Parentesi storica sui modelli di diffusione

La rappresentazione formale dei processi di diffusione si è storicamente avvalsa dei modelli epidemici presi a prestito dalla biologia, dietro l'ipotesi che un'innovazione si diffonda in un sistema sociale attraverso il meccanismo della comunicazione, in modo simile a come un'epidemia si diffonde attraverso il contagio fra persone. La prima modellazione di un processo di diffusione è collegata all'equazione logistica, formulata per la prima volta dal matematico belga P.F. Verhulst nel 1838 e usata originariamente nel campo delle scienze naturali per descrivere processi di crescita, come la diffusione di una malattia. Nel 1925 il biologo Pearl ha richiamato l'attenzione sul fatto che tutti i processi di crescita possono essere adeguatamente descritti attraverso questa equazione, che dà origine alla tipica curva a S. Devezas (2005) ha sottolineato il fatto che l'equazione logistica rappresenta uno degli strumenti previsivi tecnologici più potenti, quasi una "legge naturale" di diffusione dell'innovazione, grazie al suo successo nel rappresentare dinamiche di cambiamento all'interno di mercati e settori.

2.2.1 Dall'equazione logistica al modello di Bass

Il modello di Bass (1969), introdotto da Frank Bass, professore di marketing, rappresenta la più famosa ed utilizzata evoluzione dell'equazione logistica. Con questo modello, Bass ha dimostrato sia teoricamente sia empiricamente che le vendite cumulate di un nuovo prodotto descrivono per loro natura una curva a S. Come dichiarato nell'articolo "A New Product Growth for Model Consumer Durables" (Bass, 1969), il modello di Bass era stato concepito originariamente solo per beni durevoli; successivamente, è stato dimostrato che questo modello può essere applicato anche ai servizi.

Il modello di Bass e le sue estensioni sono una classe di modelli che ha un rilievo di spessore nel marketing strategico, da cui trae origine e ispirazione primaria. Tuttavia, questi modelli sono disponibili per l'utilizzo anche in contesti applicativi lontani da quello del marketing, come, ad esempio, la descrizione dei processi diffusivi di tipo epidemiologico. Tale separazione nei contesti applicativi è solo

apparente, poiché è possibile riconoscere un punto di contatto tra varie tematiche molto diverse tra loro proprio nel meccanismo fondamentale che opera con continuità all'interno dei sistemi socio-economici, biologici o fisici: il linguaggio interno o locale di comunicazione e controllo (Guseo, 2004).

Nei prossimi paragrafi, la trattazione dei modelli di diffusione viene effettuata con opportuno riferimento al contesto epidemiologico di interesse, ossia la diffusione di Covid-19 in Italia.

2.3 Modello di Bass

Il modello di Bass (BM) descrive il ciclo di vita di un'innovazione, caratterizzato da 4 fasi: introduzione, crescita, maturità e declino. Assumendo che l'innovazione di interesse sia il nuovo coronavirus SARS-CoV-2, le 4 fasi possono essere così interpretate:

1. **Introduzione:** il nuovo virus entra in un particolare sistema sociale (in questo caso, la popolazione italiana) e comincia a diffondersi in tale sistema.
2. **Crescita:** superata la fase iniziale, il virus cerca di diffondersi sempre più velocemente tra la popolazione.
3. **Maturità:** in questa fase viene raggiunto il numero massimo di nuovi contagiati per unità di tempo.
4. **Declino:** si registra un calo, più o meno lento, del numero di nuovi contagiati.

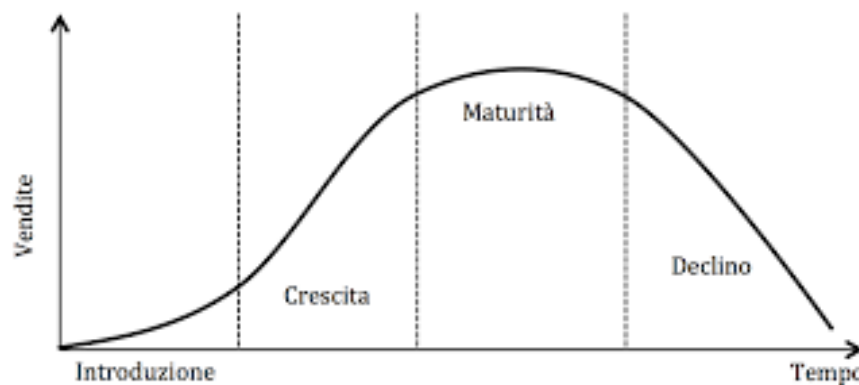


Figura 2.2: Ciclo di vita di un'innovazione (fonte: Kerin A. et al., 2010)

Dal punto di vista formale, il modello di Bass è definito da un'equazione differenziale del primo ordine:

$$z'(t) = \left(p + q \frac{z(t)}{m}\right)(m - z(t)) \quad (2.1)$$

Con riferimento alla diffusione di Covid-19, $z'(t)$ corrisponde al numero di nuovi contagiati al tempo t , mentre $z(t)$ rappresenta il numero totale di contagiati al tempo t . Il parametro m è relativo al “mercato potenziale” e, in questo caso, corrisponde al numero massimo di persone contagiate durante il ciclo di vita del virus, mentre il “mercato residuo”, $(m - z(t))$, rappresenta il numero di persone non ancora contagiate al tempo t e dipende chiaramente dal valore della stima di m .

Questo modello assume la presenza di due tipologie distinte di individui: gli innovatori e gli imitatori. Questa assunzione fa riferimento alla classificazione degli adottanti proposta da Rogers, con la differenza che le categorie dagli *early adopters* ai *laggards* sono state aggregate nella tipologia di individui definiti come imitatori. Il parametro associato alla componente di innovazione è p e fa riferimento a coloro che per primi sono stati infettati dal nuovo coronavirus, mentre il parametro associato alla componente di imitazione è q e fa riferimento a coloro che si sono ammalati di Covid-19 tramite il contagio da parte di soggetti infetti. Infine, la quantità $q \frac{z(t)}{m}$, definita come componente di contagio, descrive la dinamica del contagio. L'effetto del parametro q viene modulato dal rapporto $\frac{z(t)}{m}$, che rappresenta la pressione esercitata su chi ancora non è stato contagiato dal nuovo virus, tramite il contagio da parte di soggetti già infetti da Covid-19.

Ponendo $y(t) = \frac{z(t)}{m}$ si può riscrivere l'equazione 2.1 come:

$$y'(t) = (p + qy(t))(1 - y(t)) \quad (2.2)$$

Nell'equazione (2.2), $y(t)$ è una funzione di ripartizione e $y'(t)$ è la corrispondente densità. La ripartizione è riferita all'indice temporale t ed esprime, quindi, la probabilità che l'evento si sia verificato non oltre t . A partire dall'equazione (2.2), il modello di Bass può anche essere definito come un *hazard rate* particolare, ovvero:

$$\frac{y'(t)}{1-y(t)} = p + qy(t) \quad (2.3)$$

L'equazione (2.3) descrive la probabilità che un individuo sia contagiato da Covid-19 al tempo t , condizionatamente al fatto che non lo sia già stato. Tale probabilità corrisponde alla somma delle probabilità di due eventi indipendenti, p e $qy(t)$: questo comporta che non si può essere contemporaneamente innovatori e imitatori. La soluzione in forma chiusa del modello di Bass è una particolare distribuzione cumulata, definita come:

$$y(t) = \frac{1 - e^{-(p+q)t}}{1 + \frac{q}{p}e^{-(p+q)t}} \quad t > 0 \quad (2.4)$$

La proporzione di contagiati $y(t)$ fornita dall'equazione (2.4) descrive la dinamica del processo di diffusione, in funzione dei parametri p e q . È possibile passare alla rappresentazione in scala assoluta, ottenendo così il numero di contagiati $z(t)$, semplicemente moltiplicando l'equazione (2.4) per il parametro m , che agisce come parametro di scala del processo di diffusione.

$$z(t) = m \frac{1 - e^{-(p+q)t}}{1 + \frac{q}{p}e^{-(p+q)t}} \quad t > 0 \quad (2.5)$$

Uno dei grandi vantaggi del modello di Bass è la possibilità di spiegare la fase iniziale di diffusione, grazie alla presenza degli innovatori. Il modello di Bass, inoltre, è parsimonioso (è infatti caratterizzato da soli tre parametri: m , p e q), facilmente interpretabile e, per poter essere stimato, necessita dei soli dati cumulati. Questo modello, però, presenta una serie di limiti: i modelli presentati nei prossimi paragrafi sono estensioni del modello di Bass e cercano di risolvere alcuni di questi limiti.

2.4 Modello di Bass Generalizzato

Il modello di Bass Generalizzato (GBM), formalizzato nel 1994 da Bass, Krishnan e Jain, introduce una funzione di intervento $x(t) = x(t, \theta)$, $\theta \in \mathbb{R}^k$, assunta integrabile in domini limitati e non negativa.

$$z'(t) = \left(p + q \frac{z(t)}{m} \right) (m - z(t)) x(t) \quad (2.6)$$

Se $0 < x(t) < 1$ si assiste a un rallentamento del processo di diffusione, mentre un valore $x(t) > 1$ indica una sua velocizzazione. Il modello di Bass è un caso particolare del GBM, in cui $x(t)$ è uniformemente unitaria, ossia $x(t) = 1$: uno dei limiti del

modello di Bass è proprio legato al fatto che esso non tiene conto dell'effetto di variabili esogene che possono modificare la velocità del processo di diffusione. Nonostante il modello di Bass non presenti tale funzione di intervento, esso è ugualmente stimabile: la motivazione di ciò è stata chiarita nella pubblicazione "Why the Bass Model Fits without Decision Variables" (Bass, Krishnan e Jain, 1994). In questa pubblicazione viene affermato che la funzione $x(t)$ agisce sulla curva di diffusione, modificando la sua struttura temporale ma non il valore dei suoi parametri interni (m, p, q): di conseguenza, l'effetto importante di $x(t)$ è anticipare o ritardare le adozioni, ma non aumentarle o ridurle. In altre parole, la funzione $x(t)$ può rappresentare tutte quelle strategie applicate per controllare la tempistica di un processo di diffusione, ma non le sue dimensioni.

La soluzione in forma chiusa del GBM è definita come:

$$z(t) = m \frac{1 - e^{-(p+q) \int_0^t x(\tau) d\tau}}{1 + \frac{q}{p} e^{-(p+q) \int_0^t x(\tau) d\tau}} \quad t > 0 \quad (2.7)$$

che conferma la soluzione (2.5) nel caso particolare in cui $x(\tau) = 1$.

La funzione $x(t)$ può essere descritta in vari modi, a seconda delle necessità di modellazione. Nell'analisi effettuata, sono state utilizzate due particolari modellazioni di $x(t)$: l'impulso esponenziale e l'impulso rettangolare.

Una perturbazione forte e veloce può essere rappresentata mediante uno shock esponenziale, del tipo:

$$x(t) = 1 + c_1 e^{b_1(t-a_1)} I_{t \geq a_1} \quad (2.8)$$

Il parametro c_1 rappresenta l'intensità e il segno dello shock; b_1 descrive la persistenza nel tempo dell'effetto dello shock ed è generalmente negativo, ad indicare un ritorno in stazionarietà del sistema; a_1 indica il tempo di insorgenza dell'impulso. La funzione indicatrice I assume il valore 1 se l'evento indicato a deponente è verificato e 0 altrimenti.

Una perturbazione più stabile, che agisce sul processo di diffusione per un periodo relativamente lungo, può essere meglio descritta mediante uno shock rettangolare, del tipo:

$$x(t) = 1 + c_1 I_{t \geq a_1} I_{t \leq b_1} \quad (2.9)$$

Il parametro c_1 rappresenta ancora l'intensità e il segno dello shock, mentre i parametri a_1 e b_1 definiscono gli estremi di un intervallo chiuso entro il quale avviene l'impulso (con $a_1 < b_1$). La funzione indicatrice I è definita come in precedenza.

È interessante notare che la possibilità di definire una funzione flessibile $x(t)$ ha evidenziato una prospettiva molto più ampia nell'utilizzo del GBM, che può essere applicato come un efficiente strumento diagnostico per rilevare tutti i tipi di azioni esterne (le "variabili esogene" menzionate all'inizio del paragrafo) che influenzano un processo di diffusione.

2.5 Modello di Guseo-Guidolin

Una delle più importanti assunzioni del modello di Bass riguarda il mercato potenziale, che viene considerato costante per tutto il processo di diffusione. Questa ipotesi è tuttavia piuttosto limitante, in quanto la presenza di fattori endogeni o esogeni che agiscono sul processo di diffusione implica una struttura dinamica del mercato potenziale, e non statica, come invece assunto dal modello di Bass.

Il problema legato alla definizione formale di un mercato potenziale dinamico è stato affrontato in letteratura a partire dagli anni '70: alcune pubblicazioni hanno introdotto solo una modifica del mercato residuo ($m(t)-z(t)$), mentre altre hanno considerato una modifica del rapporto $\frac{z(t)}{m(t)}$. Nessuno di questi contributi, tuttavia, considera la struttura variabile del mercato potenziale come dipendente da un relativo processo di comunicazione. Con la pubblicazione "Modelling a dynamic market potential: A class of automata networks for diffusion of innovations" (Guseo e Guidolin, 2009), Guseo e Guidolin propongono un modello in cui il mercato potenziale è una funzione del tempo, $m(t)$, e presenta una struttura variabile. La variabilità del mercato potenziale è particolarmente evidente nella fase iniziale di diffusione, quando il successo di un'innovazione è ancora incerto e dipende da molti fattori.

Dal punto di vista teorico, il modello di Guseo-Guidolin, GGM, (Guseo e Guidolin, 2009) o con mercato potenziale variabile, si basa sul concetto di *absorptive capacity*,

elaborato da Cohen e Levinthal (1990). L'espressione "absorptive capacity" (letteralmente "capacità di assorbimento") si riferisce alla capacità di riconoscere il valore di una nuova informazione, di assimilarla e di utilizzarla. Gli autori sostengono che l'abilità di assimilare una novità e trarre vantaggio da essa sia associata ad una conoscenza precedente correlata: possedere un background di conoscenze pertinenti implica, dunque, una maggiore ricettività verso nuove idee. L'adozione di un'innovazione in un contesto sociale specifico può essere vista come una conferma dell'esistenza di una absorptive capacity: la capacità di assimilare e accettare una novità può essere infatti controllata semplicemente osservando il processo di acquisto. In questo senso, il mercato potenziale m può rappresentare una misura di questa capacità di assorbimento.

Per poter comprendere la capacità di assorbimento di un sistema sociale, è necessario studiare la sua struttura di comunicazione, con particolare attenzione al processo di trasferimento delle informazioni da un individuo ad un altro, visto che la capacità di assimilare un'innovazione dipende dall'accumulo di conoscenze pregresse. La struttura di comunicazione di un sistema sociale può essere vista come un insieme di collegamenti tra le unità del sistema; lo sviluppo di una conoscenza collettiva è così rappresentabile come un network in continua evoluzione in cui alcuni legami nascono, altri si sviluppano e altri ancora muoiono. Considerare il mercato potenziale $m(t)$ in funzione di questo processo di conoscenza implica la dipendenza del mercato potenziale da una rete di connessioni che cambiano nel tempo. Per quanto la teoria della diffusione dell'innovazione sia stata definita come una teoria della comunicazione, il modello di Bass non esplora in alcun modo l'effetto della comunicazione sulla fase di adozione. Questo modello è generalmente in grado di catturare il macro-comportamento delle diffusioni attraverso i suoi tre parametri (m, p, q) , ma non riesce ad analizzare le micro-interazioni sottostanti al processo di diffusione, la cui analisi è lasciata ad altri tipi di modelli.

Molte discipline scientifiche, come ad esempio la fisica, la biologia e l'ecologia, hanno sviluppato modelli per studiare l'evoluzione di sistemi complessi (come la comunicazione). Tra questi, gli automi cellulari si sono dimostrati uno strumento utile in campo economico e sociale. Un automa cellulare è costituito da un numero

finito di individui (o celle) che interagiscono in un determinato ambiente. Ogni cella può assumere uno stato particolare in base al suo stato nel periodo di tempo precedente e alle informazioni ricevute in seguito all'interazione con altre celle. L'evoluzione dello stato di una cella è controllata da una funzione chiamata "regola di transizione" che considera, in particolare, le interazioni tra celle. Il vantaggio offerto dai modelli basati sugli automi cellulari è la possibilità di osservare l'evoluzione di una certa struttura attraverso l'analisi di ogni singola interazione tra i suoi componenti; per questo motivo, gli automi cellulari si adattano bene all'analisi di cicli di vita. Il modello di Guseo-Guidolin si basa su una rete di automi cellulari (*cellular Automata Network*, NA) che descrive una rete di interazioni tra agenti di un sistema, che si scambiano informazioni riguardo a una particolare innovazione. In questo caso, l'unità di analisi per gli automi cellulari è rappresentata dalle relazioni interpersonali: si parla di "relazione standard" se due agenti diversi interagiscono tra di loro, mentre si parla di "relazione riflessiva" se un singolo agente opera l'auto-comunicazione. Lo stato di una relazione può essere attivo o inattivo: si suppone che l'attivazione possa avvenire attraverso la comunicazione istituzionale o attraverso un processo di passaparola che include sia la comunicazione interpersonale sia segnali dovuti a precedenti adozioni. Qualunque sia la fonte originale (comunicazione istituzionale o passaparola), il valore e l'affidabilità delle informazioni sono controllati dagli individui attraverso connessioni instaurate tra gli stessi: si parla di "processo di conferma", il quale rappresenta una fase necessaria affinché le informazioni diventino conoscenza acquisita, cioè consapevolezza e persuasione ad adottare. Tuttavia, può anche succedere che una relazione non rimanga attiva: questo accade o con un naturale e autonomo processo di decadimento o con un passaparola negativo, a causa di resistenze agli effetti dell'innovazione. Questa reazione è tipica di quelle situazioni in cui l'innovazione crea insoddisfazione ed è caratterizzata da performance inadeguate.

Oltre all'introduzione di un mercato potenziale variabile, l'altra grande novità introdotta da Guseo e Guidolin con il GGM è la modellazione della fase di comunicazione, che viene formalizzata separatamente rispetto alla fase di adozione ma che co-evolve nel tempo con quest'ultima.

Nel caso in cui gli effetti di comunicazione siano persistenti, ossia non si verificano effetti di decadimento né passaparola negativi, allora il GGM definisce il mercato potenziale come:

$$m(t) = K \sqrt{\frac{1 - e^{-(p_c + q_c)t}}{1 + \frac{q_c}{p_c} e^{-(p_c + q_c)t}}} \quad (2.10)$$

dove i parametri p_c e q_c rappresentano rispettivamente la componente innovativa e imitativa del processo di comunicazione, mentre K è il parametro di scala del processo ed è tale per cui $\lim_{t \rightarrow \infty} m(t) = K$, situazione in cui tutti coloro che sono stati informati diventano potenziali acquirenti. L'equazione (2.10) conferma il ruolo fondamentale assunto dalla fase di comunicazione nella creazione, nel tempo, del mercato potenziale.

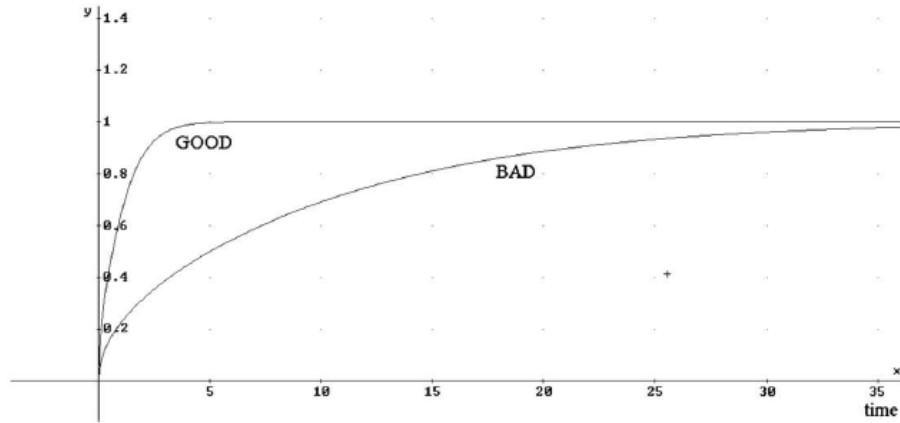


Figura 2.3: Due mercati potenziali normalizzati ($K = 1$) dipendenti dal tempo (x).

Comunicazione buona (*Good*): $p_c = 0,15$, $q_c = 0,90$.

Comunicazione insoddisfacente (*Bad*): $p_c = 0,01$, $q_c = 0,06$.

(fonte: Guseo e Guidolin, 2009)

Dalla figura 2.3 si osserva che, per valori elevati dei parametri p_c e q_c , il mercato potenziale tende asintoticamente a K molto velocemente; al contrario, valori bassi di questi parametri influenzano negativamente il mercato potenziale, che raggiunge il suo livello asintotico molto più lentamente.

Posto il vincolo $z(t) = 0$ per $t \leq 0$, il GGM assume la forma di un'equazione differenziale di primo grado:

$$z'(t) = m(t) \left\{ -r_s \frac{z(t)}{m(t)} + \left(p_s + q_s \frac{z(t)}{m(t)} \right) \left(1 - \frac{z(t)}{m(t)} \right) \right\} x(t) + z(t) \frac{m'(t)}{m(t)} \quad (2.11)$$

dove $z'(t)$ e $z(t)$ hanno lo stesso significato presentato in precedenza (si veda §2.3), p_s e q_s rappresentano la componente innovativa e imitativa del processo di adozione e sono analoghi ai parametri di innovazione e imitazione del modello di Bass, mentre r_s spiega la presenza di un possibile effetto di decadimento. La funzione $m(t)$ rappresenta il mercato potenziale dinamico mentre $x(t)$ è la funzione di intervento introdotta con il modello di Bass Generalizzato (si veda §2.4). Si noti che la quantità $z(t) \frac{m'(t)}{m(t)}$ può aumentare o diminuire a seconda del segno di $m'(t)$.

Se si considera il modello nella sua forma semplificata, ossia senza effetto di decadimento ($r_s = 0$) e senza l'intervento di azioni esogene ($x(t) = 1 \forall t$), la soluzione in forma chiusa dell'equazione (2.11) è:

$$z(t) = K \sqrt{\frac{1 - e^{-(p_c + q_c)t}}{1 + \frac{q_c}{p_c} e^{-(p_c + q_c)t}} \frac{1 - e^{-(p_s + q_s)t}}{1 + \frac{q_s}{p_s} e^{-(p_s + q_s)t}}} \quad (2.12)$$

Questo modello definisce chiaramente come la diffusione di un'innovazione sia il risultato della sinergia tra due forze: la comunicazione e l'adozione. Relativamente al contesto di interesse, le due fasi possono essere così interpretate: la fase di comunicazione rappresenta la fase di contagio, ossia quella fase in cui un individuo viene infettato da un soggetto positivo e sta sviluppando la malattia. La fase di adozione, invece, può essere interpretata come il momento in cui il soggetto, dopo essere stato sottoposto al tampone, riceve una diagnosi positiva. In questo senso, il lockdown attuato in Italia a partire dal 9 marzo è servito a ridurre la stima del parametro m , ossia il numero totale di contagiati da Covid-19 durante il ciclo di vita del nuovo coronavirus; l'utilizzo della mascherina anche successivamente al termine del lockdown (4 maggio) è servito, invece, a minimizzare la *fase di contagio* tra soggetti, dal momento che il virus si trasmette attraverso i *droplets*.

Il GGM assume che la fase di comunicazione non sia osservabile: non si hanno informazioni su come gli individui comunicano tra loro e, con questo modello, non si è interessati a determinare forme particolari assunte dal network di connessioni che si creano tra individui. Ciò che interessa è, invece, il modo in cui questa fase latente di comunicazione influisce concretamente sul mercato potenziale nel tempo. Il modello di Bass è un caso particolare del GGM: se la componente di

comunicazione assume valore pari a 1, allora si ritorna all'equazione (2.5). La complessità del GGM, data dall'introduzione di un mercato potenziale dinamico e dalla co-presenza di due fasi (comunicazione e adozione) spesso porta a risolvere problematiche tipiche del modello di Bass, come la sovrastima nella prima parte del ciclo di vita e la sottostima delle ultime osservazioni.

2.5.1 Posizionamento temporale delle fasi di contagio e di diagnosi di positività

Guseo e Guidolin hanno dimostrato con la pubblicazione "Market potential dynamics in innovation diffusion: Modelling the synergy between two driving forces" (Guseo e Guidolin, 2011) che è possibile fornire un ordinamento alle fasi di comunicazione e adozione. In questo paper viene, infatti, affermato che non necessariamente, a differenza di quanto si possa pensare, la comunicazione precede la fase di adozione. Il modello con potenziale variabile può essere riscritto nella seguente forma:

$$z(t) = K W(t; p_c, q_c, p_s, q_s) = K \sqrt{F(t; p_c, q_c)} G(t; p_s, q_s) \quad (2.13)$$

dove $W(t; p_c, q_c, p_s, q_s)$ è una funzione di ripartizione, prodotto di due ripartizioni, $\sqrt{F(t; p_c, q_c)}$ e $G(t; p_s, q_s)$, che descrivono rispettivamente i processi di contagio e di diagnosi di positività. A partire dall'equazione (2.13), il numero di nuovi contagiati, $z'(t)$, è dato da:

$$z'(t) = K W'(t; p_c, q_c, p_s, q_s) = K w(t; p_c, q_c, p_s, q_s) \quad (2.14)$$

ovvero, in forma esplicita:

$$w(t) = w(t; p_c, q_c, p_s, q_s) = \frac{1}{2} F(t)^{-\frac{1}{2}} G(t) f(t) + \sqrt{F(t)} g(t) \quad (2.15)$$

con $f(t) = F'(t)$ e $g(t) = G'(t)$, dove $f(t)$ e $g(t)$ rappresentano rispettivamente la densità della fase di contagio e di diagnosi di positività. L'equazione (2.15) può essere riscritta nel seguente modo:

$$w(t) = \frac{1}{2} F(t)^{-\frac{1}{2}} G(t) f(t) + \sqrt{F(t)} g(t) = k_1(t) + k_2(t) \quad (2.16)$$

per evidenziare la separazione del processo di diffusione in due fasi $k_1(t)$ e $k_2(t)$, trasformazioni monotone di $f(t)$ e $g(t)$ rispettivamente.

Per poter posizionare nel tempo le due densità $f(t)$ e $g(t)$, è necessario definire due variabili casuali, T_F e T_G , con funzioni di ripartizione $F(t)$ e $G(t)$. A questo punto, si possono definire alcuni indici di posizione relativi alle variabili casuali T_F e T_G , come la moda, t^+ , la mediana, $t^{0.5}$, e la media, $\bar{t}$:

$$t_F^+ = \frac{\ln(q_c/p_c)}{p_c + q_c} \quad (2.17)$$

$$t_G^+ = \frac{\ln(q_s/p_s)}{p_s + q_s} \quad (2.18)$$

$$t_F^{0.5} = \frac{1}{p_c + q_c} \ln \left(2 + \frac{q_c}{p_c} \right) \quad (2.19)$$

$$t_G^{0.5} = \frac{1}{p_s + q_s} \ln \left(2 + \frac{q_s}{p_s} \right) \quad (2.20)$$

$$E(T_F) = \bar{t}_F = \frac{1}{q_c} \ln \left(1 + \frac{q_c}{p_c} \right) \quad (2.21)$$

$$E(T_G) = \bar{t}_G = \frac{1}{q_s} \ln \left(1 + \frac{q_s}{p_s} \right) \quad (2.22)$$

Se $0 < p_c < q_c$ e $0 < p_s < q_s$, gli indici di posizione appena presentati hanno un ordine crescente: $t^+ < t^{0.5} < \bar{t}$. Questo ordinamento tra indici di posizione può far pensare all'esistenza di un analogo ordinamento stocastico tra $k_1(t)$ e $k_2(t)$: è infatti possibile ottenere informazioni sul collocamento temporale di $k_1(t)$ e $k_2(t)$ analizzando il confronto a coppie tra i corrispondenti indici di posizione, riferiti a $f(t)$ e a $g(t)$. In particolare, si osserva che, se $t_F^+ < t_G^+$, $t_F^{0.5} < t_G^{0.5}$ e $\bar{t}_F < \bar{t}_G$, allora si è nel caso più tipico, ossia quello in cui, con riferimento al Covid-19, il contagio precede la diagnosi di positività, cioè il contagio è il fattore che rende possibile l'inizio della diffusione del virus. Viceversa, se $t_F^+ > t_G^+$, $t_F^{0.5} > t_G^{0.5}$ e $\bar{t}_F > \bar{t}_G$, allora si è nella situazione inversa e meno comune, in cui la diagnosi di positività è ciò che rende manifesta la diffusione del virus e il contagio mantiene il trend nel tempo.

2.6 Metodi di stima e identificazione statistica: regressione non lineare

I modelli presentati nei paragrafi precedenti sono di tipo non lineare. Un modello non lineare, nei parametri, è rappresentabile come somma di due componenti:

$$Z(t) = f(\beta, t) + \varepsilon(t) \quad (2.23)$$

dove $Z(t)$ è la risposta del sistema dipendente dal tempo t (nel caso dei modelli di diffusione, $Z(t)$ corrisponde alla serie di dati cumulati), $f(\beta, t)$ è la componente deterministica, funzione reale nota dell'incognito parametro multiplo $\beta \in \mathbb{R}^k$ e del tempo t . La seconda componente, $\varepsilon(t)$, è definita come un processo stocastico che rappresenta il disturbo residuale. Le assunzioni sulla struttura di $\varepsilon(t)$ possono essere graduate in modo gerarchico. Una prima assunzione forte è collegata alla corretta identificabilità media del modello per cui si richiede che gli errori siano di media nulla, ossia $E(\varepsilon(t)) = 0$. Questa prima caratterizzazione è sufficiente per l'introduzione del criterio di stima robusto più tradizionale: il criterio dei minimi quadrati. Le assunzioni sul processo $\varepsilon(t)$ possono essere estese comprendendo il vincolo di varianza omoschedastica, $\text{Var}(\varepsilon(t)) = \sigma^2$, e di incorrelazione tra disturbi distinti, $\sigma_{\varepsilon(t), \varepsilon(t')} = 0$, $t \neq t'$. Queste tre condizioni congiunte su $\varepsilon(t)$ caratterizzano il processo cosiddetto *white noise*.

Si consideri ora un modello del tipo:

$$y_i = f(x_i; \vartheta^*) + \varepsilon_i, \quad i = 1, 2, \dots, n \quad (2.24)$$

con $E(\varepsilon_i) = 0$, x_i vettore di $k \times 1$ variabili esplicative, y_i la risposta del sistema e ϑ^* l'incognito vero parametro ϑ . Si indica con $\hat{\vartheta}$ una stima secondo i minimi quadrati dell'ignoto parametro ϑ^* ottenuta minimizzando la devianza $S(\vartheta)$.

$$S(\vartheta) = \sum_{i=1}^n (y_i - f(x_i; \vartheta))^2 \quad (2.25)$$

A differenza del caso lineare, $S(\vartheta)$ può avere vari minimi locali, oltre che al minimo assoluto $\hat{\vartheta}$. Quando ogni $f(x_i; \vartheta)$ è differenziabile in ϑ , allora $\hat{\vartheta}$ soddisfa il seguente sistema di equazioni normali:

$$\frac{\partial S(\vartheta)}{\partial \vartheta_r} \Big|_{\hat{\vartheta}} = 0, \quad r = 1, 2, \dots, p \quad (2.26)$$

Si supponga, per comodità, di porre $f_i(\vartheta) = f(x_i; \vartheta)$ e si indichi con $f(\vartheta)$ il vettore $(f_1(\vartheta), f_2(\vartheta), \dots, f_n(\vartheta))'$. Sia:

$$F.(\vartheta) = \frac{\partial f(\vartheta)}{\partial \vartheta'} = \left[\frac{\partial f_i(\vartheta)}{\partial \vartheta_r} \right], \quad i = 1, \dots, n, \quad r = 1, \dots, p \quad (2.27)$$

la matrice jacobiana $n \times p$ delle derivate parziali prime di f rispetto a ϑ' . Talvolta, per semplicità, si pone $F. = F.(\vartheta)$. Passando alla notazione vettoriale, la devianza può essere riscritta come:

$$S(\vartheta) = [y - f(\vartheta)]' [y - f(\vartheta)] \quad (2.28)$$

Derivando la (2.28) rispetto a ϑ ed eguagliando a zero si hanno le usuali equazioni normali:

$$-2F.'(y - f(\vartheta)) = 0 \quad (2.29)$$

ovvero

$$F.'(y - f(\vartheta)) = 0 \quad (2.30)$$

cui corrisponde una soluzione $\hat{\vartheta}$.

La soluzione per via analitica di queste equazioni non lineari può essere difficoltosa; per questo motivo si utilizzano procedure iterative, in cui la sequenza viene troncata dopo un numero finito di iterazioni, N , e si accetta ϑ_N come approssimazione di ϑ . Un esempio di metodo di stima che si basa sulla metodica appena descritta è il metodo di Gauss-Newton, che considera inizialmente un vettore di parametri ϑ_a , ritenuto una buona approssimazione di ϑ . Tramite lo sviluppo di Taylor troncato al primo ordine del vettore $f(\vartheta)$, in un intorno del punto ϑ_a , si giunge ad una possibile stima di ϑ :

$$f(\vartheta) \approx f(\vartheta_a) + F_{.a}(\vartheta - \vartheta_a) \quad (2.31)$$

La devianza si approssima in termini lineari:

$$\begin{aligned} S(\vartheta) &= [y - f(\vartheta)]' [y - f(\vartheta)] \\ &\approx [y - f(\vartheta_a) - F_{.a}(\vartheta - \vartheta_a)]' [y - f(\vartheta_a) - F_{.a}(\vartheta - \vartheta_a)] \\ &= f[z - F_{.a}\beta]' [z - F_{.a}\beta] \end{aligned} \quad (2.32)$$

dove $z = y - f(\vartheta_a)$ e $\beta = (\vartheta - \vartheta_a)$.

Il minimo in $\hat{\beta}$ si consegue per $\hat{\beta} = (F_{.a}' F_{.a})^{-1} F_{.a}' z$ e quindi:

$$\hat{\beta} = (\vartheta_b - \vartheta_a) = (F_{.a}' F_{.a})^{-1} F_{.a}' (y - f(\vartheta_a)) = \delta_a \quad (2.33)$$

In altri termini, la seconda approssimazione è del tipo:

$$\vartheta_b = \vartheta_a + \delta_a \quad (2.34)$$

Un altro metodo numerico di stima, che è quello utilizzato nell'analisi presentata in questa relazione, è il metodo di Levenberg-Marquardt. Questo metodo introduce una sostanziale modifica all'algoritmo di Gauss-Newton, eliminando eventuali fonti di singolarità dovute alla matrice $F_{.a}' F_{.a}$: se questa matrice non è singolare, non

è invertibile e quindi la soluzione non può essere calcolata. Il nuovo passo di aggiornamento introduce un'opportuna matrice diagonale a rango pieno:

$$\delta_a = (F'_{.a} F_{.a} + \eta_a D_a)^{-1} F'_{.a} (y - f(\vartheta_a)) \quad (2.35)$$

dove D_a indica una matrice non singolare e η_a viene modificato in base al valore della devianza $S(\vartheta)$: se quest'ultima aumenta, allora anche il valore di η_a aumenta, mentre se $S(\vartheta)$ diminuisce, η_a diminuisce a sua volta.

2.7 Adattamento del modello ai dati e confronto tra modelli annidati

Un indicatore di performance usuale per questa classe di modelli è l'indice di determinazione, R^2 , calcolato come:

$$R^2 = \frac{SST - SSE}{SST} = \frac{\sum_{i=1}^n (y_i - \bar{y})^2 - \sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (2.36)$$

dove SST rappresenta la devianza totale, SSE la devianza residua e $\hat{y}_i$, $i = 1, \dots, n$ sono i valori previsti secondo il modello prescelto. Il supporto dell'indice R^2 è compreso tra 0 e 1; nel caso dei modelli di diffusione, questo indice assume sempre valori particolarmente elevati, tipicamente superiori a 0,95. Ciò è riconducibile al fatto che questi modelli sono caratterizzati da una forma a S e vengono adattati a dati che, per loro natura, hanno questa forma; l'andamento dei dati viene dunque colto molto bene e questo porta, di conseguenza, ad una devianza residua quasi irrilevante, se confrontata con la devianza totale.

A partire dall'indice di determinazione, è possibile valutare la performance di un modello esteso, m_2 , confrontato con uno nidificato, m_1 : a tale scopo, si può utilizzare il rapporto di correlazione multipla parziale al quadrato, $\tilde{R}^2$:

$$\tilde{R}^2 = \frac{SSE_{m_1} - SSE_{m_2}}{SSE_{m_1}} = \frac{R_{m_2}^2 - R_{m_1}^2}{1 - R_{m_1}^2} \quad (2.37)$$

dove $R_{m_i}^2$, $i = 1, 2$ è l'indice di determinazione del modello m_i . Anche $\tilde{R}^2$, così come R^2 , è definito da 0 a 1. L'indice $\tilde{R}^2$ ha il suo diretto corrispettivo nel rapporto F, definito come:

$$F = \frac{\tilde{R}^2(n-v)}{(1-\tilde{R}^2)u} \quad (2.38)$$

dove n è il numero di osservazioni, v è il numero di parametri nel modello esteso m_2 e u è l'incremento di parametri che si ha passando dal modello ridotto m_1 al modello esteso m_2 . Un valore critico usuale del rapporto F è 4: ciò significa che valori del rapporto F superiori a 4 portano a concludere che il miglioramento apportato dal modello esteso è statisticamente significativo.

2.8 Analisi dei residui e statistica di Durbin-Watson

Se si applica il criterio di stima dei minimi quadrati, seguendo un modello non lineare (2.23), è cruciale procedere con l'analisi dei residui: essa, infatti, permette di verificare l'adeguatezza del modello scelto per stimare le componenti deterministiche e, quindi, di giustificare l'uso di tale modello a fini previsivi. Se il modello risulta correttamente specificato, allora i residui sono generati da un processo *white noise*, il quale assume che tali residui siano a media nulla, omoschedastici e incorrelati tra loro. Per valutare la struttura dei residui, uno dei grafici più utilizzati è il diagramma di dispersione, che riporta in ordinata i residui, solitamente normalizzati o studentizzati, e in ascissa i valori stimati della variabile dipendente. Se il modello stimato è soddisfacente, i residui si distribuiscono in una nuvola di punti disposta attorno alla media nulla.

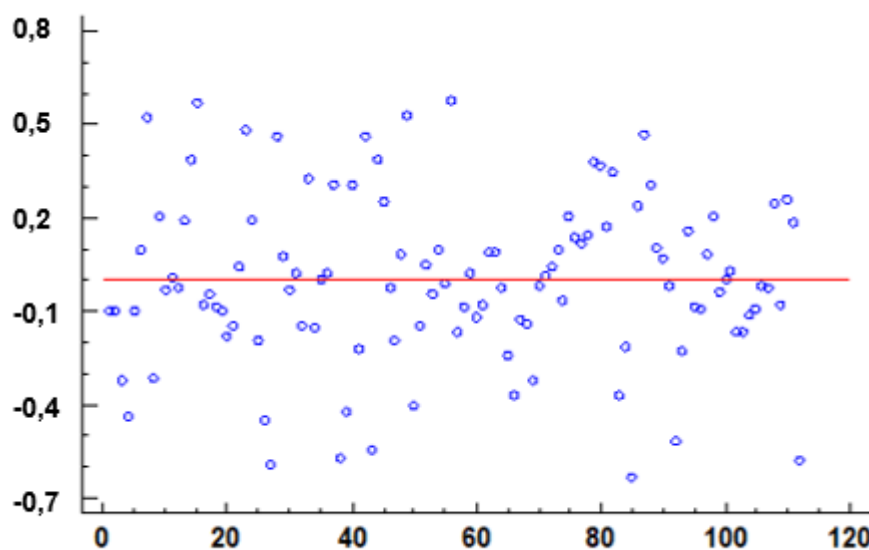


Figura 2.4: Diagramma di dispersione di residui incorrelati

Un altro grafico utilizzato per valutare la bontà dei residui è il correlogramma, che rappresenta graficamente la funzione di autocorrelazione (ACF) dei residui: se le assunzioni fatte sui residui vengono rispettate, allora tutti i coefficienti di autocorrelazione stimati sono interni alle bande di confidenza, che presentano un intervallo del tipo $\left[-\frac{z}{\sqrt{n}}, \frac{z}{\sqrt{n}}\right]$. Nella figura 2.5, tale intervallo è delimitato dalle linee blu: si osserva che tutte le autocorrelazioni sono interne alle bande di confidenza, confermando l'incorrelazione dei residui. Il valore z è solitamente posto pari a 1,96, ossia il valore corrispondente al 95° percentile della normale standard. Se, invece, alcune autocorrelazioni stimate risultano esterne all'intervallo di confidenza, l'ipotesi di successione generata da un *white noise* viene respinta, confermando così la presenza di autocorrelazione dei residui.

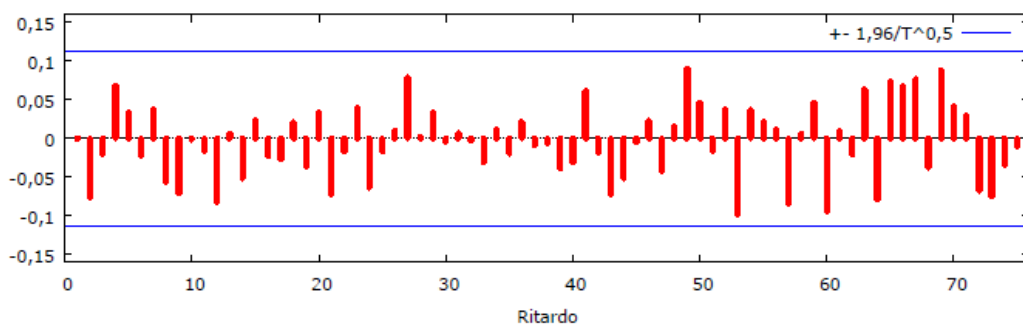


Figura 2.5: Correlogramma di residui incorrelati

Soprattutto nell'applicazione a dati reali dei modelli di diffusione, si è osservato che spesso i residui sono caratterizzati da autocorrelazione positiva: quello che si osserva graficamente in questi casi è la presenza di funzioni armoniche marcate intorno alla media nulla.

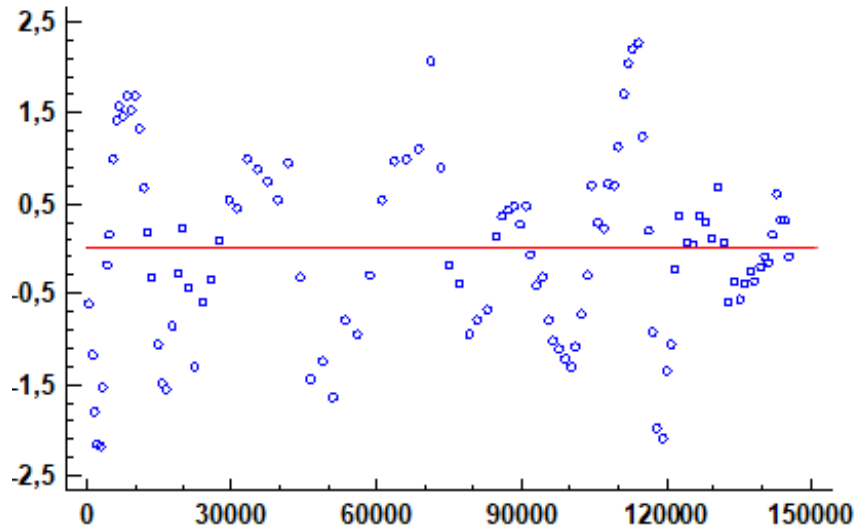


Figura 2.6: Diagramma di dispersione di residui autocorrelati positivamente

Ci sono varie forme di autocorrelazione: la più tipica è l'autocorrelazione di primo ordine, in cui il termine di errore è assunto dipendente dall'errore al tempo precedente:

$$\varepsilon_t = \rho \varepsilon_{t-1} + v_t \quad (2.39)$$

dove ρ è la correlazione tra errori consecutivi e v_t è un termine d'errore a media nulla, varianza costante σ_v^2 e indipendente dal tempo. Un test molto utilizzato per verificare la presenza di autocorrelazione di primo ordine è il test di Durbin-Watson (1950), che si basa sul seguente sistema di ipotesi:

$$\begin{cases} H_0: \rho = 0 \\ H_1: \rho > 0 \end{cases} \quad (2.40)$$

La scelta dell'ipotesi alternativa $\rho > 0$ è legata al fatto che l'autocorrelazione positiva è molto frequente nello studio dei processi di diffusione; se ritenuto di particolare interesse, è possibile considerare un'ipotesi alternativa bilaterale o unilaterale sinistra. La statistica test è definita come:

$$DW = \frac{\sum_{t=2}^T (e_t - e_{t-1})^2}{\sum_{t=1}^T e_t^2} \quad (2.41)$$

dove e_t corrispondono ai residui di regressione. Per n grande, vale la seguente relazione tra la statistica DW e $\hat{\rho}$, stima dell'autocorrelazione dei residui al lag 1:

$$DW \approx 2(1 - \hat{\rho}) \quad (2.42)$$

Dalla relazione (2.42), si ricava facilmente che $0 < DW < 4$.

Durbin e Watson sono riusciti a tabulare i limiti inferiori e superiori per i valori critici di DW, definiti rispettivamente d_L e d_U , che dipendono dalla numerosità campionaria e dal numero di variabili; per recuperare tali valori, è necessario disporre di apposite tavole. La figura 2.7 descrive come interpretare i risultati ottenuti con il test di Durbin-Watson: con riferimento al sistema di ipotesi (2.40), se $DW < d_L$, allora si rifiuta H_0 e, quindi, l'autocorrelazione positiva dei residui è statisticamente significativa. Si è impossibilitati a rifiutare H_0 quando il valore di DW è compreso tra d_U e $4-d_U$: in questo caso, si conclude che non c'è autocorrelazione tra residui. Se, invece, $d_L < DW < d_U$ o $4-d_U < DW < 4-d_L$, allora il test è inconclusivo e non si possono trarre conclusioni relativamente all'autocorrelazione dei residui. In generale, un valore di DW prossimo a 2 implica l'assenza di autocorrelazione dei residui, un valore vicino allo 0 è un indicatore di autocorrelazione positiva, mentre un valore prossimo a 4 indica la presenza di autocorrelazione negativa.

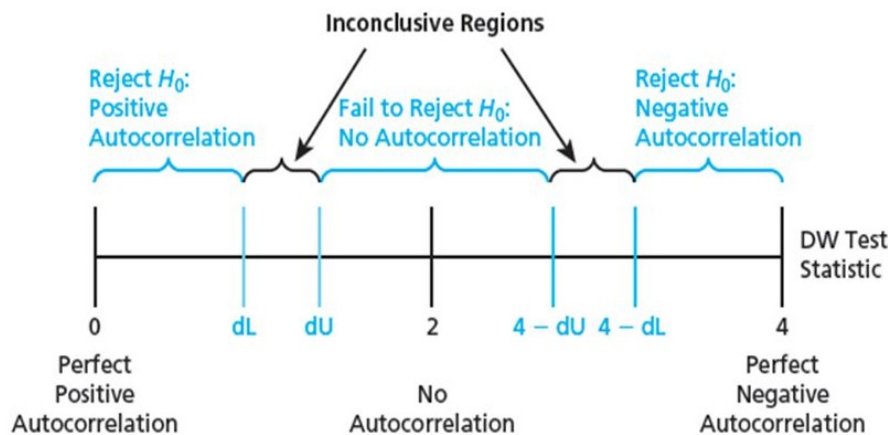


Figura 2.7: Interpretazione della statistica di Durbin-Watson

2.9 Affinamento ARMAx

Una volta diagnosticata l'autocorrelazione dei residui, questa deve essere rimossa o modellata: si tratta quindi di articolare meglio la rappresentazione formale di $\varepsilon(t)$. Una possibilità è offerta da Guseo (2004), che propone l'utilizzo di modelli ARMAx: questi modelli sono un'estensione dei modelli ARMA in quanto controllano la struttura evolutiva dell'ARMA con l'introduzione nel modello di una o più variabili

esogene (la “ x ” di ARMA x). Questo accade per il semplice motivo che le variabili esterne contengono informazioni aggiuntive rispetto al passato della serie, in grado di contribuire nella spiegazione della variabilità del modello. Nel caso in cui la serie storica sia caratterizzata da una componente stagionale, si possono utilizzare i modelli SARMA x , che estendono i modelli ARMA x includendo la stagionalità.

Si definisce un processo ARMA(p,q) come segue:

$$\phi_p(B)Y_t = \theta_q(B)a_t \quad (2.43)$$

dove a_t è un processo *white noise* e

$$\phi_p(B) = 1 - \phi_1 B - \dots - \phi_p B^p \quad (2.44)$$

$$\theta_q(B) = 1 - \theta_1 B - \dots - \theta_q B^q \quad (2.45)$$

Si definisce, per completezza, anche un processo SARMA(p,q) $\times$ (P,Q) $_s$:

$$\Phi_P(B^s)\phi_p(B)Y_t = \Theta_Q(B^s)\theta_q(B)a_t \quad (2.46)$$

dove a_t è un *white noise*, $\phi_p(B)$ e $\theta_q(B)$ assumono il significato presentato in (2.44) e (2.45) rispettivamente e

$$\Phi_P(B^s) = 1 - \Phi_1 B^s - \dots - \Phi_P B^{Ps} \quad (2.47)$$

$$\Theta_Q(B^s) = 1 - \Theta_1 B^s - \dots - \Theta_Q B^{Qs} \quad (2.48)$$

I modelli ARMA (o SARMA) possono essere utilizzati solo con serie stazionarie. Per serie non stazionarie, è possibile ricorrere a modelli che prevedono delle differenziazioni, come i modelli ARIMA(p,d,q) o SARIMA(p,d,q) $\times$ (P,D,Q). Nell’ambito del ciclo di vita di prodotto, le serie sono difficilmente stazionarizzabili tramite differenziazioni: il trend non lineare viene modellato tramite l’applicazione dei modelli di diffusione e non con procedure di Box-Jenkins come quelle appena descritte.

L’affinamento ARMA x viene definito nel seguente modo:

$$\phi(B)[w_t - c \eta(\hat{\beta}, t)] = \theta(B)a_t \quad (2.49)$$

dove $w(t)$ è la serie osservata, $\eta(\hat{\beta}, t)$ è la serie prevista secondo il modello prescelto, a_t è un processo WN, $\phi(B)$ e $\theta(B)$ sono gli usuali polinomi di ordine p e q rispettivamente. Il parametro di calibrazione c consente di verificare l’appropriatezza del modello globale $\eta(\beta, t)$ basata su una soluzione dei minimi quadrati $\hat{\beta}$. In particolare, se $c = 1$, allora il modello diventa:

$$\phi(B)\varepsilon_t = \theta(B)a_t \quad (2.50)$$

ossia si stanno modellando direttamente i residui del modello prescelto.

Tale approccio può essere esteso includendo una componente stagionale; in questo caso, l'affinamento SARMAx viene così formalizzato:

$$\Phi(B^s)\phi(B)[w_t - c \eta(\hat{\beta}, t)] = \Theta(B^s)\theta(B)a_t \quad (2.51)$$

Questa tecnica è definita a due stadi, in quanto viene applicato, in un primo momento, un modello di diffusione a una certa serie storica e, in un secondo momento, un affinamento ARMAx (o SARMAx) ai residui del modello. Il vantaggio apportato dall'utilizzo di questo affinamento è duplice: la serie dei residui viene resa WN, correggendo così l'autocorrelazione diagnosticata per i residui originari, e le oscillazioni della serie intorno al modello stimato vengono ora adeguatamente descritte. Questo approccio si rivela generalmente molto efficace e parsimonioso e permette lo svolgimento di un'analisi ancora più precisa ed accurata.

Capitolo 3

Analisi statistica della diffusione di Covid-19 in Italia

3.1 Presentazione dei dati utilizzati nell'analisi

In questo capitolo viene studiata la serie storica giornaliera dei nuovi contagiati da Covid-19 in Italia dal 21 febbraio 2020, giorno in cui sono stati registrati i primi casi di contagio nel territorio italiano, al 10 giugno 2020.

Il software statistico utilizzato per l'analisi è *Statgraphics*, mentre i modelli di diffusione applicati sono quelli illustrati nel capitolo precedente (§§2.3-5).

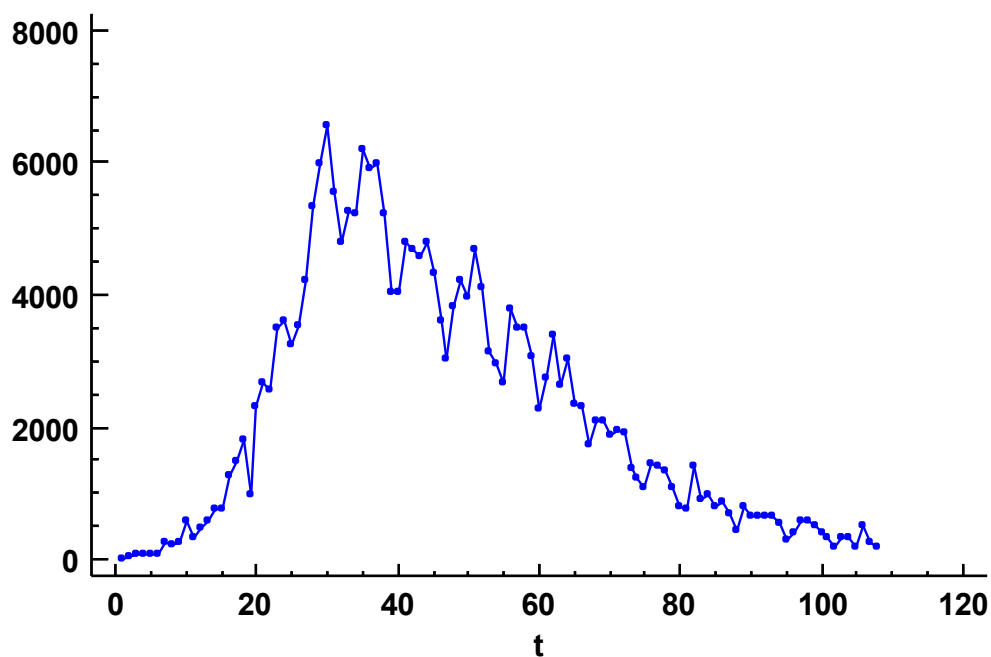


Figura 3.1: Serie giornaliera dei nuovi contagiati da Covid-19 in Italia

Come si può osservare dalla figura 3.1, il numero di nuovi contagiati è caratterizzato da un aumento progressivo nei primi giorni, che raggiunge un ritmo di crescita esponenziale dopo circa due/tre settimane. Il picco di nuovi contagiati è stato raggiunto il 21 marzo (6557 nuovi casi) e, a partire da questa data, ha inizio una discesa molto lenta; solo il 10 maggio si arriva a meno di 1000 nuovi casi. Da notare

i numerosi picchi verso l'alto e verso il basso nella parte discendente della serie, che fanno pensare alla presenza, nella serie, di una componente stagionale.

Prima di entrare nel dettaglio dell'analisi effettuata, è bene soffermarsi sulla qualità dei dati a disposizione utilizzati per l'analisi. Tali dati corrispondono ai nuovi casi giornalieri, ma essi mostrano un'immagine non del tutto corretta della diffusione del nuovo coronavirus, per vari motivi. Innanzitutto, c'è un periodo di incubazione che intercorre tra il contagio e la manifestazione dei primi sintomi; inoltre, una volta manifestati i primi sintomi ed effettuato il tampone, bisogna attendere l'esito di tale tampone, che, soprattutto nel periodo di piena emergenza (mesi di marzo e aprile), arriva anche diversi giorni dopo. In aggiunta a tutto ciò, per certi giorni mancano i dati relativi ad alcune regioni e/o province o tali dati vengono comunicati in ritardo e quindi registrati in giorni successivi. Soprattutto in una situazione di piena emergenza sanitaria come questa, non ci si può aspettare che la qualità dei dati raccolti e messi a disposizione sia ottima: è importante, però, essere consapevoli fin da subito delle caratteristiche dei dati che si vanno ad analizzare e dei limiti legati alla raccolta di tali dati.

3.2 Applicazione del modello di Bass

La serie presenta un andamento pressoché campanulare, tipico di processi di diffusione con un arco temporale limitato: questo giustifica l'utilizzo del modello di Bass, nella sua versione standard. Si è partiti dai seguenti valori iniziali:

Stime iniziali	
m	240000
p	0,001
q	0,1

Tabella 3.1: Valori iniziali dei parametri del modello di Bass

I risultati ottenuti sono i seguenti:

Parametro	Stima	Errore standard	IC inferiore	IC superiore
m	230514	1111,3	228311	232718
p	0,00209689	0,000125488	0,00184807	0,00234571
q	0,0838803	0,00203503	0,0798452	0,0879154

Tabella 3.2: Stime dei parametri del modello di Bass (livello di confidenza = 95%)

Sorgente	Somma dei quadrati	G.I.	Media dei quadrati
<i>Modello</i>	$2,84256e^{12}$	3	$9,47521e^{11}$
<i>Residuo</i>	$2,96355e^9$	105	$2,82243e^7$
<i>Totale</i>	$2,84553e^{12}$	108	
<i>Totale (Corr.)</i>	$8,45806e^{11}$	107	

Tabella 3.3: Analisi della varianza del modello di Bass

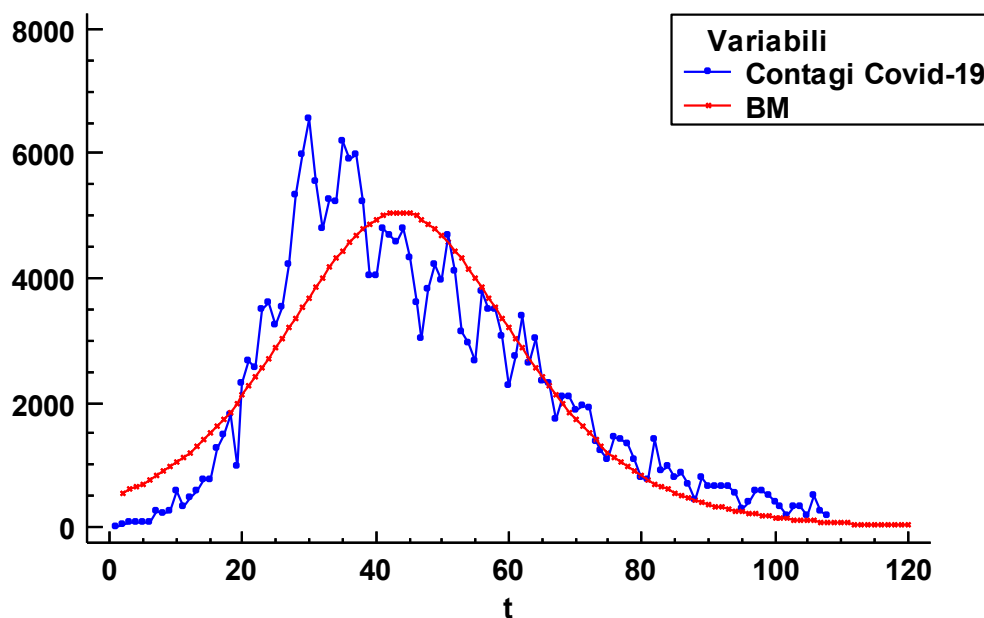


Figura 3.2: Serie giornaliera con la previsione del modello di Bass

Guardando la tabella 3.2, si nota che la stima del parametro m (230514) è minore del valore di partenza (ottenuto aumentando di poco l'ultimo valore cumulato): questo avviene perché il processo di diffusione analizzato è chiaramente nella fase di declino. Il parametro p sta ad indicare che coloro che hanno contratto per primi il virus hanno influito in maniera limitata sul processo di diffusione dello stesso. Il valore del parametro q , invece, conferma il fatto che la componente di contagio è stata determinante nella diffusione del nuovo coronavirus tra la popolazione italiana. Il coefficiente di determinazione R^2 è pari a 0,996496, ma questo non deve far pensare che il modello sia soddisfacente, come sembrerebbe confermare anche la stabilità delle stime, i cui intervalli di confidenza non comprendono lo zero. La figura 3.2 mostra, infatti, che l'adattamento può essere notevolmente migliorato: la parte iniziale della serie viene sovrastimata, il picco posticipato e la fase di declino prevista è più ripida rispetto a quanto realmente osservato.

3.3 Applicazione del modello di Bass Generalizzato

Viste le problematiche del modello di Bass in termini di adattamento ai dati, si passa ora all'utilizzo del modello di Bass Generalizzato: questo modello estende il modello di Bass introducendo una funzione di intervento, la quale include eventuali azioni esogene che hanno influito sul processo di diffusione. I prossimi 2 sottoparagrafi mostrano l'adattamento ai dati del modello di Bass Generalizzato con uno shock esponenziale (GBMe1) e con uno shock rettangolare (GBMr1).

3.3.1 Modello di Bass Generalizzato con uno shock esponenziale

Come osservato al §3.1, dopo due/tre settimane dalla diagnosi dei primi casi positivi a SARS-CoV-2 in Italia, il numero di nuovi contagiati è aumentato a livello esponenziale; si propone, quindi, la stima di un modello di Bass Generalizzato con uno shock esponenziale positivo. Di seguito vengono riportati i valori iniziali ipotizzati:

Stime iniziali	
m	230000
p	0,002
q	0,09
c_1	0,1
b_1	-0,1
a_1	20,0

Tabella 3.4: Valori iniziali dei parametri del GBMe1

Questi sono i risultati ottenuti:

Parametro	Stima	Errore standard	IC inferiore	IC superiore
m	238487	473,919	237547	239427
p	0,00090068	0,0000546817	0,000792219	0,00100914
q	0,0609813	0,00131932	0,0583644	0,0635981
c_1	3,02079	0,13238	2,75821	3,28336
b_1	-0,109238	0,0063346	-0,121802	-0,096673
a_1	19,1875	0,3367	18,5197	19,8554

Tabella 3.5: Stime dei parametri del GBMe1 (livello di confidenza = 95%)

Sorgente	Somma dei quadrati	G.l.	Media dei quadrati
Modello	$2,84543e^{12}$	6	$4,74238e^{11}$
Residuo	$1,02706e^8$	102	$1,00692e^6$
Totale	$2,84553e^{12}$	108	
Totale (Corr.)	$8,45806e^{11}$	107	

Tabella 3.6: Analisi della varianza del GBMe1

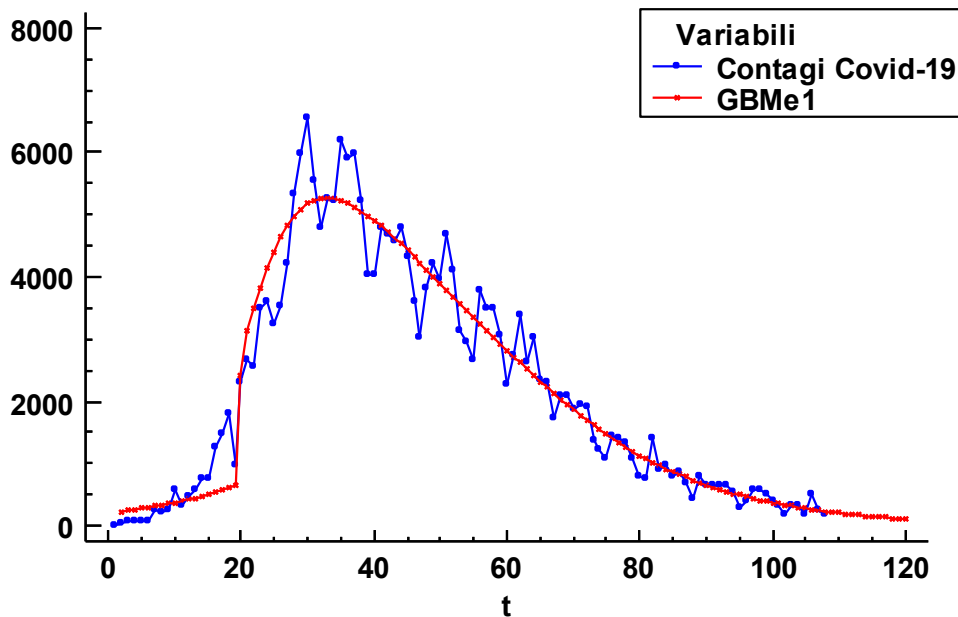


Figura 3.3: Serie giornaliera con la previsione del modello di Bass Generalizzato con uno shock esponenziale

Tutti i parametri del modello sono significativi, compresi i parametri relativi allo shock: il parametro c_1 rappresenta il segno e intensità dello shock, il quale è molto intenso, come è possibile vedere dal valore stimato (3,02). Il parametro b_1 conferma che lo shock viene riassorbito nel tempo, essendo la stima di questo parametro di segno negativo, mentre il parametro a_1 indica il momento di inizio dello shock, il cui inizio è stimato attorno al 10 marzo. Tenendo conto del periodo di incubazione di questo coronavirus, è probabile che buona parte delle persone diagnosticate positive a SARS-CoV-2 il 10 marzo siano state infettate in un periodo che va dal 28 febbraio all'8 marzo (il giorno prima dell'inizio del lockdown). Questo shock sembra dunque suggerire che il confinamento obbligatorio sia stato attuato troppo tardi, dal momento che l'epidemia aveva preso una corsa irrefrenabile già dai primi di marzo. Per vedere i primi effetti positivi di questo lockdown si è dovuto aspettare

circa due settimane, quando, una volta raggiunto il picco massimo (21 marzo), il numero di nuovi casi è cominciato, seppur molto lentamente, a scendere.

Il coefficiente di determinazione del GBMe1 è 0,999879, per cui si può pensare che il modello di Bass Generalizzato con uno shock esponenziale presenti un miglior adattamento rispetto al modello di Bass, come sembra confermato anche dalla figura 3.3. Il GBMe1 è, infatti, riuscito a cogliere in maniera abbastanza soddisfacente lo shock, descrivendo in maniera altrettanto buona il picco massimo e soprattutto la fase di declino.

Per capire se il modello di Bass Generalizzato appena stimato apporta un miglioramento statisticamente significativo al modello di Bass, è possibile effettuare un confronto tra modelli annidati. Come prima cosa, si calcola l' $\tilde{R}^2$:

$$\tilde{R}^2 = \frac{R_{m2}^2 - R_{m1}^2}{1 - R_{m1}^2} = \frac{0,999879 - 0,996496}{1 - 0,996496} = 0,965 \quad (3.1)$$

dove m_2 è il GBMe1 e m_1 è il modello di Bass. Il valore assunto dall'indice $\tilde{R}^2$ è molto elevato e già questo fa presagire che il modello più complesso sia sicuramente da preferire. Si passa ora al calcolo del rapporto F:

$$F = \frac{\tilde{R}^2(n-v)}{(1-\tilde{R}^2)u} = \frac{0,965(108-6)}{(1-0,965)3} = 950,6 \quad (3.2)$$

Un valore così elevato del rapporto F porta senza dubbi a confermare la significatività dell'impulso esponenziale introdotto: per questo motivo, è quindi preferibile il modello più complesso (il GBMe1) a quello ridotto (il modello di Bass).

3.3.2 Modello di Bass Generalizzato con uno shock rettangolare

Lo shock esponenziale stimato al paragrafo precedente è risultato significativo da un punto di vista statistico. In alternativa a questo shock, è possibile ipotizzare la presenza di uno shock rettangolare di intensità positiva che comincia dopo circa due settimane e termina dopo circa cinque settimane. Questi sono i valori iniziali scelti per la stima di un modello di Bass Generalizzato con uno shock rettangolare:

Stime iniziali	
m	230000
p	0,002
q	0,09
c_1	0,1
a_1	15,0
b_1	38,0

Tabella 3.7: Valori iniziali dei parametri del GBMr1

I risultati ottenuti sono i seguenti:

Parametro	Stima	Errore standard	IC inferiore	IC superiore
m	236479	301,074	235882	237076
p	0,000813804	0,000109984	0,00059565	0,00103196
q	0,0682339	0,000525339	0,0671919	0,0692759
c_1	0,929024	0,0285397	0,872416	0,985632
a_1	12,8064	1,73653	9,36195	16,2508
b_1	36,1905	0,186945	35,8197	36,5613

Tabella 3.8: Stime dei parametri del GBMr1 (livello di confidenza = 95%)

Sorgente	Somma dei quadrati	G.l.	Media dei quadrati
Modello	2,84542e ¹²	6	4,74237e ¹¹
Residuo	1,05796e ⁸	102	1,03721e ⁶
Totale	2,84553e ¹²	108	
Totale (Corr.)	8,45806e ¹¹	107	

Tabella 3.9: Analisi della varianza del GBMr1

Anche per questo modello, tutti i parametri del modello, compresi i parametri relativi allo shock, sono significativi: il parametro c_1 è positivo, come ci si aspettava, e di intensità abbastanza elevata. I parametri a_1 e b_1 rappresentano rispettivamente il momento di inizio e di fine dello shock: lo shock stimato comincia attorno al 4 marzo e termina attorno al 27 marzo. Dalla figura 3.4 si può osservare che lo shock rettangolare stimato viene catturato abbastanza bene.

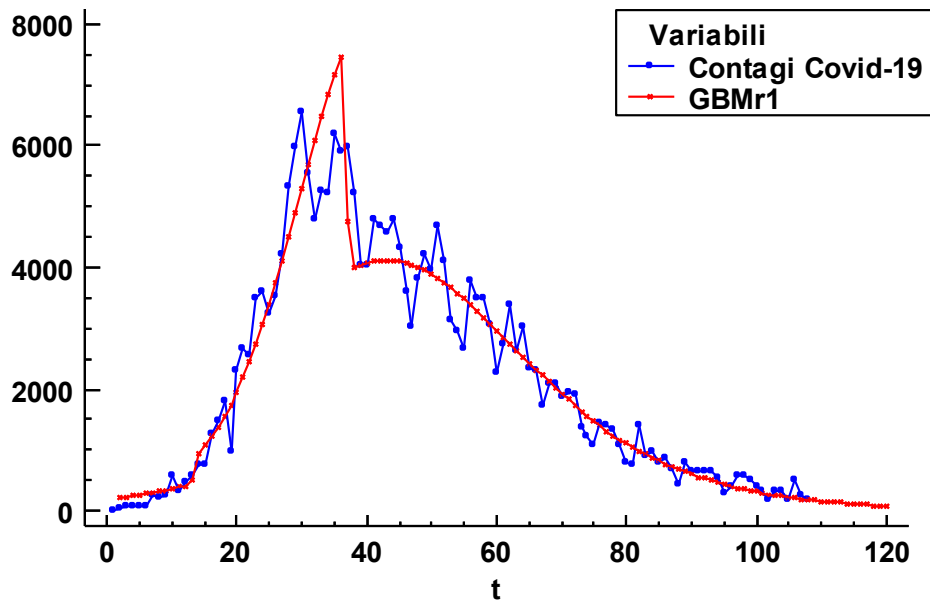


Figura 3.4: Serie giornaliera con la previsione del modello di Bass Generalizzato con uno shock rettangolare

L'interpretazione di questo shock è forse meno immediata rispetto a quella del precedente shock; è possibile, però, interpretare lo shock rettangolare in maniera simile a quanto fatto con lo shock esponenziale. Il 21 febbraio sono stati diagnosticati in Italia i primi casi affetti da Covid-19 e si può pensare che, attorno a quei giorni, il virus abbia cominciato a diffondersi sempre più velocemente, arrivando a contagiare, dopo 12 giorni, un numero elevato di persone (coerentemente col tempo di incubazione di questo coronavirus); questo ha dato inizio alla perturbazione stimata, che risulta stabile (essendo questo uno shock rettangolare) ma di intensità positiva. La fine dello shock è avvenuta qualche giorno dopo il raggiungimento del picco massimo e può essere ricollegata al lockdown: a partire dal 21 marzo (12 giorni dopo l'inizio del confinamento forzato) si è registrato, seppur in maniera molto lenta, una diminuzione del numero di nuovi contagiati che, dal 27 marzo, sono sempre stati inferiori a 6000 e, qualche giorno dopo, sempre inferiori a 5000.

L' R^2 di questo modello è uguale a 0,999875, valore pressoché simile al coefficiente di determinazione del GBMe1. Anche nel caso del modello di Bass Generalizzato con uno shock rettangolare è possibile effettuare un confronto tra modelli annidati, in cui il GBMr1 è il modello esteso, mentre il modello di Bass è quello ridotto. Si riportano i risultati dell'indice $\tilde{R}^2$ e del rapporto F:

$$\tilde{R}^2 = \frac{R_{m2}^2 - R_{m1}^2}{1 - R_{m1}^2} = \frac{0,999875 - 0,996496}{1 - 0,996496} = 0,964 \quad (3.3)$$

$$F = \frac{\tilde{R}^2(n-v)}{(1-\tilde{R}^2)u} = \frac{0,964(108-6)}{(1-0,964)3} = 919,1 \quad (3.4)$$

I valori di queste quantità sono ovviamente simili ai valori ottenuti col modello di Bass Generalizzato con uno shock esponenziale, dal momento che il coefficiente di determinazione dei due modelli è pressoché simile. Anche in questo caso è da preferire il modello esteso al modello ridotto.

3.4 Applicazione del modello di Guseo-Guidolin

I modelli di Bass Generalizzato presentati nei due precedenti sottoparagrafi costituiscono sicuramente un miglioramento significativo rispetto al modello di Bass. Si è provato, tuttavia, ad adattare alla serie anche un modello di Guseo-Guidolin, che tiene conto sia della presenza di un mercato potenziale dipendente dal tempo sia dell'evoluzione congiunta delle fasi di contagio e di diagnosi di positività. Per i valori iniziali di K , p_s e q_s sono stati utilizzati dei valori prossimi alle stime di m , p e q , rispettivamente, ottenute col modello di Bass:

Stime iniziali	
K	230000
p_c	0,001
q_c	0,1
p_s	0,002
q_s	0,1

Tabella 3.10: Valori iniziali dei parametri del GGM

Di seguito si riportano i risultati ottenuti:

Parametro	Stima	Errore standard	IC inferiore	IC superiore
K	236504	223,383	236061	236947
p_c	0,00133917	0,0000359418	0,00126789	0,00141045
q_c	0,0717059	0,00067241	0,0703723	0,0730394
p_s	0,00145037	0,0000647829	0,00132189	0,00157885
q_s	0,176185	0,00248848	0,17125	0,18112

Tabella 3.11: Stime dei parametri del GGM (livello di confidenza = 95%)

Sorgente	Somma dei quadrati	G.l.	Media dei quadrati
Modello	2,84549e ¹²	5	5,69098e ¹²
Residuo	3,87799e ⁷	103	376504
Totale	2,84553e ¹²	108	
Totale (Corr.)	8,45806e ¹¹	107	

Tabella 3.12: Analisi della varianza del GGM

R ²	0,999954
Statistica di Durbin-Watson	0,400767

Tabella 3.13: R² e statistica di Durbin-Watson del GGM

La stima del parametro K è di poco superiore all'ultimo valore cumulato della serie (234998): il GGM prevede, quindi, il termine del ciclo di vita del virus nel breve periodo. I parametri relativi alla componente innovativa, p_c e p_s , sono pressoché simili e sono molto bassi, segno che coloro che per primi sono stati contagiati e coloro che per primi sono stati diagnosticati positivi al virus del Covid-19 non hanno contribuito in maniera significativa alla diffusione del virus. Il parametro q_c rappresenta il *processo di contagio*, nel quale si trovano tutti coloro che sono stati contagiati dal virus, tra cui gli asintomatici. Il valore di q_c non è molto elevato e questo è riconducibile al fatto che questo parametro incorpora sia l'effetto del lockdown attuato in Italia sia il fatto che il modello è stato applicato ad una serie che sta registrando una fase di declino da diverse settimane. Il parametro q_s , invece, è molto elevato e questo si può spiegare col fatto che, una volta identificati i primi casi positivi al virus del Covid-19, sono stati effettuati sempre più tamponi, al fine di diagnosticare quanti più nuovi casi positivi possibili.

Il coefficiente di determinazione del GGM è molto alto e molto più elevato dello stesso coefficiente ottenuto col modello di Bass. Per verificare se il GGM apporta un miglioramento significativo al modello di Bass, si calcolano l' $\tilde{R}^2$ e il rapporto F:

$$\tilde{R}^2 = \frac{R_{m2}^2 - R_{m1}^2}{1 - R_{m1}^2} = \frac{0,999954 - 0,996496}{1 - 0,996496} = 0,987 \quad (3.5)$$

$$F = \frac{\tilde{R}^2(n-v)}{(1-\tilde{R}^2)u} = \frac{0,987(108-5)}{(1-0,987)2} = 3871,5 \quad (3.6)$$

Valori così elevati dell' $\tilde{R}^2$ e del rapporto F confermano che il miglioramento apportato dal GGM è statisticamente significativo.

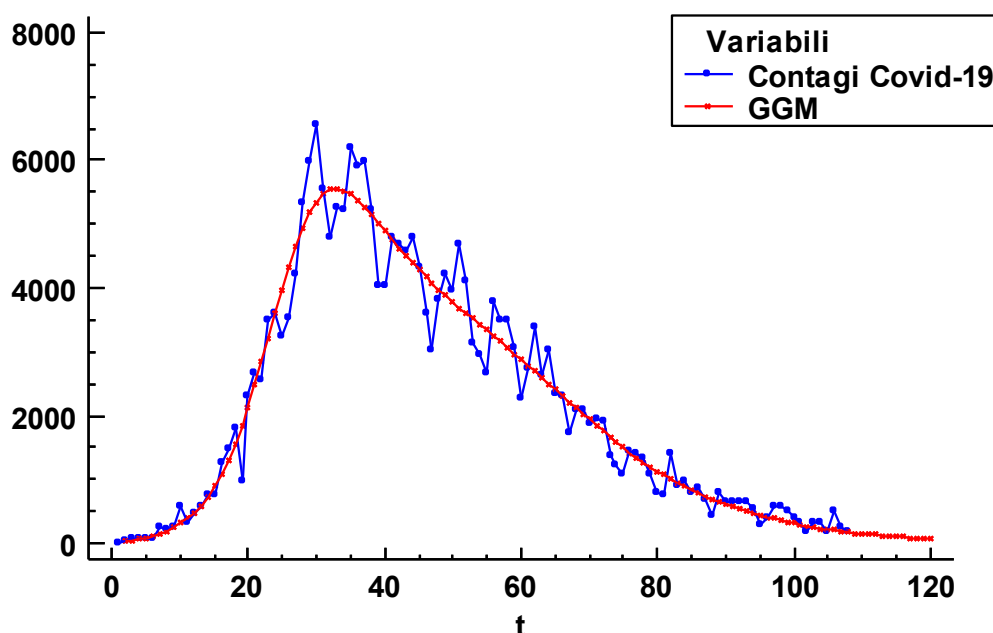


Figura 3.5: Serie giornaliera con la previsione del GGM

L'adattamento del GGM è molto soddisfacente: il modello, infatti, coglie bene fin da subito l'andamento medio della serie, non sovrastima la parte iniziale, riesce ad individuare bene il picco massimo e descrive bene anche la fase di declino. A conferma di un adattamento grafico molto soddisfacente, le stime dei parametri sono stabili e i corrispondenti intervalli di confidenza non comprendono lo zero. Per questi motivi, il GGM è stato scelto come miglior modello per descrivere la serie.

3.4.1 GGM: posizionamento temporale delle due fasi

Come già presentato da un punto di vista teorico, col GGM è possibile fornire un collocamento temporale delle fasi di contagio e di diagnosi di positività. Si procede, dunque, con il calcolo di alcuni indicatori di posizione descritti in precedenza (si veda §2.5.1). I risultati vengono riassunti nella seguente tabella:

Contagio	Diagnosi di positività
$t_F^+ = 54,494$	$t_G^+ = 27,020$
$t_F^{0.5} = 54,996$	$t_G^{0.5} = 27,112$
$\bar{t}_F = 55,770$	$\bar{t}_G = 27,289$

Tabella 3.14: Indicatori di posizione delle 2 fasi del GGM

Si nota che gli indicatori di posizione corrispondenti alla fase di diagnosi di positività sono inferiori rispetto a quelli della fase di contagio: questo risultato indica che si è nella situazione più insolita, in cui la diagnosi di positività precede il contagio. Temporalmente, gli indicatori di posizione calcolati corrispondono al 18 marzo, per la diagnosi di positività, e al 14 aprile, per il contagio. Il motivo per cui si ha un ordinamento inverso rispetto alla situazione usuale può essere legato alle caratteristiche di coloro che per primi si sono ammalati: i primi contagiati da Covid-19 erano, infatti, quasi tutti anziani e con patologie pregresse. Si può, quindi, dire che c'è stata una componente della popolazione italiana molto suscettibile al virus e questo ha fatto sì che proprio questi soggetti più deboli si siano ammalati subito. Si è ritenuto interessante effettuare alcune simulazioni, modificando in maniera coerente il valore dei parametri del GGM (ad esclusione di K , il cui valore rimane pari alla stima riportata nella tabella 3.11), per osservare il modo in cui diverse configurazioni dei parametri modificano il collocamento temporale delle fasi di contagio e di diagnosi di positività. Sono state effettuate due simulazioni, ognuna delle quali presenta un'ipotesi sottostante.

La prima simulazione suppone che siano stati effettuati molti tamponi fin dall'insorgenza dei primi casi positivi: tenendo fissi i parametri p_c e q_c , i parametri p_s e q_s sono quindi stati moltiplicati per 10 ($p_s = 1,76185$ e $q_s = 0,0145037$). Di seguito si mostrano i nuovi indici di posizione:

Contagio	Diagnosi di positività
$t_F^+ = 54,494$	$t_G^+ = 2,702$
$t_F^{0.5} = 54,996$	$t_G^{0.5} = 2,711$
$\bar{t}_F = 55,770$	$\bar{t}_G = 2,729$

Tabella 3.15: Indicatori di posizione delle 2 fasi (1^a simulazione)

Dalla tabella 3.15 si può osservare che, nel caso in cui fossero stati fatti molti tamponi fin da subito, la fase di diagnosi di positività sarebbe stata notevolmente anticipata, spostando molto in avanti la dinamica del contagio. Tramite il software *Statgraphics*, è stata effettuata una generazione di dati a partire dall'equazione in forma chiusa del GGM, utilizzando come valori dei parametri quelli assunti in questa prima simulazione. Poiché i dati generati sono cumulati, è stata effettuata

una differenziazione per ottenere i valori istantanei (corrispondenti alla curva rossa della figura 3.6).

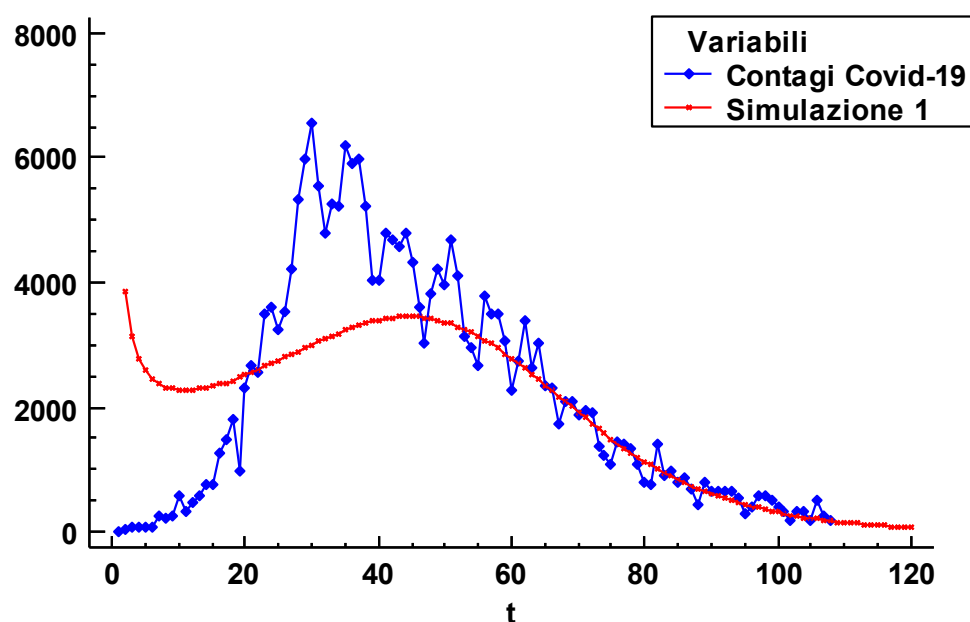


Figura 3.6: Serie giornaliera con la previsione ottenuta attraverso la prima simulazione

Si può osservare come non venga stimato un numero di nuovi casi giornalieri superiore a 4000; è interessante, però, notare il fatto che, in questa prima simulazione, il più alto numero di nuovi casi giornalieri si sarebbe registrato proprio il primo giorno (21 febbraio).

La seconda simulazione effettuata suppone una dinamica di contagio molto più veloce, ossia ipotizza una maggiore contagiosità di Covid-19: tenendo fissi i parametri p_s e q_s , i parametri p_c e q_c sono stati moltiplicati per 10 (ora $p_c = 0,0133917$ e $q_c = 0,717059$). Questi sono gli indici di posizione corrispondenti:

Contagio	Diagnosi di positività
$t_F^+ = 5,449$	$t_G^+ = 27,020$
$t_F^{0.5} = 5,500$	$t_G^{0.5} = 27,112$
$\bar{t}_F = 5,577$	$\bar{t}_G = 27,289$

Tabella 3.16: Indicatori di posizione delle 2 fasi (2ª simulazione)

Da questa seconda simulazione, si conclude che una maggiore contagiosità del virus anticipa notevolmente il processo del contagio, che, in questo caso, si colloca temporalmente prima della fase di diagnosi di positività.

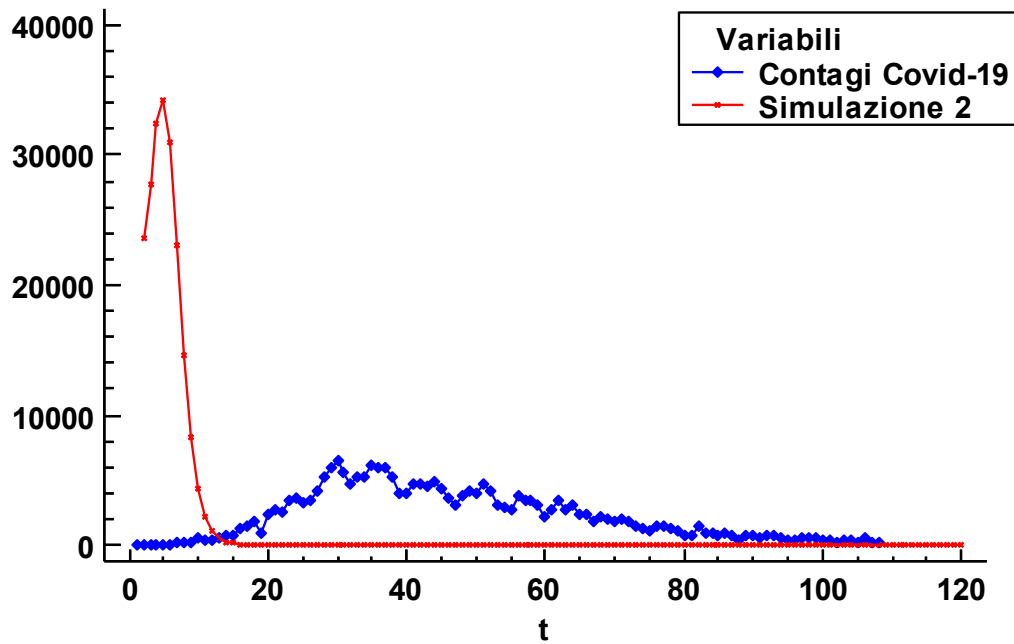


Figura 3.7: Serie giornaliera con la previsione ottenuta attraverso la seconda simulazione

Dalla figura 3.7 si osserva come, nello scenario previsto in questa seconda simulazione e fissato K (parametro di scala del processo di diffusione), si avrebbe avuto un numero elevatissimo di nuovi casi positivi nei primi giorni. Questo avrebbe portato quasi sicuramente alla saturazione del sistema ospedaliero italiano, in quanto non ci sarebbero stati posti per tutti nelle unità di terapia intensiva, dal momento che queste sono state “rafforzate” in termini di numero di posti letto e, soprattutto, di ventilatori polmonari solo a partire da marzo. Ancora una volta, aver attuato il lockdown è servito a prevenire una situazione come questa, nella quale un numero troppo elevato di persone risulta positivo al virus nello stesso momento, non rendendo così possibile garantire un posto letto in terapia intensiva a chi ne avesse avuto bisogno.

3.4.2 GGM: analisi dei residui e affinamento SARMAx

Poiché è stato scelto il GGM come modello migliore, si passa ora ad approfondire la struttura dei residui di questo modello, per verificare che tali residui siano generati da un processo *white noise*.

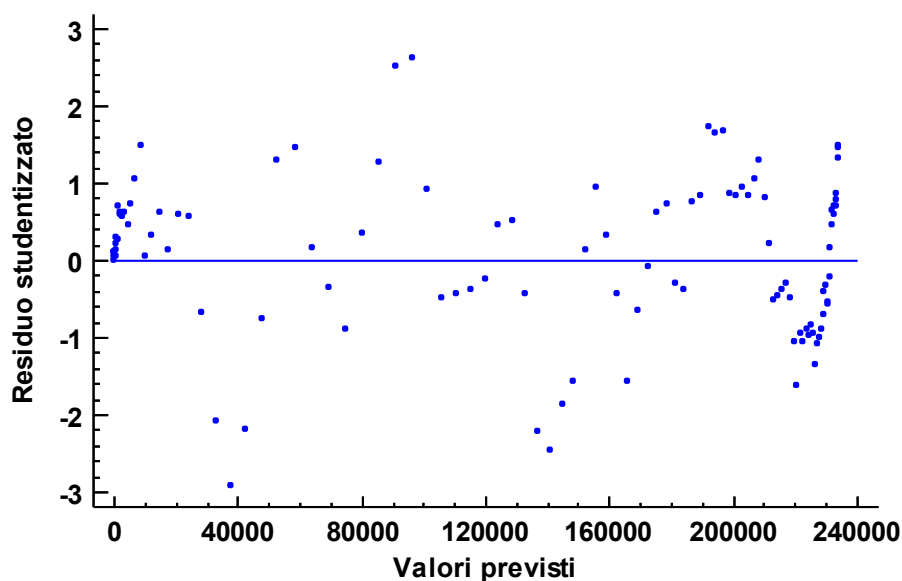


Figura 3.8: Residui studentizzati del GGM

La statistica di Durbin-Watson (0,400767), confrontata con i suoi valori critici ($d_L = 1,5909$; $d_U = 1,7841$), individua la presenza di una correlazione positiva nei residui, come si riscontra anche dalla figura 3.8. Tale correlazione positiva può essere risolta attraverso un affinamento ARMAx o SARMAx. Il miglior modello è un SARIMA (2,0,0)x(1,0,1)₇ con costante e con regressore esterno dato dalle previsioni ottenute col GGM; la componente stagionale assunta è a frequenza settimanale. Questi sono i risultati ottenuti:

Parametro	Stima	Errore standard	t	P-value
AR(1)	1,11117	0,0963424	11,5336	0,000000
AR(2)	-0,286884	0,0973895	-2,94574	0,003992
SAR(1)	0,834377	0,104421	7,9905	0,000000
SMA(1)	0,512793	0,158425	3,23682	0,001631
PRED_GGM	0,999987	0,00331485	301,669	0,000000
Media	128,318	625,553	0,205127	0,837882
Costante	3,73432			

Tabella 3.17: Stime dei parametri del modello SARMAx(2,0,0)x(1,0,1)₇

Il regressore “PRED_GGM” riportato nella tabella 3.17 corrisponde al parametro di calibrazione c ; la stima di questo parametro è associata alle previsioni ottenute col GGM e assume un valore approssimativamente unitario (0,999987), indicando un ottimo adattamento medio del modello non lineare prescelto. Dalla figura 3.9 si

osserva che l'affinamento SARMAx porta ad un ottimo adattamento, anche se è caratterizzato da una leggera traslazione a destra: in generale, si può dire che esso coglie le oscillazioni casuali che il GGM non era riuscito a catturare e risolve anche l'autocorrelazione positiva dei residui. L'incorrelazione dei residui è confermata dal correlogramma presentato nella figura 3.10: ad eccezione della nona autocorrelazione, tutte le altre autocorrelazioni sono, infatti, interne alle bande di confidenza (di livello 0,95).

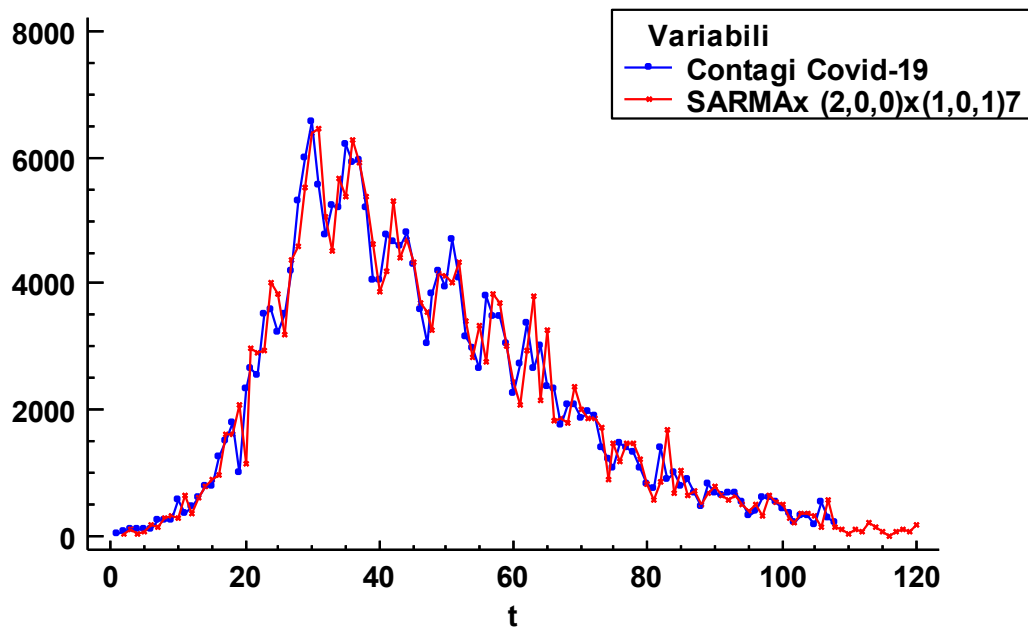


Figura 3.9: Serie giornaliera con l'affinamento SARMAx(2,0,0)x(1,0,1)₇

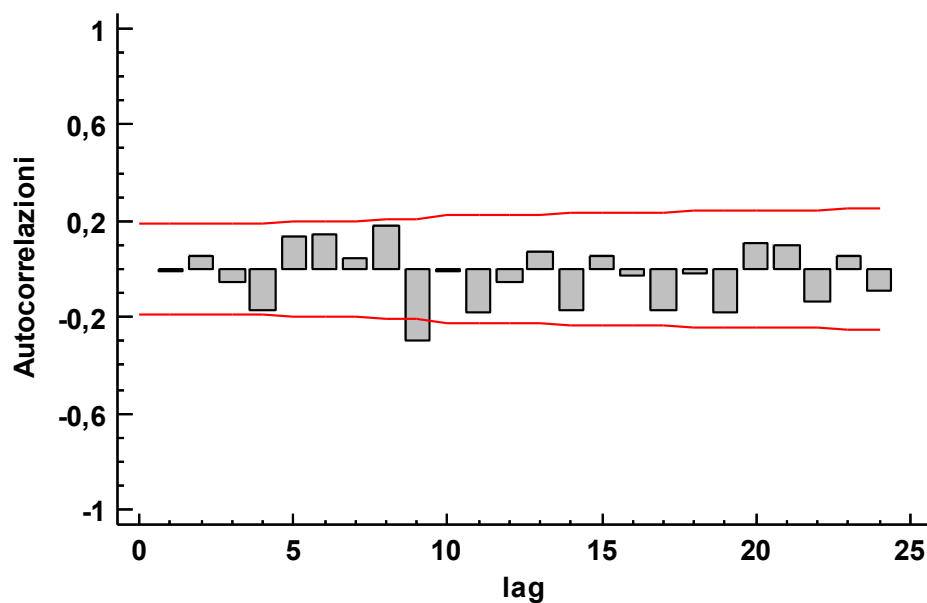


Figura 3.10: Correlogramma dei residui del SARMAx(2,0,0)x(1,0,1)₇

Conclusioni

L'obiettivo di questa relazione, attraverso l'utilizzo dei modelli di diffusione, è stato studiare il processo di diffusione di Covid-19 in Italia. Il modello che si è dimostrato più soddisfacente, in termini di adattamento alla serie analizzata, è il GGM: a partire da questo modello, è stato possibile calcolare opportuni indici di posizione, al fine di collocare temporalmente le fasi di contagio e di diagnosi di positività. Si è potuto osservare che la diagnosi di positività ha anticipato il contagio, ad indicare la presenza nella popolazione italiana di persone che, in quanto particolarmente suscettibili al nuovo coronavirus, si sono subito ammalate di Covid-19.

Relativamente al ciclo di vita del virus, al 10 giugno (ultimo giorno della serie storica analizzata) ci si trova nella fase di declino: tale declino, avente inizio approssimativamente a fine marzo, è tuttavia molto lento per i mesi di aprile e maggio, mentre raggiunge una situazione di stazionarietà nei mesi di giugno e luglio. In questi mesi, infatti, il numero di nuovi casi giornalieri è rimasto relativamente stabile, variando mediamente tra i 200 e i 400. Nel mese di agosto il numero di nuovi contagiati aumenta, superando i 1000 (22 agosto): i nuovi casi di questo mese sono stati diagnosticati principalmente mediante il *contact tracing* e lo screening su coloro che sono rientrati in Italia dopo essere stati in Grecia, Croazia, Spagna o Malta, come previsto dall'ordinanza del 12 agosto disposta dal Ministro della Salute. Questi nuovi casi sono prevalentemente asintomatici e presentano un'età mediana attorno ai 29 anni.

Se l'ipotesi di limitatezza del ciclo di vita è coerente e facilmente accettabile per un prodotto innovativo lanciato nel mercato, lo stesso non può essere assunto così immediatamente per un virus come quello responsabile di Covid-19. In questo senso, risulta più corretto affermare che il GGM ha descritto in maniera molto soddisfacente la *prima ondata di contagi*; il nuovo aumento del numero di nuovi contagiati registrato, nel mese di agosto, in Paesi europei come Spagna, Francia, Belgio e Germania fa, infatti, temere l'arrivo, in un futuro più o meno prossimo, di una seconda ondata di contagi da Covid-19. La stessa OMS avverte che questo

nuovo coronavirus potrebbe non scomparire mai, ma anzi è probabile che continui a circolare tra la popolazione mondiale ancora per molti anni. Malattie il cui agente responsabile è stabilmente presente in un determinato territorio e circola nella popolazione, manifestandosi con un numero di casi più o meno elevato ma uniformemente distribuito nel tempo, prendono il nome di endemie: esempi di malattie endemiche sono il raffreddore, il morbillo, la scarlattina e la varicella. Alcune caratteristiche proprie di SARS-CoV-2 contribuiscono a rendere Covid-19 una potenziale malattia endemica: questo virus è, infatti, molto contagioso e ha una carica virale massima all'incirca tra le 12 e le 15 ore prima della manifestazione dei primi sintomi nel soggetto contagiato. Tutto ciò, di conseguenza, porta a una maggiore difficoltà nel controllo e nel contenimento della diffusione del virus.

In generale, l'epidemiologia insegna che il modo migliore per debellare un virus è attraverso l'immunità. Un individuo diventa immune a un certo virus in due casi: se guarisce dalla malattia causata da questo virus e sviluppa gli anticorpi contro future infezioni, o se viene vaccinato contro quel particolare virus. Quando viene raggiunta una certa soglia di persone immuni a un virus, il che rende sempre più difficile la diffusione di quel virus tra la popolazione, si dice che è stata raggiunta l'immunità di gregge (*herd immunity* in inglese). Molti esperti stimano che la soglia che porterebbe al raggiungimento dell'immunità di gregge, per questo nuovo coronavirus, è superiore al 60% (percentuale corrispondente a circa 4,7 miliardi di persone). Nel caso di Covid-19 è impensabile, dunque, il raggiungimento naturale dell'immunità di gregge, perché questo porterebbe inevitabilmente alla morte di milioni e milioni di persone in tutto il mondo.

In questo senso, diventa di fondamentale importanza lo sviluppo di un vaccino contro Covid-19, per poter raggiungere l'immunità di gregge più velocemente e in modo più sicuro. Già dal mese di marzo (l'11 marzo è stata dichiarata la pandemia di Covid-19), si è assistito all'inizio di una *corsa al vaccino*, in cui molte case farmaceutiche e aziende produttrici di vaccini si sono subito rese disponibili nella produzione di un vaccino anti-Covid, una volta che lo stesso è stato sperimentato, nella sua efficacia, dal laboratorio che lo ha prodotto. Per avere un vaccino efficace, sono mediamente necessari 4 anni dall'inizio della produzione dello stesso (talvolta

si arriva anche fino a 10 anni). Per una malattia come Covid-19, queste tempistiche sono tuttavia molto lunghe. Per questo motivo, quindi, si sta cercando di sovrapporre alcune delle fasi di sperimentazione clinica del vaccino per raggiungere il traguardo in tempi più brevi (tra i 12 e i 18 mesi). Questo non è il primo vaccino realizzato contro malattie causate da coronavirus: basti pensare alla Sars (2003) e alla Mers (2012), per le quali sono stati creati diversi vaccini. Il fatto che nel passato siano stati prodotti vaccini contro altri coronavirus ha sicuramente favorito la velocizzazione nella produzione del nuovo vaccino anti-Covid, per coloro che già avevano prodotto un vaccino contro Sars o Mers, in quanto essi sono potuti partire dallo stesso, adattandolo poi a Covid-19. Pur nella consapevolezza dell'importanza di avere un vaccino in tempi brevi, questo non basta a giustificare scorciatoie talvolta pericolose per la salute: se si vuole guadagnare tempo, lo si può fare accelerando le operazioni burocratiche, ma non i controlli. Prima di poter essere distribuito su larga scala, un vaccino deve infatti riportare risultati soddisfacenti sia sul piano dell'efficacia sia su quello della sicurezza. Un vaccino deve essere innanzitutto efficace, ossia protettivo nei confronti della malattia: la protezione può esplicarsi nel senso di evitare la malattia oppure di renderla più blanda. Un secondo requisito richiesto per un vaccino è la sicurezza, cioè l'assenza di reazioni collaterali importanti o frequenti, o comunque sproporzionate ai vantaggi che l'impiego del vaccino comporta.

Finché non sarà disponibile un vaccino in grandi quantità, ogni persona può contribuire, nel proprio piccolo e in maniera responsabile, a far sì che non si formino nuovi focolai che potrebbero potenzialmente dare inizio a una nuova ondata di contagi. La prima e più importante cosa da fare è evitare la formazione di grandi assembramenti. Inoltre, se si è in un contesto in cui è difficile evitare il contatto diretto con altre persone, l'uso della mascherina, così come il lavaggio frequente delle mani o la loro disinfezione, diventano indispensabili nella misura in cui ciò permette di rallentare la diffusione del nuovo coronavirus.

Bibliografia

- [1] Bass, F.M. (1969). *A New Product Growth for Model Consumer Durables*. Management Science, 15, 215-227.
- [2] Bass, F.M., Krishnan, T.V., Jain, D.C. (1994). *Why the Bass Model Fits without Decision Variables*. Marketing Science, 13, 203-223.
- [3] Box, G.E.P., Jenkins, G.M., Reinsel, G.C., Ljung, G.M. (2015). *Time Series Analysis: Forecasting and Control*. 5th edition, John Wiley & Sons, Inc., New Jersey.
- [4] Checcacci, L. (1984). *Medicina preventiva e igiene*. 5^a edizione, Casa Editrice Ambrosiana, Milano.
- [5] Cohen, W.M., Levinthal, D.A. (1990). *Absorptive Capacity: A New Perspective on Learning and Innovation*. Administrative Science Quarterly, 35, 128-152.
- [6] Foroni, E. (2016). *Identificazione di variazioni del ciclo di vita del prodotto dovute a scelte aziendali e posizionamento nel mercato*. Università degli Studi di Padova, Tesi di Laurea Triennale in Statistica e Gestione delle Imprese.
- [7] Guidolin, M. (2007). *Aggregate and agent-based models for the diffusion of innovation*. Università degli Studi di Padova, Tesi di dottorato.
- [8] Guseo, R. (2004). *Interventi strategici e aspetti competitivi nel ciclo di vita di innovazioni*. Working Paper Series 11/2004, Department of Statistical Sciences, University of Padua.

- [9] Guseo, R., Guidolin, M. (2009). *Modelling a dynamic market potential: A class of automata networks for diffusion of innovations*. *Technological Forecasting & Social Change*, 76 (6), 806-820.
- [10] Guseo, R., Guidolin, M. (2011). *Market potential dynamics in innovation diffusion: Modelling the synergy between two driving forces*. *Technological Forecasting & Social Change*, 78 (1), 13-24.
- [11] Nuvoli, A. (2017). *Il caso delle auto elettriche: analisi e previsioni delle vendite per i modelli Renault*. Università degli Studi di Padova, Tesi di Laurea Triennale in Statistica per l'Economia e l'Impresa.
- [12] Penello, N. (2016). *Informazione, apprendimento e diffusione di un prodotto farmaceutico: il caso ITM*. Università degli Studi di Padova, Tesi di Laurea Magistrale in Economia Internazionale.
- [13] Rogers, E.M. (2003). *Diffusion of Innovations*. 5th edition, The Free Press, New York.
- [14] Verbeek, M. (2017). *A Guide to Modern Econometrics*. 5th edition, John Wiley & Sons Ltd, Chichester.

Sitografia

- <https://www.internazionale.it/notizie/2020/01/31/coronavirus-sintomi-contagio>
- <http://www.salute.gov.it/portale/nuovocoronavirus/dettaglioFaqNuovoCoronavirus.jsp?lingua=italiano&id=228>
- <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/question-and-answers-hub/q-a-detail/q-a-coronaviruses>
- <https://bestpractice.bmj.com/topics/en-gb/3000168/pdf/3000168/Coronavirus%20disease%202019%20%28COVID-19%29.pdf>
- <https://www.cdc.gov/csels/dsepd/ss1978/lesson1/section10.html>
- <https://www.mayoclinic.org/diseases-conditions/coronavirus/expert-answers/can-coronavirus-spread-food-water/faq-20485479>
- <http://www.salute.gov.it/portale/nuovocoronavirus/dettaglioFaqNuovoCoronavirus.jsp?lingua=english&id=230#4>
- https://www.iss.it/documents/20126/0/Rappporto+ISS+COVID-19+n.+25_2020.pdf/90decdd1-7c29-29e4-6663-b992e1773c98?t=1589836083759
- <https://www.grupposandonato.it/news/2020/marzo/tampone-faringeo-coronavirus>

- <https://ilbolive.unipd.it/it/news/covid19-crisanti-asintomatici-svolgono-ruolo>
- <https://jamanetwork.com/journals/jama/fullarticle/2762130>
- <https://www.businessinsider.com/coronavirus-pandemic-timeline-history-major-events-2020-3?IR=T>
- <https://www.nytimes.com/article/coronavirus-timeline.html>
- http://www.salute.gov.it/portale/news/p3_2_1_1_1.jsp?lingua=italiano&menu=notizie&p=dalministero&id=4067
- <https://www.internazionale.it/notizie/2020/03/11/pandemia-coronavirus>
- <http://sphweb.bumc.bu.edu/otlt/MPH-Modules/SB/BehavioralChangeTheories/BehavioralChangeTheories4.html>
- http://www.salute.gov.it/portale/news/p3_2_1_1_1.jsp?lingua=italiano&menu=notizie&p=dalministero&id=5018#:~:text=Ministero%20della%20Salute&text=L'ordinanza%20del%20Ministro%20della,eventuale%20contagio%20da%20Covid%2D19
- <https://www.ilpost.it/2020/08/28/coronavirus-italia-dati-settimana-4-agosto/>
- <https://theconversation.com/the-original-sars-virus-disappeared-heres-why-coronavirus-wont-do-the-same-138177>
- <https://www.washingtonpost.com/health/2020/05/27/coronavirus-endemic/>

- <https://www.nytimes.com/interactive/2020/05/28/upshot/coronavirus-herd-immunity.html>
- <https://www.netflix.com/title/81273378>
- <https://www.scienzainrete.it/articolo/speranza-del-vaccino/roberta-villa/2020-06-15?fbclid=IwAR1f5COfcITFdFVa6WI93KqF5Uw8OGtvGFgojrR5hSBto4F-7lbGill31G8>
- <https://www.mayoclinic.org/diseases-conditions/coronavirus/in-depth/herd-immunity-and-coronavirus/art-20486808>