



UNIVERSITÀ DI PISA

TERZO PROGETTO DEL CORSO DI STATISTICA

**PREVISIONE DELL'ANDAMENTO DEL MERCATO
IMMOBILIARE AMERICANO**

ANNO ACCADEMICO 2019-2020

STEFANO PETROCCHI



SOMMARIO

Introduzione	3
Scopo dell'Analisi	3
Dataset.....	3
Analisi	3
Analisi Preliminare	3
Serie	3
Autocorrelazione	3
Decomposizione.....	4
Modelli.....	4
Residui	5
Previsione.....	6
Modello Holt-Winters	6
Modelli di Autoregressione.....	8
Accuratezza.....	8
Residui	9
Conclusione	10

INTRODUZIONE

SCOPO DELL'ANALISI

Si ipotizza che l'analisi sia stata commissionata da una grossa agenzia immobiliare interessata alla previsione dell'andamento del mercato americano per i prossimi due anni:

- Lo scopo principale dell'analisi è la previsione del trend delle vendite di abitazioni monofamiliari nel mercato americano nei prossimi due anni, le previsioni servono per valutare se sia proficuo un aumento degli investimenti (come l'assunzione di nuovo personale) o se non lo sia a causa di un trend negativo nelle vendite.
- Secondariamente l'agenzia è interessata alla previsione della stagionalità delle vendite in modo da pianificare la propria campagna pubblicitaria nel periodo più adatto.

DATASET

L'analisi è stata svolta su un dataset che riguarda la vendita di abitazioni monofamiliari negli Stati Uniti, costituito da una serie storica con osservazioni mensili. La serie inizia dal gennaio 1963 e termina nell'ottobre 2019 (682 osservazioni in totale), la scala delle osservazioni è di migliaia di abitazioni vendute arrotondate all'intero.

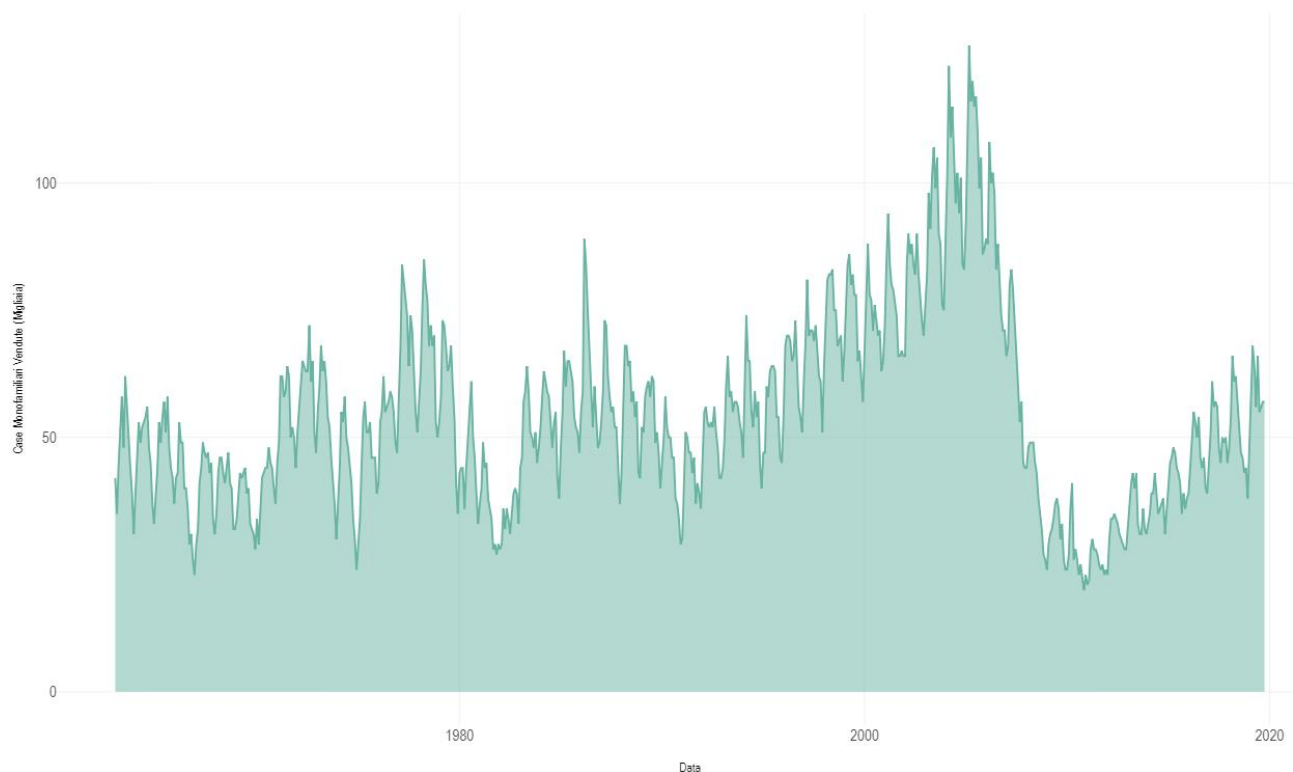
I dati sono forniti sul sito della [Banca Federale Statunitense di Sant Luis](#), il dataset in formato csv è reperibile da questo [link](#).

ANALISI

ANALISI PRELIMINARE

SERIE

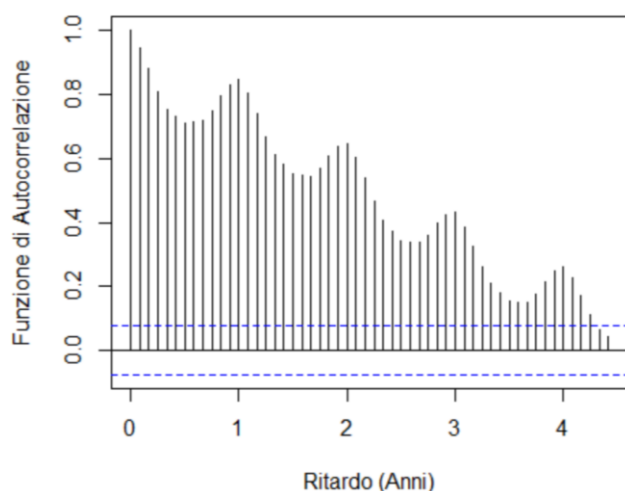
La serie sembra presentare una *stagionalità* annuale molto evidente, inoltre è possibile osservare un andamento del *trend* abbastanza irregolare. Fino agli anni 90 il numero di case vendute seguiva un andamento oscillante, negli anni 90 e nei primi anni 2000 si può notare un anomalo trend crescente fino allo scoppio della [bolla immobiliare americana](#), precursore della crisi globale iniziata nel 2008.



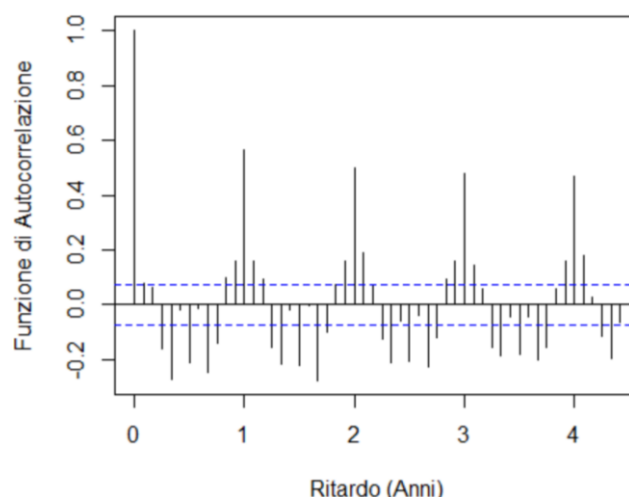
AUTOCORRELAZIONE

La funzione di **autocorrelazione** sulla serie e sulle differenze tra le osservazioni conferma chiaramente la presenza di un andamento stagionale annuale del numero di abitazioni vendute. Confrontando le *stagionalità* su anni differenti si può notare come i margini siano abbastanza ampi, ma che generalmente si ha un picco di vendite a partire da *marzo fino ad agosto* ed un calo nel periodo rimanente.

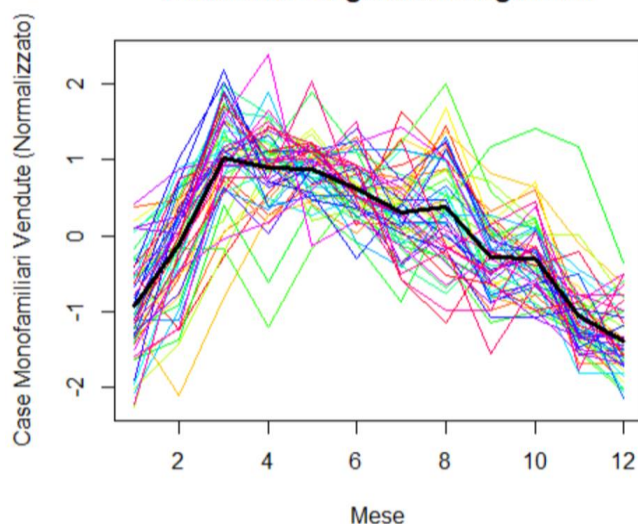
Autocorrelazione Case Vendute



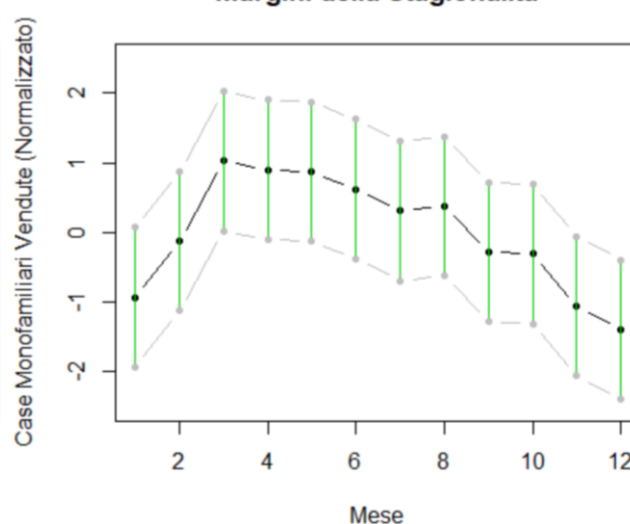
Autocorrelazione Case Vendute (Differenza)



Confronto Stagionalità negli Anni



Margini della Stagionalità



DECOMPOSIZIONE

MODELLI

STAGIONALITÀ COSTANTE ADDITIVA

Il modello di *decomposizione additiva* presenta un ottimo profilo di **trend** e una **stagionalità** annuale che conferma le previsioni dell'analisi preliminare. I **residui** sono dello stesso ordine di grandezza della *stagionalità* e presentano dei picchi in punti in cui trend delle vendite è maggiore, facendo pensare ad un profilo moltiplicativo della *stagionalità*.

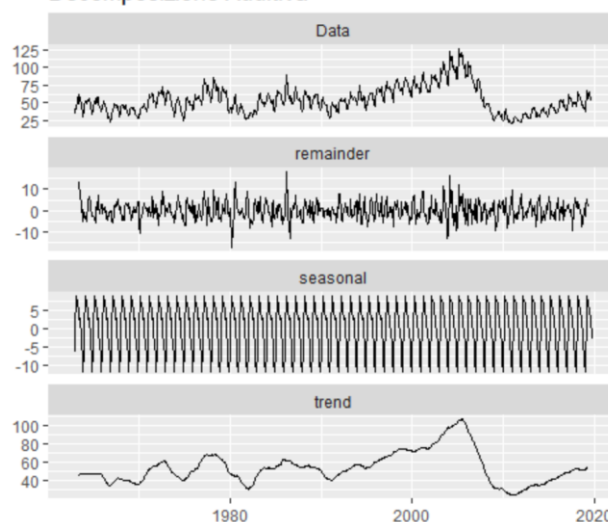
STAGIONALITÀ COSTANTE MOLTIPLICATIVA

Il modello di *decomposizione moltiplicativa* presenta il medesimo **trend** e **stagionalità** del precedente (come possibile verificare dai grafici di confronto). I **residui** sono ora di ampiezza maggiore in punti in cui trend di vendite è minore, perciò la *stagionalità* non sembra né moltiplicativa, né additiva, ma variabile.

STAGIONALITÀ VARIABILE

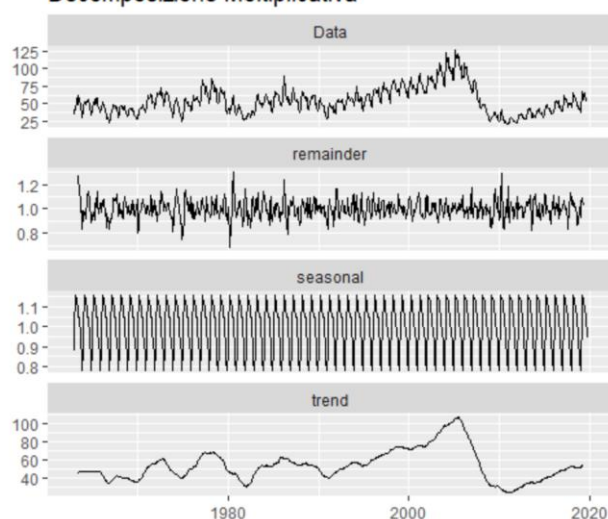
I modelli *STL additivo* e *moltiplicativo* presentano un medesimo profilo del **trend** e un grafico più smussato rispetto

Decomposizione Additiva

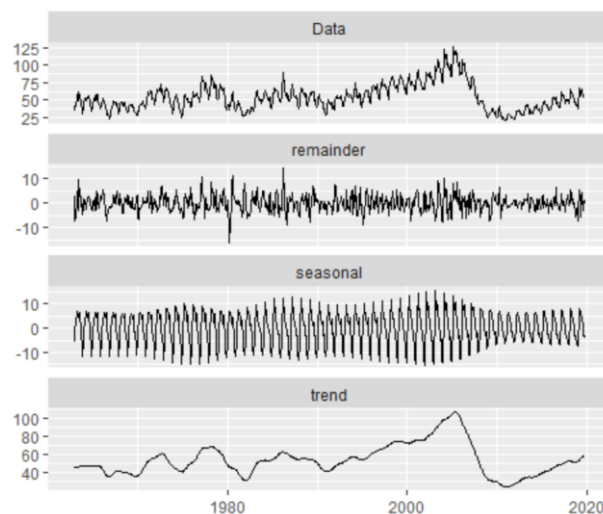


ai modelli precedenti, probabilmente a causa della **stagionalità** catturata in modo migliore. L'ipotesi viene ulteriormente confermata dal profilo dei **residui**, più uniforme e di ampiezza minore rispetto ai modelli a *stagionalità* costante.

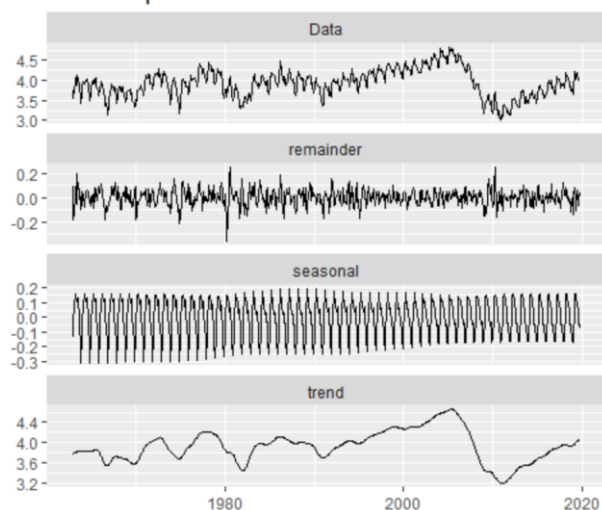
Decomposizione Moltiplicativa



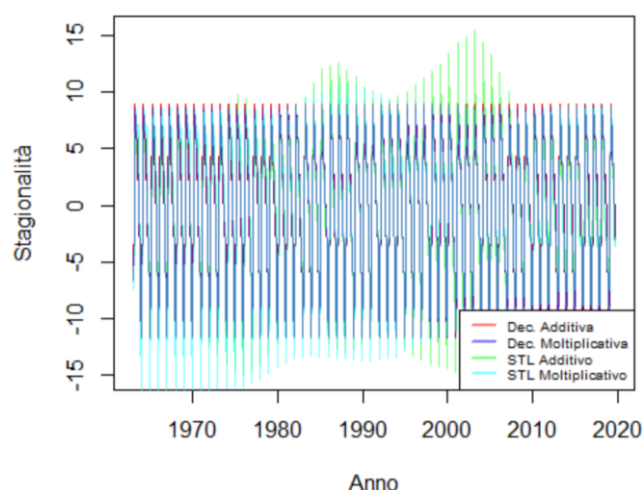
STL Additivo



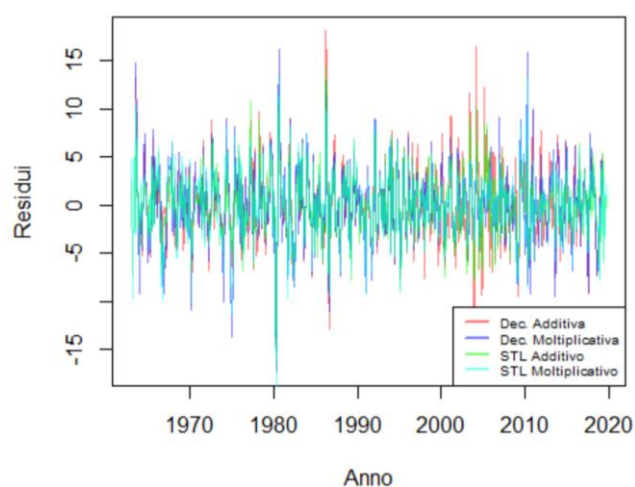
STL Moltiplicativo



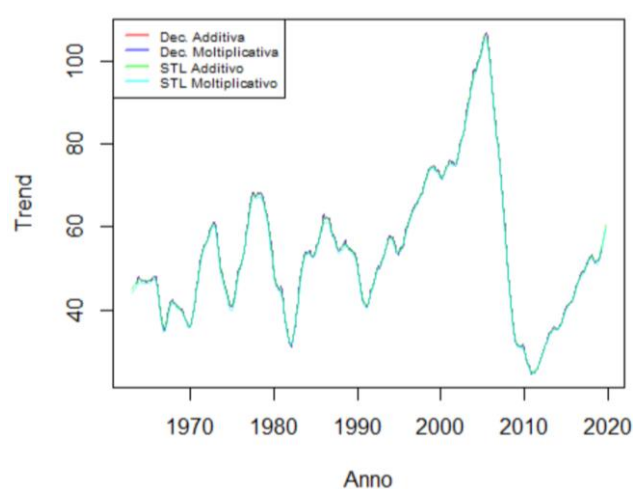
Confronto Stagionalità



Confronto Residui



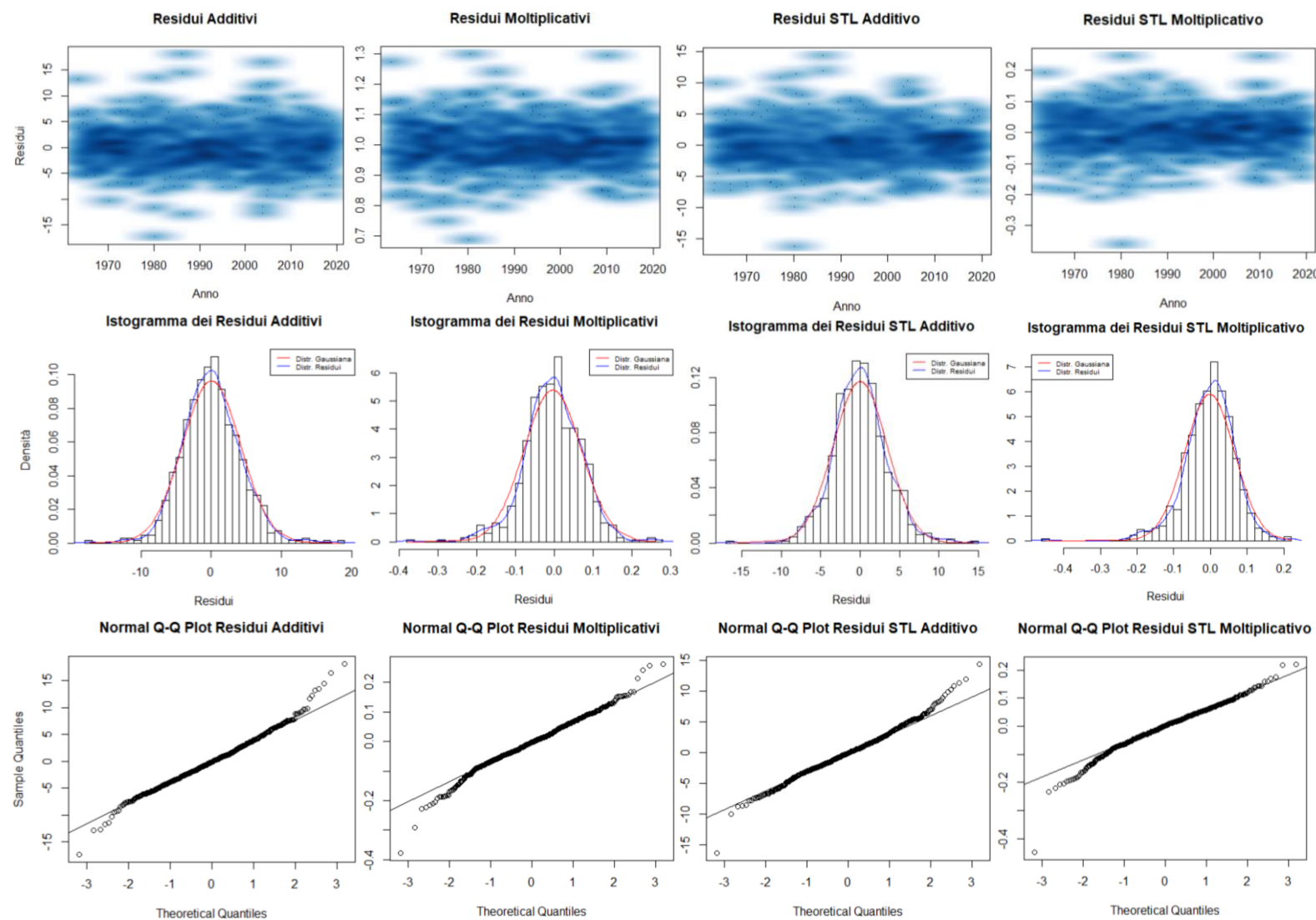
Confronto Trend



RESIDUI

DISTRIBUZIONE

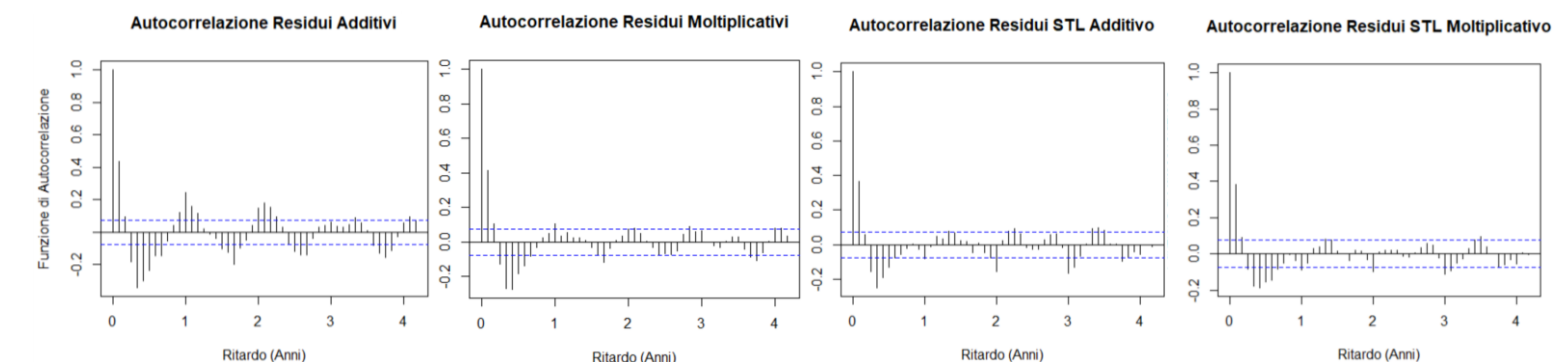
I grafici sui **residui** mostrano come tutte le distribuzioni presentino un buon andamento, senza grossi segnali di struttura residua. È interessante osservare come per tutti i modelli (particolarmente evidente nel modello STL moltiplicativo), esista un **valore anomalo** dei residui fortemente negativo, l'anomalia è probabilmente da attribuirsi alla recessione americana dei primi anni Ottanta. La **varianza spiegata** è ottima e molto simile per tutti i modelli, nell'ordine: 95,3%, 95,5%, 96,8% e 96,1%.



AUTOCORRELAZIONE

Dai grafici si può notare come il modello di *decomposizione additiva* abbia un'evidente stagionalità residua di periodo minore di un anno. Tra tutti i modelli quello *STL moltiplicativo* presenta i minori indizi di struttura residua (anche se rimane leggermente sul periodo), ciò è confermato dai valori di **deviazione standard** dell'autocorrelazione, nell'ordine: 0.253, 0.227, 0.219 e 0.216.

Date le sue caratteristiche il modello *STL moltiplicativo* sembra essere il più adatto per la decomposizione della serie e viene utilizzato come modello di confronto per i modelli predittivi.



PREVISIONE

MODELLO HOLT-WINTERS

CON PARAMETRI AUTOMATICI ($\alpha = 0.665$ | $\beta = 0$ | $\gamma = 0.49$)

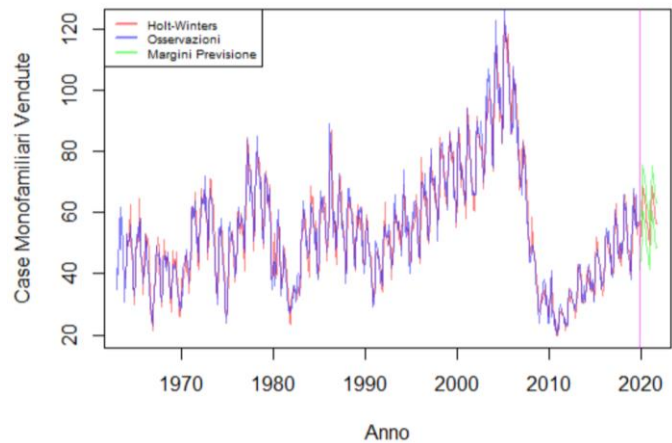
I parametri calcolati automaticamente, minimizzando lo scarto quadratico dei residui, portano ad un valore di β molto conservativo e a valori di α e γ intermedi. Il modello perciò tende a seguire molto i valori originali della serie.

Le **predizioni** mostrano una possibile stagnazione del numero di case vendute nei prossimi due anni.

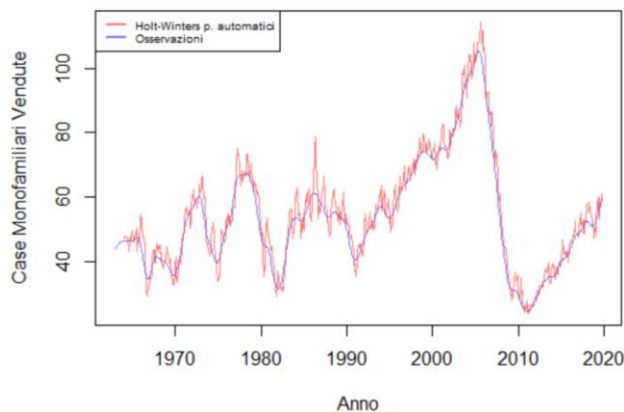
Confrontando il grafico delle *intercette* con il **trend** decomposto si può notare un andamento non molto lineare delle intercette, ma che comunque variano intorno al trend.

La **stagionalità** con *Holt-Winters* risulta avere lo stesso periodo della decomposizione, ma le ampiezze variano in modo differente.

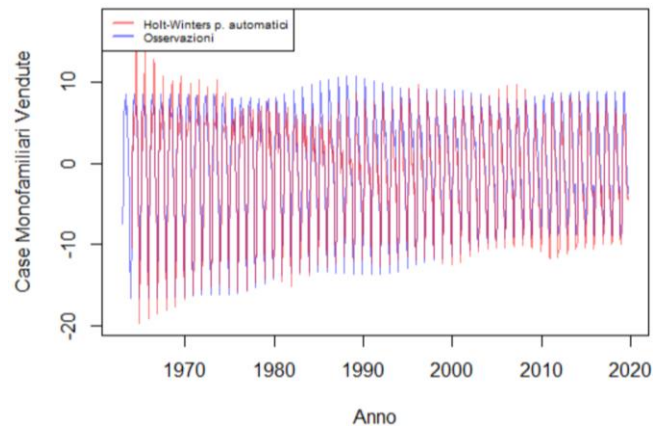
Holt-Winters con Parametri Automatici



Confronto con Trend STL



Confronto con Stagionalità STL



CON PARAMETRI SCELTI MANUALMENTE ($\alpha = 0.4$ | $\beta = 0.05$ | $\gamma = 0.1$)

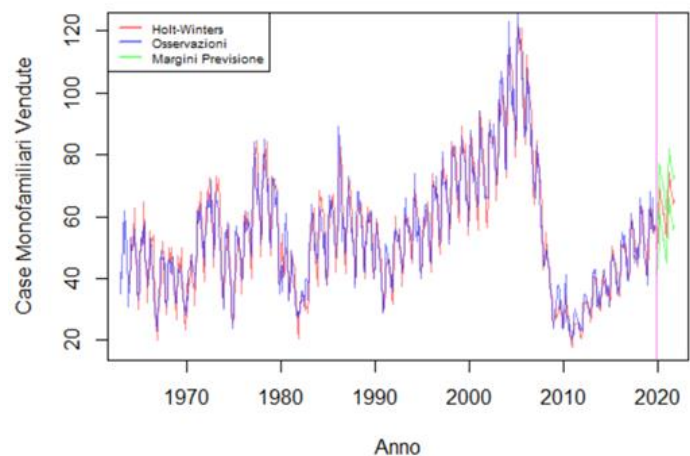
Sono stati scelti valori di α e γ più conservativi, a fronte di un valore di β più innovativo.

I parametri iniziali sono stati calcolati con una *regressione lineare* sui primi tre anni della serie, che presentano un andamento abbastanza regolare.

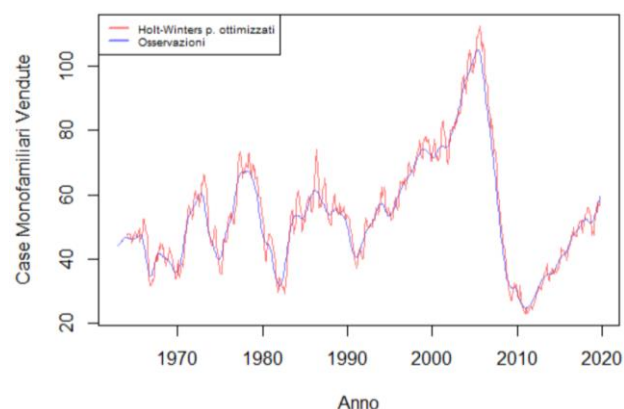
Le **predizioni** di questo modello sono più ottimistiche rispetto a quella del modello precedente, con un trend delle vendite chiaramente in aumento nei prossimi due anni.

Le **intercette** e la **stagionalità** seguono ora in maniera più accurata la *decomposizione STL* negli ultimi valori della serie, con la speranza di ottenere predizioni più accurate.

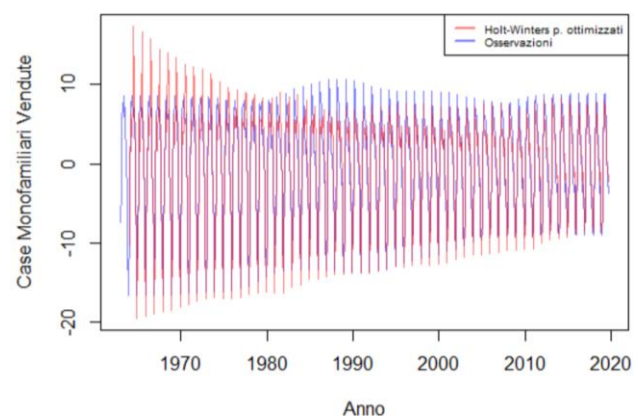
Holt-Winters con Parametri Ottimizzati



Confronto con Trend STL



Confronto con Stagionalità STL

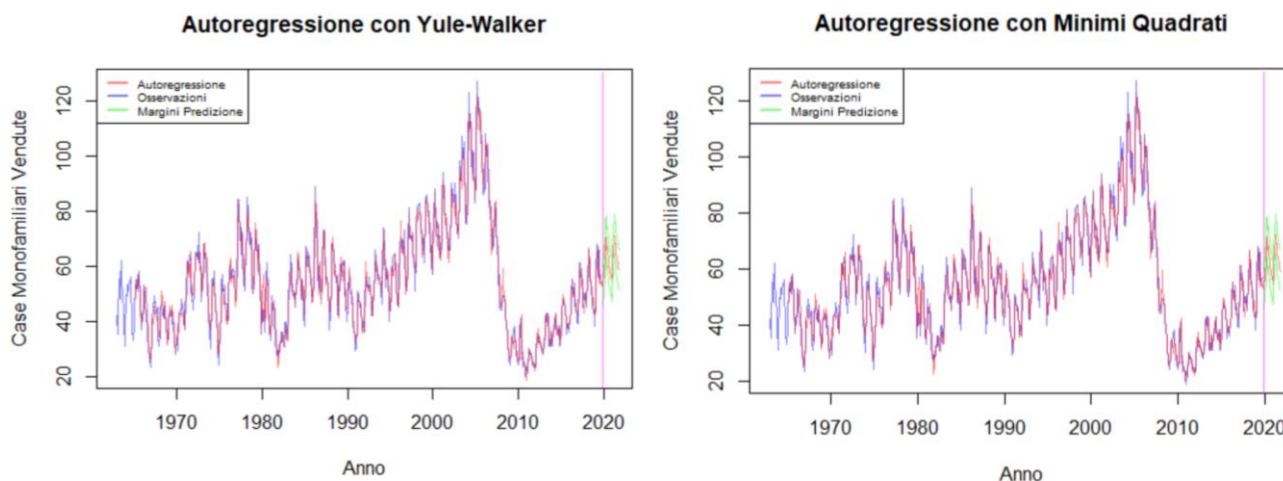


MODELLI DI AUTOREGRESSIONE

YULE-WALKER E MINIMI QUADRATI

I modelli di autoregressione calcolati tramite i metodi di *fitting* di *Yule-Walker* e *Minimi Quadrati* presentano delle differenze non significative, le analisi successive verranno perciò svolte soltanto sul modello *Yule-Walker*.

Il modello ha un andamento più regolare rispetto ai modelli di *Holt-Winters* e le **predizioni** sembrano essere e metà tra quelle precedenti, ma con un andamento positivo per i prossimi due anni.



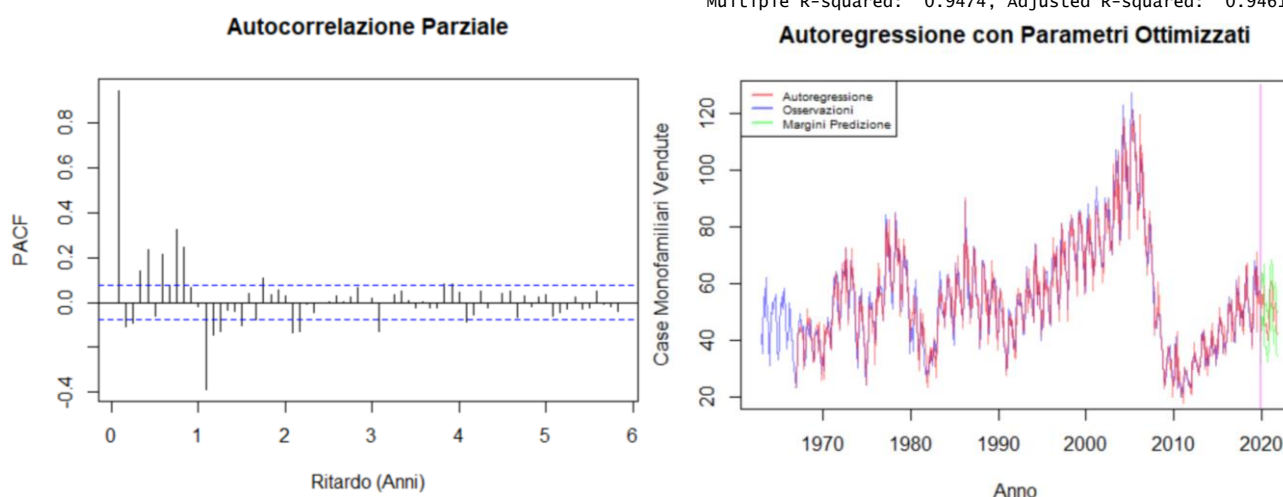
MANUALE

La *funzione di autocorrelazione parziale* evidenzia che il ritardo utile per la regressione arriva all'incirca a *26 mesi*, valore che è infatti stato preso dai modelli automatici. Volendo includere ogni possibile *lag* significativo, si può arrivare a considerare *46 mesi*, per poi mantenere solo quelli significativi per la regressione.

Le **predizioni** sui prossimi due anni risultano però essere molto più pessimistiche rispetto a quelle precedenti e il modello meno regolare, facendo dubitare della sua validità.

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	2.18888	0.66914	3.271	0.001130	**
X9	-0.15183	0.03713	-4.089	4.90e-05	***
X10	0.12088	0.04412	2.740	0.006330	**
X12	0.07045	0.02975	2.368	0.018172	*
X17	-0.06199	0.02915	-2.127	0.033828	*
X20	-0.11079	0.03253	-3.406	0.000703	***
X22	0.13583	0.03411	3.982	7.63e-05	***
X25	0.10331	0.03764	2.745	0.006224	**
X26	-0.17168	0.04483	-3.830	0.000141	***
X27	0.10654	0.04462	2.388	0.017243	*
X28	-0.10063	0.03809	-2.642	0.008454	**
X31	-0.06757	0.02963	-2.280	0.022920	*
X33	-0.19824	0.04443	-4.462	9.65e-06	***
X34	0.32566	0.03991	8.161	1.86e-15	***
X41	0.14920	0.02659	5.611	3.03e-08	***
X45	0.81212	0.02442	33.251	< 2e-16	***

Multiple R-squared: 0.9474, Adjusted R-squared: 0.9461



ACCURATEZZA

Per testare l'**accuratezza** dei modelli sono stati estratti gli ultimi *24 mesi* dalla serie originale in funzione di *testset*, come *trainingset* è stata utilizzata la parte rimanente della serie, i modelli sono stati testati con i parametri ottenuti precedentemente.

Si può notare come i modelli *Holt-Winters* (*ottimizzato* e *automatico*) e l'*autoregressione con Yule-Walker* abbiano previsioni molto simili, che rimangono all'interno dei **margini di previsione** reciproci. Il modello *autoregressivo calcolato manualmente* tende invece a sottostimare molto la serie, confermando i dubbi sull'accuratezza del modello sorti in precedenza.

Infine, si può notare come l'**errore medio** sia ottimo per tutti i modelli (escludendo il modello *autoregressivo manuale*): *HW Auto*. 4,89 mila case, *HW Ott*. 3,81 mila case, *Autoreg. YW* 3,71 mila case e *Autoreg. Manuale* con 14,37 mila case di scarto in media.

RESIDUI

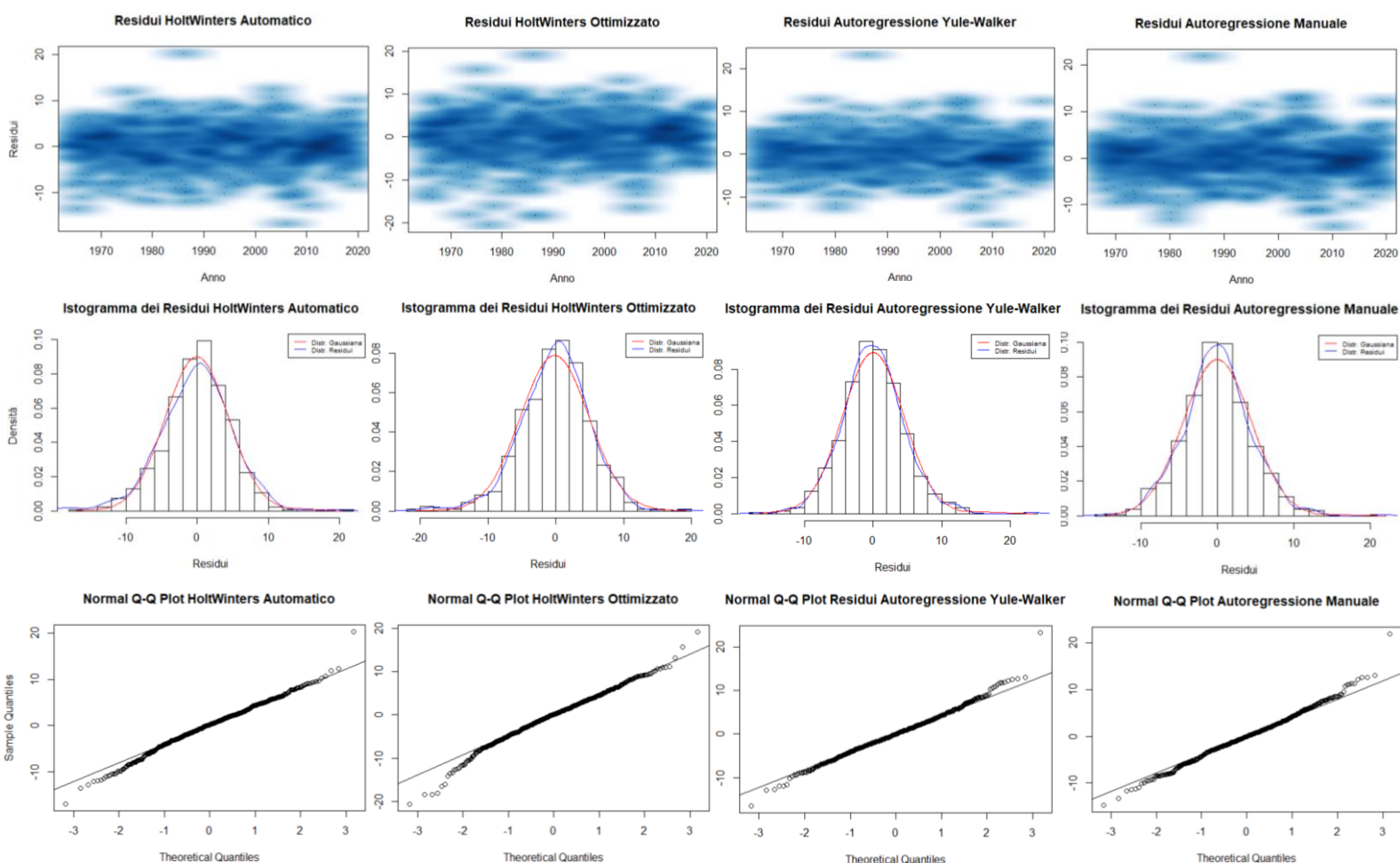
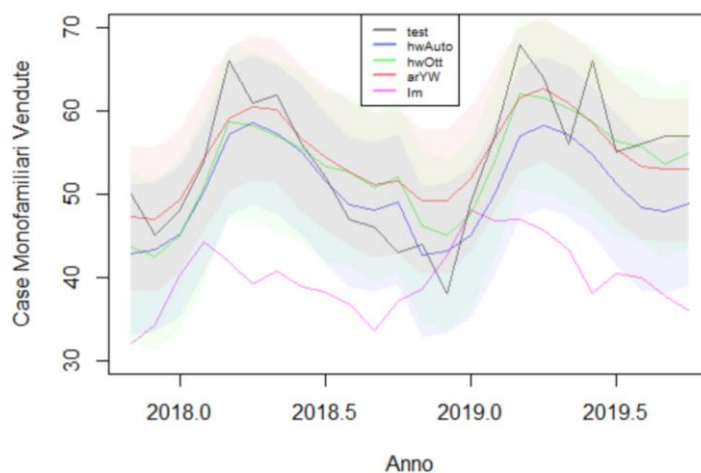
DISTRIBUZIONE

Dai grafici dei **residui** si può notare come non sembrano presenti grosse tracce di struttura residua al loro interno. In particolare, i residui dell'*autoregressione con Yule-Walker* sembrano essere ottimi. Da notare come i residui del modello di *autoregressione ottimizzato*, nonostante le scarse previsioni, risulti buono.

Si può notare come sia presente un **valore anomalo** delle previsioni nell'anno 1986, l'anomalia potrebbe essere spiegata dalla [riforma delle tasse](#) approvata in quell'anno, che ebbe un impatto consistente sul mercato immobiliare americano.

Le **varianze spiegate** non risultano significativamente diverse, con ottimi valori intorno al 94%.

Confronto Accuratezza Modelli

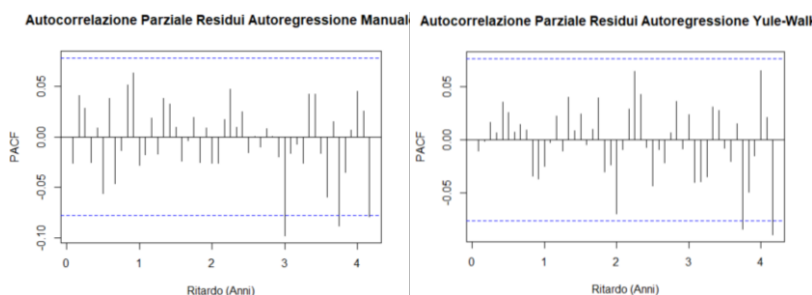


AUTOCORRELAZIONE

Le funzioni di **autocorrelazione parziale** dei residui mostrano come i *modelli autoregressivi* abbiano catturato bene la struttura della serie.

Le funzioni di **autocorrelazione** dei residui con i metodi di *Holt-Winters* mostrano la presenza di struttura residua, in particolar modo per il *modello ottimizzato*.

L'autocorrelazione dei residui dei *modelli autoregressivi* non sembrano invece evidenziare struttura residua.

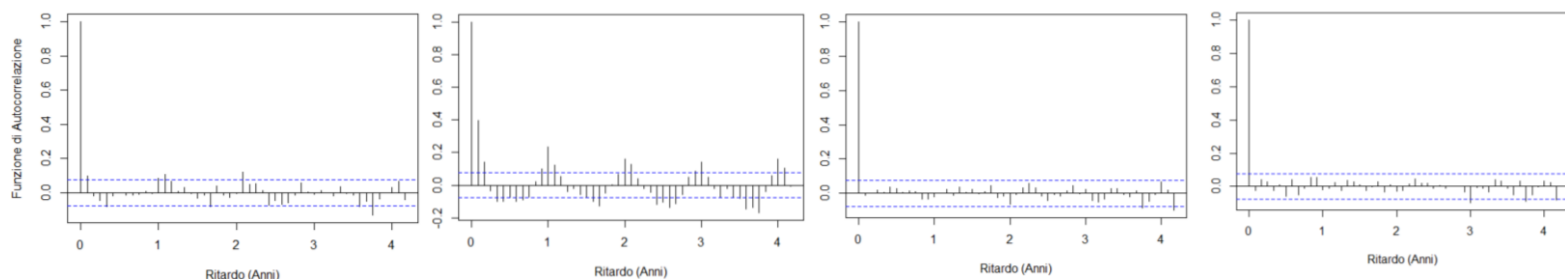


Autocorrelazione Residui HoltWinters Automatico

Autocorrelazione Residui HoltWinters Ottimizzato

Autocorrelazione Residui Autoregressione Yule-Walker

Autocorrelazione Residui Autoregressione Manuale



CONCLUSIONE

In seguito alle varie analisi il modello di *autoregressione con metodo di Yule-Walker* sembra essere quello con le migliori caratteristiche, inoltre le analisi evidenziano come le **previsioni** effettuate con questo metodo sembrano ottime.

Il **trend** della vendita di case monofamiliari nel mercato americano dei prossimi due anni, in accordo alle analisi svolte, prevede mediamente una **crescita**, anche se non è da escludere una possibile **stagnazione**.

La **stagionalità** risulta invece chiara: è presente un **aumento** di quasi **10 mila case vendute** rispetto alla media da **marzo fino a luglio**, mentre è presente un **calo** di uguale portata da **agosto a febbraio**.

Previsione

