

Alma Mater Studiorum Università di Bologna  
Polo Scientifico Didattico di Rimini  
Biblioteca Centralizzata

***Note e ricerche***  
***4/2009***

# Modelli previsionali univariati: i passeggeri negli aeroporti della Sardegna

Massimiliano Castellani\*, Pierpaolo Pattitoni†, Antonella Volgarino‡

30 luglio 2009

## Sommario

In questo lavoro confrontiamo la capacità previsionale di alcuni modelli di *Exponential Smoothing* e ARIMA stagionali. Distinguendo tra arrivi totali, nazionali e internazionali, prendiamo in esame il traffico passeggeri negli aeroporti della Sardegna nel periodo 2003-2009. Fra i modelli proposti, quelli di *Exponential Smoothing* registrano una migliore *performance* nel caso degli arrivi totali e nazionali. Il modello ARIMA, invece, offre una maggiore affidabilità nel caso degli arrivi internazionali.

*Jel classification:* C22, C53, L93

*Parole chiave:* serie storiche, previsioni, *Exponential Smoothing*, ARIMA, trasporti aerei.

*Ringraziamenti:* Desideriamo ringraziare Guido Candela e Simone Gianerini che hanno letto una versione preliminare di questo lavoro per i preziosi consigli forniteci durante la fase di stesura. Eventuali errori e omissioni rimangono tuttavia di nostra esclusiva responsabilità.

## 1 Introduzione

L'obiettivo di questo lavoro è valutare l'accuratezza delle previsioni turistiche fondate sui modelli di analisi delle serie storiche usando come dati gli arrivi nei principali aeroporti della Sardegna (Alghero, Cagliari e Olbia) dal 2003 al 2009. La scelta di utilizzare i dati degli arrivi mensili (numero di passeggeri) come *proxy* degli arrivi turistici è abbastanza frequente in letteratura, perché consente di utilizzare i dati raccolti da uno specifico Ente pubblico (ENAC) e di evitare le tipiche distorsioni dei dati registrati dagli altri soggetti privati

---

\*Università di Bologna - Dipartimento di Scienze Economiche e RCEA - m.castellani@unibo.it

†Università di Bologna - Facoltà di Economia, sede di Forlì e di Rimini - pierpaolo.pattitoni@unibo.it

‡Università di Bologna - Dipartimento di Discipline Giuridiche dell'Economia e dell'Azienda - antonella.volgarino@unibo.it.

(imprese ricettive). Inoltre, poiché gli unici mezzi di trasporto utilizzabili per raggiungere le località turistiche nelle isole sono appunto gli aerei e le navi, il numero dei turisti è facilmente approssimabile computando la percentuale (costante) di passeggeri che raggiunge le isole per motivi non turistici. Infatti, mentre il primo mezzo è utilizzato preferibilmente dai turisti che soggiornano per periodi non lunghi in *hotel*, il secondo mezzo è usato principalmente dai turisti che soggiornano per periodi non brevi in strutture ricettive non alberghiere (secondo case, *residence*, etc.). Infine, abbiamo scelto la Sardegna come oggetto di studio sia per la sua forte vocazione turistica sia per il notevole effetto sul PIL regionale della spesa dei turisti.

Pertanto l'importanza di prevedere gli arrivi è cruciale sia per gli agenti privati (*manager* delle imprese turistiche) sia per i *policy maker* (*destination manager*) che devono programmare gli investimenti, gestire il territorio e l'ambiente in generale. Queste previsioni sono ancora più importanti per le regioni fortemente dipendenti dal turismo come la Sardegna, in cui le risorse turistiche sono limitate e la capacità di carico antropico sulle risorse naturali costituisce una variabile strategica nell'offerta turistica.

Accurate forecasts are crucial for a wide range of policymaking situations, and are also necessary for efficient resource allocation in production, marketing, investment, and financial planning in the tourism industry.

Lim e McAleer [2001]

L'analisi delle serie storiche nell'approccio proposto da Box e Jenkins [1976] è stata ampiamente utilizzata per la previsione della domanda turistica. In particolare, diverse versioni dei modelli ARIMA sono state applicate in oltre due terzi degli studi per la previsione delle serie storiche. Infatti, sia il modello ARIMA sia l'ARIMA stagionale (SARIMA) vengono utilizzati in ambito turistico proprio perché la stagionalità è una caratteristica comune della domanda turistica. Esempio a questo proposito è lo studio di Vergori e Zamparini [2004] che, appunto, mette a confronto i due approcci appena descritti.

Per quanto riguarda la precisione delle previsioni, gli studi empirici presentano elementi controversi. Ad esempio, Cho [2003], esamina l'applicazione di tre tecniche di previsione delle serie storiche, ovvero l'*Exponential Smoothing* (ES), il modello ARIMA univariato e il modello a reti neurali artificiali (ANN), per prevedere la domanda di viaggio (numero di arrivi) ad Hong Kong, provenienti da sei diversi Paesi (Stati Uniti, Regno Unito, Singapore, Giappone, Taiwan e Corea del Sud). I risultati empirici suggeriscono una migliore *performance* dell'approccio ANN per tutte le serie, tranne per quella del Regno Unito, per la quale risulta migliore il metodo ES. Smeral e Wüger [2005] hanno rilevato che il modello ARIMA o SARIMA potrebbe essere superato dai modelli *Naïve*. Yorucu [2003] mette a confronto quattro metodi di previsione, cioè *Actual Static* (AS), *Double Exponential Smoothing* (DES), il metodo *Holt Winters* (HW) e i modelli ARMA, utilizzando i dati stagionali trimestrali degli arrivi turistici a Nord di Cipro e Malta, relativamente al periodo 1976-1995. Utilizzando diverse

misure di accuratezza, l'autore evidenzia come i modelli ARMA forniscano le previsioni più accurate.

Le evidenze empiriche in letteratura mostrano, quindi, come non esista in assoluto il migliore modello previsivo per le dinamiche dei flussi turistici. Questo proprio in virtù del fatto che, in generale, gli obiettivi conoscitivi di un modello previsivo sono diversi da quelli di un modello descrittivo.

Il lavoro è strutturato in cinque sezioni: nella sezione 2, seguendo l'impostazione e la notazione di Hyndman et al. [2008b], presentiamo i modelli adottati per le previsioni; nella sezione 3, attraverso indicatori descrittivi e sintetici, proponiamo alcuni fatti stilizzati relativi alla serie degli arrivi dei passeggeri negli aeroporti della Sardegna; nella sezione 4, riportiamo i principali risultati dell'analisi econometrica; nella sezione 5, infine, concludiamo.

## 2 Metodologia

Fra i diversi approcci alla previsione delle serie storiche, cinque sono comunemente adottati nella ricerca empirica [Gujarati, 2004]: (1) metodi di *Exponential Smoothing*, (2) modelli di regressione mono-equazionali, (3) modelli di regressione con equazioni simultanee, (4) modelli ARIMA (e ARIMA stagionali), e (5) modelli VAR. In questo lavoro, ci focalizziamo sui modelli della classe ES e ARIMA poiché, a fronte della semplicità di impiego che li caratterizza, si prestano bene a modellare serie storiche che esibiscono *trend* e stagionalità. Nel seguito, proponiamo una breve sintesi delle loro principali caratteristiche, ma rimandiamo a Hyndman et al. [2008b] per una trattazione approfondita.

L'espressione *Exponential Smoothing* fa riferimento a una classe di metodi previsionali adattivi in cui alle osservazioni passate di una serie storica, ordinate in senso temporale decrescente, è assegnato un peso che si riduce in modo esponenziale e che prevedono una struttura basata su un'equazione per il livello della serie ( $\ell_t$ ), una per la sua crescita ( $b_t$ ) e una per la componente stagionale ( $s_t$ ). Taylor [2003] e Hyndman e Khandakar [2008] propongono una classificazione che individua quindici metodi ES come interazione tra la componente di *trend* (assente, additiva, additiva smorzata, moltiplicativa e moltiplicativa smorzata) e la componente stagionale (assente, additiva e moltiplicativa). Per ciascuno dei quindici metodi individuano, poi, due modelli nello spazio degli stati: il primo con errori additivi, il secondo con errori moltiplicativi. E' facile dimostrare che, mentre la scelta fra errore additivo o moltiplicativo non ha conseguenze sulla previsione puntuale, la stessa scelta è importante per l'identificazione degli intervalli di previsione. Definendo con  $y_t$  il valore della serie storica al tempo  $t$  e con  $m$  la lunghezza della stagionalità, la struttura generale per ognuno dei trenta modelli (quindici con errore additivo e quindici con errore moltiplicativo) prevede un vettore delle variabili di stato  $\mathbf{x}_t = (\ell_t, b_t, s_t, s_{t-1}, \dots, s_{t-m+1})'$  che descrive il livello, il *trend* e la stagionalità della serie, e delle equazioni nello spazio degli stati:

$$\begin{aligned}
y_t &= w(\mathbf{x}_{t-1}) + r(\mathbf{x}_{t-1}) \varepsilon_t \\
\mathbf{x}_t &= \mathbf{f}(\mathbf{x}_{t-1}) + \mathbf{g}(\mathbf{x}_{t-1}) \varepsilon_t
\end{aligned}
\tag{1}$$

dove  $\varepsilon_t$  è un termine di errore indipendente e identicamente distribuito secondo una distribuzione normale con valore atteso nullo e varianza costante,  $\varepsilon_t \sim NID(0, \sigma^2)$ ,  $w(\cdot)$  e  $r(\cdot)$  sono funzioni scalari e  $\mathbf{f}(\cdot)$  e  $\mathbf{g}(\cdot)$  sono funzioni vettoriali. Per ottenere gli intervalli di previsione dei modelli proposti, è possibile simulare diversi *sample path* per  $\{y_{n+1} + \dots + y_{n+h}\}$  e trovare i percentili  $\alpha/2$  e  $1 - \alpha/2$  dei valori simulati per ciascun orizzonte temporale. Usando questo approccio, i *sample path* possono essere generati ipotizzando una distribuzione normale per gli errori (*bootstrap* parametrico), o attraverso metodi di ricampionamento (*bootstrap*). Inoltre, mentre nel caso di modelli lineari è possibile calcolare gli intervalli di previsione in modo analitico, questa soluzione non è sempre applicabile al caso dei modelli non lineari e, per questa ragione, l'approccio *bootstrap* è spesso quello preferito [Hyndman et al., 2002]. Al fine di utilizzare i modelli proposti per la previsione è necessario, quindi, (1) identificare il modello da utilizzare, (2) specificare il valore iniziale per gli elementi del vettore  $\mathbf{x}_t$  ( $\mathbf{x}_0$ ) e (3) stimare il valore dei parametri del modello. Rispetto al primo punto, Hyndman et al. [2002] propongono, come metodo di selezione fra modelli diversi, l'utilizzo di criteri basati sulla teoria dell'informazione quali l'AIC [Akaike, 1974] e il BIC [Schwarz, 1978]. Gli stessi autori, inoltre, indicano una procedura per determinare  $\mathbf{x}_0$ : gli elementi di questo vettore sono inizialmente determinati attraverso regole euristiche e poi definiti in modo più preciso durante la stima dei parametri. Rispetto al terzo punto, infine, è possibile ottenere una stima dei parametri del modello,  $\boldsymbol{\theta}$ , e dei suoi stati iniziali,  $\mathbf{x}_0$ , utilizzando come criteri per la stima il metodo della massima verosimiglianza, la varianza dei residui del modello, o una qualsiasi misura di accuratezza delle previsioni.

I modelli  $ARIMA(p, d, q)(P, D, Q)_m$  stagionali, proposti da Box e Jenkins [1976] per la previsione delle serie storiche stagionali, sono generalmente rappresentati come segue:

$$\phi_p(L) \Phi_P(L^m) (1 - L)^d [1 - L^m]^D y_t = \theta_q(L) \Theta_Q(L^m) \varepsilon_t
\tag{2}$$

dove  $m$  è la lunghezza della stagionalità,  $\varepsilon_t$  è un termine di errore  $\varepsilon_t \sim NID(0, \sigma^2)$ ,  $L$  è l'operatore ritardo,  $\phi_p(L)$ ,  $\theta_q(L)$ ,  $\Phi_P(L^m)$  e  $\Theta_Q(L^m)$  sono polinomi in  $L$  di ordine  $p$ ,  $q$ ,  $P$  e  $Q$ . Infine,  $d$  e  $D$  indicano il numero di differenze (regolari e stagionali) che assicurano la stazionarietà della serie. Per determinare l'ordine di integrazione regolare ( $d$ ) e stagionale ( $D$ ), si dovrebbero utilizzare *test* di radice unitaria:  $D$ , può essere stabilito in base all'esito del *test* di Canova e Hansen [1995] sulla serie non differenziata; mentre  $d$  può essere determinato in base al *test* di radice unitaria KPSS [Kwiatkowski et al., 1992] o ADF [Said and Dickey, 1984] sulla serie eventualmente già differenziata a livello stagionale. Dopo aver scelto il grado di differenziazione regolare ( $d$ ) e stagionale ( $D$ ) da applicare alla serie, è possibile determinare l'ordine del polinomio autoregressivo

regolare ( $p$ ) e a media mobile ( $q$ ) e le loro controparti stagionali ( $P$  e  $Q$ ) in base a criteri basati sulla teoria dell'informazione quali, per esempio, l'AIC [Akaike, 1974] e il BIC [Schwarz, 1978], applicabili se si usa il metodo della massima verosimiglianza.

L'accuratezza delle previsioni ottenute attraverso l'applicazione dei metodi proposti può essere misurata attraverso opportuni indicatori. Alcuni di questi, come il *Mean Absolute Error* (MAE) sono abbastanza intuitivi, ma dipendono dall'unità di misura della serie osservata. Un parziale miglioramento è costituito dal *Mean Absolute Percentage Error* (MAPE), che, esprimendo l'errore di previsione in termini percentuali, rende la misurazione dell'accuratezza indipendente dall'unità di misura della serie osservata. Infine, un indicatore generale, che può essere adottato virtualmente in ogni circostanza, è il MASE (Hyndman e Koehler, 2006), che ha la caratteristica di permettere il confronto tra previsioni ottenute con metodi diversi anche su serie differenti e indipendentemente dall'unità di misura in cui queste ultime sono espresse. Con tale indicatore, infatti, gli errori di previsione sono riscaldati sulla base del MAE *in-sample* riferito a un metodo *Naïve* in cui la previsione viene fatta coincidere con il valore precedente della serie.

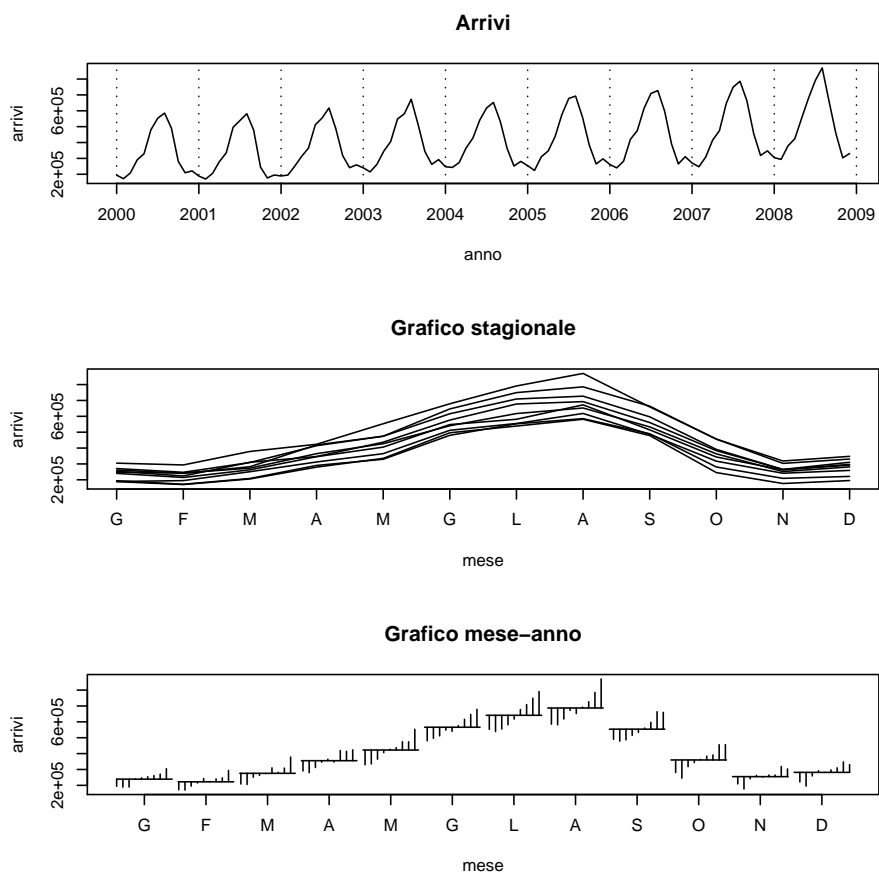
### 3 Alcuni fatti stilizzati

In questa sezione proponiamo un'analisi descrittiva della serie degli arrivi totali interstagionali presso gli aeroporti della Sardegna, allo scopo di mettere in evidenza alcuni dei fatti stilizzati che generalmente caratterizzano le serie storiche degli arrivi negli aeroporti in località a spiccata vocazione turistica. L'analisi è riferita agli arrivi totali, ma scomponendo la serie tra arrivi nazionali e internazionali, otteniamo risultati qualitativamente comparabili e quindi abbiamo deciso di non riportarli per brevità.

La figura 1 riporta, per il periodo 2000-2009, la serie storica degli arrivi totali negli aeroporti di Alghero, Cagliari e Olbia, il grafico stagionale e il grafico mese-anno. Una prima ispezione grafica della serie mette in evidenza la presenza di una marcata stagionalità. Questa evenienza è confermata, inoltre, dal grafico mese-anno e dal grafico stagionale, che, rappresentando l'andamento della serie nei singoli anni, mettono in luce alcune regolarità. In particolare, gli arrivi sono concentrati nei mesi estivi, specialmente nel mese di agosto, mentre registrano i valori minimi nei mesi invernali. Inoltre, il grafico mese-anno mostra l'esistenza di un *trend* positivo negli arrivi osservati a distanza di un anno a parità di mese. Per quanto riguarda i mesi estivi, tale circostanza potrebbe essere attribuita a un piano di sviluppo della domanda turistica promosso dalla regione Sardegna. Nei mesi invernali, invece, la stessa potrebbe dipendere, almeno in parte, dall'introduzione della cosiddetta "continuità territoriale", ossia un piano di tariffe agevolate per il trasporto aereo dei residenti.

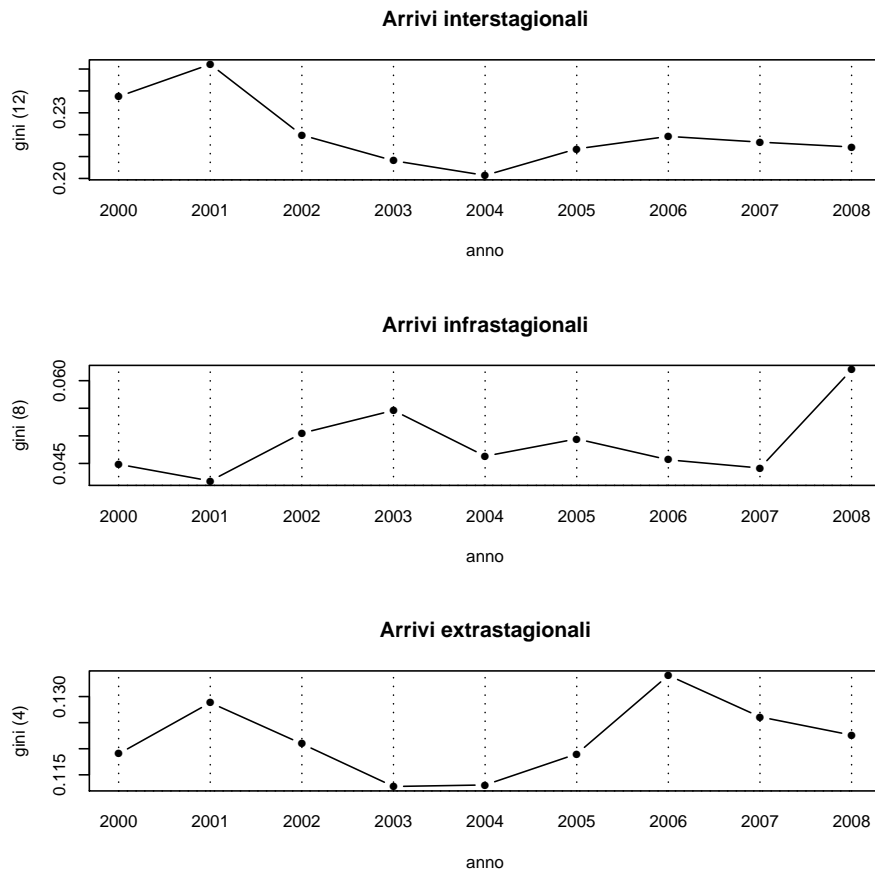
La figura 2 mostra l'evoluzione del coefficiente di Gini per gli arrivi totali interstagionali (da Gennaio a Dicembre), infrastagionali (da Giugno a Settembre) e extrastagionali (da Gennaio a Maggio e da Ottobre a Dicembre) pres-

Figura 1: Serie degli arrivi totali interstagionali  
Arrivi, Grafico stagionale e Grafico mese-anno<sup>a</sup>



<sup>a</sup>Il primo grafico fa riferimento agli arrivi totali interstagionali presso gli aeroporti di Alghero, Cagliari e Olbia nel periodo 2000-2009. Il grafico stagionale e il grafico mese-anno sono costruiti considerando le singole serie annuali nell'intero arco temporale considerato.

Figura 2: Coefficiente di Gini per gli arrivi totali  
 Arrivi interstagionali, Arrivi infrastagionali e Arrivi extrastagionali<sup>a</sup>



<sup>a</sup>La figura presenta l'evoluzione del coefficiente di Gini per gli arrivi totali interstagionali (da Gennaio a Dicembre), infrastagionali (da Giugno a Settembre) e extrastagionali (da Gennaio a Maggio e da Ottobre a Dicembre) presso gli aeroporti di Alghero, Cagliari e Olbia nel periodo 2000-2009.



so gli aeroporti della Sardegna. Coerentemente con le nostre attese, nel caso degli arrivi infrastagionali e extrastagionali i coefficienti di Gini assumono un valore minore rispetto a quello calcolato per gli arrivi interstagionali. Questa evidenza, tipica per le località turistiche, è da attribuire a una maggiore omogeneità della distribuzione degli arrivi intrastagionali e extrastagionali rispetto agli arrivi interstagionali. Sottolineiamo, inoltre, un andamento decrescente nel tempo del coefficiente di Gini per gli arrivi interstagionali. Questo suggerisce una riduzione della stagionalità nel tempo, dovuta, probabilmente, all'implementazione, da parte dell'amministrazione pubblica in Sardegna, di politiche di destagionalizzazione.

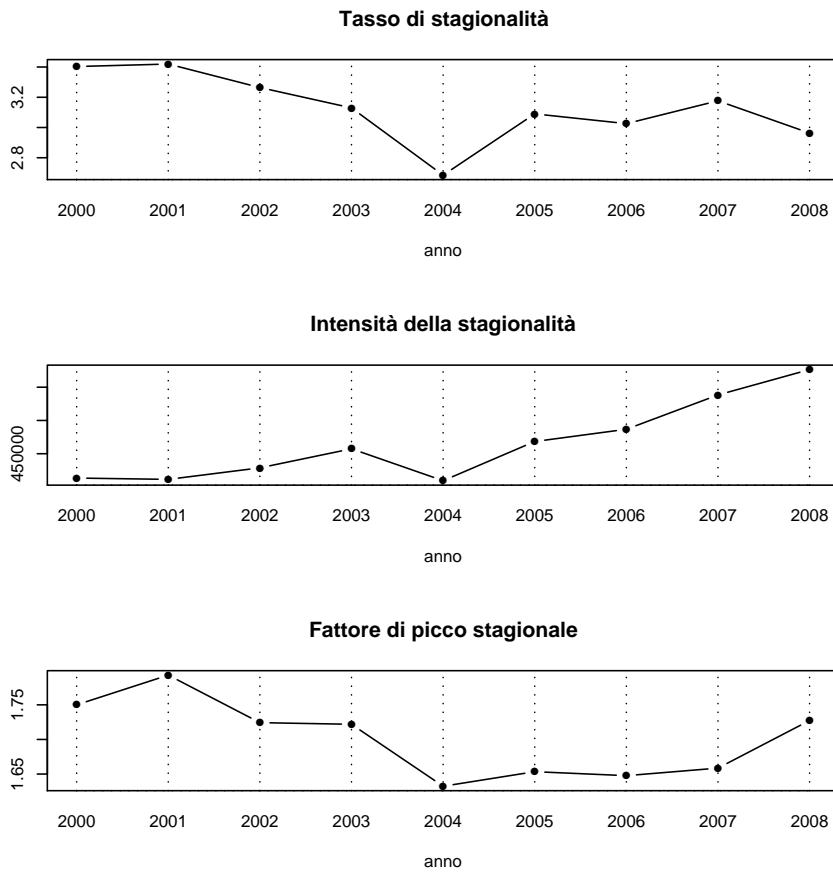
La figura 3 presenta alcuni indici di stagionalità. In particolare, riportiamo il tasso di stagionalità, calcolato come rapporto tra il numero massimo e minimo di arrivi durante l'anno considerato, l'intensità della stagionalità, ottenuta dalla loro differenza, e il fattore di picco stagionale, che coincide con il rapporto tra il numero massimo e medio di arrivi durante l'anno. Gli indici proposti denotano un andamento annuo decrescente fino al 2004 e crescente in seguito. Tale evidenza suggerisce che ci sia stato un cambiamento nelle politiche di destagionalizzazione. Tuttavia, occorre una certa cautela prima di trarre conclusioni dall'osservazione di questi indici, poiché queste misure considerano solo gli estremi della distribuzione degli arrivi e non l'intera distribuzione.

## 4 Risultati

In questa sezione, presentiamo i risultati dell'applicazione dei modelli previsionali alle serie degli arrivi interstagionali totali, nazionali e internazionali presso gli aeroporti della Sardegna. Innanzitutto, precisiamo che, considerando le conseguenze dell'undici settembre sul traffico aereo mondiale, abbiamo deciso, in via prudenziale, di limitare il nostro campione al periodo 2003-2009. Inoltre, al fine di valutare la capacità previsionale dei modelli stimati, abbiamo suddiviso l'arco temporale considerato in una finestra di stima (2003-2008), in cui selezioniamo il modello, e una finestra di previsione (2008-2009), in cui confrontiamo i valori previsti con quelli effettivamente realizzati.

La tabella 1 riporta i risultati dell'applicazione dei metodi ES alle serie degli arrivi totali, nazionali e internazionali. I modelli riportati, stimati con il metodo della massima verosimiglianza, sono selezionati in base al BIC [Schwarz, 1978]. Nel caso degli arrivi totali e degli arrivi nazionali, il modello selezionato prevede una componente erratica e una componente stagionale moltiplicativa, ma non una componente di *trend*. Indicando con  $\mu_t = E(y_t | \mathbf{x}_{t-1})$  la previsione a un passo di  $y_t$  e con  $\varepsilon_t = \mu_t - y_t$  l'errore di previsione, le equazioni nello spazio degli stati per quest'ultimo modello sono:

Figura 3: Indicatori di stagionalità per gli arrivi totali interstagionali  
 Tasso di stagionalità, Intensità della stagionalità e Fattore di picco stagionale<sup>a</sup>



<sup>a</sup>La figura presenta l'evoluzione del tasso di stagionalità, dell'intensità della stagionalità e del fattore di picco stagionale per gli arrivi totali interstagionali presso gli aeroporti di Alghero, Cagliari e Olbia nel periodo 2000-2009. Il tasso di stagionalità è calcolato come rapporto tra il numero massimo e minimo di arrivi durante l'anno considerato. L'intensità della stagionalità è ottenuta, invece, dalla loro differenza. Il fattore di picco stagionale, infine, coincide con il rapporto tra il numero massimo e medio di arrivi durante l'anno.

Tabella 1: Exponential Smoothing

Arrivi <sup>a</sup>	Totali	Nazionali	Internazionali
Errore	Moltiplicativo	Moltiplicativo	Moltiplicativo
Trend	-	-	Additivo
Stagionalità	Moltiplicativa	Moltiplicativa	Moltiplicativa
$\alpha$	0.5146	0.4930	0.4589
$\beta$	-	-	0.0461
$\gamma$	0.0001	0.0001	0.0001
AIC	1425.33	1375.87	1346.88
BIC	1454.66	1405.19	1380.39
<i>mae</i>	18147.59	12763.67	49409.63
<i>mape</i>	3.86	3.79	35.38
<i>mase</i>	0.22	0.22	1.89

<sup>a</sup>La tabella presenta i risultati dell'applicazione del metodo di *Exponential Smoothing* per la previsione degli arrivi passeggeri totali, nazionali e internazionali presso gli aeroporti di Alghero, Cagliari e Olbia. Il periodo 2003-2009 è suddiviso in una finestra di stima, 2003-2008, e una finestra di previsione, 2008-2009. Il modello riportato per ogni serie è selezionato attraverso il *Bayesian Information Criterion*. Per ogni modello sono prospettate le stime dei parametri di *smoothing*, ottenute con il metodo della massima verosimiglianza, l'AIC e il BIC. In calce alla tabella, infine, sono presentate alcune misure di accuratezza (MAE, MASE e MAPE) riferite agli errori di previsione *out-of-sample*. Non riportiamo gli *standard error* dei coefficienti in quanto, nel nostro lavoro, i modelli proposti hanno scopo puramente previsivo.

$$\begin{aligned}
\mu_t &= \ell_{t-1} s_{t-m} \\
\ell_t &= \ell_{t-1} (1 + \alpha \varepsilon_t) \\
s_t &= s_{t-m} (1 + \gamma \varepsilon_t)
\end{aligned} \tag{3}$$

dove per gli arrivi totali i coefficienti di *smoothing*  $\alpha$  e  $\gamma$  assumono il valore di 0.5146 e 0.0001, mentre per gli arrivi nazionali gli stessi sono pari a 0.4930 e 0.0001. Nel caso degli arrivi internazionali, il modello selezionato, che prevede una componente di errore e di stagionalità moltiplicativa e una componente di *trend* additiva, è caratterizzato dalle seguenti equazioni nello spazio degli stati:

$$\begin{aligned}
\mu_t &= (\ell_{t-1} + b_{t-1}) s_{t-m} \\
\ell_t &= (\ell_{t-1} + b_{t-1}) (1 + \alpha \varepsilon_t) \\
b_t &= b_{t-1} + \beta (\ell_{t-1} + b_{t-1}) \varepsilon_t \\
s_t &= s_{t-m} (1 + \gamma \varepsilon_t)
\end{aligned} \tag{4}$$

dove i coefficienti di *smoothing*  $\alpha$ ,  $\beta$  e  $\gamma$  assumono rispettivamente i valori 0.4589, 0.0461 e 0.0001.

Prima di selezionare i modelli della classe ARIMA più idonei a descrivere e prevedere le serie degli arrivi, consideriamo gli esiti di un'analisi preliminare. In primo luogo, al fine di individuare l'eventuale presenza di eteroschedasticità nelle serie considerate e decidere, quindi, una loro opportuna trasformazione per stabilizzare le fluttuazioni della varianza, sottoponiamo i residui delle regressioni ausiliarie per lo studio del *trend* di ciascuna serie (arrivi totali, nazionali e internazionali) al *test* di Breusch e Pagan [1979] e rifiutiamo, in tutti i casi, l'ipotesi di eteroschedasticità. Tuttavia, l'ispezione grafica della serie degli arrivi internazionali evidenzia un incremento nel tempo delle oscillazioni della serie che rende conveniente una sua trasformazione logaritmica. In un secondo momento, l'esito del *test* di Canova e Hansen [1995] suggerisce di propendere, in tutte le circostanze, per l'adozione di *dummy* stagionali piuttosto che per una radice unitaria stagionale. Infine, per verificare l'eventuale presenza di *trend* regolare stocastico o deterministico, consideriamo, oltre alle indicazioni emerse dalle regressioni ausiliarie per lo studio del *trend*, l'esito del *test* KPSS [Kwiatkowski et al., 1992]: nel caso degli arrivi totali e internazionali riscontriamo la presenza di un *trend* lineare, ma, dopo averne tenuto conto, non rifiutiamo l'ipotesi di stazionarietà; per gli arrivi nazionali, invece, pur non evidenziando la presenza di un *trend* deterministico nella serie, rifiutiamo l'ipotesi di stazionarietà della stessa. In quest'ultimo caso, dunque, stimiamo il modello nelle differenze prime. Considerando il numero ridotto di osservazioni nel periodo di stima, limitiamo a tre ritardi l'ordine massimo della memoria deterministica (AR) e stocastica (MA) regolare e a due ritardi quella delle loro controparti stagionali (SAR e SMA), selezionando per le previsioni, nella classe dei modelli considerata, quelli che rendono minimo il BIC [Schwarz, 1978].

Tabella 2: ARIMA

Arrivi <sup>a</sup>	Totali		Nazionali		Internazionali	
const.	248557	(12498)	-	-	9.90	(0.07)
trend	1668	(266)	-	-	0.02	(0.00)
Gen	-35434	(9229)	-36110	(5706)	-0.03	(0.04)
Feb	-58078	(10979)	-53414	(5168)	-0.21	(0.04)
Mar	-6048	(11771)	-14247	(6155)	0.19	(0.05)
Apr	83047	(12143)	40319	(6063)	0.79	(0.05)
Mag	146954	(12299)	64689	(6413)	1.23	(0.05)
Giu	286920	(12322)	180955	(6270)	1.39	(0.05)
Lug	366789	(12232)	236435	(6365)	1.52	(0.05)
Ago	404251	(11998)	263967	(5933)	1.56	(0.05)
Sett	270759	(11524)	144362	(6011)	1.47	(0.05)
Ott	83355	(10577)	26561	(4843)	0.90	(0.04)
Nov	-32791	(8539)	-26474	(5482)	-0.14	(0.03)
AR(1)	0.56	(0.11)	-0.58	(0.10)	0.70	(0.10)
SAR(1)	-	-	-	-	-0.44	(0.13)
log	no		no		si	
d	0		1		0	
D	0		0		0	
AIC	1368.68		1282.59			
BIC	1400.09		1309.60		-49.07	
<i>mae</i>	33947.40		14354.60		0.27	
<i>mape</i>	6.31		3.75		2.29	
<i>mase</i>	0.41		0.25		0.80	

<sup>a</sup>La tabella presenta i risultati dell'applicazione del metodo ARIMA per la previsione degli arrivi passeggeri totali, nazionali e internazionali presso gli aeroporti di Alghero, Cagliari e Olbia. Il periodo 2003-2009 è suddiviso in una finestra di stima, 2003-2008, e una finestra di previsione, 2008-2009. Il modello riportato per ogni serie è selezionato attraverso il *Bayesian Information Criterion*. Per ogni modello sono prospettate le stime dei parametri, ottenute con il metodo della massima verosimiglianza, gli *standard error* fra parentesi, l'eventuale trasformazione logaritmica della serie, l'ordine di integrazione regolare (d) e stagionale (D), l'AIC e il BIC. I coefficienti in corsivo non risultano significativamente diversi da zero considerando una dimensione del *test* pari al 10%. In calce alla tabella, infine, sono presentate alcune misure di accuratezza (MAE, MASE e MAPE) riferite agli errori di previsione *out-of-sample*.

Nella tabella 2, presentiamo i risultati dei modelli ARIMA applicati alle serie degli arrivi totali, nazionali e internazionali. Prima di interpretare i risultati, sottolineiamo che il *test* di Ljung e Box [1978] sui residui dei modelli stimati non evidenzia, in nessun caso, presenza di correlazione seriale. Inoltre, il *test* di Wilk e Shapiro [1965] porta a rifiutare, solo nel caso degli arrivi internazionali, l'ipotesi di processo generatore gaussiano per gli errori di regressione, suggerendo una certa prudenza nell'interpretare i risultati di questo modello.

Nel caso degli arrivi totali, sia il *trend* lineare sia le *dummy* stagionali (ad eccezione di quella associata al mese di marzo), risultano statisticamente significativi. Il modello prevede, inoltre, una memoria deterministica positiva al primo ritardo. Il modello per gli arrivi nazionali, stimato nelle differenze prime per tenere conto della presenza di un *trend* stocastico, pur essendo caratterizzato da *dummy* stagionali statisticamente significative, presenta, a differenza del primo, un effetto autoregressivo di primo ordine negativo. Infine, il modello per gli arrivi internazionali, stimato nei logaritmi per stabilizzare la variabilità della serie, esibisce un *trend* lineare e *dummy* stagionali statisticamente significativi (ad esclusione della *dummy* relativa al mese di gennaio). Per quest'ultimo modello, tuttavia, oltre alla memoria deterministica positiva osservata al primo ritardo, riscontriamo anche un effetto autoregressivo stagionale negativo di primo ordine.

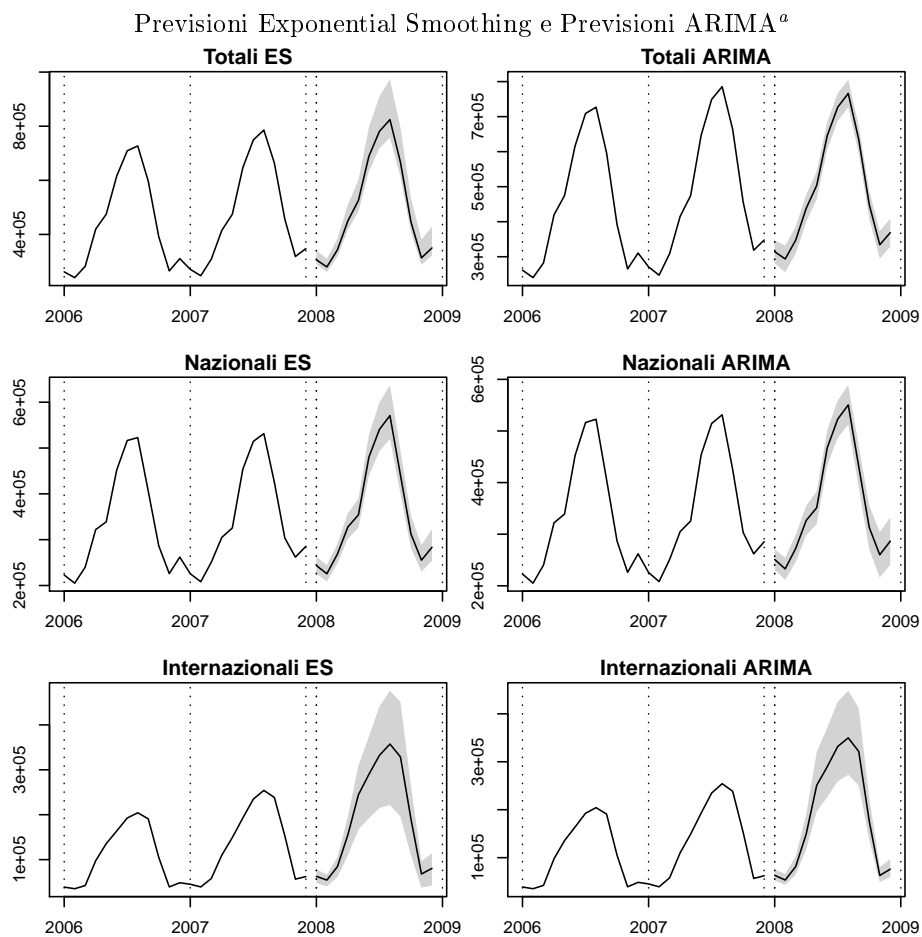
Per confrontare la capacità previsionale dei modelli ES e ARIMA, osserviamo il MASE, indicatore che Hyndman e Koehler [2006] propongono come *standard* fra le misure di accuratezza delle previsioni. Nel caso degli arrivi totali e nazionali, riscontriamo una migliore *performance* dei modelli ES, mentre, considerando gli arrivi internazionali, il modello ES ha una ridotta capacità previsionale. Confrontando le previsioni di quest'ultimo modello e quelle del modello ARIMA, opportunamente riconvertite nei livelli, con la serie degli arrivi internazionali, lo scostamento dalle effettive realizzazioni è minore nel secondo caso. Per gli arrivi internazionali, quindi, propendiamo per l'utilizzo del modello ARIMA, in quanto sembra offrire una maggiore affidabilità nelle previsioni. Segnaliamo infine che, come misura di controllo, abbiamo trasformato anche le serie degli arrivi totali e nazionali nei logaritmi e ristimato i modelli. Tuttavia, nonostante una struttura di ritardi nei modelli ARIMA molto più articolata, la capacità previsa degli stessi è solo parzialmente migliorata.

La figura 4, infine, riporta i grafici delle previsioni per il periodo 2008-2009 ottenute rispettivamente con i modelli ES (prima colonna) e ARIMA (seconda colonna). Oltre alle previsioni puntuali, nei grafici è riportata un'area ombreggiata che rappresenta gli intervalli di previsione al 95%: nel caso ES, gli stessi sono ottenuti attraverso il metodo *bootstrap*; nel caso degli ARIMA, invece, attraverso il tradizionale approccio analitico.

## 5 Conclusioni

In questo lavoro abbiamo valutato l'accuratezza delle previsioni turistiche basate sui modelli di analisi delle serie storiche, scegliendo come dati gli arrivi nei principali aeroporti della Sardegna (Alghero, Cagliari e Olbia) dal 2003 al 2009. In

Figura 4: Previsioni



<sup>a</sup>I grafici presentano le previsioni degli arrivi totali, nazionali e internazionali presso gli aeroporti di Alghero, Cagliari e Olbia e i rispettivi intervalli al 95% (area ombreggiata), ottenute sulla base dei modelli ES e ARIMA presentati nelle tabelle 1 e 2.

particolare, utilizzando il numero dei passeggeri come *proxy* degli arrivi turistici, abbiamo testato diversi modelli della classe *Exponential Smoothing* e ARIMA.

In primo luogo, un'analisi descrittiva ha permesso di evidenziare anche per la serie degli arrivi in Sardegna alcuni fatti stilizzati tipici delle località a spiccata vocazione turistica. Attraverso un'ispezione grafica preliminare e diversi indici di stagionalità, abbiamo rilevato la presenza *trend* e stagionalità nelle serie osservate. Inoltre, abbiamo riscontrato una riduzione della stagionalità nel tempo, dovuta, probabilmente, all'implementazione, da parte della regione Sardegna, di politiche di destagionalizzazione attuate mediante un piano di sviluppo della domanda turistica e attraverso l'introduzione della cosiddetta "continuità territoriale".

In un secondo momento, abbiamo considerato diversi modelli della classe ES e ARIMA per la previsione degli arrivi, distinguendo fra arrivi totali, nazionali e internazionali. Per quanto riguarda i modelli ES, nel caso degli arrivi totali e nazionali, il modello selezionato prevede una componente erratica e una componente stagionale moltiplicativa, ma non una componente di *trend*. Gli arrivi internazionali, invece, sono caratterizzati da una componente di errore e di stagionalità moltiplicativa e una componente di *trend* additiva. Per quanto attiene, invece, i modelli della classe ARIMA, nel caso degli arrivi totali il modello prevede una memoria deterministica positiva al primo ritardo. Il modello per gli arrivi nazionali, stimato nelle differenze prime per tenere conto di una radice unitaria regolare, presenta, invece, un effetto autoregressivo di primo ordine negativo. Infine, il modello per gli arrivi internazionali, stimato nei logaritmi per stabilizzare le fluttuazioni della serie nel tempo, esibisce una memoria deterministica positiva al primo ritardo e un effetto autoregressivo stagionale negativo di primo ordine.

Considerando la capacità previsionale dei modelli selezionati, nel caso degli arrivi totali e nazionali, riscontriamo una migliore *performance* dei modelli ES, mentre, per gli arrivi internazionali, il modello ARIMA sembra offrire una maggiore affidabilità nelle previsioni.

I modelli utilizzati per l'analisi empirica costituiscono, dunque, validi strumenti per la previsione degli arrivi. Previsioni accurate sono utili per la definizione di politiche economiche da parte dei soggetti pubblici (Stato, Regione, etc.) e per una valutazione accurata delle diverse opzioni di investimento da parte dei soggetti economici privati (imprese turistiche, imprese di trasporto, etc.). Inoltre per i soggetti pubblico-privati, come l'ENAC o le società aeroportuali, tali modelli potrebbero rappresentare utili ed efficaci strumenti di pianificazione e controllo delle politiche di trasporto aereo. Tuttavia, è noto che la frequenza di una serie storica influenza la scelta del modello. Se questo è il caso degli arrivi in Sardegna, ricerche future potrebbero verificare se modelli non lineari siano più appropriati quando si analizzano dati giornalieri.



## Riferimenti bibliografici

- H. Akaike. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19:716–723, 1974.
- G.E.P. Box and G.M. Jenkins. *Time series analysis: forecasting and control*. Holden-Day, San Francisco, Calif., revised edition, 1976. Holden-Day Series in Time Series Analysis.
- T. Breusch and A. Pagan. A simple test for heteroscedasticity and random coefficient variation. *Econometrica*, 47:1287–1294, 1979.
- F. Canova and B.E. Hansen. Are seasonal patterns constant over time? A test for seasonal stability. *Journal of Business & Economic Statistics*, 13:237–252, 1995.
- V. Cho. A comparison of three different approaches to tourist arrival forecasting. *Tourism Management*, 24:323–330, 2003.
- D.N. Gujarati. *Basic econometrics*. McGraw-Hill, 2004.
- R.J. Hyndman and A.B. Koehler. Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22(4):679–688, 2006.
- R.J. Hyndman, A.B. Koehler, R.D. Snyder, and S. Grose. A state space framework for automatic forecasting using exponential smoothing methods. *International Journal of Forecasting*, 18(3):439–454, 2002.
- R.J. Hyndman, A.B. Koehler, K. Ord, and R.D. Snyder. *Forecasting with exponential smoothing: the state space approach*. Springer, 2008b.
- Hyndman, R.J. and Khandakar, Y. Automatic time series forecasting: the forecast package for R. *Journal of Statistical Software*, 27(3):1–22, 3 2008. ISSN 1548-7660. URL <http://www.jstatsoft.org/v27/i03>.
- D. Kwiatkowski, P.C.B. Phillips, P. Schmidt, and Y. Shin. Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of econometrics*, 54(1-3):159–178, 1992.
- C. Lim and M. McAleer. Forecasting tourist arrivals. *Annals of Tourism Research*, 28(4):965–977, 2001.
- GM Ljung and GEP Box. On a measure of lack of fit in time series models. *Biometrika*, 65(2):297–303, 1978.
- S.E. Said and D.A. Dickey. Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika*, 71(3):599–607, 1984.
- G. Schwarz. Estimating the dimension of a model. *The Annals of Statistics*, 6: 461–464, 1978.

- E. Smeral and M. Wuger. Does complexity matter? Methods for improving forecasting accuracy in tourism: The case of Austria. *Journal of Travel Research*, 44(1):100–110, 2005.
- J.W. Taylor. Exponential smoothing with a damped multiplicative trend. *International journal of Forecasting*, 19(4):715–725, 2003.
- C.S. Vergori and L. Zamparini. Tourism seasonality in a peripheral italian region: a statistical and econometric analysis. *Working Paper, University of Lecce*, 2004.
- M.B. Wilk and S.S. Shapiro. An analysis of variance test for normality (complete samples). *Biometrika*, 52(3-4):591–611, 1965.
- V. Yorucu. The analysis of forecasting performance by using time series data for two mediterranean islands. *Review of Social, Economic & Business Studies*, 2:175–196, 2003.