Il ruolo delle aspettative nelle scelte dei consumatori

Lorenzo Scotellaro

a Karin e Filippo

$$\lim_{mioAmore \to [K \cup F]} = \infty$$

Indice

Introduzione…….… pag. 1

Capitolo 1

1 Il problema analizzato, la presentazione dei dati grezzi….. pag. 3

1.1 Gli impianti di risalita nelle stazioni sciistiche:
lo stato dell'arte, cenni……. pag. 3

1.2 I dati……. pag. 4

1.3 L'analisi……... pag. 5

1.4 I modelli……... pag. 8

1.5 Le implicazioni economiche…… pag. 8

1.6 La metodologia…… pag. 9

Capitolo 2

2 Modelli per l'accesso agli impianti…….. pag. 11

2.1 Modelli per variabili binarie…… pag. 12

2.1.1 I modelli logit, probit e gompit…… pag. 13

2.1.2 Selezione dei repressori e bontà dei modelli...…….pag. 14

2.2 Applicazione del modello logit-probit-gompit al caso
di studio: apertura impianti…….. pag. 14

2.2.1 Le variabili……... pag. 14

2.3 Il modello esplicativo …….. pag. 17

2.3.1 Stima del modello…… pag. 17

2.3.2 I parametri……..... pag. 19

2.4 I modelli previsionali…… pag. 20

2.4.1 Verifica delle condizioni di stazionarietà…… pag. 20

2.4.2 Stima del modello sul campione completo…… pag. 21

2.4.2.1 Valutazione della previsione……... pag. 22

2.4.2.2 Expletation-Prediction Table……... pag. 23

2.4.3 Stima del modello sul campione ridotto……. pag. 24

2.4.3.1 Valutazione della previsione…… pag. 25

Conclusioni……. pag. 27

Appendice…… pag. 29

Bibliografia…….. pag. 31

Introduzione

Ognuno di noi, nella propria quotidianità, compie numerose scelte di acquisto. Al mattino, appena svegli, possiamo optare se fare colazione a casa o al bar; nel caso decidessimo di andare al bar potremmo comprare il caffè o il cappuccino; successivamente in edicola abbiamo la possibilità di acquistare un quotidiano economico o sportivo. Ogni giorno procediamo così fino a sera quando scegliamo quanta energia elettrica consumare per leggere un buon libro.

Le nostre scelte d'acquisto sono dettate dai bisogni e questi, a loro volta, sono influenzati sia da fattori soggettivi, quali gusti e preferenze personali, sia da circostanze oggettive sulle quali non abbiamo facoltà di intervento.

D'altra parte chi vende propone prodotti e servizi in base alle richieste dei consumatori e alle stime dei loro bisogni: nel mercato vengono esitati d'estate i gelati, d'inverno la cioccolata calda. L'offerta e la domanda di beni e servizi sono soggette a un enorme numero di variabili; tali variabili hanno un peso economico che può essere analizzato attraverso strumenti matematici.

La teoria microeconomica spiega che lo scambio di beni e servizi avviene nel punto di incontro tra funzione di domanda e funzione di offerta[1], detto punto di equilibrio. Utilizzando 2 sole variabili, prezzi e quantità, tale modello illustra il comportamento degli agenti economici nel mercato sulla base di ipotesi e asserzioni idealizzate e astratte. Queste ristrettezze da un lato sono necessarie ai fini dell'analisi teorica ponendo le basi e lo slancio per ulteriori studi, dall'altro rendono il modello difficilmente adattabile a situazioni concrete.

Secondo Milton Friedman (1953) nel suo "*La metodologia dell'economia positiva*" afferma che i modelli prescindono dalla validità delle assunzioni per cui: è impossibile ottenere un realismo assoluto, il realismo delle assunzioni non è un indice della validità della teoria e soltanto il potere predittivo è importante[2]; quindi, secondo questa impostazione, l'efficacia previsionale dei modelli economici è la prova della loro

[1] Riassunto in Perloff, J.M. (2001). *Microeconomia, cap. 2.3*

[2] Come riassunto da Guala, F. (2006). *Filosofia dell'economia, cap. V*

efficienza. Ciò risulta maggiormente vero se le variabili sono indipendenti dall'intervento umano (come ad esempio le condizioni meteo) perché tale intromissione potrebbe rendere instabili le relazioni tra di esse. Ma per ottenere modelli utili a capire e prevedere gli andamenti di un preciso mercato dobbiamo restringere il campo di analisi, dalla generalità al caso specifico.

Domanda e offerta sono mosse da serie di cause di varia natura. Tra queste esistono fattori ambientali che influenzano i comportamenti degli individui, spesso non previsti dai tradizionali modelli teorici studiati in economia. Perciò, oltre le teorie basilari utili a focalizzare alcune idee su precisi fenomeni, esistono infiniti modelli economici le cui variabili non sono quantificabili in termini di prezzi o quantità, ma che comunque hanno influenze dominanti.

L'analisi qui proposta indaga i fattori esogeni alla tradizionale sfera economica che influenzano il mercato di riferimento di un particolare prodotto, quantificandone l'incidenza tramite la formulazione del processo generatore dei dati. I modelli econometrici creati daranno informazioni sulle cause alla base dei comportamenti di domanda e offerta; inoltre, con le medesime modalità, si cercherà di prevedere il futuro comportamento degli agenti.

Il presente lavoro è la prima parte di un progetto più ambizioso; per il completamento dell'opera si rimanda a futuri studi. In questa sede, si approfondirà unicamente la sezione relativa all'offerta.

Capitolo 1

L'analisi proposta

1 Il problema analizzato

Per comprendere e valutare il ruolo assunto da variabili esogene alla classica sfera economica negli equilibri di mercato, è stato scelto un tipo di prodotto la cui vendita è influenzata da fattori di carattere ambientale misurabili su scala ordinale. Variabili di questo tipo sono certamente i dati relativi al clima; inoltre essi danno la possibilità di poter quantificare la loro stima a priori nella mente del consumatore: un utente accorto tiene sicuramente conto delle previsioni meteo divulgate dai mass media nei bollettini e aggiusta, di conseguenza, le sue scelte.

Uno tra i settori più condizionati nel brevissimo periodo dal clima è quello del turismo invernale sciistico. Per questi motivi il caso di studio è il numero di vendite di skipass e, più in particolare, di skipass giornalieri, i cui consumi risentono maggiormente del tempo atmosferico.

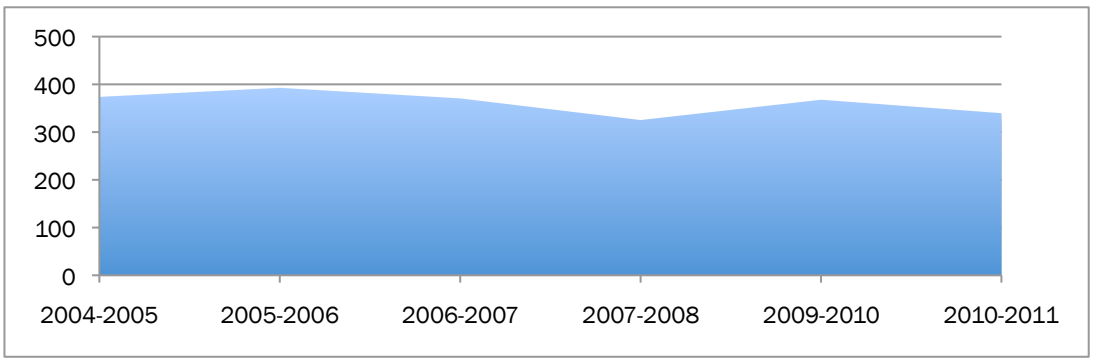
Con queste premesse, l'analisi proposta studia l'incidenza delle variabili meteo (temperatura, pressione atmosferica, situazione nivologica, previsioni ecc.) su offerta e domanda del mercato degli skipass.

1.1 Gli impianti di risalita nelle stazioni sciistiche: lo stato dell'arte, cenni

Gli impianti di risalita (sciovie, seggiovie, funivie ecc.) per definizione sono strutture per il trasporto funicolare terrestre di persone lungo i pendii montani. Sono costituiti essenzialmente da un anello di fune metallica generalmente in moto unidirezionale continuo, teso fra due stazioni terminali e mantenuto ad altezza prestabilita dal terreno; alla fune sono fissati dei dispositivi per il trasporto degli sciatori che, dalla stazione di valle, giungono alla stazione a monte. Il titolo di viaggio di questo particolare mezzo pubblico è lo skipass, una tessera elettronica che permette l'apertura dei cancelli situati agli accessi.

Le montagne italiane sono costellate di stazioni sciistiche attrezzate con questo tipo di strutture, ma per analizzare con precisione il fenomeno di studio, ovvero l'incidenza di determinate varianti sul mercato degli skipass giornalieri, si è dovuto circoscrivere l'indagine ad un unico comprensorio sciistico: *Passo Rolle – San Martino di Castrozza*. In questa ski area, composta da 60 km di piste fra i 1.404 e i 2.357 metri di altitudine, quasi nove turisti su dieci sono italiani e l'andamento delle presenze turistiche durante la stagione sciistica, nell'intervallo considerato 2004-2011 è per lo più costante con due lievi flessioni nelle stagioni 2007-2008 e 2010-2011.

Figura 1: Presenze turistiche invernali. San Martino e Passo Rolle, 2004-2010[3]



ascisse: stagione turistica invernale

ordinate: presenze turistiche totali (scala 1:1.000)

1.2 I dati grezzi

Il database è stato creato utilizzando le serie storiche giornaliere fornite dall'ufficio skipass del comprensorio San Martino - Passo Rolle per quel che riguarda le vendite, dall'ente regionale Meteotrentino per i dati meteorologici e dall'APT di San Martino di Castrozza per i prezzi. Il lavoro risultante è una matrice di 2.934 righe corrispondente ai *giorni i* e 56 colonne di variabili[4]. Le lacune sulle serie, come ad esempio la mancanza di dati sulla temperatura di alcune giornate dovuta a malfunzionamenti temporanei nei sistemi elettronici di monitoraggio, sono state integrate con la media aritmetica tra il periodo immediatamente antecedente e quello immediatamente successivo.

[3] Movimento turistico certificato dalle strutture turistico-ricettive. Fonte: Apt San Martino di Castrozza (i dati relativi il biennio 2008-2009 non sono disponibili).

[4] Le voci dell'intero database sono riportate in appendice. Sempre nell'appendice sono presenti gli indirizzi dei protocolli per accedere al file immesso in rete.

Per avere un campione omogeneo, l'intervallo considerato va dal 6 ottobre 2004 al 5 ottobre 2011 per un totale di 2.556 giorni.

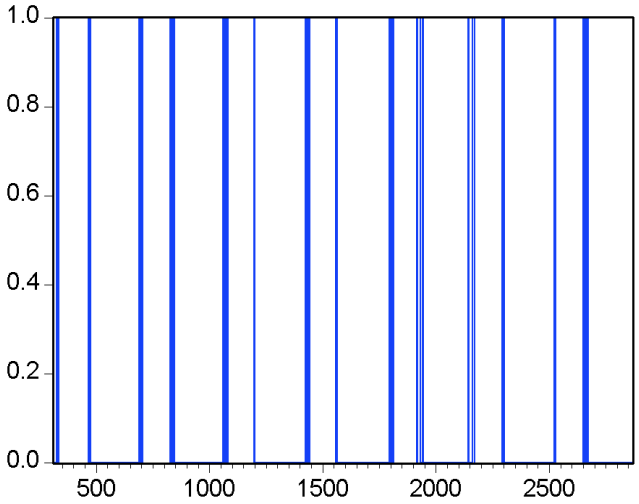
1.3 L'analisi

L'analisi proposta è divisa in due parti. La prima parte studia il lato dell'offerta, la seconda la quantità di equilibrio di mercato.

Gli skipass vengono emessi quando gli impianti di risalita sciistici sono in funzione e ciò accade solo in determinati periodi. Dato che l'offerta non è continua per tutto l'arco dell'anno, ma è vincolata da stagionalità e dalle condizioni climatiche, l'indagine iniziale riguarda i periodi di apertura degli impianti, ricercando i fattori che la determinano. Caratteristica principale della serie storica relativa all'apertura impianti *Ap* (Figura 2) è la sua binarietà, in altre parole essa assume solo due valori: 0 se gli impianti sono chiusi e 1 se aperti.

$$Ap_t = \begin{cases} 1 & se\, aperto \\ 0 & se\, chiuso \end{cases}$$

Figura 2: Linea grafica della distribuzione della variabile "apertura impianti". San Martino e Passo Rolle, 6 ottobre 2004 - 5 ottobre 2011[5]



asse delle ascisse: giorni [311 = 6 ottobre 2004; 2866 = 5 ottobre 2011]

asse delle ordinate: 1 = impianto aperto; 0 = impianto chiuso

Dalla tabulazione relativa a questa variabile (Tabella 1) si nota che gli impianti sono aperti il 36,23% delle giornate e chiusi il restante 63,77%.

Tabella 1: Tabulazione variabile "apertura impianti"

Value	Count	Percent	Cumulative Count	Cumulative Percent
0	1630	63.77	1630	63.77
1	926	36.23	2556	100.00
Total	2556	100.00	2556	100.00

Come anticipato nell'introduzione e a inizio capitolo la seconda parte dell'opera, rimandata a futuri studi, riguarderà la variabile spiegata "numero di vendite di skipass giornalieri nel giorno *t*" (Figura 3); le quantità vendute e rilevate di questo particolare titolo di viaggio indicano il punto di incontro tra la funzione di offerta, di cui si avanzerà una stima in seguito, e la funzione di domanda[6].

Ciò premesso, verranno stimate le quantità di equilibrio (skipass venduti) condizionate

[5] Fonte: Ufficio skipass di San Martino di Castrozza.

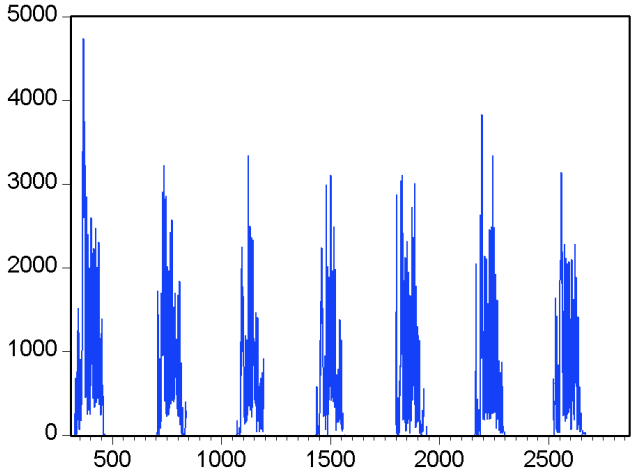
[6] La funzione di domanda è ardua da formulare in quanto composta dalla somma di tutte le funzioni di domanda della moltitudine dei potenziali consumatori, ciascuno con le proprie preferenze.

all'apertura degli impianti, al fine di studiare i fenomeni che le influenzano.

Il campione, in questo caso, è ulteriormente ristretto alle 926 giornate in cui gli impianti furono aperti.

Nel periodo considerato si sono avute in totale 687.273 vendite di skipass, con media giornaliera di 724,2, un massimo di 4.735[7] e 11 minimi di 0 (Figura 3). La macroclasse (Tabella 2) che registra il maggior numero di osservazioni è la prima (da 0 a 999) con il 73,22%.

Figura 3: Linea grafica della distribuzione della variabile "vendite di skipass". San Martino e Passo Rolle, 6 ottobre 2004 - 5 ottobre 2011.[8]



asse delle ascisse: giorni [311 = 6 ottobre 2004; 2866 = 5 ottobre 2011]

asse delle ordinate: numero di skipass giornalieri venduti

Tabella 2: Tabella di divisione della variabile "vendite di skipass" in 5 macroclassi

Value	Count	Percent	Cumulative Count	Cumulative Percent
[0, 1000)	678	73.22	678	73.22
[1000, 2000)	163	17.60	841	90.82
[2000, 3000)	68	7.34	909	98.16
[3000, 4000)	16	1.73	925	99.89
[4000, 5000)	1	0.11	926	100.00
Total	926	100.00	926	100.00

[7] Considerando la serie "vendite di skipass" nei soli intervalli di apertura, tale valore massimo è anche l'unica osservazione anomala che presenta uno scostamento positivo dal terzo quartile superiore a 3 volte il range interquartile; la seconda osservazione in ordine decrescente di grandezza è di 3.827 presenze turistiche giornaliere.

[8] Fonte: Ufficio skipass di San Martino di Castrozza.

Tabella 3: Misure della variabile "vendite skipass"

Mean	742.1955
Median	439.5000
Maximum	4735.000
Minimum	0.000000
Std. Dev.	782.4278
Skewness	1.475903
Kurtosis	4.837107
Jarque-Bera	466.3999
Probability	0.000000
Sum	687273.0
Sum Sq. Dev.	5.66E+08
Observations	926

1.4 I modelli

I modelli impiegati nell'analisi sono di due tipi: esplicativi e previsionali.

Come possono componenti meteo e stagionali condizionare le aperture degli impianti sciistici di risalita e le vendite di skipass? In che misura? La risposta a queste domande è data dai modelli esplicativi che, a seguito dell'analisi di regressione sulle osservazioni giornaliere e i dati cronometrici, restituiscono il valore dei parametri atti a quantificare l'effetto delle succitate variabili sull'apertura degli impianti a fune e sul numero vendite di skipass.

Differentemente, per fare previsione sugli eventi prima spiegati si utilizzano i modelli previsionali: facendo inferenza sui medesimi dati del modello esplicativo ritardati sui periodi si potranno pronosticare con margini di errore certi i possibili andamenti delle variabili spiegate.

1.5 Le implicazioni economiche

I modelli dinamici qui stimati, oltre ad avere valenza teorica, possono influenzare le scelte degli agenti che operano nel mercato e quindi avere implicazioni economiche.

Le possibili applicazioni sono molteplici. Ad esempio la previsione del periodo di apertura impianti potrebbe dare indicazione del grado di allerta ai servizi di pronto intervento influendo sull'ottimizzazione delle risorse pubbliche; il pronostico del numero di vendite di skipass potrebbe interessare sia gli esercenti del luogo, sia gli enti

preposti alla gestione dei flussi turistici in quanto avrebbero la possibilità di prevedere le presenze turistiche giornaliere (che non sono segnalate dalle strutture turistico-ricettive) in modo da adeguare la loro offerta con l'attuazione di politiche commerciali idonee.

1.6 La metodologia

Tutti i processi generatori dei dati riportati sono il risultato di una serie di inferenze fatte sulle variabili del database. Tali DGP (*data generating process*) sono quelli che hanno soddisfatto più degli altri gli obbiettivi di significatività dei parametri e, al contempo, coerenza delle stime di questi con il tipo di osservazioni.
Tutte le elaborazioni sono state prodotte con il sofware applicativo EViews 4.

Capitolo 2

I modelli

2 Modelli per l'accesso agli impianti

Il primo modello studia la possibilità che gli impianti sciistici di risalita siano operativi e, quindi, che le conseguenti vendite di skipass possano verificarsi. La variabile spiegata in questo modello è l'apertura degli impianti nel giorno t (Ap_t), ovvero la variabile che restituisce il valore 0 se gli impianti sono chiusi oppure il valore 1 se sono in funzione.

$$Ap_t = \begin{cases} 1 & se\, aperto \\ 0 & se\, chiuso \end{cases}$$

Le variabili esplicative sono legate al periodo dell'anno (mesi) e condizioni climatiche (temperatura, pressione atmosferica e dati nivologici) secondo un modello del tipo:

$$Ap_t = \alpha + \beta_1 x_{1t} + \beta_2 x_{2t} + ... + \beta_k x_{kt} + u_t$$

dove α rappresenta la costante, le x_{nt} rappresentano le variabili esplicative, β_n i loro coefficienti e u_t è il termine di errore.

La variabile dicotomica Ap_t, data la sua natura discreta e binaria, non si presta direttamente a un'analisi di regressione; per questo tipo di variabili l'approccio più diffuso consiste nel considerare la loro probabilità di realizzazione.

$$\Pr(Ap_t = 1) = F(\alpha + \beta_1 x_{1t} + \beta_2 x_{2t} + ... + \beta_k x_{kt})$$

2.1 Modelli per variabili binarie

Si pone il problema di quale metodo, alternativo all'OLS, usare in questo specifico caso. Una delle possibili soluzioni sviluppate in letteratura è quello di inserire nel modello una variabile latente, ad esempio Y_t^*,

$$Y_t^* = \alpha + \beta_1 x_{1t} + \beta_2 x_{2t} + ... + \beta_k x_{kt} + u_t$$

utilizzando la notazione matriciale per facilitare l'esposizione (si veda Johnstone J. (2011), *Econometric Methods*, cap.2) si ha:

$$Y_t^* = \alpha + \beta' x_t + u_t$$

dove

$\beta = (\beta_1, \beta_2, ..., \beta_k)'$, $x_t = (x_{1t}, x_{2t}, ..., x_{kt})$ e u_t è un termine di errore.

Ora mettendo in relazione a variabile latente con la variabile binaria in questo modo:

$$Ap_t = \begin{cases} 1 & Y_t^* > 0 \\ 0 & altrove \end{cases}$$

si ottiene

$$\begin{aligned} P_t &= \Pr(Ap_t = 1) \\ &= \Pr(Y_t^* > 0) \\ &= \Pr(\alpha + \beta_t' x_t + u_t > 0) \\ &= \Pr(u_t > -(\alpha + \beta' x_t)) \end{aligned}$$

Usando una distribuzione degli errori simmetrica centrata sullo 0 possiamo applicare la seguente uguaglianza (si veda Gourieroux, C. (2000), *Econometrics of Quantitative Dependent Variables*, Cambridge University Press):

$$\Pr(u_t > -(\alpha + \beta' x_t)) = \Pr(u_t < \alpha + \beta' x_t)$$

a questo punto è possibile affermare che

$$P_t = \Pr(Ap_t = 1) = \Pr(u_t < \alpha + \beta' x_t)$$

Appurato che la probabilità P_t dipende dalla scelta della distribuzione degli errori, vi sono almeno tre alternative suggerite in letteratura per proseguire l'analisi a seconda che la loro distribuzione segua la normale (Karl Fredrich Gauss), la funzione logistica (Pier Francois Verhulst) o la funzione di Gompertz (Benjamin Gompertz) a cui corrispondono nell'ordine i modelli *probit*, *logit* e *gompit*.

2.1.1 I modelli logit, probit e gompit

La funzione di questi particolari modelli dipende dalla loro distribuzione degli errori, quindi se u_t segue una distribuzione logistica, allora P_t sarà un *logit* del tipo:

$$P_t = \frac{1}{1 + \exp(-Z_t)}$$

dove $Z_t = \alpha + \beta' x_t$ e per cui

$$P_t = \frac{1}{1 + \exp(-(\alpha + \beta' x_t))}$$

nel caso u_t segua una distribuzione normale P_t sarà un *probit*:

$$P_t = \int_{-\infty}^{Z_t} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} w^2\right) \quad dw$$

mentre se gli errori seguono una distribuzione strutturata come funzione di Gompertz:

$$P_t = \exp(-(\exp(-Z_t))$$

2.1.2 Selezione dei regressori e bontà dei modelli

La scelta dei regressori nei modelli *probit-logit-gompit* è basata sulle statistiche *z-Statistic* e *z-Probability*. Entrambe si riferiscono alla significatività del coefficiente: la prima restituisce il risultato del test, quindi più questo valore si allontana da 0 maggiore è l'influenza della variabile nel modello; la seconda statistica è il livello di significatività osservato e più si avvicina a 0 e più il coefficiente è significativo.

Per testare la varianza spiegata in questo particolare tipo di modelli econometrici vengono impiegati indicatori alternativi dell' R^2 detti *pseudo-* R^2 i cui risultati possono essere interpretati in modo analogo all'indice di correlazione multipla; in particolare nel programma EViews è usato il McFadden R-squared (si veda McFadden D. (1974), *The Measurement of Urban Travel Demand* in «Journal of Public Economics», 3, 303-328).

Il sostituto del *test F* in questi modelli è la statistica rapporto di log-verosimiglianze (*LR statistic*), il cui p-value fornisce il livello del test di significatività congiunta dei regressori.

2.2 Applicazione del modello logit-probit-gompit al caso di studio: apertura impianti

In questa parte applicata viene spiegata l'offerta vista come andamento della serie binaria "apertura impianti" in dipendenza dei fattori climatici e cronometrici.

2.2.1 Le variabili

La variabile spiegata è l'apertura degli impianti nel giorno t (Ap_t). Tale variabile restituisce il valore 1 se aperto, 0 se chiuso[9].

$$Ap_t = \begin{cases} 1 & se\,aperto \\ 0 & se\,chiuso \end{cases}$$

Le variabili esplicative sono di due tipi: uno legato alla stagionalità e l'altro alle variabili meteo osservate nei giorni del periodo.

[9] La linea grafica di distribuzione della variabile è riportata in "Figura 2"

La componente stagionale è rappresentata da una variabile dummy corrispondente ai mesi "freddi" da settembre a marzo:

$$M_t = \begin{cases} 1 & se \quad t \in G \\ 0 & se \quad t \notin G \end{cases}$$

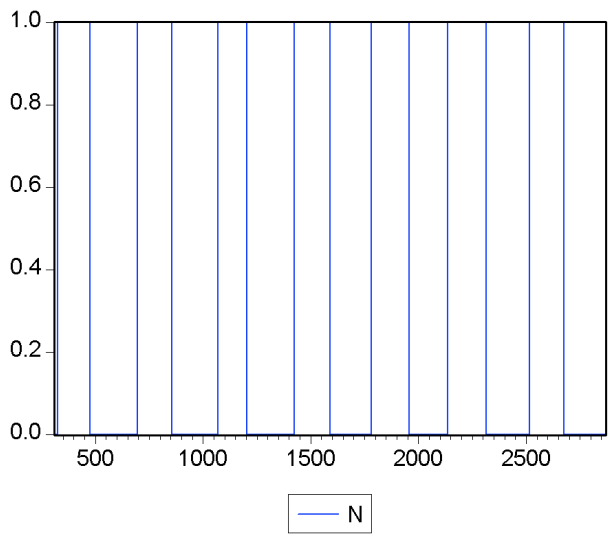
dove $G = \{$ giorni dei mesi di settembre, ottobre, novembre, dicembre, gennaio, febbraio e marzo $\}$.

I dati relativi al meteo nel modello di seguito proposto sono:

- N_t una dummy che esprime la presenza di neve (1) o meno (0)
- $X22_t$ l'altezza del manto nevoso in centimetri
- PA_t la pressione atmosferica espressa in *bar*.

Le cui distribuzioni grafiche sono rispettivamente:

Figura 5: Linea grafica della distribuzione della variabile *N*. Passo Rolle, 6 ottobre 2004 - 5 ottobre 2011.[10]

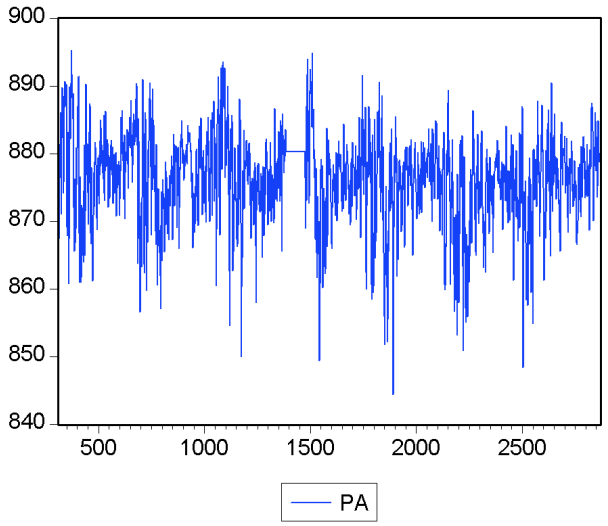


asse delle ascisse: giorni [311 = 6 ottobre 2004; 2556 = 5 ottobre 2011]

asse delle ordinate: 1 = presenza di neve; 0 = mancanza di neve

[10] Fonte: Provincia Autonoma di Trento, ente Meteotrentino.

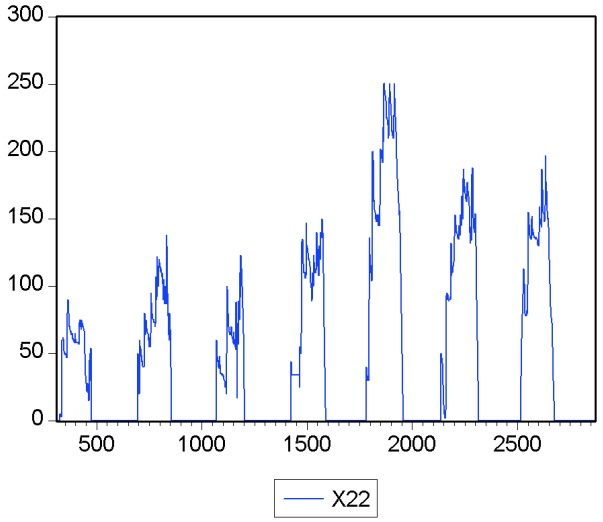
Figura 6: Linea grafica della distribuzione della variabile *PA*". San Martino di Castrozza, 6 ottobre 2004 - 5 ottobre 2011.[11]



asse delle ascisse: giorni [311 = 6 ottobre 2004; 2866 = 5 ottobre 2011]

asse delle ordinate: pressione atmosferica misurata in *bar*

Figura 7: Linea grafica della distribuzione della variabile *X22*Passo Rolle, 6 ottobre 2004 - 5 ottobre 2011.[12]



asse delle ascisse: giorni [311 = 6 ottobre 2004; 2866 = 5 ottobre 2011]

asse delle ordinate: altezza neve espressa in centimetri

[11] Fonte: Provincia Autonoma di Trento, ente Meteotrentino.

[12] Fonte: Provincia Autonoma di Trento, ente Meteotrentino.

La matrice di correlazione qui riprodotta indica il grado di dipendenza lineare reciproca delle variabili in esame.

Tabella 4: matrice di correlazione lineare variabili

	M	N	X22	PA
M	1.000000	0.418507	0.342156	0.028296
N	0.418507	1.000000	0.799592	-0.137173
X22	0.342156	0.799592	1.000000	-0.218251
PA	0.028296	-0.137173	-0.218251	1.000000

Le stime dei coefficienti danno un'idea del nesso logico tra le osservazioni: la presenza di neve è correlata alla sua altezza e ai mesi freddi, la stessa ha invece correlazione negativa con la misura della pressione atmosferica (il bel tempo fa sciogliere la neve e non ne fa produrre di nuova), mentre è neutrale la relazione lineare tra mesi freddi e pressione atmosferica.

2.3 Il modello esplicativo

Data la natura delle variabili si è scelto un modello *logit*[13].

2.3.1 Stima del modello

A seguito delle precedenti argomentazioni è risultato il modello :

$$P_t = \frac{1}{1+\exp(-Z_t)}$$

dove $Z_t = \alpha + \beta_1 M_t + \beta_2 N_t + \beta_3 X22_t + \beta_4 PA_t$.

Il suo esito di stima in EViews è il seguente:

[13] E' noto e riscontrato in letteratura l'impiego di modelli logit nei casi di studio di variabili legate alla situazione nivologica come, ad esempio, nell'articolo di Sohn, K. T., J. H. Lee e Y. S. Cho "*Ternary Forecast of Heavy Snofall in the Honam Area, Korea*" pubblicato su «Advances in atmosferic science», vol. 26 no. 2, 2009, pag. 327-332.

Tabella 5: Rappresentazione tabellare dell'equazione stimata esplicativa dell'offerta

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	-68.88540	11.48126	-5.999810	0.0000
M	3.259064	0.228300	14.27537	0.0000
N	5.857991	0.539827	10.85160	0.0000
X22	0.023161	0.002500	9.266175	0.0000
PA	0.068949	0.012973	5.314939	0.0000

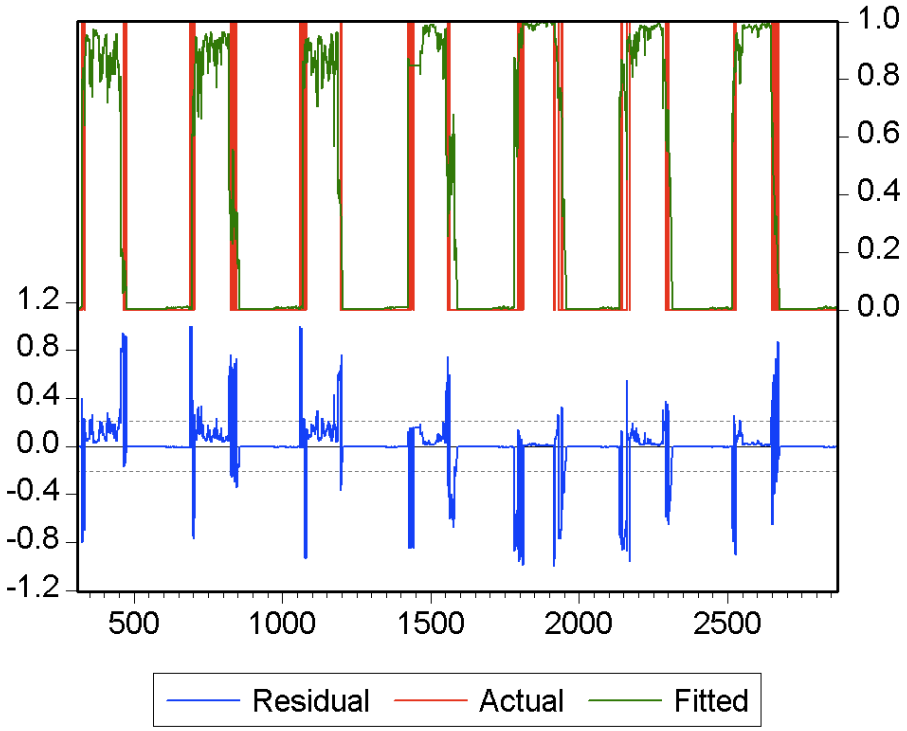
Mean dependent var	0.362285	S.D. dependent var	0.480755
S.E. of regression	0.209264	Akaike info criterion	0.294725
Sum squared resid	111.7122	Schwarz criterion	0.306161
Log likelihood	-371.6589	Hannan-Quinn criter.	0.298872
Restr. log likelihood	-1673.468	Avg. log likelihood	-0.145406
LR statistic (4 df)	2603.618	McFadden R-squared	0.777911
Probability(LR stat)	0.000000		

Obs with Dep=0	1630	Total obs	2556
Obs with Dep=1	926		

Gli indici di bontà delle stime citati al paragrafo 2.1.2 ci dicono che il *p-value* dell'ipotesi di non significatività congiunta dei coefficienti è zero e che la varianza spiegata dal modello è abbastanza alta.

Di seguito è riportato il grafico con i residui (in blu):

Figura 8: Grafico con i residui dell'equazione stimata esplicativa dell'offerta



asse delle ascisse: giorni [311 = 6 ottobre 2004; 2866 = 5 ottobre 2011]

asse delle ordinate: probabilità di apertura impianti

2.3.2 I parametri

La procedura di specificazione ha prodotto un modello econometrico che spiega in che modo, dandone misura, variabili di carattere meteorologico e stagionale incidono sull'apertura degli impianti sciistici di risalita.

I coefficienti stimati restituiscono valori concettualmente coerenti con il tipo di variabile esplicativa a cui fanno riferimento, ovvero soddisfano le ipotesi logiche causa-effetto attese. In particolare: il periodo invernale, la presenza di neve, l'altezza del manto nevoso e la maggior pressione atmosferica (segnale di buone condizioni meteo) hanno impatto positivo sui periodi di apertura.

Tramite il *Categorical Regressor Stats* (Tabella 6) si possono visualizzare in EViews le medie dei valori assunti dalle variabili nel caso gli impianti siano chiusi (0) o aperti (1).

Tabella 6: Categorical Regressor Stat dell'equazione stimata esplicativa dell'offerta

Variable	Dep=0	Mean Dep=1	All
C	1.000000	1.000000	1.000000
M	0.396319	0.906048	0.580986
N	0.123926	0.995680	0.439750
X22	8.707362	106.7354	44.22144
PA	876.7078	875.5490	876.2879

2.4 I modelli previsionali

I modelli previsionali ripropongono le stesse variabili del precedente esplicativo, ma ritardate di alcuni periodi; l'unica serie su cui non vengono applicati ritardi è quella relativa ai mesi "freddi" (*M*), perché ciò non avrebbe senso in quanto data certa a priori.

2.4.1 Verifica delle condizioni di stazionarietà

Prima di produrre un modello dinamico si deve verificare se le serie presentano radici unitarie. Tale verifica avviene utilizzando l'*Augmented Dickey-Fuller Unit Root Test* la cui ipotesi nulla è non stazionarietà. Il ritardo considerato è un periodo e il livello di confidenza è stabilito a 0.1.

I risultati sono esposti nelle seguenti tabelle:

Augmented Dickey-Fuller Unit Root Test on A

ADF Test Statistic	-7.106833	1% Critical Value*	-3.4360
		5% Critical Value	-2.8632
		10% Critical Value	-2.5677

Augmented Dickey-Fuller Unit Root Test on N

ADF Test Statistic	-3.788022	1% Critical Value*	-3.4360
		5% Critical Value	-2.8632
		10% Critical Value	-2.5677

Augmented Dickey-Fuller Unit Root Test on X22

ADF Test Statistic	-2.845912	1% Critical Value*	-3.4360
		5% Critical Value	-2.8632
		10% Critical Value	-2.5677

Augmented Dickey-Fuller Unit Root Test on PA

ADF Test Statistic	-20.12601	1% Critical Value*	-3.4360
		5% Critical Value	-2.8632
		10% Critical Value	-2.5677

I valori critici al livello stabilito non superano mai quello del test, quindi le serie risultano stazionarie.

2.4.2 Stima del modello sul campione completo

Le variabili di questo primo modello previsionale sono le medesime di quello esplicativo, con un'unica differenza: quelle climatiche sono ritardate di 5 periodi *t*.

Tabella 7: Rappresentazione tabellare dell'equazione previsionale stimata dell'offerta

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	-56.33250	12.19148	-4.620644	0.0000
M	3.895445	0.237981	16.36874	0.0000
N(-5)	5.145532	0.344631	14.93054	0.0000
X22(-5)	0.019896	0.002438	8.160624	0.0000
PA(-5)	0.055367	0.013799	4.012254	0.0001

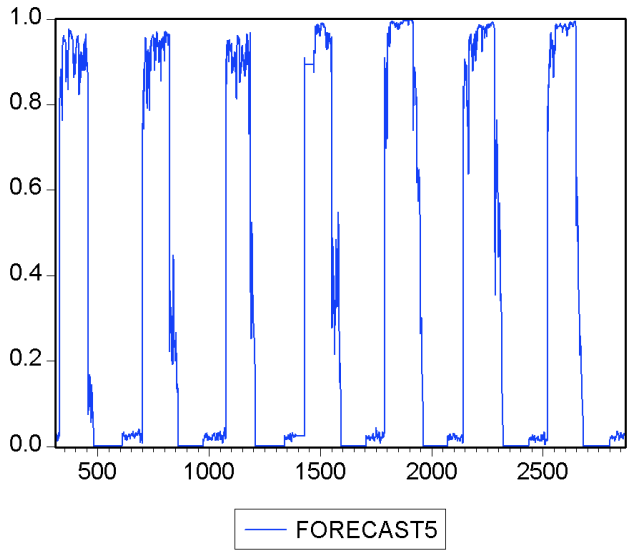
Mean dependent var	0.362285	S.D. dependent var	0.480755
S.E. of regression	0.211016	Akaike info criterion	0.305289
Sum squared resid	113.5901	Schwarz criterion	0.316725
Log likelihood	-385.1587	Hannan-Quinn criter.	0.309436
Restr. log likelihood	-1673.468	Avg. log likelihood	-0.150688
LR statistic (4 df)	2576.619	McFadden R-squared	0.769844
Probability(LR stat)	0.000000		

Obs with Dep=0	1630	Total obs	2556
Obs with Dep=1	926		

2.4.2.1 Valutazione della previsione

Di seguito è riprodotta la tabella di valutazione fatta con la procedura di EViews "Forecast":

Figura 9 e Tabella 8: Previsione con modello *logit* ritardato di 5 periodi della probabilità di apertura degli impianti e esito della relativa valutazione con la procedura "Forecast" sul totale delle osservazioni



Forecast: FORECAST5
Actual: A
Forecast sample: 311 2866
Included observations: 2556

Root Mean Squared Error	0.210809
Mean Absolute Error	0.086355
Mean Abs. Percent Error	4.317760
Theil Inequality Coefficient	0.180501
Bias Proportion	0.000000
Variance Proportion	0.047163
Covariance Proportion	0.952837

asse delle ascisse: giorni [311 = 6 ottobre 2004; 2866 = 5 ottobre 2011]

asse delle ordinate: probabilità di apertura impianti

Le grandezze delle statistiche *Roots Mean Squared Error* e *Mean Absolute Error* dipendono direttamente dalla variabile dipendente e servono per comparare differenti modelli stimati sulle medesime serie. Nel nostro caso per ritardi inferiori (-3, -4) si ottengono risultati analoghi, mentre per ritardi superiori (-6, -7) tali errori risultano entrambi maggiori; ciò conferma che il ritardo di 5 periodi è il più valido per fare previsione nel brevissimo periodo.

Le altre due statistiche hanno scala fissa e, in particolare, il *Theil Inequality Coefficient* ha scala 0-1, dove 0 è la previsione perfetta; nella tabella tale coefficiente è di 0,18, abbastanza vicino al *fit*.

Le ultime tre proporzioni, che sommano a 1, sono la scomposizione dell'errore quadratico medio del modello di previsione e corrispondono:

- la *Bias Propotion* è indicatore della distanza della media della previsione da quella della serie,
- la *Variance Propotion* è indicatore della distanza della varianza della previsione da quella della serie,
- la *Covariance Propotion* è la misura dei rimanenti errori asistemici della previsione.

Ciò detto i bassi valori delle prime due proporzioni sono sintomo di bontà della previsione.

Per esaminare le proprietà e le performance del modello binario si utilizza l'*Expectation-Prediction (Classification) Table*.

2.4.2.2 Expectation-Prediction Table

In questa tabella vengono evidenziati le rilevazioni corrette, quelle non corrette e le loro percentuali rispetto al totale.

Il modello previsionale stimato restituisce la probabilità che gli impianti siano aperti, ovvero calcola la $\Pr(Ap_t = 1)$ in funzione delle variabili esplicative ritardate. Considerando le rilevazioni effettive della serie Ap, è possibile misurare quanto questa previsione di probabilità sia precisa.

Nel procedimento di valutazione si fissa una soglia di prossimità sotto la quale si prevede la chiusura degli impianti; quindi se $\Pr(Ap_t = 1)$ è minore o uguale al livello stabilito la previsione è 0 (impianto chiuso), mentre se maggiore al livello è 1 (impianto aperto). Dando uguale peso ai due unici possibili valori assunti da Ap la soglia è 0,5.

Il passaggio successivo consiste nel contare il numero di volte in cui il modello ha predetto in modo corretto, cioè le volte in cui nel modello previsionale la $\Pr(Ap_t = 1) > 0{,}5$ quando la serie Ap assume valore 1 e le volte in cui la previsione $\Pr(Ap_t = 1) \leq 0{,}5$ quando la serie storica da 0.

Il risultato è la prima matrice 2 per 2 in alto a destra nella seguente tabella:

Tabella 9: Expectation-Prediction Table dell'equazione stimata previsionale dell'offerta

	Estimated Equation			Constant Probability		
	Dep=0	Dep=1	Total	Dep=0	Dep=1	Total
P(Dep=1)<=C	1542	74	1616	1630	926	2556
P(Dep=1)>C	88	852	940	0	0	0
Total	1630	926	2556	1630	926	2556
Correct	1542	852	2394	1630	0	1630
% Correct	94.60	92.01	93.66	100.00	0.00	63.77
% Incorrect	5.40	7.99	6.34	0.00	100.00	36.23
Total Gain*	-5.40	92.01	29.89			
Percent Gain**	NA	92.01	82.51			

Dalla somma delle prime 2 righe nasce la terza con i totali.

L'altra matrice a destra della predetta mostra i risultati senza l'utilizzo del modello previsionale e con un'unica probabilità costante stimata come segue: numero delle osservazioni con valore 1 al numeratore, il totale delle osservazioni al denominatore. Per cui:

$$\Pr(Ap = 1) = \frac{926}{2556} \approx 0{,}3622$$

I risultati dell'*Expletation-Prediction Table* mostrano che le previsioni del modello proposto sono risultate esatte 2394 volte su 2556, corrispondente al 93,66 percento del totale. Inoltre dal confronto delle due matrici si evidenzia una maggiore precisione guadagnata attraverso l'utilizzo di una funzione di previsione quantificabile nel 29,89 percento in più di risposte corrette rispetto a quanto si otterrebbe semplicemente applicando il modello con probabilità costante di cui sopra.

2.4.3 Stima del modello sul campione ridotto

Al fine di verificare ulteriormente il potere predittivo dell'impostazione analitica adottata, si ripropone la medesima regressione del precedente modello previsionale, con le stesse variabili e stessi ritardi, ma fatta su un campione ridotto di un periodo annuale: i 5 periodi di durata di un anno ciascuno si riducono a 4; il nuovo campione va dal 6 novembre 2004 al 5 novembre 2010 per un totale di 2.191 osservazioni. In questo modo

sarà possibile controllare tramite la procedura *Forecast* fatta per l'ultima annualità precedentemente tolta, che va dal 6 novembre 2011 al 5 novembre 2011, se criteri di analisi fin qui usati hanno effettiva efficacia di previsione sui periodi futuri in cui le variabili assumono valori ipoteticamente ancora ignoti.

Tabella 10: Rappresentazione tabellare dell'equazione previsionale stimata dell'offerta con campione ridotto

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	-53.99523	12.78450	-4.223491	0.0000
M	3.912427	0.257314	15.20487	0.0000
N(-5)	4.990170	0.347390	14.36474	0.0000
X22(-5)	0.019463	0.002596	7.498290	0.0000
PA(-5)	0.052863	0.014479	3.651112	0.0003

Mean dependent var	0.361022	S.D. dependent var	0.480407
S.E. of regression	0.216168	Akaike info criterion	0.323121
Sum squared resid	102.1486	Schwarz criterion	0.336110
Log likelihood	-348.9787	Hannan-Quinn criter.	0.327868
Restr. log likelihood	-1432.923	Avg. log likelihood	-0.159278
LR statistic (4 df)	2167.889	McFadden R-squared	0.756457
Probability(LR stat)	0.000000		

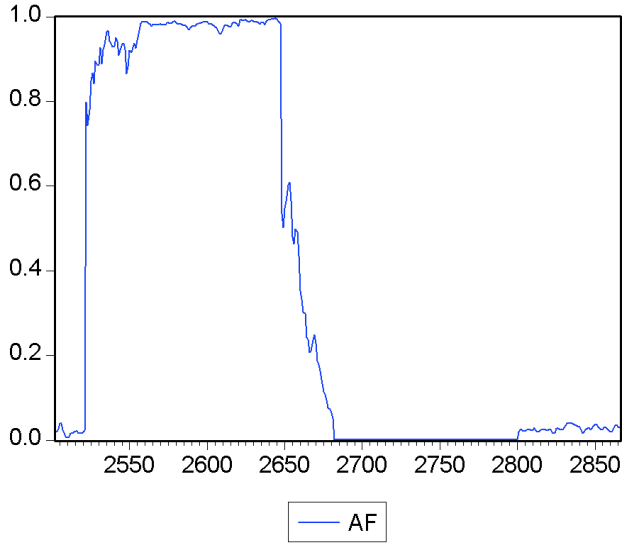
Obs with Dep=0	1400	Total obs	2191
Obs with Dep=1	791		

I risultati della tabella sono analoghi a quelli nella medesima rappresentazione della stima del modello sul totale delle osservazioni considerate (Tabella 7) con una leggera perdita di varianza spiegata.

2.4.3.1 Valutazione della previsione

Nella seguente tabella è esposta la valutazione della previsione fatta sul periodo escluso dalla precedente regressione:

Figura 10 e Tabella 11: Previsione con modello *logit* ritardato di 5 periodi della probabilità di apertura degli impianti e esito della relativa valutazione con la procedura "Forecast" per il periodo dal 6 ottobre 2010 al 5 ottobre 2011



Forecast: AF
Actual: A
Forecast sample: 2502 2866
Included observations: 365

Root Mean Squared Error	0.175326
Mean Absolute Error	0.065502
Mean Abs. Percent Error	3.431659
Theil Inequality Coefficient	0.148038
Bias Proportion	0.000319
Variance Proportion	0.047930
Covariance Proportion	0.951751

asse delle ascisse: giorni [2502 = 6 ottobre 2010; 2866 = 5 ottobre 2011]
asse delle ordinate: probabilità di apertura impianti

Dal confronto dei rispettivi esiti delle procedure di valutazione *Forecast*, anche se diversamente impostate, emergono forti analogie tra i due modelli previsionali. Si evince la conferma che la regressione su queste variabili così ritardate è valida per prevedere gli andamenti futuri della variabile "apertura impianti".

Data l'affinità dei risultati delle valutazioni delle due regressioni, sarà preso in considerazione come modello previsonale definitivo il primo in quanto ottenuto da un campione più completo e aggiornato.

Conclusioni

L'obbiettivo dello studio effettuato è individuare e valutare le variabili di natura climatica che caratterizzano il mercato degli skipass giornalieri.

A seguito dell'analisi econometrica delle variabili che determinano l'apertura degli impianti si è formulata un'equazione esplicativa dell'offerta con la quale è possibile calcolare la probabilità che gli skipass vengano esitati sul mercato sulla base di dati cronometrici e di dati relativi a fenomeni climatici.

$$P_t = \frac{1}{1+\exp\left(-\left(-68{,}885+3{,}259M_t+5{,}857N_t+0{,}023X22_t+0{,}068PA_t\right)\right)}$$

In cui M_t è una variabile che indica se il mese del giorno t è un mese "freddo" (considerando freddi i mesi da settembre a marzo), N_t indica la presenza o meno di neve, $X22$ è l'altezza del manto nevoso espressa in centimetri e PA_t la pressione atmosferica espressa in *bar*.

Per quantificare gli effetti delle singole variabili esplicative sull'apertura degli impianti di risalita e dare un ordine in termini di significatività si confronta la statistica *z-Statistic* dei coefficienti stimati.

Il fattore con maggior incidenza sull'offerta di skipass è la stagionalità rappresentata dai mesi "freddi", seguita in ordine dalla presenza di neve e dall'altezza del manto nevoso, infine troviamo la pressione atmosferica sinonimo di bel tempo.

Una volta individuato il modello teorico è stato creato il modello previsionale semplicemente ritardando le tutte le variabili esplicative di 5 periodi[14] ottenendo una funzione di probabilità simile alla esplicativa, ma con variabili e coefficienti diversi:

$$P_t = \frac{1}{1+\exp\left(-\left(-56{,}332+3{,}895M_t+5{,}145N_{t-5}+0{,}019X22_{t-5}+0{,}055PA_{t-5}\right)\right)}$$

[14] Si è scelto il ritardo di 5 periodi in quanto, a seguito dei test, si è dimostrato il più valido per fare previsione.

Sostituendo le opportune variabili del giorno *t* si ottiene la probabilità che gli impianti siano aperti nel giorno *t*+5, ossia con 5 giorni di anticipo, tramite la funzione:

$$P_{t+5} = \frac{1}{1+\exp\left(-\left(-56{,}332 + 3{,}895 M_{t+5} + 5{,}145 N_t + 0{,}019 X22_t + 0{,}055 PA_t\right)\right)}$$

Grazie all'applicazione del metodo econometrico l'analisi effettuata ha dato i risultati sperati: si è dimostrato che l'offerta di skipass giornalieri è soggetta oltre che alla stagionalità, anche alle variabili meteorologiche. Inoltre, sulla base di questi risultati, è possibile fare previsioni sul comportamento degli agenti che operano in questo mercato. Si è così verificata l'influenza di variabili di varia natura sull'offerta ed è stato possibile prevederne le conseguenze.

Ottenuto il modello relativo all'offerta, il prossimo passo dello studio riguarderà il punto di equilibrio tra domanda e offerta, ovvero il numero di vendite di skipass giornalieri nel giorno *t* (V_t) condizionate alla probabilità di apertura degli impianti (P_t) qui stimata, secondo la formula:

$$V_t \middle| P_t = \alpha + \beta_1 x_{1t} + \beta_2 x_{2t} + \ldots + \beta_k x_{kt} + u_t .$$

Appendice

Variabili del database

n	Data	Aperto	Vendite
N	D	A	Y

Giorno		
settimanale	di apertura	annuale
X1	X2	X3

Mese	Stagio-nalità	Meteo
X4	X5	X6

previsioni meteo				
0	-1	-2	-3	-4
X7	X8	X9	X10	X11

anno	prezzi	Condizioni del tempo	Nuvolosità	Visibilità	Vento in quota 1
X12	X13	X14	X15	X16	X17

Vento in quota 2	Temp. aria Mean	Temp. aria (°C) Min	Temp. aria (°C) Max	Escursione	Altezza man-to nevoso
X18	X19	X20	X21	ES	X22

Altezza neve fresca	Densità della neve	Temp. neve a 10 cm	Temp. neve a 30 cm	Penetrazione della sonda	Strato neve superfciale
X23	X24	X25	X26	X27	X28

Rugosità superficiale	Brina di superficie	Gennaio 1	Febbraio 2	Marzo 3	Aprile 4
X29	X30	M1	M2	M3	M4

Maggio 5	Giugno 6	Luglio 7	Agosto 8	Settembre 9	Ottobre 10
M5	M6	M7	M8	M9	M10

Novembre 11	Dicembre 12	Presenza neve	Pressione atmosferica	lunedì 1	martedì 2
M11	M12	N	PA	G1	G2

mercoledì	giovedì	venerdì	sabato	domenica
3	4	5	6	7
G3	G4	G5	G6	G7

Il database su file Excell è liberamente scaricabile collegandosi al *link*:

<https://docs.google.com/open?id=0Bw5Fbpbt261sd3c4OWJxdnF0OUU>

Inoltre è disponibile in rete il file elaborato con il programma EViews 4:

<https://docs.google.com/open?id=0Bw5Fbpbt261seHlKcGFkOVphM0k>

Bibliografia

Amisano, G. (2004), *Elementi di Econometria*, Mondadori.

Friedman, M. (1953), *The Methodology of Positive Economics*, in *Essays in Positive Economics*, University of Chicago Press (trad. it. *La metodologia dell'economia positiva*, in *Metodo, consumo e moneta*, il Mulino, 1996).

Gourieroux, C. (2000), *Econometrics of Quantitative Dependent Variables*, Cambridge University PressGreene, W. H. (1997), *Econometric Analysis*, Prentice Hall.

Guala, F. (2006), *Filosofia dell'economia*, il Mulino.

Maddala, G. S. (1983), *Limited-Dependent and Qualitative Variables in Economics*, Cambridge University Press.

McFadden, D. (1974), *The Measurement of Urban Travel Demand* in «Journal of Public Economics», 3, 303-328.

Perloff, J.M. (2001), *Microeconomia*, Apogeo.

Piccolo, D. (1998), *Statistica*, il Mulino.

Johnstone, J. (1966), *Econometric Metods*, Mc Graw Hill (trad. it. *Econometrica*, Franco Angeli)

Sartore, D. (2011), materiale del corso Introduzione all'econometria.

Sohn, K. T., J. H. Lee, e Y. S. Cho "*Ternary Forecast of Heavy Snofall in the Honam Area, Korea*" in «Advances in atmosferic science», vol. 26 no. 2, 2009, pag. 327-332.

AA. VV. (1994), *Eviews 4 User's Guide*. Quantitative Micro Software.

AA. VV. (1994), *Eviews 4 Command an Programming References.* Quantitative Micro Software.