Statistica per l'impresa

8. Valutazione statistica delle prestazioni delle imprese

Indici di bilancio

Analisi di bilancio: elaborazioni volte a caratterizzare la situazione (reddituale, patrimoniale...) in cui l'azienda si trova, mediante il calcolo di opportune misure di performance:

- struttura del capitale
- struttura finanziaria
- situazione finanziaria
  - ROD (oneri finanziari/debiti)
- redditivita'
  - ROE
  - ROA
  - ROS

che vengono dette *indici di bilancio*

Il bilancio e le aree gestionali

Concetti da ricordare:

- Stato patrimoniale
- Conto economico
- Nota integrativa

- Gestione operativa (caratteristica)
- Gestione non caratteristica
  - extracaratteristica
  - straordinaria

Può essere necessario *riclassificare* il bilancio secondo un certo schema interpretativo per evidenziare certi aspetti di interesse mediante il calcolo degli indici opportuni.

Scopi dell'analisi statistica

Come si posiziona l'azienda rispetto a:

- Valori teorici (standard)
- Valori medi

Sono necessarie banche dati aziendali:

- CERVED
- AIDA
- ...

e certi requisiti di qualita' dei dati:

- comparabilita'
- accuratezza
- pertinenza
- completezza

Valori medi e il bilancio somma

Siano per ciascuna impresa i di un gruppo di n imprese x_i un certo indice di bilancio, es. ROA: $x_i = \frac{B_i}{A_i}$ Per sintetizzare la tendenza centrale della distribuzione dell'indice sulle n imprese si possono calcolare:

- media semplice del ratio:

$$\bar{x} = \frac{\sum_{i=i}^{n} x_i}{n}$$

- media ponderata del ratio:

$$\bar{x}_P = \frac{\sum_{i=i}^{n} x_i A_i}{\sum_{i=i}^{n} A_i}$$

La media ponderata del ratio equivale al ratio che si otterrebbe dal *bilancio aggregato*, o *bilancio somma*, ottenuto sommando le poste contabili di tutte le imprese.

Particolarità empiriche dei dati di bilancio

L'analisi statistica dei bilanci ha due ordini di peculiarità:

- caratteristiche delle distribuzioni empiriche dei *ratio*
  - presenza di numerosi *outlier* (generalmente determinati da valori molto piccoli della posta al denominatore, pertanto la ponderazione li "annulla")
  - asimmetria (e troncamento): l'ipotesi di normalità è spesso poco plausibile
  - popolazione eterogenea caratterizzata da *sottogruppi*: spesso conviene identificare un sottoinsieme omogeneo come popolazione di riferimento
- poste che possono assumere valori positivi o negativi
  - l'indice risultante è *positivo* quando numeratore e denominatore sono *concordi*...
  - ... pertanto un ROE positivo può anche risultare dalla coesistenza di una perdita di esercizio con un capitale netto negativo!

Il benchmarking

Può essere interessante considerare, anziché quella relativa ai valori medi, la posizione di un indice aziendale nella distribuzione degli indici relativi a tutte le altre imprese confrontandolo con i *quantili* della distribuzione, spesso usando i *percentili*.

I risultati ottenuti, per esempio, da imprese che si situano sul 90 percentile del ROA possono costituire valori obiettivo da raggiungere. Se è possibile osservarle e studiarle, le prassi organizzative e gestionali di tali imprese eccellenti costituiranno delle *best practices*.

Il *benchmarking* consiste appunto in

- misurare le prestazioni "eccellenti" di un certo mercato (*benchmarking quantitativo*)
- analizzare (e seguire) le *best practices* che hanno permesso di raggiungerle (*benchmarking qualitativo*)

Analisi statistica multivariata degli indici di bilancio

Consideriamo *simultaneamente*

- vari indici di bilancio
- di varie (numerose) imprese

I *metodi statistici multivariati* servono per l'analisi *descrittiva* simultanea di più variabili e più unità.
Essendo generalmente metodi *computazionalmente intensivi*, hanno visto uno sviluppo considerevole in anni recenti, nell'ambito del fenomeno detto *data mining*: l'esplorazione e analisi statistica di "grandi" moli di dati. Ne vedremo due:

- l'*analisi in (delle) componenti principali* è una tecnica di *riduzione del numero delle variabili*
- l'*analisi dei gruppi*, o analisi dei *cluster*, permette di classificare le unità in gruppi omogenei rispetto alle variabili considerate.

Analisi delle componenti principali

Obiettivo: sintetizzare e ridurre i molteplici indici di bilancio disponibili a un numero "ragionevole" di componenti *incorrelate* e capaci di *spiegare la maggior parte della varianza complessiva*. p variabili $x_1, \ldots, x_p$, osservate su n unità. Costruiamo $k < p$ variabili, dette *componenti principali*

- combinazione lineare delle p variabili originali
- incorrelate tra loro

Con p variabili si possono ricavare al massimo p componenti principali (CP). La generica CP è così determinata:

$$C_h = a_{h1}x_1 + a_{h2}x_2 + \ldots + a_{hp}x_p$$

Il *punteggio*, o *score*, dell'i-esima unità sulla h-esima CP è

$$C_{ih} = \sum_{j=1}^{p} a_{hj}x_{ij}$$

Determinare i coefficienti delle CP

I coefficienti a_{hj} sono funzione dei dati. Essi sono determinati in modo che le CP siano:

- tra loro incorrelate
- ordinate gerarchicamente secondo la varianza:

$$var(C_1) \geq var(C_2) \geq \ldots \geq var(C_p)$$

- riproducono la varianza totale delle variabili originarie:

$$\sum_{j=1}^{p} var(x_j) = \sum_{h=1}^{p} var(C_h)$$

I coefficienti a_{hj} sono determinati tramite una tecnica di *rotazione degli assi fattoriali*.

Efficacia e obiettivi

I principali obiettivi dell'analisi sono dunque:

a) individuare la posizione di un'unità di osservazione (impresa) rispetto alle altre

b) comunicare in modo compatto i risultati dell'analisi statistica

La capacità delle CP di ridurre le p variabili originarie dipende da:

a) varianza delle variabili originarie: le x_j con la massima varianza assumeranno peso maggiore nelle varie CP (opportuna la *standardizzazione*?)

b) correlazione tra le variabili originarie: tanto più le p variabili originarie sono correlate, tanto più la loro varianza potrà essere riprodotta da un numero limitato di CP

Le fasi dell'analisi in CP

Le fasi dell'analisi in CP:

1. selezione delle unità (imprese che vogliamo comparare con la "nostra")
2. scelta delle variabili ed eventuale standardizzazione
3. calcolo delle CP (= dei coefficienti a_{hj})
4. scelta delle componenti da considerare: si prendono le prime k CP in modo che queste
   a) spieghino complessivamente almeno, per esempio, il 70% della varianza totale
   b) spieghino ciascuna più della varianza spiegata media
5. interpretazione delle componenti principali e conseguente loro utilizzo nell'analisi del posizionamento dell'impresa

Analisi cluster

Identificare *gruppi omogenei di unità* rispetto a p variabili di classificazione (di *imprese* rispetto a *indici di bilancio*). Si distinguono metodi:

- gerarchici: i gruppi provengono dall'aggregazione progressiva di altri gruppi
- non gerarchici: i gruppi provengono direttamente dall'aggregazione di unità

Metodi gerarchici:

- divisivi: partendo da un unico gruppo di n unità le si suddividono in gruppi fino a ottenere n gruppi ciascuno di 1 unità
- agglomerativi: da n unità si aggrega in sottogruppi fino ad arrivare a un solo gruppo

- i metodi non gerarchici prefissano il numero g dei gruppi da ottenere
- nei metodi gerarchici il numero "ottimale" di gruppi g è il risultato della procedura agglomerativa

Analisi cluster gerarchica

Si parte da una matrice di misure di *dissimilarità* (per variabili quantitative: *distanza*) a coppie. Dato un insieme di p variabili osservate sulle n unità, cosicché $\mathbf{x}_i = [x_{i1}, x_{i2}, \ldots, x_{ip}]$, prese due unità U_i, U_j, la distanza

$$d(U_i, U_j) = d(\mathbf{x}_i, \mathbf{x}_j)$$

è tale che:

1. identità: $d(U_i, U_j) = 0$ se e solo se $\mathbf{x}_i = \mathbf{x}_j$
2. non negatività: $d(U_i, U_j) \geq 0$ per ogni i, j
3. simmetria: $d(U_i, U_j) = d(U_j, U_i)$ per ogni i, j
4. triangolarità: $d(U_i, U_j) \leq d(U_i, U_w) + d(U_w, U_j)$ per ogni i, j, w

Alcune definizioni di distanza

Alcune definizioni di distanza:

- distanza euclidea:

$$D(U_i, U_j) = \sqrt{\sum_{k=1}^{p}(x_{ik} - x_{jk})^2}$$

- distanza city-block:

$$D_M(U_i, U_j) = \sum_{k=1}^{p}|x_{ik} - x_{jk}|$$

La distanza city-block è detta anche *distanza di Manhattan*
Gli *indici di distanza* soddisfano solo le $1 - 3$. Esempio: $D^2(U_i, U_j)$ viene spesso impiegato nelle analisi cluster. D'ora in poi parleremo genericamente di indici.

Fasi dell'analisi cluster

Le fasi fondamentali dell'analisi cluster:

1. scelta delle unità
2. scelta delle variabili di classificazione (ed eventuali trasformazioni)
3. scelta dell'indice di distanza
4. scelta del criterio gerarchico agglomerativo
5. interpretazione dei gruppi individuati

Aspetti da considerare

Scelta delle variabili

- Le variabili devono essere espresse nella stessa unità di misura
- Quelle con la varianza più elevata avranno più peso (opportuno standardizzare?)
- Variabili tra loro molto correlate avranno grande peso nel calcolo della distanza: una soluzione è utilizzare le CP anziché le variabili originarie

Scelta dell'indice di distanza

- L'indice scelto determina il risultato della procedura
- La distanza euclidea (e il suo quadrato) applicati alle componenti principali godono di interessanti proprietà

Scelta dell'algoritmo di raggruppamento

- L'algoritmo definisce come calcolare le distanze tra unità e *gruppi*, e tra i gruppi

La logica del raggruppamento gerarchico

Con n unità, l'algoritmo opera in $n-1$ passi:

1. si uniscono le due unità meno distanti, ottenendo il gruppo G_1
2. si ricalcola la matrice di distanza tra le unità rimaste e il nuovo gruppo
3. su uniscono le due unità *o gruppi* meno distanti ottenendo il gruppo G_2
4. ...
5. al passo $n-1$ si uniscono i due gruppi, o il gruppo e l'unità, rimasti nell'unico gruppo G_{n-1}

Il calcolo delle distanze tra gruppi

L'algoritmo (o criterio) agglomerativo stabilisce come calcolare le distanze tra gruppi:

- metodo del *legame singolo*: la distanza $d(G_i, G_j)$ è uguale alla *minima* distanza tra unità appartenenti una al primo e una al secondo gruppo
- metodo del *legame completo*: la distanza $d(G_i, G_j)$ è uguale alla *massima* distanza tra unità appartenenti una al primo e una al secondo gruppo
- il metodo della *distanza media* prende la media di tutte le distanze tra le possibili coppie di unità appartenenti una al primo e una al secondo gruppo
- il *metodo di Ward* si basa sulla scomponibilità della devianza tra *intragruppo* e *intergruppo*, e procede ad aggregare la coppia che comporta il minimo incremento della devianza interna

Interpretazione

Una volta individuati i gruppi, si vorrà:

- verificare quanto essi siano
  - omogenei al loro interno
  - eterogenei tra loro
- individuare le variabili che maggiormente contribuiscono a differenziare (*discriminare*) i gruppi

Sarà utile a questi fini

- studiare le distribuzioni univariate degli indici di bilancio all'interno di ogni gruppo
- scomporre la devianza totale (*sst*) in
  - devianza intragruppo (*ssw*)
  - devianza intergruppo (*ssb*)

  per valutare la quota attribuibile alla differenza tra gruppi (*ssb*/*sst*)
- rappresentare ciascun gruppo in base alle medie degli indici di bilancio (*centroidi*)

Il rischio di insolvenza

Supponiamo di disporre di informazioni di bilancio consistenti in p indici e di volerla utilizzare per valutare il rischio di insolvenza (*analisi fondamentale)*. L'analisi statistica permette di

- Descrivere come si manifesta e si sviluppa una crisi aziendale
- Predire il livello di rischio associato a una determinata configurazione degli indici di bilancio

A scopi predittivi, si costruiscono modelli statistici capaci di

- identificare gli indici che possono segnalare una crisi di insolvenza
- produrre un punteggio sintetico:
  - uno *score* oppure
  - una *probabilità di insolvenza*

  che sintetizzi l'informazione rappresentata dai p indici di bilancio

Interpretazione

L'individuazione degli indici premonitori richiede di disporre di un campione che contenga sia *imprese insolventi* che *imprese sane*.
Lo scopo predittivo dell'analisi discriminante (AD) servirà per *classificare* a priori le imprese, di cui non si conosce l'appartenenza, nel gruppo delle imprese insolventi o in quello delle imprese sane:

- il comportamento passato di un gruppo di imprese con certe caratteristiche viene utilizzato per prevedere quello futuro di altre imprese con caratteristiche simili
- l'AD consente di sintetizzare in un solo indice (*score* o *punteggio discriminante*) il profilo dell'impresa espresso dai p indici
- tale punteggio *discrimina* se un'impresa appartenga a un gruppo o all'altro

Fasi dell'analisi discriminante

La predisposizione di un modello per la previsione delle insolvenze aziendali richiede:

- individuazione della popolazione di riferimento e bipartizione nei due gruppi (sane/insolventi)
- formazione dei due campioni (sane/insolventi)
- selezione delle variabili
- scelta di una regola classificatoria, definizione e stima della funzione discriminante
- determinazione del valore critico (*cut-off point*) che, associato alla funzione discriminante, permetta di separare i due gruppi (sane/insolventi)
- validazione della regola stimata

Selezione dei due campioni di imprese

La costruzione di un modello di previsione dell'insolvenza si basa su dati *retrospettivi*

- Selezione della popolazione di riferimento e bipartizione dei gruppi
  - per un insieme di imprese il più possibile omogenee (merceologicamente)
  - si raccolgono informazioni ai tempi t e $t - d$
  - si individua un criterio non ambiguo per identificare le imprese insolventi
- Formazione dei due campioni (sane/insolventi)
  - le imprese insolventi sono in genere assai meno numerose, pertanto di solito si considerano tutte quelle per le quali sono disponibili i dati in $t - d$
  - le imprese sane sono più numerose; tra queste si sceglie il campione delle imprese *sane* rispettando opportune politiche di *bilanciamento* (spesso si ricorre alla stratificazione) per individuare un gruppo di controllo che sia omogeneo alle imprese insolventi

Selezione delle variabili

Generalmente si utilizzano tre tipi di variabili:

a. indici di bilancio
b. variabili derivate dagli indici di bilancio
   - variazione media
   - varianza
c. variabili macro

Scelta della regola classificatoria

Consideriamo una popolazione di imprese composta da due sottopopolazioni o gruppi *disgiunti ed esaustivi* G_1 e G_0, dove G_1 è il gruppo delle aziende insolventi e G_0 di quelle sane al tempo t.
Supponiamo di poter osservare

- al tempo $t - d$
- su ogni unità i

le p variabili (gli indici di bilancio). Sia x_i il vettore $1 \times p$ di tutte le variabili relative all'i-esima unità.
A priori, nell'AD normale si ipotizza che X_i sia la determinazione di una variabile $p-$variata la cui distribuzione è condizionata al gruppo:

$$X|G_j \sim f(x; \mu_{G_j}, \Sigma);\ j = 0, 1$$

dove $f(.)$ è la distribuzione congiunta normale multivariata.

L'AD normale

La regola classificatoria della *massima verosimiglianza* prevede che:

- se $f(x_i|G_1) > f(x_i|G_0)$ (ovvero: *è più probabile osservare congiuntamente le caratteristiche dell'impresa i-esima se questa appartiene a* G_1) allora l'unità i-esima è assegnata a G_1
- altrimenti l'unità i-esima è assegnata a G_0

Equivalentemente, applicando il logaritmo (trasformazione monotona), la regola di assegnazione a G_1 diventa $ln(f(x_i|G_1)) > ln(f(x_i|G_0))$, da cui ponendo $h(x_i) = ln(f(x_i|G_1)) - ln(f(x_i|G_0))$ la regola diviene:

- se $h(x_i) > 0$ allora l'unità i-esima è assegnata a G_1
- altrimenti l'unità i-esima è assegnata a G_0

$h(x_i)$ è detta *funzione discriminante* di massima verosimiglianza per il cutoff 0.

Esempio: AD normale univariata

Nel caso normale univariato ($p = 1$) e con *varianza uguale tra i gruppi*, la funzione discriminante diviene la variabile stessa ($h(x_i) = x_i$) e la regola di massima verosimiglianza diviene:

- se $x_i > x^*$ allora l'unità i-esima è assegnata a G_1
- altrimenti l'unità i-esima è assegnata a G_0

Si può dimostrare come il cutoff in questo caso sia

$$x^* = \frac{\mu_{G_0} + \mu_{G_1}}{2}$$

La funzione discriminante lineare

Si noti come (in questo caso dove $p = 1$, ma anche per p arbitrario) l'AD produca sempre uno *score* scalare: la differenza fra le log-verosimiglianze.

- Nel caso generale in cui le p variabili casuali abbiano varianze e covarianze uguali tra G_1 e G_0, si ottiene la *funzione discriminante lineare*

$$Z = a_1 X_1 + a_2 X_2 + \ldots + a_p X_p$$

  dove le variabili X_k sono espresse in scarti dalla media
- Altrimenti, se le varianze fossero diverse, si produrrebbe una funzione discriminante *quadratica*: operativamente, la procedura sarebbe simile.

Come osservato, l'AD produce uno *score* scalare anche per p arbitrario. In altre parole, l'AD riduce lo spazio delle p variabili di classificazione ad una sola variabile, lo score discriminante.

Estensione dell'AD in senso Bayesiano - 1

Supponiamo che i gruppi abbiano numerosità (molto) differenti: ciò è tipico del nostro caso (ci sono in genere molte più imprese sane che insolventi).
Poniamo ora che la probabilità *a priori* che l'unità i-esima cada in G_1 sia ϕ_1. Dal Teorema di Bayes segue che ora la probabilità *a posteriori* che l'unità i-esima provenga da G_1 *condizionatamente all'aver osservato* x_i è

$$P(G_1|x_i) = \frac{f(x_i|G_1)\phi_1}{f(x_i)}$$

dove $f(x_i|G_1)$ è la verosimiglianza di x_i condizionatamente a G_1 e $f(x_i)$ è la densità marginale:

$$f(x_i) = f(x_i|G_0)\phi_0 + f(x_i|G_1)\phi_1;\ \ \phi_0 + \phi_1 = 1$$

Estensione dell'AD in senso Bayesiano - 2

In generale, per un numero arbitrario di gruppi $J+1:\ 0,1,\ldots,j,\ldots,J$,

$$P(G_j|x_i) = \frac{f(x_i|G_j)\phi_j}{f(x_i)}$$

e la densità marginale

$$f(x_i) = \sum_{j=0}^{J} f(x_i|G_j)\phi_j;\ \sum_{j=0}^{J} \phi_j = 1$$

si può quindi definire la seguente regola classificatoria:

- se $P(G_1|x_i) > P(G_0|x_i)$ allora l'unità i-esima è assegnata a G_1
- altrimenti l'unità i-esima è assegnata a G_0

Si vede come nel caso l'appartenenza ai gruppi sia a priori equiprobabile, questa equivalga alla regola della massima verosimiglianza.

L'AD logistica - 1

Alternativamente, si può specificare direttamente $P(G_1|x_i)$ e poi stimarla: nell'AD *logistica* si pone:

$$P(G_1|x_i) = \frac{e^{\beta_0+\beta_1 x_{i1}+\ldots+\beta_p x_{ip}}}{1 - e^{\beta_0+\beta_1 x_{i1}+\ldots+\beta_p x_{ip}}}$$

da cui si ottiene (in termini delle cosiddette *odds ratio*)

$$\frac{P(G_1|x_i)}{P(G_0|x_i)} = \frac{P(G_1|x_i)}{1 - P(G_1|x_i)} = e^{\beta_0+\beta_1 x_{i1}+\ldots+\beta_p x_{ip}}$$

e, passando ai logaritmi,

$$ln(\frac{P(G_1|x_i)}{P(G_0|x_i)}) = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_p x_{ip}$$

L'AD logistica - 2

A questo punto la regola classificatoria diventa:

- se $logit(x_i) = ln(\frac{P(G_1|x_i)}{P(G_0|x_i)}) > 0$ allora l'unità i-esima è assegnata a G_1
- altrimenti l'unità i-esima è assegnata a G_0

siccome $logit(x_i)$ è funzione monotòna di $P(G_1|x_i)$; questa diventa la funzione discriminante lineare, associata al cutoff 0, e $z = logit(x_i)$ è lo score discriminante.

L'AD logistica si configura quindi come una particolare analisi di regressione, dove la variabile dipendente è la probabilità di appartenere al gruppo G_1.
In particolare, le variabili esplicative vengono qui viste come dati osservati anziché come variabili aleatorie, con importanti conseguenze:

- non è richiesta la normalità
- possono anche essere di carattere qualitativo

Validazione e impiego della regola classificatoria

Una volta stimata la regola classificatoria, bisogna valutarne le qualità.

Come per tutti i modelli, si può distinguere tra analisi *interna* ed *esterna*:

- L'analisi *interna* riclassifica le unità del campione, valutando a quale dei due gruppi il modello *assegnerebbe* l'unità e confrontandolo con la realtà: si identificano così
  - proporzione di unità correttamente riclassificate nel gruppo 0 o 1
  - proporzione di *falsi positivi* (unità di G_0 classificate erroneamente in G_1)
  - proporzione di *falsi negativi* (unità di G_1 classificate erroneamente in G_0)
- L'analisi *esterna* prevede di usare due campioni:
  - uno che viene usato per stimare il modello (*training set*)
  - uno di cui sono noti gli esiti ma che non è stato usato per stimare il modello (*test set*)

L'analisi *esterna* è più indicata per valutare la capacità predittiva del modello.

Variabili dipendenti binarie

Vari esempi in cui la variabile dipendente è binaria (dicotomica): si vuole modellizzare

- perché certe aziende pagano dividendi e altre no
- quali fattori deteriminano il default sul debito sovrano
- perché certe aziende si finanziano con capitale proprio e altre con capitale di debito

In tutti questi casi, la variabile dipendente può essere rappresentata in forma binaria/logica/booleana: insomma come 0 oppure 1

Il modello di probabilità lineare

Il primo approccio che viene naturale è noto come *modello di probabilità lineare*

- è basato sull'ipotesi che la probabilità di un evento sia funzione lineare di un insieme di regressori

$$P_i = p(y_i = 1) = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + \cdots + \beta_k x_{ki} + u_i$$

- Non potendo osservare le probabilità, si prendono come variabile dipendente i rsultati osservati, y_i (una serie di zeri e uni)
- Questo è un modello lineare, che si può stimare con gli OLS, includendo regressori di qualsiasi tipo
- I valori stimati rappresentano le probabilità $y_i = 1$ per ogni osservazione i.

Il modello di probabilità lineare

- i $\hat{\beta}_{OLS}$ possono essere interpretati come l'incremento di probabilità che $y = 1$ per una variazione unitaria di un dato regressore, tenendo costanti tutti gli altri
- supponiamo ad esempio di voler modellare la probabilità che un'azienda i paghi dividendi ($p(y_i = 1)$ in funzione della capitalizzazione di mercato (x_{2i}, misurata in milioni di dollari), stimando:

$$\hat{P}_i = -0.3 + 0.012 x_{2i}$$

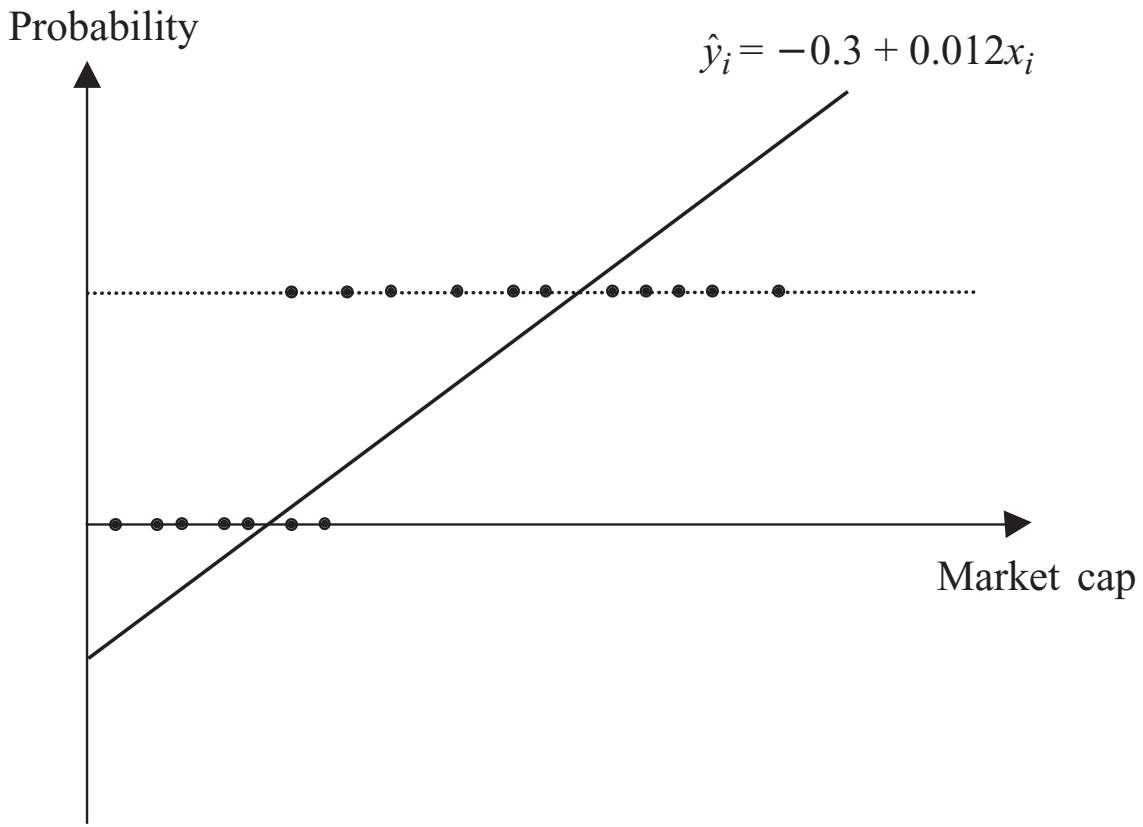
  dove $\hat{P}_i$ denota la probabilità stimata per l'azienda i.
- Questo modello suggerisce che per ogni \$1m di incremento dimensionale, la probabilità che l'azianda paghi dividendi cresce dello 0.012 (o 1.2%).

Il modello di probabilità lineare (Cont'd)

- Un'azienda con capitalizzazione di mercato di \$50m avrà una probabilità $-0.3+0.01250=0.3$ (o 30%) di pagare dividendi.

Problemi del modello di probabilità lineare

- Graficamente, la situazione si piò rappresentare come segue



Problemi del modello di probabilità lineare

Per tutta la sua semplicità di impiego e intuitività, il LPM può produrre stime inaccettabili:

- per ogni azienda dalla capitalizzazione inferiore a \$25m, la probabilità stimata è negativa, mentre se essa è superiore a \$88m la probabilità è maggiore di 1.
- Chiaramente, si tratta di valori inaccettabili per una probabilità, che deve per definizione essere compresa in (0,1).
- Una soluzione ovvia è il troncamento delle probabilità previste a 0 o 1, cosicché, per esempio, una probabilità di -0.3 verrebbe posta uguale a 0 e, rispettivamente, una probabilità di 1.2 a 1.
- Tuttavia, due conseguenze di tale troncamento:
  - troppe osservazioni concentrate sugli estremi 0 e 1
  - non è plausibile stimare probabilità di 0 (impossibilità pratica) o 1 (pratica certezza)

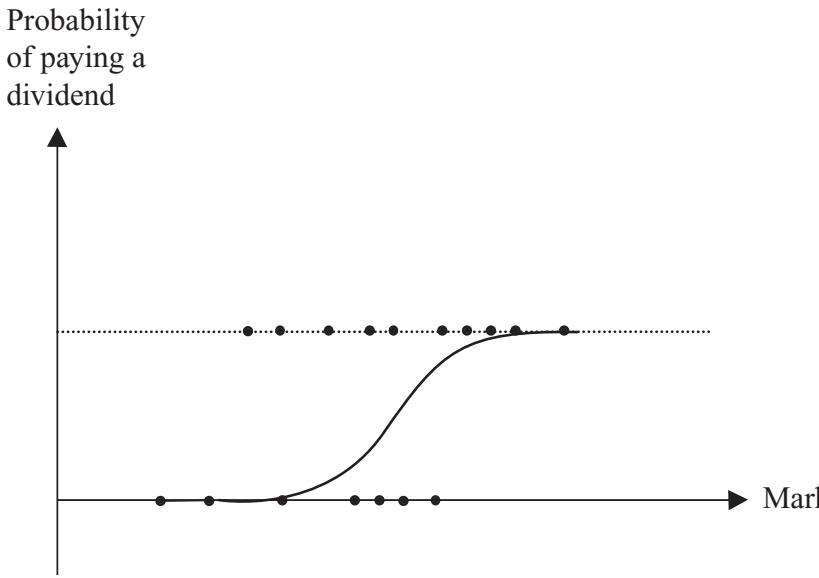
Problemi del modello di probabilità lineare

Inoltre, il LPM soffre di unteriori problemi dal punto di vista econometrico:

- Se y assume due soli valori, sotto l'ipotesi di X non stocastica anche gli errori assumono due soli valori
  - pertanto non si può assumere che l'errore sia normalmente distribuito.
- Inoltre, dato che l'errore cambia sistematicamente con le X, esso sarà eteroschedastico
  - pertanto bisogna sempre usare stimatori per ES *robusti*.

Il modello Logit (e il Probit)

- I modelli di tipo *Logit* (e gli analoghi *Probit*) superano le limitazioni del LPM, "costringendo" le stime ad assumere valori "plausibili".
- Essi trasformano le stime del predittore lineare $X\hat{\beta}_{OLS}$ verso l'intervallo (0,1) per mezzo di una *funzione link*
- Visivamente, il modello stimato avrà la forma di una curva a "S" anziché quela di una retta (come era per il LPM).



Il modello Logit

- Il modello *Logit* è così chiamato perché usa la cumulata di una distribuzione logistica per trasformare il predittore lineare verso il dominio (0,1).
- Grazie alla trasformazione logistica, 0 e 1 sono asintoti per i valori stimati e perciò la probabilità prevista non sarà mai esattamente 0 o 1, per quanto possa avvicinarsi.
- Il modello Logit è non lineare, non è linearizzabile tramite trasformazioni e perciò non può essere stimato con gli OLS.
- La stima richiede di usare il metodo della Massima Verosimiglianza (ML).

Interpretazione dei parametri

- Gli ES e i t-ratios vengono calcolati automaticamente dal software e permettono di condurre test di significatività al "solito" modo
- tuttavia, l'interpretazione dei coefficienti come derivate parziali di y rispetto a x non è più possibile:
- affermare che un incremento unitario in, per esempio, x_{2i} produca un incremento $\beta_2\%$ nella probabilità che $y_i = 1$ (come nel caso del LPM) non è più corretto.
- Infatti, la forma funzionale nnon è più quella del LPM: $P_i = \beta_1 + \beta_2 x_{2i} + u_i$ bensì $P_i = F(x_{2i})$ dove F rappresenta la cumulata della funzione logistica.

Interpretazione dei parametri

- Per ottenere il rapporto tra variazioni in x_{2i} e P_i (la *derivata parziale*), dobbiamo differenziare F rispetto a x_{2i}: tale derivata è $\beta_2 F(x_{2i})$.
- Quindi, un incremento unitario in x_{2i} è associato a un incremento $\beta_2 F(x_{2i})$ nella probabilità che $y_i = 1$.
- Queste derivate parziali sono note come *effetti marginali*, e sono, come osservato, *funzioni di* x_{ki}.
- Un modo per presentare tali quantità in modo simile a quello del modello classico e del LPM (dove sono costanti) è quello di calcolarle in corripondenza dei valori medi $\bar{x}_k$.

Bontà di adattamento nel modello Logit

- Seppure il loro calcolo è tecnicamente possibile, le "solite" misure di adattamento come RSS ed R^2, queste non hanno significato in questo contesto.
- R^2, se calcolato nel consueto modo, sarà fuorviante perché $\hat{y}$ può assumere tutti i valori tra 0 e 1, mentre y è binaria.
- Pertanto, se $y_i = 1$ e $\hat{P}_i = 0.8$, il modello ha previsto "bene", fatto che non verrà colto pienamente dall'R^2.
- Si impiegano comunemente due misure di bontà di adattamento:
  - La percentuale di y_i previsti correttamente
  - Lo 'pseudo-R^2' di McFadden, definito come uno meno il rapporo tra la log-verosimiglianza del modello contro quella del modello "vuoto", contenente solo l'intercetta (v. l'F-test).