# Statistica per l'impresa

Variabili dipendenti binarie

# 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:

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

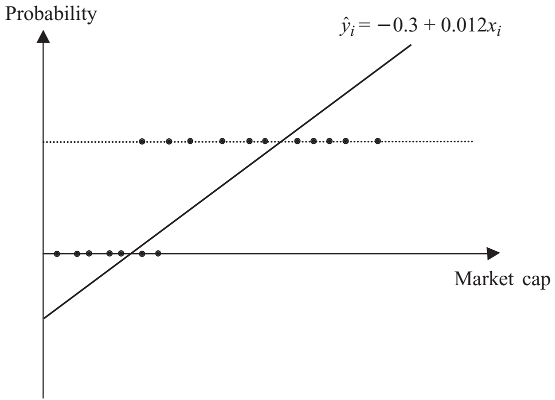
  dove $\hat{P}_i$ denota la probabilità stimata per l'azienda $i$.

- Questo modello suggerisce che per ogni \$1m di incremento dimensionale, la probabilità che l'azianda paghi dividendi cresce dello 0.012 (o 1.2%).

- 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)

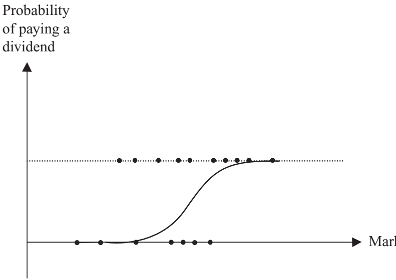
# Problemi del modello di probabilità lineare

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).