$$y_t = B_1 + B_2 x_T + u_t$$

t perché abbiamo una serie di osservazioni nel modello e non una sola. Può essere l'istante temporale. E' un numero va che da 1 ad N.

ut è il termine di errore, non osservata. Chiamata anche disturbo. E' un'imprevedibilità. Come la rappresentiamo? Con una variabile casuale. Rende il modello stocastico (e non deterministico) → quando c'è una variabile casuale il modello è stocastico.

Yt è la variabile dipendente, nel modello reddito è il consumo

Xt è la variabile esplicativa, dipendente

B1 e B2 sono coefficienti non osservati ma stimabili. B1 è l'intercetta, la costante e b2 è il coefficiente ad angolare

Assumiamo che l'aspettativa del termine ut sia zero, solo in questo modo possiamo definire in modo univoco e identificare i parametri b1 e b2. Valore atteso ut=0.

Spieghiamo il modello di regressione in termini finanziari prendendo il modello del Capital asset price model: cerca di spiegare l'eccesso di rendimento di un titolo. Mette come variabile dipendente l'eccesso di rendimento del titolo tesla e lo fa dipendere dall'eccesso di rendimento del portafoglio di mercato. L'eccesso di rendimento del portafoglio di mercato x è l'indice nasdaq. Si va a stimare il b del titolo, andiamo a vedere se il titolo è difensivo o aggressivo rispetto l'indice di mercato. Se b2 > 1 è particolarmente speculativo, ogni volta che c'è un piccolo eccesso di rendimento il titolo tesla ha un'amplificazione dell'eccesso di rendimento (in modo maggiore). Se b2<1 è difensivo, ha una ripercussione sul titolo più piccola. Il modello implica che l'intercetta è zero, quindi il b1 non è statisticamente diverso da 0 (non possiamo escludere che il vero parametro possa essere uguale a 0).
Un altro modello può essere cercare la probabilità che il rendimento della borsa di Milano oggi siano -10. Andiamo a creare un campione, una raccolta di dati per definire una variabile casuale caratterizzata dalla media dei rendimenti della borsa di Milano e vedere qual è la probabilità di osservare un -10.

$y_t = B_1 + B_2 x_T$ → è la funzione di regressione (senza disturbo)

Pensiamo a stimare un modello come il CAPM, stimiamo il b2 e vediamo se il titolo che abbiamo di fronte sia più o meno un titolo speculativo. E' un modello che è una specie di realtà virtuale. Con il CAPM abbiamo la necessità di misurare l'eccesso di rendimento, una possibilità quindi per verificare e stimare il CAPM è scaricare i prezzi, calcolare i rendimenti e andare a realizzare un modello lineare, questo è un modello di regressione. Un elemento essenziale è quello di capire dove preferire i dati, yahoo finance mette una vasta quantità di dati a disposizione.

Un altro caso è stimare i prezzi delle abitazioni. Si considerano i fattori che possono influenzare i prezzi (come i metri quadrati). Abbiamo delle variabili esplicative, come il numero di bagni e la dimensione, e una variabile dipendente che è il prezzo. Prima di capire nei dettagli come sia possibile utilizzare lo strumento di regressione dobbiamo introdurre il modello.

$$y_t = B_1 + B_2 x_T + u_t$$

La u rappresenta la nostra parziale conoscenza del comportamento umano. Imprevedibile perché non abbiamo conoscenza su tutti i possibili eventi verificabili. (ricordiamo che x e y sono osservabili)

Una variabile casuale è il modo in cui viene matematicamente viene rappresentata l'incertezza rispetto un determinato evento, è una funzione che può essere di due categorie: discreta (quando assume solo valori discreti) e continua (i valori sono definiti lungo la linea dei numeri reali). Ad ogni possibile valore assunto dai valori casuali possiamo associare una probabilità.

VARIABILI CASUALI DISCRETE

- La variabile discreta generica può assumere un numero finito di valori (x1,x2)
- La somma delle probabilità deve essere uguale a 1. $\sum_{i=1}^{N} P(x_1) = 1$
- Ad ogni variabile casuale discreta associo una probabilità.

VARIABILI CASUALI CONTINUE

- Viene associata alla variabile casuale una funzione di distribuzione cumulata f(x) e una densità (dx).
- La cumulative distribution function mi dice la probabilità che la variabile casuale assuma un valore uguale o più piccolo di un numero determinato).

  $$f(x) = P_r(x \leq x)$$

- La variabile dipendente è di solito continua. Nella tavola distribuzione standard la probabilità che la curva normale standard sia inferiore o uguale a 0 è del 50%

Quali regole segue qualsiasi distribuzione di probabilità?

- Non esiste una probabilità negativa. Devono essere comprese tra 0 e 1.
- La p di osservare l'insieme vuoto è 0 e La p di osservare tutti i possibili eventi è 1
- Se abbiamo due eventi disgiunti e vogliamo calcolare la probabilità che si verificano, questa è data dalla somma delle probabilità che si verifichi uno + che si verifichi l'altro.
- La funzione di probabilità cumulata tende a 0 se x tende a – infinito
- Tende a 1 se x va verso infinito (consideriamo tutti i possibili valori)
- La funzione cumulata è incrementale (increasing function)
- Per una funzione continua se calcoliamo la p di un singolo punto è 0. Se vogliamo calcolare data la probabilità di una normale standard di osservare esattamente il valore 0, la probabilità è 0. Dobbiamo avere almeno un intervallo

La funzione di densità

- E' la derivata della funzione cumulata = f(x)
- f(x) = F'(x)

- $\int_{-\infty}^{\infty} \quad (x)d = \int_{-\infty}^{\infty} \quad F'(x)d = F(\infty) - \mathrm{F}(-\infty) = 1$

- $Pr(a \leq x \leq b) = F(b) - F(a)$ = $\int_a^b f(x)\,dx$

- Se PDF è una normale standard → $2\pi^{-1/2} exp\left(-\frac{1}{2}x^2\right)$

Variabile di tipo binario: assume solo due valori, 0 e 1. La p di avere 0 è p e la probabilità di avere 1 è la sua complementare (1-p).

F(x) $\begin{cases} 0 & con\ x < 0 \\ P & con\ 0 \leq x < 1 \\ 1 & con\ x > 1 \end{cases}$

Che cosa sono i momenti di una distribuzione? E' un modo di caratterizzare la variabile casuale. Abbiamo il primo momento, secondo, terzo e di ordine superiore.

Il valore atteso di una variabile casuale = è un modo per caratterizzare ciò che succede a quella variabile nel caso in cui possiamo osservare molteplici osservazioni. E' rappresentato con la lettera E.

VALORE ATTESO VARIABILE CONTINUA

$$E(x) = \int_{-\infty}^{\infty} x \cdot f(x)\,dx$$

VALORE ATTESO VARIABILE DISCRETA

$$E(x) = \sum_{i=1}^{n} p(x_1)x_1$$

MOMENTO DI ORDINE K

$$Mk(x) = \int_{\infty}^{\infty} x^k f(x)\,dx$$

(momento di ordine 1 di x= all'integrale tra meno inf e piu di x elevato a k per fx in dx) E' quello che abbiamo visto per la variabile casuale. Non è centrato, per esserlo bisogna sottrarre il valore atteso prima dell'elevamento a potenza. Il momento 1 è il valore atteso.

Il secondo momento di una distribuzione vuol dire m con k = 2.
Il secondo momento centrato è la varianza.

$$VAR(x) = \int_{\infty}^{\infty} (x-u)^2 f(x)dx$$

Questi momenti sono momenti non centrati. Il momento è centrato quando gli sottraiamo il valore atteso (per il momento superiore al primo, sottraiamo il valore atteso dalla x prima di effettuare l'elevamento a potenza). Nel caso continuo:

$$\int_{-\infty}^{\infty} (x-\mu)^k f(x)\, dx$$

u= aspettativa di x (il valore atteso)

Nel caso discreto: $u = E\left(x - E(x)\right)^k$ = $\sum_{i:1}^{N} P(x_1)(x-\mu)^k$

Skewness → momento di ordine 3 (denota la simmetria e l'asimmetria di una distribuzione). Ad esempio la gaussiana è una distribuzione simmetria (quanta massa c'è a destra ce né a sinistra). Se la skewness è uguale a 0 la distribuzione è simmetrica.

Kurtosis → momento di ordine 4(massa di probabilità presenti nelle code della distribuzione, p di osservare valori estremi). Se è =3 è una normale distribuzione, se è maggiore a 3 c'è una probabilità più elevata rispetto alla distribuzione normale di avere valori estremi) Quanto è probabile avere valori outlier (estremi).

Lo stesso calcolo che abbiamo fatto per il caso continuo lo facciamo per il discreto

La varianza è il secondo momento centrato di una distribuzione.

La media è l'aspettativa di una variabile casuale o di una distribuzione

La varianza è una misura di dispersione (se è 0 c'è una costante non una variabile casuale)

Deviazione standard = radice della varianza. Assume solo valori positivi

Standard error: stima della deviazione standard.

DENSITA' DISTRIBUZIONE GAUSSIANA

$$f(x) = \frac{1}{\sqrt{\sigma 2\Pi}} exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right);\ -\infty < x < \infty$$

When x has a normal distribution with mean μ and variance σ 2 , we write

x ~ N (μ, σ 2) → distribuzione come una normale standard con media e varianza.

SECONDO MOMENTO NON CENTRATO: integrale +∞ −∞ x^2 fX(x) dx

L'area che c'è sotto la standard normal è esattamente pari a 1

Il secondo momento non centrato = 1. Il valore atteso al quadrato = 0 (perché il valore atteso è 0), e la varianza è pari a 1. Di solito rappresentiamo la variabile come x distribuita come una gaussiana con media zero e varianza 1. In una standard normal la probabilità di avere un numero maggiore di zero positivo è 0,5 e di avere un numero compreso tra meno 2 e 2 è 0,96%. STUDIARE PER VERO FALSO

$$E(x)^2 = 0$$

$$VAR(x) = E(x^2) - E(x^2) = 1$$

$$p_x(x \geq 0) = 0{,}5$$

$$p_r(-2 \leq x \leq 2) \approx 0{,}96$$

Ogni volta che abbiamo il termine di errore anche la y diventa una variabile casuale.

DISTRIBUZIONE MULTIVARIATA

Ci spostiamo dal considerare una singola variabile casuale a più di una variabile casuale.
La cumulata è F di x1 x2, è la pr che x1 assuma un valore più piccolo uguale di un x1, intersezione che la seconda variabile casuale (X2) sia più piccolo di x2. PFD se esiste è la derivata della funzione congiunta rispetto la prima variabile casuale e rispetto la seconda variabile casuale.
Se abbiamo due variabili casuali abbiamo una distribuzione di probabilità CONGIUNTA (non una sola come accade con una variabile casuale).
Abbiamo quindi integrale doppio. (11)

Se le due variabili casuale sono INDIPENDENTI (totalmente) si può fattorizzare la distribuzione congiunta perché non hanno comportamento congiunto. (12). A volte si parla di indipendenza lineare (la covarianza è uguale a 0 quindi sono indipendenti linearmente, MA non si esclude che hanno una dipendenza di tipo quadratico. Per questo specifichiamo che le variabili sono totalmente indipendenti, e per questo si può fattorizzare).

La prob condizionata ci permette di capire data l'osservazione di una variabile cosa succede all'altra variabile. Pensiamo ad esempio, io devo uscire di casa e non so se prendere l'ombrello. Il cielo è grigio, il fatto di sapere che c'è quella manifestazione induce ad avere una probabilità mutata sul fatto che possa piovere e quindi prendere l'ombrello. Se non avessi finestre, non ho l'info legata alla pioggia, quindi ci sarebbero molteplici elementi da considerare. Conoscere come è la condiziona atmosferica modifica le probabilità di prendere l'ombrello → probabilità condizionata. (simbolo intersezione vuol dire che sono condizionate).

Pr(A ∩ B) = Pr(B)Pr(A | B) Il simbolo u contrario vuol dire che sono condizionate.

Il fatto di avere questa conoscenza parziale che mi aiuta a definire meglio la probabilità dell'altra variabile è estremamente importante, lo vedremo con il modello di regressione lineare.

Come esiste la probabilità condizionata, esiste anche l'aspettativa condizionata. Due proprietà importanti sono legate all'aspettativa condizionata

Legge aspettative reiterate: se devo fare un'aspettativa condizionata, (valore atteso di x1 dato x2) vado ad applicare su questa aspettativa di nuovo l'operatore atteso, ottenendo l'aspettativa incondizionata. (se c'è la barra è condizionata, se non c'è è incondizionata).

E (E (X1|X2)) = E (X1)

Seconda proprietà: riguarda l'aspettativa condizionata con all'interno una funzione di tipo deterministico (non ci sono variabili casuali all'interno, tutto è perfettamente calcolabile perché non c'è aleatorietà, al contrario degli elementi stocastici che hanno variabili aleatorie).

E (X1h(X2)|X2) = h(X2)E (X1|X2)

Interpretiamo il modello di regressione in termini di aspettativa condizionata

Assumiamo che il mio errore sia indipendente dalla variabile esplicativa del modello (x). Se è indipendente è vero che l'aspettativa condizionata è pari a 0. Questo ci permette di interpretare il modello come aspettativa di yt dato il set informativo identificato da una sola variabile xt (asp cond di y dato x)

$E(y_t \mid x_t) = \beta_1 + \beta_2 x_t + E(u_t \mid x_t) = \beta_1 + \beta_2 x_t$

(la seconda parte assume valore 0)

E' un modello per descrivere l'aspettativa condizionata della variabile dipendente dato il valore della variabile indipendente

Il prezzo di un'abitazione che ha metri quadri 100 sarà come valore atteso b1+b2*100.

Nei modelli di regressione è fondamentale parlare dell'insieme di variabili che noi mettiamo sul condizionamento. Dopo la barra noi abbiamo solo un termine, ma potremmo condizionare anche rispetto al numero di bagni, al numero di piani dell'abitazione ecc.. Descriviamo un comportamento medio rispetto a delle variabili.

Le variabili sono di due tipi: ESOGENE e ENDOGENE. (esogene: trae il suo valore al di fuori del modello che stiamo considerando) Nel nostro modello le variabili sono ESOGENE (se no ci sfuggono, se la x dipende dalla y e la y dipende dalla x abbiamo un problema). Nel set informativo noi mettiamo variabili esogene. (il consumo dipende dal reddito, il reddito è considerato esogeno).

Il termine di errore →

1) il disturbo è ut in forma generica (u1,u2,un) → differenti realizzazioni della variabile casuale disturbo. Sono indipendenti? Se si, significa che i nostri errori sono distribuiti in modo indipendente. → IID (La prima I rappresenta l'indipendenza degli errori). La prob di avere un numero pos o neg è sempre indipendente rispetto le estrazioni precedenti. La seconda I sta per Identicaly : il disturbo uno è distribuito come il due? La distribuzione del primo disturbo è standard normal (0,1=?) , la seconda anche? Se si, allora sono distribuiti in modo identico. Se la distribuzione del primo è standard normal, la seconda è sempre gaussiana ma con media 0 e varianza 2, siccome cambia la varianza non sono distribuite identiche. La nostra ipotesi II è che sono indipendenti e distribuiti nello stesso modo.

Nelle serie storiche c'è correlazione seriale, non c'è indipendenza quindi la prima I va via. → SERIAL CORRELAZIONE

La seconda I per errori identicamente distribuiti può sparire quando la distribuzione degli errori cambia, ci possono essere dei periodi in cui la varianza può essere maggiore (il consumo di alcune famiglie con reddito basso ha meno varianza rispetto a quelle con reddito più alto). → ETEROSCHEDASCHICITA'

Qualche volta possiamo avere sia seria correlation sia eteroschedaschicità (non c'è IID)

B2 COME VARIA Y AL VARIARE DI UNA PICCOLA QUANTITA' DI X2. Non è l'elasticità!

B1 è il valore di y quando x2 la variabile esplicativa è = 0.

Creiamo il modello scaricando R studio (o hello word r)

Quando ci riferiamo al modello di regressione lineare si intende la relazione tra la variabile dipendente y e il coefficiente (Beta). Se questa relazione è lineare il modello di regressione è lineare. Non si guarda la relazione con la variabile esplicative x. Ci sono anche modelli non lineare (esponenziali) che tuttavia possono diventare lineari prendendo il logaritmo.

21 settembre

Supponiamo che teta, l'angolo che si forma dal primo vettore al secondo sia diverso da 0, quindi non sono parallele. Partendo sempre da 0 possiamo disegnare il secondo vettore x2. Quello che ci interessa è calcolare il risultato di una combinazione lineare dei nostri vettori (b1*x1+x2*b2). Il metodo grafico dice: dopo aver disegnato b1*x1 sull'asse x e il punto b2*x2 nell'asse y, disegniamo il vettore segmento in modo che sia parallelo all'asse y. Costruiamo cosi un parallelo gramma. Il vettore somma sarà la diagonale di questo parallelogramma.
Adesso abbiamo nella sostanza visto partendo da x1 e x2 come si fa a fare la somma, questi elementi sono sostanzialmente coefficienti del regressore 1 e del regressore 2. Possiamo quindi disegnare il nostro modello lineare dopo aver definito cosa si intende per indipendenza lineare. Geometricamente non lo vediamo, lo vediamo da un punto di vista di definizione. E' quando abbiamo dei vettori che sono tutti linearmente indipendenti. Non esiste quindi una combinazione lineare di questi vettori che sia in grado di generare uno dei vettori che noi abbiamo a disposizione. Le x che mettiamo come variabili esplicative le mettiamo in modo che ci diano delle info, perché se una è semplicemenTe una combinazione lineare dei vettori non dandoci informazioni è inutile. Se non c'è un'indipendenza lineare esiste una combinazione lineare dei vettori che ho ognuno moltiplicato per b la cui somma sia uguale a zero, cioè che produca un vettore fatto da tutti 0. Noi vogliamo però info piena, quindi inserire regressori che producono utilità nell'analisi.

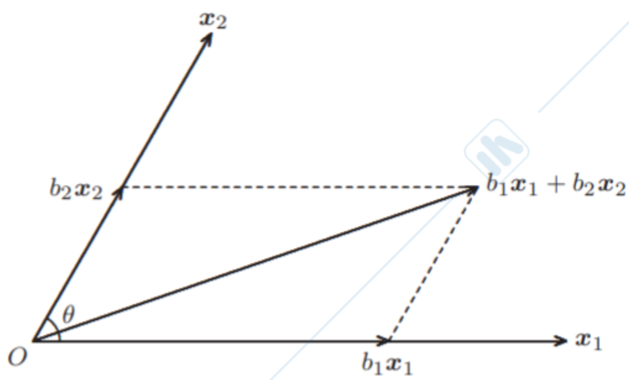

Figure 2.8 A 2-dimensional subspace

Il nostro modello è y=xb + u in forma matriciale. La x può essere partizionata in modo tale da avere i vettori che la costituiscono. La nostra X è fatta da k variabili indipenenti quindi vettori, il cui primo può essere un vettore yota (fatto da tutti 1) e questo ci permette di fare l'annotazione da 1 a k. Nel modello semplice abbiamo solo il primo vettore, ovvero la X corrisponde a un semplice vettore x1. Quando moltiplichiamo x * b nel nostro modello abbiamo il prodotto tra la matrice partizionata (x1,x2,xk) e un vettore colonna)(b1,b2,bk) che sono i nostri coefficienti. Il prodotto di questo in forma vettoriale per questo vettore colonna di beta ritorniamo al modello in forma algebrica (x1*b2+xk*bk). = Sommatoria da 1 a k di bi*xi

Questa è una combinazione lineare delle x. E' una combinazione lineare che fa parte del sottospazio delta x dove delta x è lo spazio vettoriale creato ai vettori x1 fino a xk. Quindi stiamo dicendo che i vettori x1 xk costituiscono la base di uno spazio in cui andremo a posizionare x*b.

Una delle condizioni del metodo dei momenti era che il regressore fosse ortogonale agli errori: a livello di dgp deve valere l'aspettativa del prodotto tra la x il regressore e il termine di errore pari a 0 che si traduce in ipotesi sul campione. La media di x per l'errore è pari a 0. Questo da un punto di vista matriciale è equivalente a dire che il prodotto scalare di x trasposto per il residuo sia uguale a zero. E' il punto cruciale legato alla perpendicolare. XT(y − Xβˆ) = 0.

Come lo vediamo graficamente

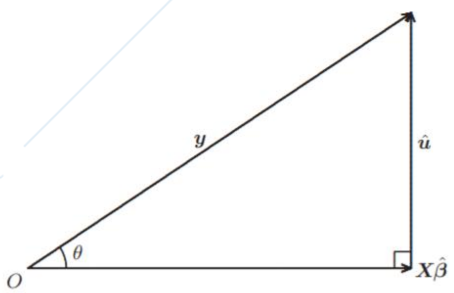

Figure 2.10 Residuals and fitted values

Abbiamo un vettore x * b cappello che rappresentiamo orizzontalmente, abbiamo poi sempre con partenza da 0 il vettore y che è il vettore contenente le osservazioni sulla variabile dipendente. L'angolo tra y e yb è teta. Quando noi andiamo a fare la regressione, questa è la proiezione di y su xb cappuccio. Di fatto noi andiamo a selezionare quel beta cappuccio. Solo avendo la linea perpendicolare tra y e xb siamo sicuri di avere un punto di minimo (a differenza del grafico di prima in cui non stavamo stimando i regressori e consideravamo i beta veri e non i beta normali, avendo il segmento obliquo non minimizzavo).

U cappello è il vettore dei residui. Geometricamente si traduce nel segmento perpendicolare rispetto a xb cappello.

Quando noi stimiamo il modello noi abbiamo di fatto una x (se abbiamo i regressori abbiamo più x). Abbiamo la y. Il nostro obiettivo quando stimiamo un modello significa andare a scegliere il beta cappello. Il metodo dei momenti e dei minimi quadrati ci dice che andiamo a scegliere quel beta cappello che garantisca la perpendicolarità tra y e xb perché è la distanza minima tra il modello che stiamo generando e le osservazioni sulla y. Proiettare y su xb o regredire y su xb è la stessa cosa: vuol dire tracciare una perpendicolare.

U cappello è anche chiamato residui della regressione. In un certo senso è la variabilità che il mio modello non è in grado di riprodurre.

Immaginiamo di avere due regressori x1 e x2. Quello che conta è la lunghezza, il punto di partenza è indifferente immaginiamo sia il punto O. Come faccio a disegnare la nostra regressione? Dal punto O facciamo partire il vettore y fino al punto B. Prima abbiamo definito il subspace, che in questo caso è rappresentato dal vettore x1 x2, ovvero la base del nostro spazio vettoriale. In mezzo posizioneremo la combinazione lineare tra x1 e x2. Noi prendiamo il B lo iettiamo sul piano costituito dal vettore x1 e x2. Lo regrediamo, lo prendiamo dal punto b e tracciamo una linea perperndicolare e determiniamo un punto A all'interno del piano. Che cos'è il punto A? e' x*b cappuccio. E' il punto in cui b cappuccio minimizza i residui perché c'è ortogonalità. Andiamo nella figura due: da cos è dato il punto A? è la combinazione di beta cappuccio x1 + beta cappuccio 2 * x2.