1 Stima di un modello panel

1.1 Tipi di modello

I **dati panel**, sono osservazioni raccolte da variabili che si spostano sia nel tempo che tra individui, sia Y l'insieme dei dati, l'osservazione $Y_{i,t}$ si riferisce all'i-esima persona al t-esimo tempo, dove $i \in 1...N$ e $t \in 1...T$.
Se $N < T$ si parla di **pooled time series**, se $N > T$ si parla di **dati longitudinali**. Per indicare questo modello avrò come notazione $Y_{i,t} = \alpha_i + \sum_{r=2}^{K} \beta_{r,i} * X_{r,i,t} + U_{i,t}$.

Posso aggregare i per ogni individuo il suo tempo ottenendo così $Y_i = \alpha_i * \overline{1} + X_i' * \beta_i + U_i$, dove non ho più come risultati un unico valore, ma un vettore Y_i di dimensione $Tx1$ e la matrice X_i di dimensione $T*(k-1)$ e β_i ha dimensione $(k-1)*1$, unico valore singolo è α_i perchè la costante che distingue gli individui.
Possiamo aggregare ancora ottenendo $Y = X' * \delta + U$, dove Y_i ha dimensione $TN * 1$, la matrice X di dimensione $KT * KN$ ed è una matrice "diagonale" dove sulla "diagonale" ha tutte le sottomatrici X_i, δ ha dimensione $NK * 1$.

1.2 Stimatore SURE

Se vogliamo stimare il parametro δ devo imporre delle ipotesi su di esso:

- Non sistematicità degli errori cioè $E(U) = 0$.
- La matrice X non è stocastica.
- A livello di singolo individuo gli errori sono omoschedastici e incorrelati nel tempo.
- Se vediamo invece tra gli individui notiamo che sia $E(UU') = \phi = \Sigma_n \otimes I_t$, dove $\otimes$ prodotto di kronecker cioè ogni elemento della matrice di sinistra viene moltiplicata per l'intera matrice di destra.
  Otteniamo così che $\phi = \begin{bmatrix} \sigma_{1,1}^2 * I_t & \sigma_{1,2}^2 * I_t & ... & \sigma_{1,n}^2 * I_t \\ \sigma_{2,1}^2 * I_t & \sigma_{2,2}^2 * I_t & ... & \sigma_{2,n}^2 * I_t \\ ... \\ \sigma_{n,1}^2 * I_t & \sigma_{n,2}^2 * I_t & ... & \sigma_{n,n}^2 * I_t \end{bmatrix}$, dove $\sigma_{i,i}^2$ varianza dell'individuo i, $\sigma_{i,k}^2$ covarianza dell'individuo i con l'individuo k allo stesso tempo t. Notiamo che nella diagonale ci sono matrici diagonali essendo varianza di individui con incorrelazione temporale, ma tutte diverse tra loro visto che dipendono da $\sigma_{i,i}^2$ che varia, quindi noto che ho **eteroschedasticità** degli errori. Notiamo che gli individui sono correlati contemporaneamente ma incorrelati nei tempi diversi.
  Se tale varianze non variassero avremmo che c'è omoschedasticità, in questo caso possiamo usare OLS ma è una situazione assurda.

Date queste ipotesi avendo correlazione anche solo se contemporanea e errori eteroschedastici stimeremo δ con un particolare GLS cioè $\delta = (X'\phi^{-1}X)^{-1}X' * \phi^{-1}Y$, tale stime è detta SURE (semingly unrelated regression estimator). Però la matrice ϕ non è nota, allora si stima.

Sia preso un individuo i $Y_i = X_i' * \delta_i + U_i$, dove $E(UU') = \sigma_i^2 * I_t$ ho che $\sigma_i^2 = \frac{\hat{U}'\hat{U}}{T-K}$, in questo modo ottengo tutte le matrici sulla diagonale di ϕ, se poi voglio $\sigma_{i,j} = \frac{\hat{U_i}'\hat{U_j}}{T-K}$ e così ottengo anche tutte le altre matrici che compongono ϕ.
Ottengo i residui $\hat{U_i} = Y_i - \hat{Y_i} = Y_i - x_i * \delta_{i,ols}$, quindi per stimare le componenti di ϕ uso comunque la stima OLS.
In questo modo posso rendere operativo lo stimatore SURE come $\delta = (X'\hat{\phi}^{-1}X)^{-1}X' * \hat{\phi}^{-1}Y$.

Lo stimatore δ **è uguale in OLS e in SURE solo se i regressori sono uguali**, prendiamo in esempio un Var(1) bivariato ho: $\begin{bmatrix} Y_{1,t} \\ Y_{2,t} \end{bmatrix} = \begin{bmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{bmatrix} * \begin{bmatrix} X_{1,t-1} \\ X_{2,t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{bmatrix}$ in questo caso SURE= OLS se $X_{1,t-1} = X_{2,t-1}$. Tutta via anche se i regressori sono uguali ma impongo per esempio che $a_{1,2} = a_{2,1}$ allora tali modelli non coincidono più e dobbiamo stimare sempre attraverso il SURE.

Tale affermazione è generalizzabile per n elementi ed è dimostrabile. Prendiamo un modello $Y = X' * \delta + U$, dobbiamo dimostrare che $X_i = ... = X_n = \tilde{X}$ cioè è scrivibile come matrice:
$\begin{bmatrix} Y_{1,t} \\ ... \\ Y_{n,t} \end{bmatrix} = \begin{bmatrix} \tilde{X} & 0 & ... & 0 \\ 0 & ... & \tilde{X} & ... \\ 0 & ... & 0 & \tilde{X} \end{bmatrix} * \begin{bmatrix} \delta_1 \\ ... \\ \delta_n \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ ... \\ \epsilon_n \end{bmatrix}$. Noto quindi che $X = I_n \otimes \tilde{X}$.
Dunque $\delta = (X'\hat{\phi}^{-1}X)^{-1}X'\hat{\phi}^{-1}Y = ((I_n \otimes \tilde{X})' * (\Sigma_n \otimes I_t)^{-1}(I_n \otimes \tilde{X}))^{-1} * (I_n \otimes \tilde{X})' * (\Sigma_n \otimes I_t)^{-1} * Y$. Vedo che $((I_n \otimes \tilde{X})' * (\Sigma_n \otimes I_t)^{-1}(I_n \otimes \tilde{X}))^{-1} = (I_n \Sigma_n^{-1} I_n) \otimes (\tilde{X}' I_t \tilde{X}) = \Sigma_n^{-1} \otimes \tilde{X}'\tilde{X}$, sostituisco e vedo che $\delta = I_n \otimes (\tilde{X}'\tilde{X})^{-1}\tilde{X}'Y$, cioè è un vettore che si ottiene da una matrice matrice diagonale (dim NK NT) dove le matrici sulla diagonale sono $(\tilde{X}'\tilde{X})^{-1}\tilde{X}'$. Quindi abbiamo dimostrato che SURE è uguale OLS anche in presenza di correlazioni istantanee tra gli individui, ovviamente i regressori devono essere liberi.

1.3 Stima per dati longitudinali

Ora se immaginiamo che $N > t$, quindi nel caso di dati lungitudinali, la stima di δ diventa più complessa.
Se ipotiziamo che per stimare δ usiamo OLS ho $\delta_{ols} = (X'X)^{-1}X'Y$, dobbiamo stimare usando un inversione di matrice di dimensione $Nt * Nk$ ciò porta a errori di calcolo e dobbiamo ridurre il numero di parametri. Per farlo ipotiziamo che:

- Gli slopes cioè i B_i siano costanti per ogni i, quindi $B_i = \begin{bmatrix} B_1 \\ ... \\ B_k \end{bmatrix}$ e le osservazioni Y_i vengono differenziate solo dai α_i (simili modello cumulato). In questa formulazione dobbiamo stimare (N+k-1) parametri.
  Il modello per esteso è quindi $Y_{i,t} = \alpha_i + \sum_{r=2}^{K} \beta_r * X_{r,i,t} + U_{i,t}$ se raggruppiamo porta a $Y_i = \alpha_i * J + X_i^* * B + U_i = X_i * \delta_i + U_i$ dove $\delta_i = \begin{bmatrix} \alpha_i \\ B_1 \\ ... \\ B_k \end{bmatrix}$, se unisco tutto e lo vedo in maniera matriciale vedo che $\begin{bmatrix} Y_1 \\ ... \\ Y_n \end{bmatrix} = \begin{bmatrix} J & 0 & ... & X_1^* \\ 0 & J & ... & X_i^* \\ 0 & ... & J & X_n^* \end{bmatrix} * \begin{bmatrix} \alpha_1 \\ ... \\ \alpha_n \\ B_1 \\ ... \\ B_k \end{bmatrix} + \begin{bmatrix} U_1 \\ ... \\ U_n \end{bmatrix}$. Dove la prima parte della matrice X ha dimensione NT*N e l'altra NT*(K-1), vedo quindi che DIM(x)=NT*(N+K-1).
  La matrice X può essere scritta anche come $(I_n \otimes J, X^*)$, che può essere tradotta come una variabile dummy dove la colonna J-esima diventa formata anche da 1 se l'individuo che trattiamo è J.
- Se $N > t$ come da assunzione c'è eteregeneità tra gli individui che non è osservata quindi va modellata e va fatto attraverso α_i e ipotizziamo che tale **eterogeneità sia correlata alle X**.
- X non stocastica.
- $E(UU') = \sigma_u^2 * I_{nt}$, quindi errori incorrelati ed omoschedastici.

Date queste assunzioni usiamo OLS per stimarlo, cioè $\hat{\delta}_{ols} = (X'X)^{-1}X' * Y$, ciò deriva dalle condizioni del primo ordine quindi derivando ho che $X'X * \delta = X'Y$, detto sistema normale. Vedo che $DIM(X'X) = (N + K - 1) * (N + K - 1)$, quindi risolvo parzialmente il problema, ora devo agire solo se N è molto grande, ed uso la formula dell'inversa partizionata.
In forma matriciale il sistema normale ho $\begin{bmatrix} (I_n \otimes J)'(I_n \otimes J) & (I_n \otimes J)'X^* \\ X^{*'}(I_n \otimes J) & X^{*'}X^* \end{bmatrix} * \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} (I_n \otimes J)' \\ X^{*'} \end{bmatrix} * Y$.
Applico l'inversa partizionata ho $\hat{B} = (X^{*'}DX^*)^{-1}X^{*'}DY$ dove DIM(D)=NT*NT.

Ora per stimare l'altro so che $\hat{\alpha} \otimes J = CY - CX^* \hat{B}$ dove $\hat{\alpha} \otimes J = \begin{bmatrix} \hat{\alpha}_1 \\ ... \\ \hat{\alpha}_1 \\ ... \\ \hat{\alpha}_n \end{bmatrix}$, dove DIM(C)=NT*NT.

Le matrici C e D sono viste anche come $C = I_n \otimes C_t$ e $D = I_n \otimes C_t$ (between).
So anche che $C_t = JJ'/T = \frac{1}{T} \begin{bmatrix} 1 & ... & 1 \\ 1 & ... & 1 \end{bmatrix}$, quindi se Z vettore ho $C_t * Z$ ottengo un vettore uguale con tutti i valori uguali alla media del valori di Z, e $D_t = I_t - C_t$ (within) quindi $D_t * Z$ calcola il vettore dagli scarti dalla media.
Le matrici C_t e D_t hanno diversi proprietà:

- Sono simmetriche cioè $C_t = C_t'$ e $D_t = D_t'$ visto che $C_t(\frac{1}{T}JJ')' = \frac{1}{T}JJ'$ e di conseguenza anche D_t lo è.
- Idem potenti cioè $C_t * C_t = C_t$ e anche D_t questo perchè $\frac{1}{T}JJ' * \frac{1}{T}JJ' = \frac{1}{T^2}JJ'JJ'$ e $J'J = T$ quindi dimostrato
- Ortogonali tra loro quindi $D_t * C_t = 0$ e $(I_t - C_t) * C_t = 0$.

Ora ricordiamoci che $\hat{B} = (X^{*'} D X^*) X^{*'} * DY$ dalle proprietà di prima è uguale a $(X^{*'} D' D X^*) X^{*'} * D'DY$ chiamo $\tilde{X}^* = DX^*$ e $\tilde{Y} = DY$ ricordo che D vettore degli scarti dalla media, cioè $\hat{B}$ è un OLS con vettori gli scarti dalla media ed è detto stimatore within.

Ora ritorniamo a $\hat{\alpha} \otimes J = CY - CX^* \hat{B} = CY - CX^* \hat{B}$ cioè faccio una differenza tra medie temporali tale trasformazione è detta between.
Tale stima non è diversa da quella fatta in maniera classica cioè $\overline{X} = CX^*$. Se prendo il modello classico con scarti dalla media ho $Y_t - \overline{Y} = \alpha + B * X_t + U_t - \overline{Y}$ so che $\overline{Y} = B * \overline{X} + \alpha$ ottengo quindi $Y_t - \overline{Y} = B * (X_t - \overline{X}) + U_t$, cioè uguale al modello classico. Notiamo quindi che il calcolo di α è contemporanea a quello di B ma fatto in maniera indiretta.

1.3.1 Test sugli effeti fissi

$H_0 = \alpha_1 = ... = \alpha_n = \alpha$ vs H_1 gli effetti fissi sono liberi. Tale test ha un ipotesi nulla congiunta lineare con N-1 restrizioni, per verificarlo uso il Test F. In generale vale che $S_f = \frac{RSS_r - RSS_u / N_{restrizioni}}{RSS_u / (N_{oss} - N_{par})}$ e sotto H_0 si dispone come $F_{N_{res}, N_{oss} - N_{par}}$.
Per costruire il test abbiamo quindi $M_r = Y_{i,t} = \alpha + \sum_{r=2}^{K} \beta_{r,i} * X_{r,i,t} + U_{i,t}$ e $M_l = Y_{i,t} = \alpha_i + \sum_{r=2}^{K} \beta_{r,i} * X_{r,i,t} + U_{i,t}$ da qua calcolo i residui ed ottengo che la statistica test è $S_f = \frac{RSS_r - RSS_u/(N-1)}{RSS_u/(N*T-N-K+1)}$.

1.4 Esempio pooled Time Series

Se per esempio prendiamo un dataset Pooled Time-Series, quindi usiamo lo stimatore SURE per i dati. Vogliamo verificare che l'andamento dei nostri dati sia del tipo parabolico concavo, visto che secondo la teoria di Kuzntes le emissioni di CO_2 aumentano all'aumentare del reddito fino ad un livello X e poi superato tale livello di reddito le emmissioni diminuiscono.
Il dataset è formato da tre paesi con rilevazioni in 144 mesi.
Per il calcolo usiamo i logaritmi dei dati così da vedere il cambiamento percentuale rispetto a prima, e costruiamo il modello del tipo $Y_{i,t} = B_{1,i} + B_{2,i} * X_{i,t} + B_{3,i} * X_{i,t}^2 + U_{i,t}$ verificando se $B_2 > 0$ e $B_3 < 0$.
So che lo stimatore per δ è consistente ma distorto quindi posso applicare la teoria assintotica, cioè posso valutare se ho zero slopes attraverso un test chi quadro.
Queste informazioni ci vengono date attraverso lo stimatore "reg y1 x11 x12" oppure come "sureg y1 x11 x12".
Posso verificare anche se i residui sono correlati immediatamente attraverso il test di Bresh-Pagan cioè ho H_0 tutti indipendente VS H_1 almeno uno correlato, se fosse così comunque non invaliderebbe il modello visto che con lo stimatore SURE trattiamo questi problemi.

Operativamente posso calcolare anche l'iterative SURE, cioè la versione algoritmica. So che nello

stimare $\delta = (X'\hat{\phi}^{-1}X)^{-1}X' * \hat{\phi}^{-1}Y$ uso $\hat{\phi}$ che è stimata con OLS, ora che ho una prima stima di $\hat{\delta}$ posso ristimare $\hat{\phi}$, avendo così un nuovo valore da usare per stimare $\hat{\delta}$, questa operazione viene ripetuta fin che $|\hat{\delta}^n - \hat{\delta}^o| < tol$.
Con i comandi di stata con Predict Y1hat, eq(Y1) predico i valori quindi calcolo i valori stimati con la tecnica $Y = X\hat{B}$. Dai risultati riportati di stata vediamo si che è una funzione concava ma che non ha una forma parabolica o che comunque il massimo non viene raggiunto così facilmente.

Potremmo analizzare anche il caso in cui abbiamo molti paesi e pochi dati temporali quindi in questo caso $N > T$ cioè dei dati panel. Creeremo il modello $Y_{i,t} = \alpha_i + B_2 * X_{i,t} + B_3 * X_{i,t}^2 + U_{i,t}$ ho qundi B_2 e B_3 comuni questo implica che il massimo avrà sempre la stessa ascissa e ordinata diversa.

1.5 Esempio panel longitudinali

Se voglio analizzare un insieme di aziende e vedere come la produzione è legata con i costi, ma ho pochi dati storici. Per prima cosa il dataset su stata è rappresentato in maniera diversa infatti ho i dati raccolti per tipo di impresa.
Inoltre avrò un modello $COST_{i,t} = \alpha_i + B_2 * OUTPUT_{i,t} + U_{i,t}$ che verrà confrontato con un modello pooling semplice $COST_{i,t} = \alpha + B_2 * OUTPUT_{i,t} + U_{i,t}$, per ricordarmi il suo residuo dopo la regressione uso il comando Scalar RSSR=e(RSS).

In forma matriciale vado a stimare $\begin{bmatrix} COST_1 \\ ... \\ COST_4 \end{bmatrix} = \begin{bmatrix} J & 0 & ... & OUTPUT_1^* \\ 0 & J & ... & OUTPUT_i^* \\ 0 & ... & J & OUTPUT_4^* \end{bmatrix} * \begin{bmatrix} \alpha_1 \\ ... \\ \alpha_4 \\ B_1 \end{bmatrix} + \begin{bmatrix} U_1 \\ ... \\ U_4 \end{bmatrix}$.

Il problema sta che non ho delle dummy e devo generarle uso quindi "gen D1 = 0" (STATA ha salvato la lunghezza del dataset è 40) e poi "replace D1=1 if Firm==1" e le unisco con "list D1-D4". Così facendo posso calcolare δ le opzioni sono 2:

- Se N non è tanto grande posso calcolarlo attraverso reg ma mettendo l'opzione ", noconst" visto che se aggiungo la costante si crea collinearità.
- Se N è grande uso l'inversione partizionata quindi avendo $\hat{\alpha} \otimes J = CY - CX^*\hat{B} = CY - CX^*\hat{B}$ e $\hat{B} = (X^{*'}DX^*)X^{*'} * DY$ calcolo DX^* e DY sapendo che sono gli scarti dalla medie e faccio una regressione su quelli, successivamente calcolo le medie per i gruppi singoli CY e CX cosi da calcolare una matrice che ripete T volte i nostri alpha.

Per il test empirico sugli effetti fissi posso ricostruirmelo a mano ottenendo la statistica test e poi usare il comando "Ftail(3,35,Ftest)" oppure uso il comando "Test D1=...=D4" calcola test F.

Esistono comandi automatici che ci permettono di calcolare automaticamente le stime beetween e within. In particolare ho che con il comando XTSET firm time, indico che un dataset panel longitudinale, tale comando ci dice che è "strongly balance" cioè non ci sono missing values.
Successivamente con il comando "xtreg cost output, fe" faccio una regressione sui data trasformati come prima. Con il comando "fe" fixed effect ci ci trova il modello con una sola costante che deriva da $\mu_i + \alpha_. = \alpha_i$ e la costante del modello di regressione è $\alpha_.$.
Per ottenere tale parametro la trasformazione che effettua Stata diversa dalla nostra, infatti noi per trovare i valori within usiamo $Z_{i,t}^w = Z_{i,t} - Z_{i,.}$ (con il punto indico la media), il softwear fa $Z_{i,t}^w = Z_{i,t} - Z_{i,.} + Z_{.,.}$, se applico questa trasformazione al modello $Y_{i,t} = \alpha_i + B * x_{i,t} + U_{i,t}$ ottengo che $Y_{i,t} - Y_{i,.} + Y_{.,.} = B(x_{i,t} - x_{i,.} + x_{.,.}) + \alpha_. + (U_{i,t} - U_{i,.} + U_{.,.})$ e da qua posso stimare tramite OLS α e B.

Un altra diversità si ha guardando gli errori $U_{i,t}$ che in stata sono chiamati $e_{i,t}$ quindi la loro varianza è $sigma_e^2$ (si salva come scalar $SU = e(sigma_e)$), inoltre stat indica i valori μ_i come U e da qua ne deriva anche il nome della sua varianza. Con queste informazioni possiamo vedere quanto incide la variabilità degli effetti fissi sulla variabilità totale ed ho $\rho = \frac{\hat{\sigma}^2(\mu_i)}{\hat{\sigma}^2(\mu_i) + \hat{\sigma}^2(U)}$.
Altra informazione in outpur è $Cor(\mu_i, \hat{Y})$, o se B è uno scalare $Cor(\mu_i, \hat{B})$, ed è giusto che sia diversa da 0 da ipotesi. Ora calcolo μ_i con "xtpred mihat, u" con questa calcola i residui che ci permettono di ottenere gli α_i. Con questi valori posso costruire il test F di prima però ponendo $H_0 = \mu_1 = ... = \mu_4 = 0$ ed è uguale che porre $H_0 = \alpha_1 = ... = \alpha_4 = \alpha$ ha quindi lo stesso test.

2 Modello panel ad effetti random o error component

Dato il modello $Y_{i,t} = \alpha_i + \sum_{r=2}^{K} \beta_r * X_{r,i,t} + U_{i,t}$, se scorporo α_i ottengo $Y_{i,t} = \alpha + \sum_{r=2}^{K} \beta_r * X_{r,i,t} + U_{i,t} + \mu_i$, se impongo che μ_i sia una variabile aleatorie dove $E(\mu_i) = 0$, $Var(\mu_i) = \sigma^2$ quindi omoschedastica e $Cov(\mu_j, \mu_i) = 0$ se $j \neq i$ allora l'espressione può essere raccolta e formare $Y_{i,t} = \alpha + \sum_{r=2}^{K} \beta_r * X_{r,i,t} + V_{i,t}$ dove $V_{i,t} = U_{i,t} + \mu_i$.
Questo modello per funzionare ha bisogno di un ipotesi fondamentale cioè che $Cor(B, V_{i,t}) = 0$ perchè in caso contrario avremmo che lo stimatore trovato è inconsistente. Se aggreghiamo i dati trovo che $Y_{i,t} = \alpha * j + X_i^* * B + \mu_i * j + U_i = X_i * \delta + V_i$.

Dobbiamo trovare lo stimatore per δ che dipenderà dallo studio della matrice di varianza covarianza dell'errore V_i, trovo $E(V_i V_i') = E((\mu_i * J + U_i)(\mu_i * J + U_i)') = E(\mu_i^2 * J * J' + \mu_i * U_i * J' + \mu_i * J * U_i + U_i * U_i') = \sigma_\mu^2 * JJ' + \sigma_U^2 * I_t$ visto che $E(\mu_i U_i) = 0$ essendo indipendenti, ho quindi $E(V_i V_i') = \begin{bmatrix} \sigma_\mu^2 + \sigma_U^2 & \sigma_\mu^2 & ... & \sigma_\mu^2 \\ \sigma_\mu^2 & \sigma_\mu^2 + \sigma_U^2 & ... & \sigma_\mu^2 \\ \sigma_\mu^2 & \sigma_\mu^2 & ... & \sigma_\mu^2 + \sigma_U^2 \end{bmatrix} = Vx$ noto quindi che c'è omoschedasticità essendoci stessa varianza nella diagonale principale, gli errori tra loro sono autocorrelati infatti non è una matrice diagonale ma tale autocorrelazione è costante.
Se ci focalizziamo sull'autocorrelazione infatti essendo costante per tutti gli elementi è indipendente dal tempo, quindi diversa dalle solite serie storiche AR o MA.
Se tratto come AR, ho $V_{i,t} = \phi_i * V_{i,t-1} + u_{i,t}$ avremmo che è omoschedastico avendo sulla matrice principale tutti 1, ma l'autocorrelazione decresce esponenzialmente in base all'aumentare del ritardo, per far si che sia stazionario ϕ_i. Idem se trattato da MA, $V_{i,t} = \theta_i * u_{i,t-1} + u_{i,t}$ ho omoschedasticità ma la matrice ha valori solo nei ritardi 1 e al di fuori vale sempre 0.

Ora sapendo come è fatta la la matrice $E(V_i V_i')$ so che stimerò δ con FGLS, avrò quindi $\hat{\delta} = (X_i' * V^{-1} * X_i)^{-1} X_i' * V^{-1} * Y_i$, dobbiamo trovare la trasformazione V^{-1} che ci permette di trattare δ come OLS.
So che $V = \sigma_\mu^2 * JJ' + \sigma_u^2 * I_t = T\sigma_\mu^2 * \frac{JJ'}{T} + \sigma_u^2 * I_t - \sigma_u^2 * C_t + \sigma_u^2 * C_t = \sigma_u^2 * D_t + (T\sigma_\mu^2 + \sigma_u^2) C_t$ come ultima riscrittura ho $\sigma_u^2 * D_t + \sigma_1^2 * C_t$ quindi come inversa ho $\frac{C_t}{\sigma_1^2} + \frac{D_t}{\sigma_u^2}$.
Identificato cos'è V dobbiamo trovare una trasformazione V^* idem potente da poter applicare, sapendo che C_t e D_t sono idem potenti e ortogonali allora tale trasformazione è $V^* = \frac{C_t}{\sigma_1} + \frac{D_t}{\sigma_u}$ come prima tale equazione può essere vista come somma pesata di trasformazione between con tarsformazione within.
La matrice V^* può essere anche vista come $\frac{C_t}{\sigma_1} + \frac{I_t}{\sigma_u} - \frac{C_t}{\sigma_u} = \frac{I_t}{\sigma_u} + (\frac{1}{\sigma_u} - \frac{1}{\sigma_1})C_t$, se raccolgo $\frac{1}{\sigma_u}$ ottengo $\frac{1}{\sigma_u}(I_t - \theta * C_t)$ dove $\theta = 1 - \frac{\sigma_u}{\sigma_1}$. Ecco quindi che se σ_μ tende a 0 allora avrò che $\sigma_1 = \sigma_u$ quindi $\theta = 0$ cioè non ho trafsormazione quindi uso un modello di Pooling semplice, vice versa se σ_u tende a 0 ho che $\theta = 1$ cioè un trasformazione within ottenendo un simil modello ad effetti fissi.

Posso scrivere quindi che $V^* * Y = \tilde{Y}$ e $V^* * X = \tilde{X}$ ottengo quindi lo stimatore OLS trasformato per $\delta = (\tilde{X}'\tilde{X})^{-1} * \tilde{X}'\tilde{Y}$. Inoltre se notiamo il parametro $\frac{1}{\sigma_u}$ è di scala perchè viene eliso nelle trasformazioni.

2.0.1 Generalizzazione

Se accorpiamo il modello iniziale otteniamo che $\begin{bmatrix} Y_1 \\ ... \\ Y_n \end{bmatrix} = \begin{bmatrix} X_1 \\ ... \\ X_n \end{bmatrix} * \delta + \begin{bmatrix} V_1 \\ ... \\ V_n \end{bmatrix}$ in questo caso ho che $E(V_e' V_e) = I_n \otimes V = \phi$ quindi lo stimatore è $\delta = (X' * \phi^{-1} * X)^{-1} X' * \phi^{-1} * Y$.

Dobbiamo quindi stimare ϕ ma per farlo devo stimare V e per farlo devo stimare le due varianze che lo compongono σ_1 e σ_u.
σ_u è ricavabile da un modello pannel ad effetti fissi infatti se prendo la RSS cioè la somma degli errori $U_{i,t} = Y_{i,t} - \hat{\alpha}_i - \sum_{r=2}^{K} \hat{B}_r * X_{r,i,t}$ per ottenere $\hat{\sigma}_u^2 = \frac{RSS}{NT-(N+K-1)}$.
Per calcolare σ_1 stimo il modello cross section con trasformazione between quindi $Y_{i,.} = \alpha + \sum_{r=2}^{K} \hat{B}_r * X_{r,i,.} + \mu_i + U_{i,.}$, ora voglio calcolare $Var(\mu_i + U_{i,.}) = Var(\mu_i) + Var(U_{i,.}) = \sigma_\mu^2 + \sigma_u^2/T = \sigma_1^2/T$, quindi ora trovo RSS del modello sopra descritto cioè la somma dei residui, ho quindi $\sigma_1^2 = \frac{RSS}{N-K} * T$

2.1 Esempio effetti random stata

Ritornando al dataset costi ed output, devo stimare il random effect e per farlo uso il comando "xtreg cost output, re theta", l'output proposto è quello classico, con in aggiunta il test di wald che è la versione generica del test F zero slopes usato perchè usiamo un GLS e non un OLS.
Altra cosa che si può notare è il valore della correlazione che è uguale a 0 in questo caso, e lo diamo per assunzione fatta dal modello, infatti ho $COR(\mu_i, X) = 0$ questo perchè trattiamo μ_i come errori e se fossero correlati lo stimatore diventerebbe inconsistente.
Infine ci viene detto che la stima per $\theta = 0.85$ cioè vince la componente within nella somma pesata, questo ci dice che probabilmente gli errori non sono random ma fissi, questo si può anche notare dalle stime dei coefficienti molto simili nei due modelli effetti fissi e random effect. Si può notare anche perchè $\sigma_\mu^2 >> \sigma_u^2$. Un'altra osservazione che conferma che in questo caso il modello FE ed RE danno risultati simili, è che nel modello FE la $cor(X_i, \mu_i) = 0.02$ quindi molto simile a 0 cioè quello che vogliamo nel modello RE.
Infine noto che $\sigma_u^2 = sigma_e$ nei due modelli è uguale ed è ovvio visto che gli errori vengono trattati allo stesso modo.
Se uso il comando "xtreg cost output, be"(regressione cross-section) ottenendo σ_1 che è $SD(U_i + avg(e_i)) * \sqrt{T}$ che posso usare per calcolare manualmente θ.
Se voglio ottenere gli stessi risultati della regressione di prima posso usare anche la trasformazione V^* su Y e X, tale trasformazione include anche la costante devo quindi regredire anche su 1-θ.
Se voglio manualmente ottenermi σ_1, non facciamo inferenza solo su 4 dati ma su i 40 dati solamente ripetuti 10 alla volta, in questo modo possiamo anche non moltiplicare per T per ottenere σ_1.

2.2 Test di Hausman per la correlazione

Per controllare in maniera rigorosa quale dei due modelle descrive meglio il fenomeno usiamo il Test di Hausman. Tale test è genericamente applicato avendo un modello $Y = Z * B + U$. Se ho $Cor(Z, U)$ il modo più efficiente per stimare B è OLS, se invece la correlazione è diversa tale stimatore è inconsistente e allora dovrò usare delle variabili strumentali X tale che $Cor(X, U)$ e stimo $B_{iv} = (x'z)^{-1}x'y$ tale stimatore è consistente, inoltre è sempre giusto da usare ma nel caso uno è meno efficiente.
Questo test ha come H_0: Cor(z,u)=0 (preferisco B_{ols}) vs H_0: $Cor(z, u) \neq 0$ (preferisco B_{iv}), ho come statistica test $T = (B_{iv} - B_{ols})'\phi^{-1}(B_{iv} - B_{ols}))$ dove $\phi = Cov(B_{iv}) - Cov(B_{ols})$, e sotto H_0 si dispone come una χ^2_{k-1}, se invece trattono una cosa unidimensionale avrò che $T = \frac{(B_{iv}-B_{ols})^2}{Var(B_{iv})-Var(B_{ols})}$

In quest'ottica è nato il test di Hausman per dati panel. Infatti sia FE: $Y_{i,t} = \alpha_i + \sum_{r=2}^{k} \hat{B}_r * X_{r,i,t} + U_{i,t}$ e modello RE: $Y_{i,t} = \alpha + \sum_{r=2}^{k} \hat{B}_r * X_{r,i,t} + \mu_i + U_{i,t}$, testo quindi che $H_0 : Cor(\mu_i, X) = 0$ sia B_{fe} che B_{re} sono consistenti ma il più efficiente è B_{re} vs H_1 solo che B_{fe} è consistente.
La statistica test unidimensionale è $= \frac{(B_{fe}-B_{re})^2}{Var(B_{fe})-Var(B_{re})}$ e se multidimensionale $T = (B_{fe}-B_{re})'\phi^{-1}(B_{fe}-B_{re}))$ dove $\phi = Cov(B_{fe}) - Cov(B_{re})$.
Su Stata, uso "xtreg cost output, fe" "est sto ef" "xtreg cost output, re" "hausman ef".

2.3 Stima degli effetti individuali

Per stimarli ricordiamo sempre un approccio generale che usiamo nel caso di $Y = X * B + U$ dove $1 \dots T$ sono le osservazioni del sample e $T + 1 \dots T + G$ osservazioni che vogliamo prevedere, se U è errore classico (omoschedasticità) posso stimare B con OLS, posso prevedere in maniera ottima come $\hat{Y} = X * \hat{B}$ e so che $E(U * U_f') = 0$.
Se però l'errore non fosse classico allora avrei dei problemi nella stima che risolverei usando un B_{gls} e dato $\Omega = E(U * U')$ e $Q = E(U * U_f')$ allora posso prevedere come $\hat{Y} = X * \hat{B_{gls}} + Q'\Omega^{-1} * \hat{U}_{gls}$.

Nel contesto dei dati panal ho $Y_i = X_i * \delta + v_i$ $1 \dots T$ sono le osservazioni del sample e voglio prevedere l'osservazione $T + 1$, so che v_i non è un errore classico quindi uso GLS. Quindi per prevedere calcolo $\Omega = V$ poi $Q = E(v * v_f') = \sigma_u^2 * J'$, la previsione sarà quindi $Y_{i,t+1} = X_{t+1} * \sigma_{gls} + \sigma_\mu^2 * J' * V^{-1} * \hat{v}_{i,gls}$ che ricordandoci quanto vale V si espande a $X_{t+1} * \sigma_{gls} + \sigma_\mu^2 * J' * (\frac{C_t}{\sigma_1^2} + \frac{D_t}{\sigma_u^2}) * \hat{v}_{i,gls}$.

Se ci concentriamo sulla prima equazione ho $\frac{\sigma_\mu^2}{\sigma_1^2} * J' * C_t * \hat{v}_i$ sapendo che $C_t = J * J'/T$ semplifico ed ottengo $\frac{\sigma_\mu^2}{\sigma_1^2} * J' * \hat{v}_i$ ed è la stima di μ_i, noto se $\theta = 1$ sono simili alla stime con effetti fissi.

2.4 Test Breush-Pagan

E' un test per vedere se gli effetti individuali esistono, cioè valuta $H_0 = \sigma_\mu^2 = 0$ quindi $\sigma_1 \to \sigma_u$ quindi $\theta = 0$ cioè che in realtà il modello è un pooling semplice. La statistica test BP si dispone sotto H_0 come un χ_1^2, si calcola su stata come "quietly xtreg cost output, re" "xtest0" "xtpred migls, u".

3 Modello two-way

Serve per descrivere dati con un profondità temporale e spaziale, che hanno un eterogeneità anche temporale che vogliamo studiare, avrò quindi che la mia serie sarà influenzata da due tipi di dummy D_i che spiega l'etereogenità individuale, D_t che spiega l'eterogeneità temporale.
Tale etereogeneità può essere catturata sia da un modello ad effetti fissi che da un modello ad effetti random, come nella one way.

3.1 Effetti fissi

Nel modello one-way stata calcola una costante comune la α di $\alpha_i = \alpha + \mu_i$, infatti calcola $Y_{i,t} - Y_{i,.} + Y_{.,.} = \alpha_i + \sum_{r=2}^k B_r * X_{r,i,t} + U_{i,t} - Y_{i,.} + Y_{.,.}$ che semplificando otteniamo $= \alpha + \sum_{r=2}^k B_r * (X_{r,i,t} - X_{r,i,.} + X_{r,.,.}) + U_{i,t} - U_{i,.} + U_{.,.}$, nel caso di due dimensioni abbiamo che il modello trasformato con within è $Y_{i,t} - Y_{.,t} - Y_{i,.} + Y_{.,.} = \alpha_i + \lambda_t + \sum_{r=2}^k B_r * X_{r,i,t} + U_{i,t} - Y_{i,.} - Y_{.,t} + Y_{.,.}$, so che

- $Y_{i,.} = \alpha_i + \lambda_. + \sum_{r=2}^k B_r * X_{r,i,.} + U_{i,.}$
- $Y_{.,t} = \alpha_. + \lambda_t + \sum_{r=2}^k B_r * X_{r,.,t} + U_{.,t}$
- $Y_{.,.} = \alpha_. + \lambda_. + \sum_{r=2}^k B_r * X_{r,.,.} + U_{.,.}$

ed unendoli quindi si semplifica ed ottengo $Y_{i,t} - Y_{.,t} - Y_{i,.} + Y_{.,.} = \sum_{r=2}^k B_r * (X_{r,i,t} - X_{r,i,.} - X_{r,.,t} + X_{r,.,.}) + U_{i,t} - U_{i,.} - U_{.,t} + U_{.,.}$.
Per calcolare quindi gli effetti individuali calcolo OLS del modello trasformato poi ottengo $\alpha_i = (Y_{i,.} - Y_{.,.}) - \sum_{r=2}^k B_r * (X_{r,i,.} - X_{r,.,.})$ e $\lambda_t = (Y_{.,t} - Y_{.,.}) - \sum_{r=2}^k B_r * (X_{r,.,t} - X_{r,.,.})$

3.2 Effetti random

Se immaginiamo di decomporre $\tilde{\alpha}_i = \alpha + \alpha_i$ e $\tilde{\lambda}_i = \lambda + \lambda_i$ creiamo il solito modello ad effetti random ho $Y_{i,t} = \alpha + \lambda + \sum_{r=2}^k B_r * X_{r,i,t} + U_{i,t} + \lambda_i + \alpha_i$, se ipotizziamo che sia α_i,λ_i e U_i siano errori classici quindi a media nulla, omoschedastici e incorrelati, e che $Cor(X, \alpha_i) = Cor(X, \lambda_i) = 0$ posso stimare con consistenza attraverso il solito stimatore GLS.
Per completare l'operazione devo trovare $\tilde{V}$ tale da creare la trasformazione corretta so che è $E(\tilde{V}\tilde{V}') = \begin{bmatrix} \sigma_\alpha^2 + \sigma_\lambda^2 + \sigma_u^2 & \sigma_\alpha^2 + \sigma_\lambda^2 & ... \\ \sigma_\alpha^2 + \sigma_\lambda^2 & \sigma_\alpha^2 + \sigma_\lambda^2 + \sigma_u^2 & ... \\ \sigma_\alpha^2 + \sigma_\lambda^2 & ... & \sigma_\alpha^2 + \sigma_\lambda^2 + \sigma_u^2 \end{bmatrix}$, ed è facile da utilizzare.
Se invece con alcuni regressori sono correlati quindi ho $Cor(X, \alpha_i) \neq 0$ o $Cor(X, \lambda_i) \neq 0$, per calcolare i B_r devo usare delle variabili strumentali incorrealte agli errori che mi permettano di stimare i coefficienti.
I modelli con possibile correlazione sono due:

- Modello esteso in tutte e due le dimensioni quindi ho eterogeneità sia in N che in T.
- Modello esteso solo ai regressori che variano rispetto agli individui.

Dato un modello $Y = X_1 * B_1 + X_2 * B_2 + Z_1 * \delta_1 + Z_2 * \delta_2$ dove X_1, X_2 hanno eterogeneità in tutte e due le dimensioni, Z_1, Z_2 invece hanno solo la parte cross sezionale, inoltre i dati in X_2, Z_2 sono correlati con gli errori, ciò ci rende obbligatorio stimarli in maniera alternativa. Abbiamo due metodi di stima:

- Consistente o a due stadi. Questa tecnica ha come primo passaggio la stima dei B tramite stima within (FE), quindi usando la trasformazione D, nel modello $Y = XB + Z * \delta + V$ considerando quindi $Z * \delta$ degli errori, in questo modo i B_w ottenuti sono consistenti, con questa stima calcolo i residui $Y - X\hat{B}$.
Come secondo passaggio calcolo le stime between degli errori cioè $C'(Y - X\hat{B}) = C'(Z * \delta + V)$, sapendo che Z varia solo su i, ho $= Z*\delta + C'V)$, so anche che $V = \alpha \otimes J + U$ ho $= Z*\delta + \alpha \otimes J + U$, sapendo che $Cor(V, Z) \neq 0$ devo usare una variabile strumentale Z_1 essendo incorrelata strumenta se stessa, invece Z_2 deve essere per forza strumentata da $C * X_1$, uso la sua forma between per far combaciare le dimensioni di X_1 e Z_2.
Uso poi stimatore IV o $2SLS$ a seconda se il numero di variabili strumentali è uguale o no alle strumentate.

- Efficiente o ad uno stadio, ho il modello $Y = XB + Z * \delta + V$ dove $E(V * V') = \phi$ è una matrice di errori non classici quindi devo trovare la trasformazione GLS tale che $V^*V^{*'} = \phi^{-1}$, applicando questa trasformazione ottengo $V^*Y = V^*XB + V^*Z\delta + V^*V$, dopo aver applicato tale trasformazione devo trovare gli strumenti che mi permettanno di stimare B e δ:
  - Per la variabile X_1 ho $D * X_1$, ho usato una trasformazione della variabile anche se potevo usare X_1 per rendere indipendente tale strumento dagli altri così da aumentarne l'efficienza.
  - Per X_2 ho $D * X_2$ in questo modo elimina gli effetti individuali che cuasano il problema di correlazione.
  - Per Z_1 uso come strumento se stesso.
  - Per Z_2 uso $C * X_1$ che è giusto perchè è correlata con Z_2 ma non e correlata con gli errori per definizione.

  Ora per scegliere se IV o GIVE, vedo il numero di strumenti se il numero è uguale alle variabili allora uso IV se maggiore uso GIVE.

4 Modelli Panel dinamici

E' un tipo di modelli che evolve nel tempo infatti ho $Y_{i,t} = X_{i,t}\delta + \phi * Y_{i,t-1} + \epsilon_t$ dove $\epsilon_t = \mu_i + U_{i,t}$, noto quindi che è un modello AR(1) con dei regressori statici, dove X è una variabile deterministica. Se ipotizziamo che $\delta = 0$ ho un AR(1), e se $|\phi| < 1$ ho un modello stazionario, ora dobbiamo stimare ϕ considerando la stazionarietà, un semplice OLS non è corretto infatti ho correlazione tra errori e regressori, infatti $Y_{i,t}$ è correlato con μ_i, quindi se voglio avere uno stimatore consistente devo usèrò uno stimatore IV.

Inizialmente posso pensare di trasformare i dati in modo che la correlazione scompaia, posso provare ad usare una trasformazione within D, ottengo così $Y_{i,t} - Y_{i,.} = \phi(Y_{i,t-1} - Y_{i,.}) + U_{i,t} - U_{i,.}$, noto però che la correlazione rimane infatti $U_{i,.}$ è presente in $Y_{i,.}$.
Posso provare una seconda trasformazione con l'operatore differenza prima Δ, lo applico ed ho $Y_{i,t} - Y_{i,t-1} = \phi(Y_{i,t-1} - Y_{i,t-2}) + U_{i,t} - U_{i,t-1}$, anche quà però rimane la correlazione tra $U_{i,t-1}$ e $Y_{i,t-1}$, quindi devo per forza applicare IV.
Mantengo comunque l'operatore differenza Δ perchè spesso rende stazionaria la serie.

Gli stimatori a variabili strumentali per questo problema sono

- **Anderson-Hsiao** dato il modello $\Delta * Y_{i,t} = \phi * \Delta * Y_{i,t-1} + \Delta * U_{i,t}$, lo stimatore che troveremo ha due caratterstiche: è di tipo IV quindi il numero di strumenti è il numero di variabili, e secondo ignora la struttura dell'errore.
L'errore che otteniamo è un MA(1), infatti $\Delta * U_{i,t} = U_{i,t} - \psi * U_{i,t-1}$ dove $\psi = 1$ quindi ha come matrice di varianza e covarianza $G = \begin{bmatrix} \psi^2+1 & \psi & 0 & ... & 0 \\ \psi & \psi^2+1 & \psi & 0 & ... \\ 0 & ... & 0 & \psi & \psi^2+1 \end{bmatrix} = \begin{bmatrix} 2 & -1 & 0 & ... & 0 \\ -1 & 2 & -1 & 0 & ... \\ 0 & ... & 0 & -1 & 2 \end{bmatrix}$,
noto inoltre che visto che $|\psi| \geq 1$ tale serie non è invertibile.
Questo metodo ci dice che per strumentare $\Delta Y_{i,t-1}$ devo usare $Y_{i,t-2}$ e così via infatti tale strumento è indipendente dagli errori $U_{i,t-1}$ ed è correlata con $\Delta Y_{i,t-1}$, tale soluzione non è l'unica

ma è la più efficiente perchè perdo meno dati, infatti con questa strumentale posso usare $T-2$ dati se per esempio scelgo $i, t-3$ potrei usare solo $T-3$ e così via, ed inoltre indebolisce il legame tra strumentale e strumentata. In stata tale operazione si fa con il comando "xtivreg".

- Modello **Arellano-Bond**: in questo caso vengono utilizzati due strumenti cioè il modello $GIVE$ e viene tenuto conto della struttura dell'errore.
  Il modello che prendiamo in considerazione è sempre $Y_{i,t}-Y_{i,t-1} = \phi(Y_{i,t-1}-Y_{i,t-2})+U_{i,t}-U_{i,t-1}$, come nel caso di prima uso le osservazioni che partono da $t=3$, in questo caso lo strumento che uso per descrivere quando $t=3$ è $Y_{i,1}$, infatti non ho nessun errore correlato, se invece valuto $t=4$ uso 2 variabili strumentali $Y_{i,1}$ e $Y_{i,2}$, e lo faccio ripetutamente per tutti i tempi, costruisco poi la matrice strumentale come $W_i = \begin{bmatrix} Y_{i,1} & 0 & ... & 0 & 0 & 0 & 0 \\ 0 & Y_{i,1} & Y_{i,2} & ... & 0 & 0 & 0 \\ ... \\ 0 & 0 & 0 & ... & Y_{i,1} & ... & Y_{i,T-2} \end{bmatrix}$, tale matrice ha dimensioni $(T-2) * C$ dove $C = \sum_{j=1}^{T-2} J$, posso raccogliere queste matrici nella matrice di sistema in $W = \begin{bmatrix} w_1 \\ ... \\ W_N \end{bmatrix}$.
  Sapendo che strumentali w usare e sapendo che il sistema cross section è $\Delta Y_i = \phi \Delta Y_{i,-1} + \Delta U_i$ e che $VAR(\Delta U_i) = G$ e che il sistema raccolto totale è scritto come $\Delta Y = \phi \Delta Y_{-1} + \Delta U$ ne ricavo la sua matrice di sistema $I_n \otimes G$. Applico la trasformazione w' alla variabili ottenendo il modello $w'\Delta Y = \phi w' \Delta Y_{-1} + w' \Delta U$, quindi devo calcolare $E(w'\Delta U * \Delta U' * w)$ così da poter calcolare lo stimatore GLS, tale varianza è deterministica e vale $w' I_n \otimes G w$, sapendo ciò ottengo che $\hat{\phi}_{gls} = (\Delta Y'_{-1} w (w' I_n \otimes G w)^{-1} w' \Delta Y_{-1})^{-1} \Delta Y'_{-1} w (w' I_n \otimes G w)^{-1} \Delta Y$.
  Non è detto però che il dataset abbia la varianza teorica ecco perchè calcolo $\hat{\phi}_{gls}$ in maniera teoria ottengo i residui e ricalcolo una seconda volta $\hat{\phi}_{gls}$ solamente usando una varianza empirica e non teoria.
  Per valutare la significatività di $\hat{\phi}$ devo usare gli standard error corretti.

4.1 Esempio su stata

Come dataset prendiamo "panel2dyn" composto da tre variabili produzione, capitale, lavoro ed ho come funzione di produzione una coubb-dougls. Analizziamo le variazioni percentuali quindi usiamo i logaritmi dei vari valori, quindi la nostra funzione è $log(y) = \alpha + B_l * log(lavoro) + B_k * log(capitale)$, inoltre $B_l + B_k$ decidono il tipo di produzione di scala.
Stimo per primo il modello statitico ad effetti random "xtreg y l k" noto che ha rendimenti decrescenti di scala, però vedo che ho una super significatività quindi probabilmente il processo è integrato.
Posso fare anche un test di Breush-Pagan da controllare se gli effetti individuali esistono, vedo che rifiuto H_0 quindi esistono tali effetti, "xttest0".
Posso stimare anche il modello FE, noto che stimandolo ha una correlazione altissima, qundi il modello RE non è molto sensato

Ora stimiamo il modello a regressori dinamici uso "xtivreg y (L.y=L2.y) L K, fd" questa è la formula per calcolare lo stimatore di Anderson-Hsiao, quindi creo il modello $\Delta Y_{i,t} = \alpha + \phi * \Delta Y_{i,t-1} + B_l * \Delta L_{i,t} + +B_k * \Delta K_{i,t} + \Delta U_{i,t}$ e sto strumentando la variabile $\Delta Y_{i,t-1}$ con $Y_{i,t-2}$ come da teoria. I risultati però sono insoddisfacenti infatti ho che i coefficienti B sono negativi e non significativi e ciò non è concorde alla teoria economica, inoltre il ϕ è in modulo maggiore di 1 ciò indica non stazionarietà della serie stimata, questi problemi possono derivare da una mal-specificazione del modello o di variabili strumentali sbagliate.
Posso fare un analisi di robustezza cioè controllare se i risultati con "xtivreg y (L.y=L3.y) L K, fd" sono sensati, in questo caso lo sono e risolvono tutti i problemi trovati in precedenza.

Uso quindi per rimediare lo stimatore di Arellano-Bond che sicuramente funziona meglio avendo più strumenti la dimensione degli strumenti è $(T-2) * C$, si fa "xtabond y l k", nel nostro caso essendo $T=10$ ho 36 strumenti per $\Delta Y_{i,t-1}$ poi ne ho altri 3 per stimare α, B_l, B_k.
Posso usare anche il metodo two step per valutarlo, usando il comando "xtabond y l k, twostep",

meglio ancora se aggiungo "robust", perchè così mi calcola automaticamente la significatività corretti dei regressori.
I modelli dinamici sono utili in particolare se T diventa grande.

5 Modelli per variabili qualitative

In questa parte tratteremo principalmente le cross section, in particolare vogliamo stimare $P = F(X,B)$, dove X_i sono le variabili esplicative e B_i sono i regressori, U_i è l'errore classico, Y_i invece è la risposta che può essere solo dicotomica.
So che $E(Y) = P(Y=1)*1 + P(Y=0)*0 = P(Y=1)$ inoltre so che $E(Y_i) = F(X_iB) = P(Y=1) = P_i$.

5.1 Linear probability model

Come prima trattazione si vede un modello naive, infatti dico che $Y = X_iB + U_i$ quindi $P_i = X_iB$, questa formulazione non impone che la probabilità stia tra 0 ed 1, infatti potenzialmente ha un range infinito e questo è un problema. ho varie proprietà:

- $E(U_i) = 0$, infatti se $Y_i = 1$ ho che $U_i = 1 - X_iB$, se $Y_i = 0$ ho che $U_i = -X_iB$ so che $P_i = X_iB$ quindi $E(U_i) = P_i * (1-P_i) - P_i(1-P_i) = 0$.
- $Var(U_i) = E(U_i^2) = P_i*(1-P_i)^2 + P_i^2(1-P_i) = P_i(1-P_i)$ noto quindi che ho eteroschedasticità tra gli elementi.
- Questo modello può esseere utile per fare un analisi preliminare e capire quali regressori sono utili e quali no.

Possiamo trattare l'eterosschadisticità con una trasformazione GLS $V = \frac{1}{P_i(1-P_i)}$, ho quindi $\frac{Y}{P_i(1-P_i)} = \frac{B_0}{P_i(1-P_i)} + \frac{B_1*X_1}{P_i(1-P_i)} + ... + \frac{U_i}{P_i(1-P_i)}$.

5.2 Modelli probit e logit

Per ovviare ai problemi del LPM ovvero:

- La continuità e l'illimitatezza del range della variabile risposta.
- Non interpretabilità dei coefficienti e minor robustezza nelle previsioni.

vengono introdotte le trasformazioni logit e probit, essi essendo delle funzione cumualate di distribuzione stanno sempre tra 0 e 1. Ho quindi:

- **Probit**, ho come F la funzione di probabilità di una normale standard, quindi $P_i = \Phi(X_iB) = F(X_iB) = \int_{-\infty}^{X_iB} \frac{e^{-\theta^2/2}}{\sqrt{2\pi}} d\theta$.
- **Logit**, ho come $P_i = \Lambda(X_iB) = F(X_iB) = \frac{e^{X_iB}}{1+e^{X_iB}}$, è importante e facile calcolare l'odds ratio cioè $\frac{P_i}{1-P_i} = e^{X_iB}$ e valutare anche i logODDs cioè X_iB che essendo lineare sono facili da trattare. E' preferito al modello probit perchè non ha integrali da calcolare.

Per questi modelli per stimare la B devo usare delle stime di massima verosimiglianza $L(B|Y_1...Y_n, X) = \prod_{i=1}^{n} P_i^{Y_i} * (1-P_i)^{1-Y_i}$ oppure identicamente massimizzare la log verosimiglianza, trovo così $\hat{B}_{ml}$, purtroppo non ho una soluzione chiusa quindi utilizzerò dei metodi numerici per trovare i miei coefficienti. Nei metodi numerici si rishica di trovare un massimo locale ecco perchè si provano molti set iniziali così da trovare quello realmente massimo.

Questi due modelli hanno anche delle giustificazioni in senso economico:

- Random utility model (RUM): questo modello si basa sull'idea che il consumatore massimizza la sua utilità e tale utilità ha forma $U_{i,j} = S_{i,j} + \epsilon_{i,j}$, avendo questo modello e voglio stimare $P(Y_i = 1) = P(U_{i,1} > U_{i,0}) = P(S_{i,1} + \epsilon_{i,1} > S_{i,0} + \epsilon_{i,0}) = P(\epsilon_i < S_{i,1} - S_{i,0})$, noto quindi la mia distribuzione finale è $F(S_{i,1} - S_{i,0})$ se vedo $S_{i,1} = \alpha_1 + X_{i,1}^*\gamma$ allora noto che ogni probabilità ha una caratteristica per ogni individuo, ciò rende di difficile interpretazione il coefficiente.

- Variabile Latente (LVM): Sia $Y_i = 1$ se $Y_i^* > \lambda$ e $Y_i = 0$ se $Y_i^* < \lambda$, se ipotizziamo che $Y_i^* = X_i * B + U_i$, per modellare in questo caso so che $P(Y_i = 1) = P(X_i * B + U_i > \lambda) = P(U_i > \lambda - X_i * B)$ che per simmetria diventa $P(U_i < X_i * B - \lambda)$ se dico che $\lambda = 0$ allora ho $F(X_i * B)$ quindi anche questa ha effetti individuali.

Se voglio interpretare gli effetti marginali cioè $\frac{\delta Y_i}{\delta X_{i,r}}$ che di solito nei modelli lineari coincidono con B_r il coefficiente, nel caso probit e logit no. Se analiziamo vediamo che:

- Linear probability model ho che $Y_i = X_i * B + U_i$ ho che $E(Y_i) = P(Y_i = 1) = X_i * B$ allora vedo facilmente che l'effetto marginale è $\frac{\delta E(Y)}{\delta X_{i,r}} = B_r$
- Probit sia $Y_i = \Phi(X_i * B) + U_i$ ho che $E(Y) = \Phi(X_i * B)$ quindi $\frac{\delta E(Y)}{\delta X_{i,r}} = \frac{\delta \Phi(X_i * B)}{\delta X_i * B} * \frac{\delta X_i * B}{\delta X_{i,r}} = \phi(X_i * B) * B_r$, non ha un interpretazione semplice ed ha un effetto marginale che varia per ogni individuo.
- Logit sia $Y_i = \Lambda(X_i * B) + U_i$ ho $E(Y) = \Lambda(X_i * B)$ allora come prima $\frac{\delta E(Y)}{\delta X_{i,r}} = \frac{\delta \Lambda(X_i * B)}{\delta X_i * B} B_r$, dove però $\frac{\delta \Lambda(X_i * B)}{\delta X_i * B} = \Lambda(X_i * B)(1 - \Lambda(X_i * B))$.

Per ovviare al calcolare ed interpretare moltissimi effetti marginali diversi si sono accorpati in due metodi:

- Calcolo n effetti diversi e ne faccio la media quindi $\sum_r^n \frac{\delta E(Y)}{\delta X_{i,r}} / n$, ha un problema di fatica computazionale.
- Effetto marginale della media dei regressori, quindi prima calcola la media e poi calcolo un unico effetto, è più efficiente computazionalmente.

5.3 Indicatori di bontà per modelli dicotomici

Se ho dati continui per le regressioni di solito usiamo come indice di bontà di adattamento ai dati $R^2 = 1 - Rss/Tss$, in questo caso con modelli dicotomici ho uno pseudo $R^2 = 1 - \frac{log(L_1)}{log(L_0)}$ dove L_1 ed L_0 sono le likelyhood del modello $M_1 = F(X_i * B) + U_i$ e $M_0 = F(B) + U_i$, quindi un modello creato da noi contro un modello formato da una sola costante. Questo quasi R^2 è compreso tra 0 e 1 con casi limite 0 cioè il modello scelto è indifferente da una costante, oppure se è 1 in quel caso trattando verosimiglianza vuol dire che non c'è incertezza e quindi non è interessante da studiare.
I valori $log(L_1) = MAX(\sum_{i=1}^n Y_i * log(P_i) + (1 - Y_i) * log(1 - P_i))$ ed il suo argmanx è B_{ml}, invece $log(L_0)$ si calcola più facilmente infatti so che $Y_i = F(B_1) + U_i$ quindi $\hat{P}$ è costante ed è uguale a $\frac{N_1}{N}$ cioè casi favorevoli fratto casi totali, quindi $log(L_0) = N_1 * log(\frac{N_1}{N}) + N_0 * log(\frac{N_0}{N})$ tale formula è deterministica.

5.4 Previsioni per modelli logit

Sapendo che $F(X_i * B_{ml}) = P_i = E(Y_i)$ che sta tra 0 e 1 invece la risposta e dicotomica quindi non ha molto senso usare come errore empirico $\hat{U}_i = Y_i - \hat{Y}_i$, allora devo inserire una regola di classificazione tipicamente se $P_i > 0.5$ allora $\hat{Y}_i = 1$ se no $\hat{Y}_i = 0$, in questo modo posso costruire sensatamente un $MSFE = \sum_{i=1}^n \hat{U}_i^2 / N$.
Se vogliamo fare un test dobbiamo trovare un modello di riferimento dove calcolare $MSFE0$, come benchmark uso il modello M_0 in questo modo se $\hat{P} > 0.5$ avrò sempre $\hat{Y}_i = 1$ se no $\hat{Y}_i = 0$, quindi il suo $MSFE0 = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 / N$ se $\hat{P} > 0.5$ avrò $MSFE0 = 1 - \hat{P}$ se no $MSFE0 = \hat{P}$.

5.5 Esempio in stata

Ho un datatset con 32 righe e come variabile target Grade che è una variabile dicotomica, poi ho GPA, TUCE (conoscenza pregressa della materia), PSI (metodi di apprendimento alternativi).

Come primo modello possiamo trattare il linear probability model, si fa come una normalissima regressione, prevediamo con "predict PI-LM" e poi "sum PI-LM "i dati vediamo che il range non è tra 0 e 1 quindi utilizziamo un altra tecnica.

Usiamo il probit, quindi usiamo il comando "probit Grade GPA ...". Ci vengono fornite informazioni del tipo:

- Line-ratio test che è un f-test zero slopes nel caso dicotomico.
- Pseduo R^2.
- Stima automaticamente B_{ml}.
- Ci da in automatico anche il valore di $log(L_1)$

Notiamo inoltre che i coefficienti non sono troppo diverso dal linea probability model.
Con i dati a disposizione posso calcolare $\hat{P} = \frac{N_1}{N}$, lo faccio come "sum grade if grade==1" così ottengo N_1 oppure uso "probit grade", in questo modo calcolo $\hat{P}$ e poi posso ricavarmi manualmente pseduo R^2.
Posso anche prevedere le probabilità future con il comando "Gen PI-p=Normal(Bcons+Bgpa*Gpa+...)" oppure potevo fare "predict xib-p, xb".
Se volessi calcolare gli effetti marginali ho due strade:

- Media degli effetti marginali, cioè calcolo ogni effetto marginale per una determinate feuteres che so essere $\phi(X_i * B) * B_r$ e poi ne faccio la media.
  Nel nostro caso faccio "gen mfxz = normalden(xib-p)*Bgpa" e poi ne faccio la media di quei valori.
- Effetto nella media dei regressori. Calcolo la media con "sum Gpa" "gpax = r(mean)", poi uso la formula di prima "gen mfxz = normalden(bcons+Bgpa*gpax+...)*Bgpa".
  Queste operazioni sono evitabili se usiamo il comando "MFX", questa funzione ci fa vedere anche che PSI non è una variabile continua quindi il suo effetto marginale ha interpretazione diversa, infatti non indico la derivata ma $\Phi(X * B + B_{psi}) - \Phi(X * B)$.

Ora per prevedere mettiamo un treshold cioè se $P_i > 0.5$ allora $Y_i = 1$ come da teoria così poi riusciamo a calcolare U_i^2 e quindi MSFE e confrontarlo con MSFE0, quindi faccio "predict pi" e salvo le probabilità poi "gen =0" ed infine "replace yihat =1 if $pi > 0.5$", creo i residui con "gen Uhat=grade-yihat" e "gen uhat2=$uhat^2$" poi calcolo "gen msfe0=11/32" come da teoria l'errore è la media delle osservazioni positive.

Tutte queste operazioni si possono rifare chiamando il link e mettendo il *logit*, e valori dei coefficienti trovati sono ovviamente diversi e non ha senso confrontarli, però ha senso confronta la significatività dei coefficienti, gli effetti marginali, gli pseudo R^2 anche se spesso nei due modelli sono molto simili.
Per valutare anche l'eventuale MSFE, discretizzando le probabilità sempre con la stessa treshold.

6 Scelte multiple

6.1 Scelte ordinate

Se ho come variabile target una molteplicità di risposte con uno specifico ordine $M_1 < ... < M_j < ... < M_k$ allora le modello attraverso le variabili latenti.
Sia $Y_i^* = X_i * B + \epsilon_i$ dove i sono gli individui e invece Y_i^* è la vera motivazione, diciamo che $Y_i = 1$ se $\gamma_0 < Y_i^* \leq \gamma_1$ e $Y_i = 2$ se $\gamma_1 < Y_i^* \leq \gamma_2$ e così via fino a $Y_i = k$ se $\gamma_{k-1} < Y_i^* \leq \gamma_k$, questi dati possono essere **normalizzati** se imponiamo che $\gamma_0 = -\infty$ e $\gamma_1 = 0$ e $\gamma_k = \infty$.
Questa visione mi impone di stimare con la massima verosimiglianza non solo i B ma anche le varie soglie.
Se imponiamo che $k = 3$ e che l'errore si dispone come una normale, allora ottengo che $P(y_i = 1) = P(Y_i^* < \gamma_1) = P(X_i * B + \epsilon_i < \gamma_1) = \Phi(-X_i * B)$, $P(y_i = 3) = P(Y_i^* > \gamma_2) = P(X_i * B + \epsilon_i > \gamma_2) = 1 - \Phi(\gamma_2 - X_i * B)$ infine $P(y_i = 2) = 1 - 1 + \Phi(\gamma_2 - X_i * B) - \Phi(-X_i * B) = \Phi(\gamma_2 - X_i * B) - \Phi(-X_i * B)$.
Noto quindi la log massima verosimiglianza è $\sum_i^n 1_{y_i = j} log(P_i(y_i = j))$.

Da questo modello è possibile sapere a priori il segno degli effetti marginali esclusivamente alle scelte estreme 1 e k. Infatti so che $\Phi(-X_i * B)$ essendo probabilità è monotona quindi se B è positivo e X_i

aumenta allora la probabilità diminuisce, ha quindi un effetto marginale negativo, lo stesso ragionamento vale ma al contrario nel caso della categoria K.
Abbiamo visto prima che abbiamo adottato delle normalizzazioni, ma si può invece che imporre $\gamma_1 = 0$, si impone $B_1 = 0$ infatti vedo che $P(y_i = 1) = \Phi(\gamma_1 - B_1 - X_i^* * B)$ notiamo che il cambiamento apportato non cambia per la verosimiglianza.

6.1.1 Esempio in stata

Se prendiamo il dataste "pension.dta" è di tipo cross sezionale con 226 individui e 19 colonne, la target è "pctstck" con livello 0,50,100, cioè il numero di stock presente nei fondi pensioni, e 18 variabili tra cui la dummy "choice" che ti dice se il piano è stesso personalizzato o no.
Per fine descrittivo faccio una regressione, posso calcolarne i valori fittati, notiamo che gli unici valori significativi sono "choice", "age", "profitshr", inoltre posso calcolare l'effetto marginale della differenza di "Choice" di un individuo tipo creato da noi facendo la differenza delle probabilità calcolati nei due modi.

Ora possiamo regredire con "probit pctstck ...", noto che ho significatività simili al modello di prima, noto anche che stata impone $B_1 = 0$ non $\gamma_1 = 0$, calcolo i valori soglia come "/cut1" e "/cut2".
Posso calcolare anche in questo caso l'effetto marginale di choice che è simile al modello precedenti, lo calcolo attraverso le stime delle 3 probabilità per le tre categorie "predict P1 P2 P3" nei due casi poi ne calcolo la differenza di valore atteso.
Un'altra similitudine con il modello di prima si può trovare calcolando i valori attesi cioè "gen pctstockopm=0*P1+50*p2+100*p3" poi vedo la correlazione tra questi valori attesi e quelli reali e noto che valuta molto male i dati essendo incorrelati, proprio come il modello precedente, posso calcolare anche la correlazione tra i valori attesi del modello ordinati e quello lineare e notare che è molto alta indicandone ancora la somiglianza.

6.2 Scelte multiple non ordinate

Sia Y_i una dummy multinomiale se usiamo il modello random utility ho $U_{i,j} = S_{i,j} + \epsilon_{i,j}$ allora avrò $P(Y_i = j) = P(U_{i,j} = Max(U_{i,1}...U_{i,m}))$ oppure lo vedo come $P(S_{i,j} + \epsilon_{i,j} = Max_{k \neq j}(S_{i,j} + \epsilon_{i,j}))$, per modellare tale probabilità devo fare due assunzioni:

- $\epsilon_{i,j}$ sono indipendenti tra di loro, ed è una restrizione molto forte.
- $\epsilon_{i,j}$ si distribuisce come una logweibul quindi $F(t) = e^{-e^{-t}}$,ed è utile per modellare gli eventi estremi.

Date queste restrizioni allora $P(Y_i = j) = \frac{e^{S_{i,j}}}{e^{S_{i,1}} + ... + e^{S_{i,m}}}$, dove $S_{i,j}$ è la parte regressiva che può essere di due tipi:

- Conditional logit se $X_{i,j} * B$ cioè i regressori cambiano sia per individuo che per scelta ma il coefficiente rimane costante.
- Multinomial logit se $X_i * B_j$ cioè i regressori cambiano solo per individuo ed il coefficiente per scelta.

Se avessi $X_i * B$ avrei che le $P(Y_i = j) = 1/m$ cioè costanti ed è spesso usato come modello di riferimento.
Spesso si ipotizza di avere anche una scelta normalizzata cioè che $S_{i,1} = 0$ questo da origine a $P(Y_i = j) = \frac{e^{S_{i,j}}}{1 + ... + e^{S_{i,m}}}$.

Proprietà del modello:

- Se $\epsilon_{i,j}$ è normale allora P è ricavabile nel metodo classico con $M - 1$ integrali spessi risolti in maniera numerica.
- Se $\epsilon_{i,j}$ è indipendente allora $U_{i,j}$ è indipendente, questa assunzione non è realistica nella realtà infatti se presenti delle scelte simili tali saranno correlate. Tale modello è detto Indipendenza per alternative irrilevanti.

Un modo per correggere l'errore di indipendenza è il modello nasted logit, cioè posso strutturare le scelte su più livello così da renderle realmente indipendenti. Prima abbiamo un livelo S con due scelte e quindi calcoliamo $P(Y_i = S)$ che viene stimato con un logit binario, poi nel livello inferiore abbiamo J scelte per ogni s quindi avrò $P(Y_i = j|Y_i = 1)$.

6.2.1 Esempio in stata

Abbiamo il dataset "Keane.dta" nella target ho 3 variabili scuola, disoccupato, lavoro, dove useremo diversi regressori per predirlo. La regressione multinomial è "mlogit status educ ... if year==87, b(1)", dove b(1) indica che normalizziamo per la scelta 1, ha un output molto simile al logit.
Posso calcolarne gli effetti marginali con "mfx predict(oucome(2))" ed ha segnati di sensata interpretazione.
Posso fare "Test [2]exp [2]exp2" per valutare se i coefficienti di questi due regressori sono uguali a 0, invece con "Test [2]exp=[2]exp2" valuto se sono uguali tra di loro, se voglio calcolare il modello ristretto uso "Constraint 1 Test [2]exp=[2]exp2" e poi "mlogit status educ .., constraint(1)".

Per il conditional logit, cioè $X_{i,j}B$ ho come dataset Choice.dta, per regredire uso "clogit choice ..., group(Id)" anche questo ha un output molto simile al logit, vediamo con "predict phat" che predice mallissimo ed era ovvio dato R^2 basso.

Se voglio unire caratteristiche che variano tra individui e inidividui e scelte faccio "aclogit choice dealer, case(id) alternatives(car) casevars(sex income)", dove quelle che variano per solo la persona sono sex ed income, per testare la bontà delle previsioni posso vedere quanto correlano con le risposte reali, noto però che è molto bassa e quindi il modello non è molto buono, si poteva intuire anche da un R^2 molto basso .
Se volessi testare se sono utili i coefficienti aggiunti nell'ultimo modello potrei fare "Test [Japan]sex ... [Europe]cons", tale test rifiuta l'ipotesi nulla tale che gli ultimi due modelli siano uguali significativamente.

6.3 Modello nested logit

Abbiamo visto prima che per risolvere il problema delle alternative irrilevanti abbiamo diviso in layer le scelte, dove il secondo dipendeva dal primo.
In particolare date le scelte $U_{i,j} = S_{i,j} + \epsilon_{i,j}$ dove nel caso nested ho $S_{i,j} = X_{i,j} * B + Z_i\gamma$ dove però i non è l'individuo ma il numero del layer e j è la scelta che varia da $1...N_i$, noto quindi che la scelta è influenzata anche della posizione di densità in cui si trova, poi abbiamo la variabile $\epsilon_{i,j}$ che è indipendente ed ha distribuzione per gestire i valori estremi.
Ora se vogliamo modellare la probabilità di scelta devo ricordarmi che $P_{i,j} = P_{j|i} * P_i$, ma noi sappiamo:

- $P_{j|i} = \frac{e^{X_{i,j}*B+Z_i*\gamma}}{\sum_k^{N_i} e^{X_{i,k}B+Z_i\gamma}}$ che semplificando otteniamo $\frac{e^{X_{i,j}B}}{\sum_k^{N_i} e^{X_{i,k}B}}$ vediamo quindi che tali scelte sembrano indipendenti da i.
- $P_i = \frac{\sum_j^{N_i} e^{X_{i,j}*B+Z_i*\gamma}}{\sum_m^C \sum_k^{N_i} e^{X_{m,k}B+Z_m\gamma}}$, noto che in questo caso non si elidono.

Se chiamiamo $I_i = log(\sum_k^{N_i} e^{X_{m,k}B})$ allora posso scrivere $P_{j|i} = \frac{e^{X_{i,j}B}}{e^{I_i}}$ e $P_i = \frac{e^{I_i+Z_i*\gamma}}{\sum_m^C e^{I_m+Z_m\gamma}}$.
Vedo quindi che P_i e $P_{j|i}$ sono collegate da I_i, noto anche che la stima di $P_{j|i}$ è un conditional logit quindi facile da stimare, dopo aver stimatore I_i se inserisce nella seconda equazione che è un multinomial logit così da stimare γ e ricostruirmi senza problemi $P_{i,j}$.

7 Modelli per variabili dipendenti limitate

Se siamo in un ambito di target continue, cross-sezionalità allora per trattare una possibile limitatezza di variabili continue ho:

- A censura: cioè $Y_i = Max(Y_i, C)$ o $Y_i = Min(Y_i, C)$, in questo caso però i dati vengono sempre osservati solo che in output quelli sopra o sotto una certa soglia vengono buttati, utile sopratutto per problemi di ottimo con condizioni di kuntacher, gli esempi più famosi sono quelli microeconomici di offerta limitata e i derivati di tipo europeo.

- Troncamento: cioè $Y_i = Max(Y_i, C)$ o $Y_i = Min(Y_i, C)$ però i dati se non superano la soglia non vengono neanche registrati, in questo caso come quello sopra $C = 0$ tipicamente.

7.1 Censura

Se prendiamo un modello classico del tipo $Y_i = X_i * B + \epsilon_i$ allora sappiamo che lo stimatore OLS è adatto a stimare B senza vere distorsioni o inconsistenze, se però ha forma $Y_i = Max(Y_i, 0)$ ho che $E(Y_i) = E(Y_i|Y_i > 0) * P(Y_i > 0)$ (elido già gli 0) e ricordandoci la forma di Y_i ho $E(X_i * B + \epsilon_i|X_i * B + \epsilon_i > 0) * P(X_i * B + \epsilon_i > 0)$ quindi ho $(X_i * B + E(\epsilon_i|\epsilon_i > -X_i * B)) * \Phi(X_i * B)$(simmetria della normale), ho quindi un Bias ed è $E(\epsilon_i|\epsilon_i > -X_i * B)$ se ϵ_i è $N(0, 1)$ allora tale Bias è uguale $\frac{\phi(-X_i*B)}{1-\Phi(-X_i*B)}$ o se non fosse normalizzato $\sigma * \frac{\phi(-X_i*B/\sigma)}{1-\Phi(-X_i*B/\sigma)} = \sigma * \frac{\phi(X_i*B/\sigma)}{\Phi(X_i*B/\sigma)}$.
Se lo sostituisco nel valore atteso originale ottengo $(X_i * B + \sigma * \frac{\phi(X_i*B/\sigma)}{\Phi(X_i*B/\sigma)}) * \Phi(X_i * B/\sigma)$ ho quindi la distorsione come $\sigma * \phi(X_i * B/\sigma)$.

Per risolvere ciò creo il modello $Y_i = X_iB * \Phi(X_i * B/\sigma) + \sigma * \phi(X_i * B/\sigma) + \eta_i$, usiamo quindi per stimare con OLS un modello aumentato.
Purtroppo però Φ e ϕ dipendono direttamente da B, quindi devo stimare B nell'equazione originale con un probit, può essere visto come uno schema di censura semplificato dove ho 0 lascio 0 se ho dei valori diversi metto 1, in questo poso minimi la logverosimiglianza di un modello probit ottenendo così $\frac{B}{\sigma}$, non riesco a separarli perchè non ho abbastanza informazioni.
Trovati questi valori li sostituisco in $Y_i = X_iB * \Phi(X_i * B/\sigma) + \sigma * \phi(X_i * B/\sigma) + \eta_i$, che con un OLS, ed avendo più informazioni riesco a stimare B e σ, questo è un approccio chiamato minimi quadrati corretti.

Esiste un modello alternativo che fa uso di un solo stadio e si chiama Tobit 1, per costruire tale modello si usa la massima verosimiglianza, ad essa contribuisce i valori noti di Y_i^* e di 0, stimeremo 0 con un probit e Y_i^* con un modello lineare.
Sappiamo che la funzione di logverosimiglianza del modello è $l(\theta) = \sum_{i \in Y_i < 0} ln(P(Y_i^* = 0)) + \sum_{i \in Y_i > 0} ln(f(Y_i^*|Y_i > 0) * P(Y_i > 0)) = \sum_{i \in Y_i < 0} ln(1 - \Phi(X_i * B/\sigma)) + \sum_{i \in Y_i > 0} ln(\sigma * \phi(Z_i))$ dove $Z_i = (Y_i - X_iB)/\sigma$, visto che $f(Y_i^*|Y_i > 0) = \frac{f(Y_i^*)}{P(Y_i>0)}$ funzione di distribuzione di normale.

7.2 Troncamento

Come detto prima nel troncamento non posso osservare tutte le variabili $Y_i < 0$, quindi se andiamo a calcolare il valore atteso ho $E(Y_i^*|(Y_i > 0) = E(X_iB + \epsilon_i|X_iB + \epsilon_i > 0) = X_iB + E(\epsilon_i|\epsilon_i > -X_iB) = X_iB + \frac{\phi(X_iB/\sigma)}{\Phi(X_iB/\sigma)}$, quindi se come prima volessi trovare il modello corretto avrei che $Y_i = X_iB + \frac{\phi(X_iB/\sigma)}{\Phi(X_iB/\sigma)} + \eta_i$, e posso stimarlo tramite due stadi come fatto in precedenza, oppure usare un approccio di massima verosimiglianza, cioè ho $log(L) = \sum_i log(f_i|Y_i > 0)$ per bayes $= \sum_i log(\frac{f(y_i)}{P(Y_i>0)}) = \sum_i log(f(y_i)) - log(\Phi(X_i * B/\sigma)$, dove $f(y_i) = (\phi(Z_i)/\sigma)$ è normale.

Entrambi i modelli presi in considerazioni non ammettono che le motivazioni alle decisioni sia note, quindi che le X_i siano diversi sia per scelte che per individuo, dobbiamo modellare X_iB diverse per ogni tipo di scelta.
Per risolvere ciò si sono individuati due modelli uno per le probabilità e una per il livello Y_i, per esempio se prendiamo l'offerta di lavoro $w_i = X_{i,1}B_1 + \epsilon_{i,1}$ e $h_i = X_{i,2}B_2 + \epsilon_{i,2}$, gli errori provengono da una normale bivariata con covarianza diversa da 0, vedo che i regressori del salario sono diversi se sono in troncamento ho se $h_i > 0$ allora $w_i > 0$ se $h_i = 0$ $w_i = NA$ e se voglio calcolare $E(w_i|h_i > 0) = X_{i,1}B_1 + \sigma_{1,2}\frac{\phi(X_{1,2}B_2)}{\Phi(X_{1,2}B_2)}$.
Uno potrebbe pensare che le due scelte siano scorrelate, per dimostrare che non è così, ipotizziamo la creazione del salario di riserva $w_i^r = Z_i\gamma + \eta_i$ se $w_i > w_i^r$ allora $h_i = 1$ se no 0, quindi $h_i = w_i - w_i^r = X_{1,i}B_1 + \epsilon_{i,1} - Z_i\gamma - \eta_i$ se mettiamo $\epsilon_{i,2} = \epsilon_{i,1} - \eta_i$ allora essendo fortemente legati non possono essere indipendenti.

7.3 Esempio su stata di censura

Il dataset è Mrol.dta, per prima cosa faccio una regressione quindi un linea probability model, e prova a valutare le significatività dei coefficienti.
Come seconda cosa stimiamo un logit e con il comando "estat classifcation" calcolo la confusion matrix (utile per capire dove fare censura delle variabili).
Faccio censura quindi "Hours=(0,ore di lavoro)", come prima stima faccio una regressione, e per valutarne la bontà valuto la correlazione tra reali e i valori fittati.
Come seconda applico "tobit hours ..., LL(0)" questa è una regressione di massima verosimiglianza, e mi salvo "scalar sigma=sqrt(-b(/var(e.hours)))".
Come terzo applico lo stimatore a due stadi quindi ottengo le stime preliminari di B/σ con "gen dummy=0" "replace dummy=1 if hours>0", "quietly probit dummy ..." e poi "predict xb-probit, xb" così mi salvo i valori stimati e metto "gen fi=normlden(xb-probit)" e "gen FI=normal(xb-probit)" infine "gen li=fi/FI" così posso unirli nella regressione come "reg hours... li"

7.4 Esempio su stata di troncamento

Se volessi usare il troncamento ho "troncreg hours..., ll(0)" oppure facciamo come prima la regressione in due stadi. Posso calcolare il salario di riserva come "Heckman lwage edu..., select(inlf= edu exper ...)" dove inlf è un dummy ottenuta per regressioni su variabili precedenti.
Calcolo il mils-ratio che è un indicatore che se non significativo ti indca che $X_1 = X_2$ quindi sono stati mal specificati