

Introduzione alle serie storiche

Definizione: E' una collezione di osservazioni indicate come X_t dove X è l'osservazione indicizzata al tempo t che va da 1 a n , tali osservazioni sono limitate. Da un punto di vista statistico è la parte finita delle realizzazioni di un processo stocastico.

Stazionarietà debole si dice che un processo è debolmente stazionario se momento primo e momento secondo sono invarianti rispetto a traslazioni temporali essi soddisfano le seguenti condizioni:

- Valore atteso del processo è finito ed è μ .
- La sua covarianza $E(X_t - \mu)(X_{t-j} - \mu) = \gamma_j$ essa dipende solo dalla distanza temporale dei due. Se $j=0$ abbiamo la varianza.

Una variabile economica spesso non è stazionaria infatti non ha media costante ma evolve nel tempo. Come processo stazionario per eccellenza prendiamo il white-noise u_t , un processo con media 0, varianza costante a σ^2 e covarianza 0. Se si distribuisce come I.I.D. $N(0, \sigma^2)$, è detto processo gaussiano.

Le assunzioni che si fanno su un processo stazionario non sono estendibili ad un processo non stazionario. In particolare la previsione peggiora con il passare del tempo, infatti un processo stazionario avrà dopo uno shock la tendenza a tornare alla sua media, uno non stazionario no. Inoltre gli stimatori dei minimi quadrati se sono trovati su un processo non stazionario perdono la validità asintotica, quindi i test T ed F non sono più validi.

Per **memoria** di un processo ci si riferisce al grado di dipendenza tra il processo in un certo tempo t e la sua storia passata, si ha memoria elevata se è molto influenzato dagli eventi del passato. Si parla quindi di assenza di memoria se non è influenzato dal percorso passato (markov omogeneo).

Se analizziamo un processo AR stazionario detto anche a memoria transitoria, allora l'influenza decresce in modo esponenziale quando si va indietro nel tempo.

Se invece la componente auto-regressiva contiene una radice unitaria allora non è stazionario, quindi la funzione di autocorrelazione decrescente più lentamente, perciò gli shock osservati tendono a modificare in modo permanente la serie storica, questo processo non stazionario è detto a memoria permanente.

1 Tipi di non stazionarietà

1.1 Non stazionarietà e trend deterministici

La non stazionarietà non è sempre dettata dal processo stocastico ma anche da una trend deterministico. Tale processo è detto trend-stazionario.

Per rendere la serie stazionaria va depurata dal trend, se supponiamo che la non stazionarietà è dovuta alle variazioni del livello medio della serie espresso come funzione deterministica del tempo $g(t)$, dove tale funzione individua un trend deterministico.

Sottraiamo $X_t - g(t)$ tale processo è stazionario, se prendiamo un ARMA(p,q), possiamo scrivere $X_t - g(t) = H(L) * u_t$, dove L è l'operatore ritardo $LX_t = X_{t-1}$ ed $H(L) = \frac{\theta(L)}{\phi(L)}$ sopra rappresenta la parte a media mobile e sotto quella auto-regressiva.

Posso scrivere $X_t = g(t) + H(L) * u_t$ l'ultima parte è detta componente transitoria ed è l'unica soggetta a shock aleatori, notiamo quindi che è scrivibile come somma.

1.2 Non stazionarietà e trend stocastici

1.2.1 Processi a radice unitaria

Esempio classico il random walk, processo non stazionario con una radice unitaria, si scrive come $X_t = X_{t-1} + u_t$. Notiamo che il coefficiente di X_{t-1} è uguale a 1 in modulo cioè lo rende non stazionario.

Esso può essere riscritto come $X_t = X_{t-2} + u_{t-1} + u_t$ e si può andare avanti in maniera ricorsiva come $X_t = X_0 + \sum_{j=1}^t u_j$, notiamo quindi che la componente stocastica è data dalla sommatoria degli shock, tale somma è un trend stocastico.

In questo caso per rendere la serie stazionaria usiamo l'operatore differenza una volta e la rendiamo stazionaria, cioè $\Delta X_t = X_t - X_{t-1} = u_t$, inoltre X_t è detto processo integrato di ordine 1.

In generale un processo si dice integrato di ordine S , se applicando S differenziazioni ottengo un processo stazionario.

Se invece ho un random walk con drift quindi $X_t = S + X_{t-1} + u_t$ può essere visto come $X_t = X_0 + S * t + \sum_{j=1}^t u_j$ e dopo aver differenziato una volta ottengo $\Delta X_t = S + u_t$.

1.2.2 Scomposizione di Beveridge-Nelson

E' una tecnica che si usa per i **processi a radice unitaria intorno ad una funzione deterministica**, esso divide in componenti un processo integrato di ordine 1.

Sappiamo che $\Delta X_t = H(L)u_t$ è un processo stazionario allora esso è scrivibile come $MA(\infty)$, cioè $\Delta X_t = \sum_{j=0}^{\infty} h_j * L^j * u_t = \sum_{j=0}^{\infty} h_j * u_{t-j}$, dove $h_0 = 1$ e $\sum_{j=0}^{\infty} |h_j| * j < \infty$, ciò implica che $\sum_{j=0}^{\infty} h_j^2 < \infty$.

Il polinomio $H(L)$ può essere anche scritto come $H(1) - \sum_{j=1}^{\infty} h_j + L * \sum_{j=1}^{\infty} h_j + \dots$
 $= H(1) - (1 - L) \sum_{j=1}^{\infty} h_j + \dots = H(1) - (1 - L) \sum_{j=0}^{\infty} (-\sum_{i=j+1}^{\infty} h_i) * L^j$, dove $h^*(j) = (-\sum_{i=j+1}^{\infty} h_i)$, vediamo quindi che $X_t = X_{t-1} + H(1)u_t + H^*(L) * u_t - H^*(L)u_{t-1}$, sostituendo in modo ricorsivo noto che $X_t = X_0 + H(1) * \sum_{j=1}^t u_j + H^*(L)u_t - H^*(L)u_0$, visto che $\sum_{j=0}^{\infty} (\sum_{i=j+1}^{\infty} h_i) < \infty$ allora $H^*(L) * u_t - H^*(L)u_{t-1} = \xi_t$ è un processo stazionario.

Quindi il processo attraverso questa scomposizione può essere scritto come

$X_t = X_0 + H(1) * \sum_{j=1}^t u_j + \xi_t$, e sono dette rispettivamente parte deterministica, parte permanente ed è la parte di non stazionarietà, notiamo che è un randomwalk, ed infine componente transitoria. La componente permanente indica l'effetto di uno shock sul lungo periodo.

Se invece voglio valutare l'effetto di uno shock avvenuto al tempo t nel tempo $t+k$ posso per semplicità porre $u_0 = 0$, calcolo $X_{t+k} = X_0 + H(1) * \sum_{j=1}^{t+k} u_j + H^*(L)u_{t+k}$ e faccio la derivata prima rispetto a u_t ottengo quindi $H(1) + h_k^*$, sapendo che $\sum_{j=0}^{\infty} |h_j| * j < \infty$ allora se $k \rightarrow \infty$ ho $h_k = 0$, quindi nel lungo periodo vale solo $H(1)$ detta persistenza.

Se voglio valutare invece l'effetto immediato dello shock valuterò la derivata prima nell'istante 0 ed

avrò $H(1) + h_0^*$ so che posso scriverlo come $\sum_{j=0}^{\infty} h_j - \sum_{j=1}^{\infty} h_j = 1$ perciò l'effetto contemporaneo è uguale alla perturbazione.

Un esempio valutabile è se X_t è un ARIMA(0,1,1) esso avrà quindi $\Delta X_t = u_t + \theta * u_{t-1} = (1 + \theta * l) * u_t = \Theta(L) * u_t$ visto che è un MA(1) allora $\Theta(L) = H(L)$ perciò $H(1) = 1 + \theta$ quindi l'effetto istantaneo $H(1) > 1$ se $\theta > 0$.

Un altro esempio è se X_t è un ARIMA(1,1,0) esso avrà quindi $\Delta X_t = u_t + \phi * \Delta X_{t-1}$ dove $|\phi| < 1$, ho che è un AR(1) perciò $\Delta X_t = \sum_{j=0}^{\infty} \phi^j * L^j * u_t = H(L)u_t$ se voglio $H(1)$ sapendo che $|\phi| < 1$ avrò $\frac{1}{1-\phi}$ quindi l'effetto istantaneo $H(1) > 1$ se $\phi > 0$.

Se X_t è un ARIMA(1,1,1) avrò che $\Delta X_t = u_t + \phi * \Delta X_{t-1} + \theta * u_{t-1}$ dove $|\phi| < 1$ e $|\theta| < 1$ ho che è una ARMA(1,1) che può essere anche scritto come $\Delta X_t = \frac{1+\theta}{1-\phi} * u_t = H(L)u_t$, e si trova facilmente $H(L)$, si può risolvere anche attraverso lo sviluppo in serie del MA, infine notiamo la persistenza da θ e ϕ .

1.2.3 Proprietà degli stimatori sui processi integrati di ordine 1

Non vale l'usuale teoria asintotica, esse non sono più normali quindi i valori critici usati nei test non sono più validi.

Sia AR(1) un $X_t = \phi * X_{t-1} + u_t$ e abbiamo n osservazioni temporali e che $X_0 = 0$ allora in questo caso lo stimatore dei minimi quadrati è $\hat{\phi}_n = \phi + \frac{\sum_{t=1}^n X_{t-1} * u_t}{\sum_{t=1}^n X_{t-1}^2}$ se $|\phi| < 1$, converge in maniera esponenziale $\hat{\phi}_n \rightarrow N(\phi, \frac{1-\phi^2}{n})$.

Se invece $|\phi| = 1$ è una variabile degenera dato che la varianza diventa 0 perciò è inutile per fare inferenza.

Se voglio che non degeneri devo testare la **super-consistenza**, cioè sostituire n^2 al posto di n infatti facendo ciò converge a velocità maggiore a 1, abbiamo $n(\hat{\phi}_n - \phi) \rightarrow v$ dove v è una distribuzione di Dickey-Fuller.

2 Test di radice unitaria

Capire se una serie è stazionaria o no e che tipo di non stazionarietà è importante anche dal punto di vista economico infatti gli shock dal lato della domanda e monetari hanno effetti transitori cioè una persistenza nulla, gli shock sulla domanda hanno effetti permanenti quindi una persistenza positiva.

Per capire il tipo di non stazionarietà abbiamo i test di radice unitaria detti di non stazionarietà, dei test di verifica di ipotesi che hanno come ipotesi nulla la presenza nella componente auto-regressiva di una radice unitaria, quindi la presenza di un trend stocastico.

2.1 Test di Dickey-Fuller

Per capire se un processo AR(1) si muove attorno ad una componente deterministica CD_t oppure ho una radice unitaria. Cioè $X_t - CD_t = \phi(X_{t-1} - CD_{t-1}) + u_t$, il test serve per verificare $H_0 = \phi = 1$ vs $H_1 = |\phi| < 1$ quindi che sia stazionario. Tale test ha diverse forme in base alla forma della componente deterministica.

2.1.1 Test con $CD_t = 0$

Quando $CD_t = 0$ e accetto H_0 il processo è un random-walk infatti ha forma $X_t = X_{t-1} + u_t$, se invece rifiuto H_0 ho un processo stazionario a media nulla, ciò tutta via è molto raro.

Quando andiamo a calcolare il test di D-F creiamo una regressione lineare su $X_t = \phi * X_{t-1} + u_t$. Oppure $X_t - X_{t-1} = \phi * X_{t-1} - X_{t-1} + u_t = \Delta X_t = (\phi - 1) * X_{t-1} + u_t$ se $\alpha^* = 1 - \phi$ questo porta ad $H_0 : \alpha^* = 0$ vs $H_1 : -2 < \alpha^* < 0$, questo a differenza di prima è un test ad una coda.

La statistica test utilizzata per questo caso è $\frac{\hat{\alpha}^*}{\hat{\sigma}}$ dove $\hat{\sigma}$ è l'errore standard, e sotto H_0 si dispone come una particolare Dickey-Fuller.

2.1.2 Test con $CD_t = K$

Quando $CD_t = K$ ho che $X_t - K = \phi(X_{t-1} - K) + u_t$ ho poi $X_t = S + \phi * X_{t-1} + u_t$ dove $S = k - \phi * k$, sotto H_0 abbiamo quindi con un random-walk infatti $\phi = 1$ ed ho $X_t = \phi * X_{t-1} + u_t$ se invece vale H_1 ho un AR(1) con media diversa da 0.

Anche in questo caso differenziamo e creiamo un modello di regressione del tipo $\Delta X_t = S + (\phi - 1)X_{t-1} + u_t$ cioè $\Delta X_t = S + \alpha^* X_{t-1} + u_t$.

La statistica test utilizzata è la stessa di prima, però la distribuzione sotto H_0 è diversa da quella di prima, ciò sottolinea l'importanza di CD_t .

2.1.3 Test con CD_t lineare

Quando $CD_t = S * t + K$ ho che $X_t - K - St = \phi(X_{t-1} - K + S(t-1)) + u_t$ ho poi $X_t = S^* + \gamma * t + \phi * X_{t-1} + u_t$ dove $\gamma = S(1 - \phi)$ e $S^* = K(1 - \phi) + S * \phi$. Sotto H_0 avrò un random walk con drift visto che $\gamma = 0$ e $S^* = S$, sotto H_1 avrò un processo stazionario attorno ad un processo deterministico.

Anche in questo caso creo $\Delta X_t = S^* + \gamma * t + (\phi - 1)X_{t-1} + u_t$ cioè $\Delta X_t = S^* + \gamma * t + \alpha^* X_{t-1} + u_t$. La statistica test utilizzata è uguale, però la distribuzione sotto H_0 è diversa da quella dei casi precedenti.

2.2 Test di D-F aumentato (ADF)

Spesso però la maggior parte delle serie storiche non sono descritte da degli AR(1), quindi il test capisce se è formata anche da un modello AR(P), invece l'altro test trasferisce l'autocorrelazione sui residui quindi portando dei problemi ad essi. Quindi andiamo ad aggiungere delle ulteriori differenze $\Delta X_{t-1}, \dots, \Delta X_{t-p}$.

Il modello AR(P) scritto come $\Phi(L)X_t = (1 - \phi_1 L - \dots - \phi_p L^p)X_t = u_t$, se $\rho = \phi_1 + \dots + \phi_p$ e $\zeta_j = -(\phi_{j+1} + \dots + \phi_p)$ si può dimostrare che $\Phi(L) = (1 - \rho L) - (\zeta_1 * L + \dots + \zeta_{p-1} * L^{p-1})(1 - L)$ da qua capiamo che $\Delta X_t = X_{t-1}(\rho - 1) + \zeta_1 * \Delta X_{t-1} + \dots + \zeta_{p-1} * \Delta X_{t-p+1} + u_t$ quindi ho $\rho^* = \rho - 1$, oppure vediamo $\rho^* = -(1 - \phi_1 + \dots + \phi_p)$.

Il test di ipotesi avrà $H_0 = \rho^* = 0$ si imposta la statistica test allo stesso modo di D-F $\frac{\hat{\rho}}{\hat{\sigma}}$.

Anche per il test ADF distinguiamo i 3 casi in base alla forma della componente deterministica, ed avrà anche come prima tre distribuzioni diverse e si disporrà come loro, inoltre per verificarlo useremo sempre un modello di regressione, inoltre avendo i parametri ζ posso verificare la loro significatività attraverso un test T.

3 Serie storiche multivariate

Una serie storica multivariate, a K dimensioni, è formata da K serie storiche univariate ciascuno indicizzata rispetto al tempo t.

Quindi $X_1 \dots X_n$ è la realizzazione finita di un processo stocastico multivariato con X_t l'osservazione ed esso ha k dimensioni, ogni dimensione corrisponde ad una variabile casuale che può essere trattata come un processo univariato, inoltre le variabili sono correlate tra di loro.

Essendo variabili aleatorie hanno un valore atteso $E(X_{it}) = \mu_{it}$ e funzione di covarianza $\gamma_{is}(t, t-j) = E((X_{it} - \mu_{it})(X_{st-j} - \mu_{st-j}))$. Ovviamente ciò si può raggruppare in vettori così da avere $E(X_t) = \mu_t$ e matrice di autocovarianza $\Gamma(t, t-j)$ se $J=0$ diventa matrice di varianza e covarianza.

Il processo **White noise** multivariato è un processo stocastico costituito da una successioni di variabili casuali k-dimensionali fra loro incorrelate quando sono riferite ad istanti diversi cioè $E(\mu_t) = 0$ e $\Gamma(j) = E(\mu_t * \mu'_{t-j})$, se $J=0$ avrò una matrice di varianza-covarianza Σ e $J > 0$ avrò matrice di soli 0.

Nel caso multivariato si hanno le seguenti **proprietà**:

- **Stazionarietà forte** per un processo stocastico se la funzione di ripartizione K-variata è invariante alle traslazioni temporali.
- **Stazionarietà debole** se il processo stocastico è per ogni istante di t il vettore $E(X_t) = \mu_t = \mu$ cioè non dipende dal tempo, e $\Gamma(t, t-j) = \Gamma(j)$ cioè l'autocovarianza dipende solo dallo sfasamento temporale, inoltre $\Gamma(j)$ non è simmetrica.
- **Ergodicità** se la funzione $g(X_t)$ converge in probabilità a $E(g(X_t))$.
- **Invertibile** se un elemento del processo può essere espresso in funzione del suo passato e di una componente aleatoria contemporanea incorrelata con il passato della serie (White noise). Si potrà scrivere come $X_t = f(X_{t-1} \dots, X_{t-n}, u_t)$.

3.1 Modello VARMA(P,Q)

Sia X_t vettore $X_t = c + \Phi_1 * X_{t-1} + \dots + \Phi_p * X_{t-p} + u_t + \Theta_1 * u_{t-1} + \dots + \Theta_q * u_{t-q}$ dove c è un vettore di costanti, Φ è la matrice quadrata dei componenti autoregressivi e Θ matrice quadrata dei componenti a media mobile.

Se X_t è debolmente stazionario allora il suo valore atteso sarà quindi $E(X_t) = c + \Phi_1 * E(X_{t-1}) + \dots + \Phi_p * E(X_{t-p})$ visto che il white noise ha valore atteso 0, inoltre tutti gli $E(X_{t-j})$ sono costanti vista la stazionarietà, quindi ho $\mu = c + \Phi_1 * \mu + \dots + \Phi_p * \mu$, ottengo che $\mu(1 - \Phi_1 - \dots - \Phi_p) = c$ se la matrice $(1 - \Phi_1 - \dots - \Phi_p)$ è invertibile allora $\mu = (1 - \Phi_1 - \dots - \Phi_p)^{-1} * c$ inoltre se $c=0$ avrò sicuramente che $\mu = 0$, si parla di processo a media nulla $X_t - \mu$ e sottraggo μ in tutti le variabili autoregressive, ottengo $X_t - \mu = c + \Phi_1 * (X_{t-1} - \mu) + \dots + \Phi_p * (X_{t-p} - \mu) + u_t + \Theta_1 * u_{t-1} + \dots + \Theta_q * u_{t-q}$, noto che il modello originario e quello con scarti dalla media hanno gli stessi coefficienti.

Per esempio un VARMA(1,1) è rappresentato come $X_t = c + \Phi_1 * X_{t-1} + u_t + \Theta_1 * u_{t-1}$, quindi ho una moltiplicazione matriciale.

Ora vediamo come un VARMA(P,Q) e lo scriviamo come polinomi matriciali nell'operatore ritardo L , ho quindi il processo a media nulla $\Phi(L)(X_t - \mu) = \Theta(L)u_t$, dove $(I - \Phi_1 L - \dots - \Phi_p L^p)(X_t - \mu) = (I + \Theta_1 L + \dots + \Theta_q L^q)u_t$, noto che se $p=0$ non esiste la parte autoregressiva quindi il processo è solo a media mobile e si indica come VMA(q), se $q=0$ non c'è la parte a media mobile ed ho un VAR(p).

Se ho un VAR(1) si scrive come $X_t = \Phi_1 * X_{t-1} + u_t$, inoltre se ho un VAR(p) posso sempre ricondurre a un VAR(1), ciò è dimostrabile considerando un sistema di equazioni con $\mu = 0$.

$$\begin{cases} X_t = \Phi_1 * X_{t-1} + \dots + \Phi_p * X_{t-p} + u_t \\ X_{t-1} = X_{t-1} \\ \dots \\ X_{t-p+1} = X_{t-p+1} \end{cases} \quad \text{Notiamo quindi che abbiamo aggiunto solo delle identità.}$$

$$\text{Risolvo tale sistema attraverso il calcolo matriciale come } \begin{bmatrix} X_t \\ X_{t-1} \\ \dots \\ X_{t-p+1} \end{bmatrix} = \begin{bmatrix} \Phi_1 & \dots & \Phi_p \\ I_k & 0 & 0 \\ 0 & I_k & 0 \\ 0 & 0 & I_k \end{bmatrix} * \begin{bmatrix} X_{t-1} \\ X_{t-2} \\ \dots \\ X_{t-p} \end{bmatrix} + \begin{bmatrix} u_t \\ 0 \\ \dots \\ 0 \end{bmatrix}$$

Le matrici o vettori analizzati hanno dimensione $K * p$, infatti ogni X_t ha dimensione p . Tale scrittura può essere alleggerita scrivendo $\xi_t = F * \xi_{t-1} + v_t$, vediamo quindi che è un VAR(1) quindi le proprietà di esso possono essere applicate ad un modello di ordine superiore riscrivendolo in questa forma detta **compagna**.

4 Stazionarietà dei VAR

Sia VAR(1) K-dimensionale e media nulla perciò ho $X_t = \Phi * X_{t-1} + u_t$ scritto anche come $(I_k - \Phi * L)X_t = u_t$.

Supponiamo ora che tutti gli autovalori della matrice Φ sono in modulo minori di 1, sappiamo anche che gli autovalori corrispondono ai reciproci delle radici dell'equazione caratteristica in z, ciò implica che tutte le radici siano in modulo minore di 1 e quindi stazionarietà.

Se il VAR(1) è stazionario vale che $X_t = (I_k - \Phi * L)^{-1}u_t = (I_k + \Phi * L + \dots \Phi^p * L^p) = \sum_{j=0}^{\infty} \Phi^j * u_{t-j}$ questa è la rappresentazione VMA(∞) del VAR(1) ed essa vale solo se:

- Il $\lim_{j \rightarrow \infty} \Phi^j = 0$ una matrice di 0.
- $\sum_{j=0}^{\infty} \Phi^j = (I_k - \Phi)^{-1}$ ed esiste finita.

Ciò si può dimostrare, per facilità supponiamo che Φ è diagonalizzabile, quindi scriviamo che $\Phi = P * D * P^{-1}$, dove D è una matrice diagonale con autovalori di Φ sulla diagonale principale.

Dimostriamo la condizione 1, notiamo che se eleviamo alla j Φ otteniamo $\Phi^j = (P * D * P^{-1})^j = P * D^j * P^{-1}$ di conseguenza ho che D^j è elevare alla j gli elementi diagonali, quindi se gli autovalori sono minori di 1 in modulo ho che $\lim_{j \rightarrow \infty} \Phi^j = 0$ e moltiplicata per $P * D^j * P^{-1}$ da matrice nulla.

Dimostriamo la condizione 2, so che $\sum_{j=0}^{\infty} \Phi^j = I_k + \Phi + \dots \Phi^m$ inoltre $(I_k + \Phi + \dots \Phi^m)(I_k - \Phi) = I_k - \Phi^{m+1}$ o scritta come $(I_k + \Phi + \dots \Phi^m) = (I_k - \Phi^{m+1})(I_k - \Phi)^{-1}$, sappiamo che se gli autovalori sono minori in modulo di uno, ho $\lim_{j \rightarrow \infty} \Phi^j = 0$ perciò $(I_k + \Phi + \dots \Phi^m) = (I_k - \Phi)^{-1}$ visto che è una moltiplicazione per I_k .

4.1 Teorema di stazionarietà

Se gli autovalori della matrice Φ sono interni al cerchio di raggio unitario, allora il processo è debolmente stazionario, ciò vale per un VAR(1).

Sapendo che il VAR(p) può essere scritto in forma compagna come un VAR(1) allora tale teorema si può estendere a tutti i VAR. **Corollario** Una condizione affinché un processo X_t sia stazionario è che gli autovalori della matrice F siano interni al cerchio di raggio unitario, cioè $\xi_t = \sum_{j=0}^{\infty} F^j * v_{t-j}$ questa è la rappresentazione VMA(∞) di un VAR(1).

Ora vediamo la rappresentazione di un VAR(p), K-dimensionale a media nulla, in forma di VMA(∞). Sia $(I_K - \Phi * L - \dots \Phi^p * L^p)X_t = \Phi(L)X_t = u_t$ per scriverlo in forma di VMA devo ottenere l'espressione di $X_t = \Theta(L)u_t = (I_K + \Theta * L + \dots)u_t$ vediamo che $\Phi(L)^{-1} = \Theta(L)$.

Posso ottenere le Θ_j in maniera ricorsiva da $(I_K + \Theta * L + \dots)(I_K - \Phi * L - \dots \Phi^p * L^p) = I_k$, moltiplicando vado ad elidere gli I_k ottenendo $(\Theta_1 - \Phi_1)L + (\Theta_2 - \Phi_2 - \Phi_1 * \Theta_1)L^2 + \dots = 0$, ma io so che $(\Theta_1 - \Phi_1) = 0$ quindi $\Theta_1 = \Phi_1$ e $(\Theta_2 - \Phi_2 - \Phi_1 * \Theta_1) = 0$ trovo che $\Theta_2 = \Phi_1^2 + \Phi_2$, facendo così in modo ricorsivo posso trovare la formula $\Theta_s = \Phi_1 * \Theta_{s-1} + \dots + \Phi_p * \Theta_{s-p}$.

5 Stima di un modello VAR(p)

La stima avviene tramite la stima di massima verosimiglianza. Data la correlazione tra le variabili del processo occorre utilizzare la verosimiglianza condizionata.

Si consideri una serie composta da $p+n$ vettori cioè da $X_{t-p+1}, \dots, X_t, \dots, X_n$, dove i primi p sono i valori iniziali e gli altri n sono usati per il calcolo della verosimiglianza.

La funzione di verosimiglianza condizionata $L^c(c, \Phi_1, \dots, \Phi_p, \Sigma) = \prod_{i=1}^n f_{x_{t+i}|x_{t-p+i}}$, e noi vogliamo stimare i parametri incogniti cioè una matrice di dimensione $K^*(K^*P+1)$. La stima ottenuta con il metodo della massima verosimiglianza è la stessa di quella ottenuta con il metodo dei minimi quadrati ordinari.

Consideriamo un VAR(1) a media nulla come $\begin{bmatrix} X_{t,1} \\ X_{t,2} \end{bmatrix} = \begin{bmatrix} \Phi_{1,1} & \Phi_{1,2} \\ \Phi_{2,1} & \Phi_{2,2} \end{bmatrix} * \begin{bmatrix} X_{t-1,1} \\ X_{t-1,2} \end{bmatrix} + \begin{bmatrix} u_{t,1} \\ u_{t,2} \end{bmatrix}$

Otteniamo che $\begin{bmatrix} X_{t,1} = \Phi_{1,1} * X_{t-1,1} + \Phi_{1,2} * X_{t-1,2} + u_{t,1} \\ X_{t,2} = \Phi_{2,1} * X_{t-1,1} + \Phi_{2,2} * X_{t-1,2} + u_{t,2} \end{bmatrix}$. Abbiamo n osservazioni per ogni X_i ,

avremo: $\begin{bmatrix} X_{1,2} \\ \dots \\ X_{1,n} \end{bmatrix} = \begin{bmatrix} X_{1,1} & X_{2,1} \\ \dots & \dots \\ X_{1,n-1} & X_{2,n-1} \end{bmatrix} * \begin{bmatrix} \phi_{1,1} \\ \phi_{1,2} \end{bmatrix} + \begin{bmatrix} u_{1,2} \\ \dots \\ u_{1,n} \end{bmatrix}$ e così anche per il termine X_2 , il numero di

valori da prendere se è un VAR(p) è $n-p$. In forma compatta $X_1 = X * \Phi_1 + u_1$

Perciò per trovare il vettore usando il metodo dei minimi quadrati ho $\Phi_1 = (X'X)^{-1} * X' * x_1$ e idem per Φ_2 .

5.1 Stima del numero di ritardi

La scelta dei ritardi per un test determina anche la dimensione della matrice dei Φ per capirlo usiamo due metodi:

- **Test ad esclusione ricorsiva dei ritardi** è basato sul rapporto di verosimiglianza. Ha ipotesi nulla H_0 il processo generatore dei dati è un $VAR(p_0)$, contro l'ipotesi alternativa dove ho $VAR(p_1)$ con $p_1 > p_0$.

Il test confronta il valore massimo della verosimiglianza sotto H_0 con il valore massimo di essa sotto H_1 , come $\Lambda = \frac{L^c(\Pi_0, \Sigma_0)}{L^c(\Pi_1, \Sigma_1)}$ o usiamo $w = -2\log(\Lambda)$ grazie al teorema di Wilks so che w sotto H_0 di distribuisce come una $\chi_{p_1-p_0}^2$, dove $p_1 - p_0$ è il numero di elementi vincolati, quindi avrò una matrice di dimensione $K^2 * (p_1 - p_0)$ con solo 0. Quindi si rifiuta H_0 quando w è maggiore di un quantile 1- α di una chi quadro con $p_1 - p_0$ gradi di libertà.

- **Test secondo criteri** sono di Akaike, Hannan e Quinn, Schwartz. Essi consentono di individuare il numero dei ritardi minimizzando una funzione, essa è crescente rispetto al numero dei parametri e decrescente rispetto al valore della log-verosimiglianza.

– **AIC(p)** = $\frac{-2l(\theta, p)}{n} + \frac{2pk^2}{n}$ dove p numero ritardi, k numero delle variabili.

– **HQ(p)** = $\frac{-2l(\theta, p)}{n} + \frac{2\log(\log(n))pk^2}{n}$ penalizza di più di prima il numero di variabili.

– **SC(p)** = $\frac{-2l(\theta, p)}{n} + \frac{2\log(n)pk^2}{n}$ massima penalizzazione delle variabili.

6 Applicazioni del VAR

6.1 Previsione con un modello VAR(P)

Si vuole ottenere la previsione di X_{t+1} avendo a disposizione tutte le informazioni passate fino a t , I_t . Quindi dobbiamo prevedere $X_{t+1}|I_t$ ed il miglior previsore è quello che minimizza l'errore quadratico medio della previsione cioè $MSE(\hat{X}_{t+1}|I_t) = E((X_{t+1} - \hat{X})(X_{t+1} - \hat{X})')$.

Il miglior previsore è $\hat{X}_{t+1} = E(X_{t+1}|I_t)$ (come se fosse una martingala). Avrò quindi $\hat{X}_{t+1} = c + \Phi_1 * x_t + \dots + \Phi_p * x_{t-p+1}$.

Se invece voglio prevedere X_{t+h} devo prevederlo sempre con le informazioni I_t che avevo prima, si ottiene come $\hat{X}_{t+h} = E(X_{t+h}|\hat{X}_{t+h}, \dots, I_t) = E(X_{t+h}|I_t) = \hat{X}_{t+h} = c + \Phi_1 * \hat{X}_{t+h-1} + \dots + \Phi_p * \hat{X}_{t+h-p}$.

L'errore di previsione sarà dato da $X_{t+h} - \hat{X}_{t+h}$, se usiamo la rappresentazione $VMA(\infty)$ visto che VAR(p) stazionario, posso scrivere $X_{t+h} = u_{t+h} + \Theta_1 * u_{t+h-1} + \dots + \Theta_{h+1} * u_{t-1} \dots$, se io voglio $\hat{X}_{t+h} = E(X_{t+h}|I_t)$ che vedendolo come un VMA ho $E(u_{t+h}) + \Theta_1 * E(u_{t+h-1}) + \dots + \Theta_{h+1} * u_{t-1}$, non uso più il valore atteso perchè conosco quei valori, notiamo quindi che i valori futuri sono tutti uguali al loro valore atteso quindi 0, perciò vediamo che $\hat{X}_{t+h} = \Theta_h * u_t + \dots + \Theta_{h+p} * u_{t-p}$, quindi l'errore di previsione è $u_{t+h} + \Theta_1 * u_{t+h-1} \dots + \Theta_{h-1} * u_{t+1}$, più aumenta il periodo di previsione più l'errore aumenta.

Il suo MSE sarà $(\sum_{j=0}^{h-1} \Theta_j * \Sigma * \Theta_j')$.

6.2 Analisi dinamica

Lo strumento che si utilizza per l'analisi dinamica di K variabili nel tempo è **funzione degli impulsi**. Questa funzione misura la reazione del sistema a shock improvvisi rispetto ad ognuna delle variabili. Se il VAR(p) è stazionario, allora vale il $VMA(\infty)$ quindi si scrive come $X_t = u_t + \Theta_1 * u_{t-1} + \dots$, valutiamo l'effetto di una variazione unitaria del s-esimo u al tempo t, sull'r-esimo elemento di X_{t+j} , ed è pari a $\frac{\delta X_{r,t+j}}{\delta u_{s,t}} = \gamma_{rs,j}$ è l'elemento in posizione (r,s) della matrice Θ_j .

Questo dice che esistono K^2 risposte perchè ho risposte (r,s) e (s,r) per ogni tempo.

Ora valutiamo l'effetto contemporaneo di uno shock cioè se $j = 0$, quindi $\Theta_0 = I_k$, quindi se $r=s$ allora vale 1 se invece $r \neq s$ ho che è 0 l'effetto.

6.3 Non casualità di Granger bivariata

Dice se uno scalare Y può aiutare a prevedere uno scalare X, se non può si dice che Y non causa nel senso di Granger X, cioè se per tutti gli $h > 0$, se $MSE(\hat{X}_{t+h}|X_t, \dots, X_1)$ è lo stesso di $MSE(\hat{X}_{t+h}|X_t, \dots, X_1, Y_t, \dots, Y_1)$. Se un evento Y causa un altro evento X allora l'evento Y deve precedere temporalmente l'evento X, ed esclude la casualità contemporanea.

Se $MSE(\hat{X}_{t+h}|I_t, Y_t, \dots, Y_1) < MSE(\hat{X}_{t+h}|I_t)$ allora posso dire che Y contribuisce a prevedere X.

Consideriamo un VAR(2), con K=2 e con media nulla.

$$\begin{bmatrix} X_{1,t} \\ X_{2,t} \end{bmatrix} = \begin{bmatrix} \Phi_{1,1,1} & 0 \\ \Phi_{2,1,1} & \Phi_{2,2,1} \end{bmatrix} * \begin{bmatrix} X_{1,t-1} \\ X_{2,t-1} \end{bmatrix} + \begin{bmatrix} \Phi_{1,1,2} & 0 \\ \Phi_{2,1,2} & \Phi_{2,2,2} \end{bmatrix} * \begin{bmatrix} X_{1,t-2} \\ X_{2,t-2} \end{bmatrix} + \begin{bmatrix} u_{1,t} \\ u_{2,t} \end{bmatrix}$$
 Se prendo la prima riga ottengo che $X_{1,t} = \Phi_{1,1,1} * X_{1,t-1} + \Phi_{1,1,2} * X_{1,t-2} + u_{1,t}$ vedo subito che X_2 non causa in senso di Granger visto che ho 0 nell'elemento (1,2) della matrice.

In generale posso dire che X_2 non causa X_1 se ogni matrice Φ_j è una matrice triangolare inferiore, questa è una condizione sufficiente.

La stessa condizione vale se tale processo è rappresentato in forma $VMA(\infty)$, si rappresenta come:

$$\begin{bmatrix} X_{1,t} \\ X_{2,t} \end{bmatrix} = \begin{bmatrix} \Theta_{1,1,1} & 0 \\ \Theta_{2,1,1} & \Theta_{2,2,1} \end{bmatrix} * \begin{bmatrix} u_{1,t-1} \\ u_{2,t-1} \end{bmatrix} + \dots + \begin{bmatrix} u_{1,t} \\ u_{2,t} \end{bmatrix}$$
 Vedo che la matrice è triangolare inferiore e lo è per ogni j, quindi X_2 non causa X_1 .

6.3.1 Test di non causalità di Granger

Dato un VAR(p) con K=2 dimensioni, per verificare se X_2 non causa secondo Granger X_1 si effettua un test basato sul confronto tra un modello non vincolato (M1) con equazione $X_{1,t} = \Phi_{1,1,1} * X_{1,t-1} + \Phi_{1,2,1} * X_{2,t-1} + \dots + \Phi_{1,1,p} * X_{1,t-p} + \Phi_{1,2,p} * X_{2,t-p} + u_{1,t}$, e un modello con vincoli (M0) ed ha la stessa forma di prima solo che $\Phi_{1,2,1} = \dots = \Phi_{1,2,p} = 0$, quindi ho $H_0 : X_2$ non causa secondo Granger cioè equivale a dire che M0 è giusto, contro l'ipotesi alternativa H_1 esiste almeno un j tale che $\Phi_{1,2,j} \neq 0$.

Si effettua un test F basato sulla statistica test $S = \frac{RSS_0 - RSS_1}{\frac{p}{n-2p-1}}$, dove RSS somma dei residui di regressione al quadrato. Sotto H_0 la statistica test S si distribuisce come una $F_{p,n-2p-1}$ dove p e n-2p-1 sono i gradi di libertà, ed avrà regione di rifiuto i valori particolarmente alti di S. Se rifiuto H_0 allora X_2 causa secondo Granger X_1 .

7 Cointegrazione

Sia VAR(1), bivariato e stazionario. Lo esprimiamo in forma **strutturale**, essa considera sia le relazioni contemporanee con le variabili sia le relazioni con le variabili ritardate. Si indica come :

$$\begin{bmatrix} X_{1,t} \\ X_{2,t} \end{bmatrix} = \begin{bmatrix} b_{1,0} \\ b_{2,0} \end{bmatrix} - \begin{bmatrix} b_{1,2} \\ b_{2,1} \end{bmatrix} * \begin{bmatrix} X_{1,t-1} \\ X_{2,t-1} \end{bmatrix} + \begin{bmatrix} \gamma_{1,1} \\ \gamma_{2,1} \end{bmatrix} * \begin{bmatrix} X_{1,t-1} \\ X_{2,t-1} \end{bmatrix} + \begin{bmatrix} \gamma_{1,2} \\ \gamma_{2,2} \end{bmatrix} * \begin{bmatrix} X_{1,t-1} \\ X_{2,t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{bmatrix}.$$

In notazione matriciale ho:

$$\begin{bmatrix} 1 & b_{1,2} \\ b_{2,1} & 1 \end{bmatrix} * \begin{bmatrix} X_{1,t} \\ X_{2,t} \end{bmatrix} = \begin{bmatrix} b_{1,0} \\ b_{2,0} \end{bmatrix} + \begin{bmatrix} \gamma_{1,1} & \gamma_{1,2} \\ \gamma_{2,1} & \gamma_{2,2} \end{bmatrix} * \begin{bmatrix} X_{1,t-1} \\ X_{2,t-1} \end{bmatrix} + \dots + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{bmatrix}.$$

Si passa dalla rappresentazione in forma strutturale alla ridotta mettendo la matrice inversa di B, avrò quindi $B^{-1} * B * X_t = B^{-1} * \Gamma_0 + B^{-1} * \Gamma_1 * X_{t-1} + B^{-1} * \epsilon_t$ si trasforma in $X_t = c + \Phi_1 * X_{t-1} + u_t$. Notiamo che il nostro VAR(1) ci fa stimare $K + K^2$ parametri la forma strutturale, molti di più.

Dato un sistema $\begin{bmatrix} X_{1,t} \\ X_{2,t} \end{bmatrix} = \gamma * \begin{bmatrix} X_{1,t} \\ X_{2,t} \end{bmatrix} + u_{1,t}$.

Se prendiamo la seconda equazione notiamo che è la rappresentazione di un random walk senza drift, quindi è un processo integrato di ordine 1.

Se prendiamo la prima equazione noto che è la somma di un processo non stazionario e stazionario, e da come risultato uno non stazionario I(1).

Ora se consideriamo $Z_t = X_{1,t} - \gamma * X_{2,t} = u_{1,t}$, la loro differenza da un processo stazionario.

In generale vale che se X_t è un processo $I(R)$ e Y_t è un processo $I(s)$ allora $Z_t = X_t + \pi * Y_t$ tale è un processo $I(c)$ dove $c = \max(r, s)$ se $r \neq s$ oppure $c \leq \max(r, s)$ se $r = s$. Se $c < \max(r, s)$ allora il processo è detto **cointegrato** di ordine r-c.

Se consideriamo X_t vettore k-variato, esso si dice cointegrato se sussistono due condizione:

- Ogni singola componente di X_t è un processo integrato di ordine 1.
- Esiste almeno un vettore a k-dimensionale, tale che la serie storica risultante tra $a' * X_t$, crea una serie storica stazionaria, ed a' è detto vettore di cointegrazione.

Possono esistere al massimo $h < k$ vettori di cointegrazione linearmente indipendenti.

I quali inclusi in una matrice $A(K * h)$ formano la matrice di cointegrazione, il rango di A sarà h ed è detto rango di cointegrazione.

Teorema Se A' è formato da h vettori riga linearmente indipendenti allora $A' * X_t$ è un processo stazionario h-dimensionale. Preso un'altro vettore c linearmente indipendente dalle righe di A' allora il prodotto tra $c' * X_t$ è un processo non stazionario, ciò ci dice che esistono solo h relazioni di cointegrazione tra le variabili incluse in X_t .

7.1 Modello correzione d'errore

Si applica ad un sistema cointegrato, per trattarlo in passato facevamo una differenza prima però ciò andava a perdere informazioni della serie. Ecco che si usa questo modello ed è utile per valutare le relazioni di lungo periodo tra le variabili, e per superare i problemi di stima nella non stazionarietà.

Consideriamo un VAR(1) con K=2 variabili ha struttura in forma estesa nota, $\begin{bmatrix} X_{1,t} = \gamma_{1,2} * X_{2,t} + u_{1,t} \\ X_{2,t} = X_{2,t-1} + u_{2,t} \end{bmatrix}$.

Noto che sono due processi integrati di ordine 1.

In forma strutturale matriciale ho: $\begin{bmatrix} 1 & -\gamma \\ 0 & 1 \end{bmatrix} * \begin{bmatrix} X_{1,t} \\ X_{2,t} \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} * \begin{bmatrix} X_{1,t-1} \\ X_{2,t-1} \end{bmatrix} + \begin{bmatrix} u_{1,t} \\ u_{2,t} \end{bmatrix}$.

Ora vogliamo averlo in forma ridotta quindi trovo $B^{-1} = \begin{bmatrix} 1 & \gamma \\ 0 & 1 \end{bmatrix}$, perciò VAR(1) in forma ridotta è:

$\begin{bmatrix} X_{1,t} \\ X_{2,t} \end{bmatrix} = \begin{bmatrix} 0 & \gamma \\ 0 & 1 \end{bmatrix} * \begin{bmatrix} X_{1,t-1} \\ X_{2,t-1} \end{bmatrix} + \begin{bmatrix} u_{1,t} + \gamma * u_{2,t} \\ u_{2,t} \end{bmatrix}$. Se sottraiamo X_{t-1} dai membri dell'equazione ho $\begin{bmatrix} \Delta X_{1,t} \\ \Delta X_{2,t} \end{bmatrix} = \begin{bmatrix} -1 & \gamma \\ 0 & 0 \end{bmatrix} * \begin{bmatrix} X_{1,t-1} \\ X_{2,t-1} \end{bmatrix} + \begin{bmatrix} u_{1,t} + \gamma * u_{2,t} \\ u_{2,t} \end{bmatrix}$. Quindi abbiamo un modello valutato nelle dif-

ferenze prime. Si può decomporre come: $\begin{bmatrix} \Delta X_{1,t} \\ \Delta X_{2,t} \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \end{bmatrix} * [X_{1,t-1} - \gamma * X_{2,t-1}] + \begin{bmatrix} u_{1,t} + \gamma * u_{2,t} \\ u_{2,t} \end{bmatrix}$.

Questo modello è indicato come **rappresentazione a correzione d'errore**.

Vediamo quindi che abbiamo un punto di equilibrio se $\begin{bmatrix} \Delta X_{1,t} = 0 \\ \Delta X_{2,t} = 0 \end{bmatrix}$. Chiamiamo Z_{t-1} la differenza, detto grado di squilibrio, se esso vale 0 allora ho un equilibrio, cioè se $X_{1,t-1} = \gamma * X_{2,t-1}$. Z_{t-1} per definizione di prima è una serie stazionaria del tipo white-noise, ed è l'errore che separa le due variabili dalla condizione di equilibrio. Il vettore $\begin{bmatrix} -1 \\ 0 \end{bmatrix}$ indica il peso del disequilibrio che viene corretta su X_t .

La matrice $\pi = \begin{bmatrix} -1 & \gamma \\ 0 & 0 \end{bmatrix}$ è detta matrice di cointegrazione (diversa da A che è matrice dei vettori di cointegrazione). Essa ha determinante uguale a 0, quindi ha rango 1. Se immaginiamo che le equazioni sono $\begin{bmatrix} X_{1,t} = X_{1,t-1} + u_{1,t} \\ X_{2,t} = X_{2,t-1} + u_{2,t} \end{bmatrix}$, due random-walk in questo caso ho $\pi = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$, ho rango 0. Quindi $X_{1,t}$ e $X_{2,t}$ sono si dello stesso tipo, ma non hanno alcun vettore di cointegrazione.

Consideriamo ora $\begin{bmatrix} X_{1,t} = \rho * X_{1,t-1} + u_{1,t} \\ X_{2,t} = \theta * X_{2,t-1} + u_{2,t} \end{bmatrix}$ dove i due coefficienti sono in valore assoluti minori di uno, tali equazioni sono due AR(1). Quindi la matrice $\pi = \begin{bmatrix} \rho - 1 & 0 \\ 0 & \theta - 1 \end{bmatrix}$, tale matrice ha rango 2.

Possiamo generalizzare e dire che $\Delta X_t = \pi * X_{t-1} + u_t$ se il rango di π è:

- 0 allora è composto da random walk non cointegrati.
- k allora è un processo multivariato stazionario.
- $0 < h < k$ allora è X_t è un sistema cointegrato.

Se ci focalizziamo sul caso 3, π può essere scomposta nel prodotto di due matrici, $B * A'$ ($K * h$, $h * k$), vediamo quindi che la matrice dei vettori di cointegrazione è legata con la matrice di cointegrazione. Nel caso di prima $A' = [1 - \gamma]$, visto che $R(\pi) = 1$ quindi esisteva una relazione di cointegrazione tra le variabili.

Il modello ECM $\Delta X_t = \pi * X_{t-1} + u_t$ è funzione lineare di variabili non stazionarie X_{t-1} e di u_t , ma è stazionario visto che $\Delta X_t = B * A' * X_{t-1} + u_t$ dove $A' * X_{t-1} = Z_{t-1}$ è stazionaria, cioè è possibile visto che X_{t-1} sono cointegrate e gli elementi di A' sono i coefficienti di cointegrazione.

Dalla teoria economica si sa che c'è una relazione di lungo periodo in base alla quale il livello di $X_{1,t}$ è proporzionale a quello di $X_{2,t}$, tale relazione però non è esatta e quindi per ogni t si definisce un errore stazionario, ed è dato da $Z_t = X_{1,t} - \gamma * X_{2,t} = A' * X_t$ quindi come detto prima Z_t è il disequilibrio al tempo t.

Ora ci focalizziamo sulla matrice o vettore degli aggiustamenti $B = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$ che forma la matrice π , consid-

eriamo $\begin{bmatrix} \Delta X_{1,t} = b_1 * (X_{1,t-1} - \gamma * X_{2,t-1}) + (u_{1,t} + \gamma * u_{2,t}) \\ \Delta X_{2,t} = b_2 * (X_{1,t-1} - \gamma * X_{2,t-1}) + u_{2,t} \end{bmatrix}$ quello che discrimina queste equazioni è il coefficiente b , essi misurano l'importanza relativa del disequilibrio al tempo $t-1$ per $X_{1,t}, X_{2,t}$.

Nel nostro esempio abbiamo $B = \begin{bmatrix} -1 \\ 0 \end{bmatrix}$ quindi otteniamo $\begin{bmatrix} \Delta X_{1,t} = -(X_{1,t-1} - \gamma * X_{2,t-1}) + (u_{1,t} + \gamma * u_{2,t}) \\ \Delta X_{2,t} = u_{2,t} \end{bmatrix}$,

perciò nell'ultima equazione non abbiamo alcuna importanza relativa. Inoltre vedo che il valore atteso condizionato ad I_{t-1} è $E(\Delta X_{1,t}|I_{t-1}) = -Z_{t-1}$, $E(\Delta X_{2,t}|I_{t-1}) = 0$.

Da qua abbiamo 3 casi possibili:

- Se al tempo $t-1$ non si è realizzato un disequilibrio allora $E(\Delta X_{1,t}|I_{t-1}) = 0$, visto che il disequilibrio è nullo.
- Se al tempo $t-1$ si è realizzato un disequilibrio positivo avrò che $E(\Delta X_{1,t}|I_{t-1}) = \text{variazione attesa negativa}$. Ciò ci dice che se c'è un disequilibrio positivo il sistema tende a ritornare all'equilibrio attraverso una variazione negativa.
- Se al tempo $t-1$ si è realizzato un disequilibrio negativo avrò che $E(\Delta X_{1,t}|I_{t-1}) = \text{variazione attesa positiva}$. Ciò ci dice che se c'è un disequilibrio negativo il sistema tende a ritornare all'equilibrio attraverso una variazione positiva.

Notiamo quindi che b_1 esprime velocità in cui avviene l'avvicinamento all'equilibrio, e visto che vale -1 è corretto in media in un solo periodo, se fosse stato minore del modulo di 1 i periodi aumenterebbero.

Concludiamo che un VAR(1) k -dimensionale con $h < k$ (relazione di cointegrazione) ammette la rappresentazione a correzione d'errore del tipo $\Delta X_t = B * A' * X_{t-1} + u_t$.

Si può generalizzare al caso di un VAR(p) k -dimensionale con $h < k$ (relazione di cointegrazione), scriviamo il modello $X_t = c + \Phi_1 * X_{t-1} + \dots + \Phi_p * X_{t-p} + u_t$.

Definiamo $\zeta_s = -(\Phi_{s+1} + \dots + \Phi_p)$ dove s va da 1 a $p-1$, $\zeta_0 = -(I_k - \Phi_1 - \dots - \Phi_p) = B * A'$, sottraggo X_{t-1} dal modello e vedo che $\Delta X_t = \zeta_1 * \Delta X_{t-1} + \dots + \zeta_{p-1} * \Delta X_{t-p+1} + c + \zeta_0 * X_{t-1} + u_t$.

Visto che $Z_t = A' * X_t$ è un processo stazionario h -dimensionale, allora la rappresentazione a correzione d'errore del sistema cointegrato diventa: $\Delta X_t = (\zeta_1 * \Delta X_{t-1} + \dots + \zeta_{p-1} * \Delta X_{t-p+1}) + c + B * Z_{t-1} + u_t$, dove $B * Z_{t-1}$ descrive i comportamenti di lungo periodo, mentre l'altra componente tra parentesi descrive i comportamenti di breve periodo.

Se un sistema cointegrato ha h relazioni di cointegrazione allora esistono h relazioni di lungo periodo e quindi h processi stazionari che descrivono l'andamento nel tempo dei disequilibri di queste relazioni.

Se al tempo $t-1$ i disequilibri sono diversi da 0, quindi il vettore Z_{t-1} non ha tutti elementi pari a 0, allora vi sarà un movimento nel vettore X_t che tende a riassorbire gli squilibri e questo movimento è misurato da $B * Z_{t-1}$. Una volta definiti il numero di relazioni h , si possono stimare i parametri dell'ECM, cioè le matrici $\zeta_1 \dots \zeta_{p-1}$ e le matrici B, A .

Teorema di rappresentazione di Granger esteso Si considera un processo X_t di dimensione k e integrato di ordine 1, se esistono h relazioni di cointegrazione tra gli elementi di X_t allora il processo ΔX_t è un processo stazionario, ed esiste una matrice A ($K \times h$) di rango h tale che $Z_t = A' * X_t$ ed è un processo stazionario.

Se X_t è un VAR(p) allora esiste anche B tale che $I_k - \Phi_1 - \dots - \Phi_p = -B * A'$ e $\zeta_1 \dots \zeta_{p-1}$ che verificano l'equazione $\Delta X_t = (\zeta_1 * \Delta X_{t-1} + \dots + \zeta_{p-1} * \Delta X_{t-p+1}) + c + B * Z_{t-1} + u_t$.

In definitiva processo X_t di dimensione k e integrato di ordine 1, se esistono h relazioni di cointegrazione tra gli elementi di X_t allora il sistema cointegrato ammette la rappresentazione a correzione d'errore.

7.2 Stima del ECM

Si distinguono 3 casi in base alle informazioni di cui disponiamo:

7.2.1 E' nota la matrice A dei vettori di cointegrazione

A può essere conosciuto attraverso studi economici precedenti, è la situazione meno realistica. Noi dobbiamo verificare che A sia effettivamente la matrice dei vettori di cointegrazione, eseguiamo quindi dei test:

- Test di radice unitaria per accertare che le k-variabili sono tutti processi integrati di ordine 1
- Accertato il punto precedente verifichiamo attraverso un test di radice unitaria se gli h processi di $Z_t = A' * X_t$ sono stazionari (cioè si deve rifiutare l'ipotesi nulla del test di prima).

7.2.2 Si conosce il rango di A, quindi so h il numero delle relazione di cointegrazione ma A è ignota

Conosciamo h cioè $\text{rango}(B*A') = (\zeta_0)$, quindi per stimare A e B uso la tecnica di Johansen. Parto da un ECM di un VAR(p) indicato come $\Delta X_t = (\zeta_1 * \Delta X_{t-1} + \dots + \zeta_p * X_{t-p+1}) + c + \zeta_0 * X_{t-1} + u_t$, per stimarli consideriamo due modelli di regressione:

- $\Delta X_t = \pi_0 + \pi_1 * \Delta X_{t-1} + \pi_{p-1} * \Delta X_{t-p+1} + u_t$ vediamo che c'è la componente di breve periodo del ECM.
- $X_{t-1} = \gamma_0 + \psi_1 * \Delta X_{t-1} + \dots + \psi_{p-1} * \Delta X_{t-p+1} + v_t$ vediamo che c'è la componente di breve periodo.

Con la loro regressione avrò che i residui saranno \hat{u}_t che contiene le relazione di ΔX_t al netto della componente di breve periodo, \hat{v}_t include le informazioni su X_{t-1} al netto della componente di breve periodo.

Se avessi $\hat{\zeta}_0$ esprimerebbe il legame tra ΔX_t e X_{t-1} al netto delle variabili differenziate nei ritardi, quindi $\hat{u}_t = B * A' * \hat{v}_t + \epsilon_t$ così da ottenere la stima di $\hat{\zeta}_0$, tale modello si chiama di **regressione concentrata**.

Ipotizziamo di conoscere di una stima A come \hat{A} e definito $\hat{w}_t = \hat{A}' * \hat{v}_t$, tale vettore include i regressori del modello $\hat{u}_t = B * \hat{w}_t$.

Devo trovare \hat{B} , che si trova facendo $\hat{B} = \frac{\sum \hat{u}_t * \hat{w}_t'}{n} * (\frac{\sum \hat{w}_t * \hat{w}_t'}{n})^{-1} = \frac{\sum \hat{u}_t * \hat{v}_t' * \hat{A}}{n} * (\frac{\sum \hat{A}' * \hat{v}_t * \hat{v}_t' * \hat{A}}{n})^{-1}$. Se chiamiamo la prima matrice $\hat{\Sigma}_{u,v} * \hat{A}$ e la seconda $\hat{A}' * \hat{\Sigma}_{v,v} * \hat{A}$. Perciò so che $\zeta_0 = \hat{B} * \hat{A}'$.

Ora però dobbiamo stimare A, lo facciamo attraverso il metodo della massima verosimiglianza, quindi \hat{A} è formata dagli h autovettori associati agli autovalori più grandi fra i K autovalori (tutti minori di 1) di $\hat{\Sigma}_{v,v}^{-1} * \hat{\Sigma}_{v,u} * \hat{\Sigma}_{u,u}^{-1} * \hat{\Sigma}_{u,v}$, dove $\hat{\Sigma}_{v,u} = \hat{\Sigma}_{u,v}'$.

Dopo aver ottenuto A lo inserisco per la stima di B

7.2.3 Non si conosco informazioni su A ed h

In questo caso è importante trovare h (rango di cointegrazione). Esso si individua con dei test di ipotesi su valori diversi di h, una volta identificati di ritorna al caso 2.

I test sono basati sul rapporto di verosimiglianza con H_0 ci sono h relazioni, ed è comune ai test:

- **Test della traccia** in questo caso H_1 dice che il $\text{rango}(\zeta_0) = K$, equivale a dire che è uguale al numero delle variabile, quindi sosteniamo che X_t include dei processi stazionari. Perciò non possiamo usare ECM.

Ha come statistica test $-2 * \log(\Lambda) = -n * \sum_{i=h+1}^k \log(1 - \hat{\lambda}_i)$ dove $\Lambda = \frac{\max(\text{IECM}_{H_0})}{\max(\text{IECM}_{H_1})}$, gli autovalori sono quelli della matrice di prima. Ha distribuzione non standard e i suoi valori critici sono ottenuti attraverso delle simulazioni. Il test rifiuto H_0 quando il valore è maggiore del quantile simulato $1-\alpha$.

- **Test dell'autovalore massimo** ha H_1 essa afferma che vi sono $h+1$ relazioni di cointegrazione. Ha come statistica test $-2 * \log(\Lambda) = -n * \log(1 - \hat{\lambda}_{h+1})$. Anche essa ha una distribuzione non standard, rifiuto H_0 quando il valore è maggiore del quantile simulato $1-\alpha$.