

1 Modelli previsivi ottimi

Quando si vuole fare una previsione si vuole trovare un valore puntuale alla distribuzione $Y|X$ dove $X = X_1 \dots X_n$. Per fare ciò bisogna introdurre l'errore di previsione cioè $e = Y - P(X)$ dove P è una funzione misurabile, ed è una variabile casuale, per valutare tale errore si usa la funzione di perdita cioè $l(e)$ possono essere:

- e^2 , se non è specificata è questa di default e pesa molto gli errori molto grandi.
- $|e|$ pesa in maniera lineare tutti gli errori.
- Funzioni asimmetriche più complicate. Si parla di sovra previsione se $e < 0$, e di sotto previsione se $e > 0$. Quando c'è un'asimmetria si dà più importanza alla parte con integrale più grande.

Noi vogliamo P tale che è $ARGMIN E(l(e))$, se la funzione di perdita è e^2 allora il previsore ottimo è $E(Y|X)$. Tale previsore ha delle proprietà:

- $E(aY + bY|X) = E(aY|X) + E(bY|X)$
- $E(Yg(x)|X) = E(Y|X)g(x)$
- $E(Y|X) = E(Y)$ se x e y sono indipendenti.
- Proprietà della torre $E(E(Y|X)) = E(Y)$ visto che $\int_{S_x} f(x) \int_{S_y} Yf(y|x) dy dx = \int_{S_y} \int_{S_x} Yf(x)f(y|x) dx dy = \int_{S_y} Yf(y) dy$.
- Ortogonalità $E((Y - E(Y|X))g(x)) = 0$ per qualsiasi $g(x)$ misurabile. Questo perché $E((Y - E(Y|X))g(x)) = E(E((Y - E(Y|X))g(x)|X)) = E(g(x)(E(Y|X) - E(Y|X))) = 0$
- $Var(y) = E(Var(y|x)) + Var(E(Y|X))$

Da queste proprietà riusciamo a dimostrare perché $E(Y|X)$ è il previsore che minimizza la funzione di perdita. Io voglio $P \in p$ tale che $E((Y - P)^2)$ è minimo, se aggiungo e tolgo $E(Y|X)$ ho $E((Y - E(Y|X) + E(Y|X) - P)^2) = E((Y - E(Y|X))^2 + (E(Y|X) - P)^2 - 2(Y - E(Y|X))(E(Y|X) - P))$ per ortogonalità $2(Y - E(Y|X))(E(Y|X) - P) = 0$ ho quindi una somma di quadrati che è minima se $P = E(Y|X)$.

1.1 Previsore lineare

E' un previsore non ottimo ma è molto efficiente, infatti il previsore ha forma $P(x) = B_0 + B_1 * X_1 + \dots + B_n * X_n$ e ho la funzione di perdita classica, allora se minimizzo $MSE = E((Y - P(x))^2)$ però scegliendo nella forma lineare, quindi minimizzo trovando dei B , ciò lo rende molto più veloce.

Il previsore lineare ottimo è $P(Y|X) = \mu_y + \Sigma_{y,x} * \Sigma_{x,x}^{-1}(x - \mu_x)$, tale è unico anche se X non è una matrice di rango pieno, inoltre chiamo $B_{y,x} = \Sigma_{y,x} * \Sigma_{x,x}^{-1}$ ed è una matrice non stocastica. Noto che se le variabili sono incorrelate tra loro allora previsione sarà uguale alla media di Y .

Ha le seguenti proprietà:

- Non distorto $E(Y - P(Y|X)) = 0$, è dimostrabile $E(y - \mu_y + B_{y,x}(x - \mu_x)) = E(y) - \mu_y - B_{y,x} * E(x - \mu_x) = 0$.
- Ortogonalità semplice cioè $E((Y - P(Y|X))X^t) = 0$, sostituisco e trovo $E((y - \mu_y + B_{y,x}(x - \mu_x))X^t)$ trasformo X^t in $(X - \mu_x)^t$ ho così $E((y - \mu_y)(x - \mu_x)^t) - B_{y,x} * E((X - \mu_x)(X - \mu_x)^t) = 0$.
- $MSE = \Sigma_{y,y} - \Sigma_{y,x} \Sigma_{x,x}^{-1} \Sigma_{x,y}$ si dimostra $E((y - P(Y|X))^t (y - P(Y|X))^t)$ se sviluppo ottengo $E((y - \mu_y)(y - \mu_y)^t) + B_{y,x} E((x - \mu_x)(x - \mu_x)^t) B_{y,x}^t - E((y - \mu_y)(x - \mu_x)^t) B_{y,x}^t - B_{y,x} E((x - \mu_x)(y - \mu_y)^t)$ che semplificando vedo che è uguale alla tesi.
- Linearità $P(aY + bZ|X) = P(aY|X) + P(bZ|X)$.
- Legge iterata quindi $P(Y|X) = P(P(Y|Z, x)|x)$, si dimostra sapendo che $Y = P(Y|z, x) + e$ quindi $e = Y - P(Y|z, x)$ quindi $E(e * X^t) = 0$ da qua $P(Y|X) = P(P(Y|z, x) + e|X) = P(P(Y|z, x)|X)$.

- Se $E(((y - \mu_y)(x - \mu_x))) = 0$ allora $P(Y|z, x) = \mu_y + P(Y - \mu_y|x) + P(Y - \mu_y|Z)$ questo perchè il previsore ottimo se ho valori incorrelati è $\mu_y + (\Sigma_{y,x}, \Sigma_{y,z}) * \begin{bmatrix} \Sigma_{x,x}^{-1} & 0 \\ 0 & \Sigma_{z,z}^{-1} \end{bmatrix} \begin{bmatrix} (x - \mu_x) \\ (z - \mu_z) \end{bmatrix}$ e risolvendo si ottiene la tesi.
- $P(Y|z, x) = P(Y|x) + P(Y - P(Y|X)|Z - P(Z|X)) = P(Y|x) + \Sigma_{y,z|x} * \Sigma_{z,z|x}^{-1} (Z - P(Z|X))$ dove $\Sigma_{z,z|x} = E((Z - P(Z|X))(Z - P(Z|X))^t)$, ciò ci permette di dire che $MSE_{y|z,x} = MSE_{y|x} - \Sigma_{y,z|x} * \Sigma_{z,z|x}^{-1} * \Sigma_{z,y|x}$.
- So che se y, x sono una normale multivariata allora $Y|X$ è $N(\mu_y + \Sigma_{y,x} * \Sigma_{x,x}^{-1}(x - \mu_x), \Sigma_{y,y} - \Sigma_{y,x} * \Sigma_{x,x}^{-1} * \Sigma_{x,y})$ cioè la sua media è il previsore lineare.

1.2 Esempi applicativi

- Voglio stimare una serie storica con trend e stazionarietà usando esclusivamente la regressione lineare.
Prendo il modello e uso una trasformazione logaritmica per rendere lineare la stagionalità, inoltre il logaritmo è facile da interpretare visto che rappresenta circa il tasso di crescita annuale, inoltre ci permette valutare eventualmente il logrendimenti ed è adattivo.
Per modellare la stagionalità uso le variabili dummy cioè uso gli α_i variabili per ogni mese i . Il modello creato se ho $n+1$ periodo sarà quindi $Y_t = \alpha + B_t + \delta_1 * D_1 + \dots + \delta_n * D_n + \epsilon_t$.
Se voglio far sì che tutti i regressori delle dummy siano 0 devo modificare il valore dei D_i in particolare dove c'è 0 aggiungerci $-1/n$ così da obbligare i regressori della dummy a fare 0.
Con questo metodo però se sommo tutti le n componenti ottengono il valore ipotetico che avrebbe il coefficiente della $n+1$ componente.
- Il previsore ottimo se la funzione di perdita è $|e|$ è la mediana.
- Se ho una funzione di perdita asimmetrica la funzione di perdita è uguale a ad un determinato percentile che si ricava dalla funzione. Che si ricava come $K_{max}/(K_{max} + K_{min})$. (dimostra)

2 State space form

E' una forma particolare che ci permette di racchiudere tutti i modelli lineari di serie storiche, è utile perchè ci permette di fare un inferenza efficienti sulle componenti di ciclo trend e stagionalità.

La forma è composta da:

- Equazione di misurazione/osservazione del tipo $Y_t = c_t + Z_t * \alpha_t + \epsilon_t$ dove α_t è detto vettore di stato e non ha per forza la dimensione $DIM(Y_t) = p * 1$, se imponiamo $DIM(\alpha_t) = m * 1$ ho che $DIM(Z_t) = p * m$ puntiamo a minimizzare m così siamo più veloci, e ϵ_t è un white noise, con varianza H_t che quindi può evolvere nel tempo.
- Equazione di stato o di transizione $\alpha_{t+1} = d_t + T_t * \alpha_t + \nu_t$, dove ν_t è un white noise con varianza Q_t , inoltre come notiamo α contiene molte informazioni utili come ciclo trend e stagionalità, notiamo anche che è una VAR(1). Il nostro obiettivo è stimare α per poi andare a stimare Y_t .

La forza di questo algoritmo è che la previsione della serie storica non viene divisa dalla modellizzazione della stagionalità ciò ci aiuta ad essere efficienti e coerenti.

La forma state space è completa se si specificano i primi due momenti di α_1 , mi serve quindi:

- $a_{1|0} = E(\alpha_1)$ e $P_{1|0} = E((\alpha_1 - a_{1|0})(\alpha_1 - a_{1|0})^t)$, se ho delle componenti stazionarie è molto semplice da trovare, nel caso non stazionario invece posso inserire una media arbitraria e varianza infinite, dette condizioni diffuse.
- Si chiede che ν_t e ϵ_t , siano incorrelati tra loro, anche se è possibile che sia correlati con G_t come valore ma è molto raro.
- ν_t e ϵ_t sono incorrelati anche nei t diversi tra di loro.

- $E((\alpha_1 - a_{1|0}) * \epsilon_t = 0$ quindi non c'è correlazione tra la variabile e i suoi disturbi, questo vale anche con ν_t .
- Se ν_t e ϵ_t sono $N(0, Q_t)$, $N(0, H_t)$ e α_1 è $N(a_{1|0}, P_{1|0})$, allora il processo α_{t+1} è gaussiano essendo combinazione lineare di gaussiane, idem per y_t è gaussiano essendo combinazione di gaussiane. Avrò quindi che la forma state space è detta gaussiana. In questo modo riusciamo a fare stime di massima verosimiglianza o fare inferenza. Il software non tiene in conto delle costanti.

Un sistema state space **omogeneo** nel tempo è tale se un nessuno dei suoi parametri dipendendo dal tempo quindi $Y_t = c + Z * \alpha_t + \epsilon_t$ e $\alpha_{t+1} = d + T * \alpha_t + \nu_t$, questa forma è molto semplice da trattare. Possiamo dire anche che una forma omogenea non è per forza un processo stazionario ma che un processo stazionario è per forza omogeneo.

2.1 Esempi su AR

Sia AR(1), $Y_t = K + \phi * y_{t-1} + \epsilon_t$ può essere anche vista attraverso un equazione di transizione $\alpha_{t+1} = \phi * \alpha_t + \nu_t$, ed equazione di osservazione $Y_t = \mu + \alpha_t$, la prima e l'ultima sono lo stesso metodo per scrivere le stesse equazioni.

Posso calcolare facilmente $E(\alpha_1) = 0$ e $Var(\alpha_1) = \frac{\sigma^2}{1-\phi^2}$ ciò vale essendo un AR(1).

Sia AR(2) cioè $Y_t = K + \phi_1 * y_{t-1} + \phi_2 * y_{t-2} + \epsilon_t$, in questo caso per modellarlo credo due equazioni di transizioni, $\begin{bmatrix} \alpha_{1,t+1} \\ \alpha_{2,t+1} \end{bmatrix} = \begin{bmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{bmatrix} * \begin{bmatrix} \alpha_{1,t} \\ \alpha_{2,t} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ 0 \end{bmatrix}$, con questa costruzione impongo che $\alpha_{2,t+1} = \alpha_{1,t}$.

Come equazione di osservazione avrò $Y_t = \mu + [1 \ 0] * \begin{bmatrix} \alpha_{1,t} \\ \alpha_{2,t} \end{bmatrix} + 0$, inoltre ho che la matrice varianza e covarianza di α è $Q = \begin{bmatrix} \sigma^2 & 0 \\ 0 & 0 \end{bmatrix}$ e varianza di y_t è $H = 0$. In questo caso so che $a_{1|0} = E(\alpha_1) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$,

so anche che $P_{1|0} = \begin{bmatrix} \gamma(0) & \gamma(1) \\ \gamma(1) & \gamma(0) \end{bmatrix}$, dove $\gamma(1) = E(\alpha_{1,1}\alpha_{2,1}) = E(\alpha_{1,1}\alpha_{1,0})$.

Tale modello va anche con P ritardi quindi è generalizzabile.

2.2 Esempi su MA

Se ho un MA(2) ho $Y_t = \epsilon_t + \theta_1 * \epsilon_{t-1} + \theta_2 * \epsilon_{t-2}$, esso ha equazione di transizione

$$\begin{bmatrix} \alpha_{1,t+1} \\ \alpha_{2,t+1} \\ \alpha_{3,t+1} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} * \begin{bmatrix} \alpha_{1,t} \\ \alpha_{2,t} \\ \alpha_{3,t} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ 0 \\ 0 \end{bmatrix}, \text{ da qua ne deriva che } \alpha_{1,t+1} = \epsilon_t \text{ e } \alpha_{2,t+1} = \alpha_{1,t} = \epsilon_{t-1}$$

e $\alpha_{3,t+1} = \alpha_{2,t} = \epsilon_{t-2}$, calcolo poi l'equazione di osservazione come $Y_t = \mu + [1 \ \theta_1 \ \theta_2] * \alpha_t$, la condizione iniziale essendo formato da white noise è semplice $a_{1|0} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$ e $P_{1|0} = \begin{bmatrix} \sigma^2 & 0 & 0 \\ 0 & \sigma^2 & 0 \\ 0 & 0 & \sigma^2 \end{bmatrix}$.

Si può vedere in un altro modo detta forma verticale quella di prima era forma orizzontale, ho equazione di transizione $\begin{bmatrix} \alpha_{1,t+1} \\ \alpha_{2,t+1} \\ \alpha_{3,t+1} \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} * \begin{bmatrix} \alpha_{1,t} \\ \alpha_{2,t} \\ \alpha_{3,t} \end{bmatrix} + \begin{bmatrix} 1 \\ \theta_1 \\ \theta_2 \end{bmatrix} * \epsilon_t$, notiamo che $\alpha_{3,t+1} = \theta_2 * \epsilon_t$ e $\alpha_{2,t+1} = \alpha_{3,t} + \theta_1 * \epsilon_t = \theta_1 * \epsilon_t + \theta_2 * \epsilon_{t-1}$ ed infine ho $\alpha_{1,t+1} = \alpha_{2,t} + \epsilon_t = \epsilon_t + \theta_1 * \epsilon_{t-1} + \theta_2 * \epsilon_{t-2}$, ecco che ci riconduciamo ad un processo MA e quindi avrò come equazione di osservazione $Y_t = \mu + [1 \ 0 \ 0] * \alpha_t$, ed ovviamente avrò le stesse misure di varianza di prima.

2.3 Esempi su ARMA

Prendiamo ARMA(2,2) ho $Y_t = \phi_1 * y_{t-1} + \phi_2 * y_{t-2} + \epsilon_t + \theta_1 * \epsilon_{t-1} + \theta_2 * \epsilon_{t-2}$, possiamo riassumerlo come $Y_t = \frac{\theta_q(B)}{\phi_p(B)} * \epsilon_t$ se il processo è stazionario ciò è ben definito.

Avrò equazione di transizione $\begin{bmatrix} \alpha_{1,t+1} \\ \alpha_{2,t+1} \\ \alpha_{3,t+1} \end{bmatrix} = \begin{bmatrix} \phi_1 & \phi_2 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} * \begin{bmatrix} \alpha_{1,t} \\ \alpha_{2,t} \\ \alpha_{3,t} \end{bmatrix} + \begin{bmatrix} \epsilon_t \\ 0 \\ 0 \end{bmatrix}$ ottengo che $\alpha_{2,t+1} = \alpha_{1,t}$, $\alpha_{3,t+1} = \alpha_{2,t} = \alpha_{1,t-1}$ ho quindi $\alpha_{1,t+1} = \phi_1 * \alpha_{1,t} + \phi_2 * \alpha_{1,t-1} + \epsilon_t$ cioè un processo AR(2), ho

come equazione di osservazione $Y_t = \begin{bmatrix} 1 & \theta_1 & \theta_2 \end{bmatrix} * \alpha_t$, in questo modo credo un ARMA(2,2). Questo funziona perchè ho per esempio $\alpha_{3,t+1} = \frac{B^2 * \epsilon_t}{1 - \phi_1 * B + \phi_2 * B^2}$, che sommandoli otteniamo un ARMA(2,2).

E come rappresentazione verticale ho $\begin{bmatrix} \alpha_{1,t+1} \\ \alpha_{2,t+1} \\ \alpha_{3,t+1} \end{bmatrix} = \begin{bmatrix} \phi_1 & 1 & 0 \\ \phi_2 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} * \begin{bmatrix} \alpha_{1,t} \\ \alpha_{2,t} \\ \alpha_{3,t} \end{bmatrix} + \begin{bmatrix} 1 \\ \theta_1 \\ \theta_2 \end{bmatrix} * \epsilon_t$ vediamo che $\alpha_{3,t+1} = \theta_2 * \epsilon_t$ e $\alpha_{2,t+1} = \phi_2 * \alpha_{1,t} + \theta_2 * \epsilon_{t-1} + \theta_1 \epsilon_t$ infine $\alpha_{1,t+1} = \phi_1 * \alpha_{1,t} + \phi_2 * \alpha_{1,t-1} + \theta_1 * \epsilon_{t-1} + \theta_2 * \epsilon_{t-2} + \epsilon_t$ quindi ho un ARMA(2,2). In questo caso la matrice di osservazione ho $Y_t = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} * \alpha_t$, infine ho che $Q = \begin{bmatrix} 1 & \theta_1 & \theta_2 \\ \theta_1 & \theta_1^2 & \theta_1 * \theta_2 \\ \theta_2 & \theta_2 * \theta_1 & \theta_2^2 \end{bmatrix} * \sigma^2$ ed $H = 0$, è facile da vedere che $a_{1|0} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$, invece $P_{1|0}$ è più difficile.

Tutto ciò può essere generalizzato quindi un modello ARMA(p,q) alla forma state space.

3 Modelli a componenti non osservabili

Se identifichiamo il modello come $Y_t = \mu_t + \psi_t + \gamma_t + \epsilon_t$ dove:

- μ_t è il trend che rappresenta il livello medio della serie storica in un tempo t .
- ψ_t è la componente ciclica che aiuta a modellare le fluttuazioni
- γ_t rappresenta la stagionalità molto presente se ho dei dati frequenti o comunque minore di un anno. Possono essere presenti anche più componenti stagionali.
- ϵ_t è un white noise con tutte le sue proprietà.

3.1 Trend

Un trend è detto stocastico se evolve nel tempo con incertezza. Il più usato è il **local linear trend**, è formato:

- Livello $\mu_t = \mu_{t-1} + B_{t-1} + \eta_t$ dove η_t è un white noise. Notiamo subito che è un simil random walk, dove però B_{t-1} è uno slope cioè la pendenza del trend. Tale scrittura può essere anche vista come una retta, dove B_{t-1} è un intercetta che si muove nel tempo in maniera stocastica e μ_{t-1} è un coefficiente lineare che si evolve in maniera stocastica.
- $B_t = B_{t-1} + \zeta_t$ dove ζ_t è un white noise, noto quindi che è un perfetto random walk.

Tale retta è scrivibile in maniera deterministica come $\mu_t = \mu_0 + B * t$, posso scriverla anche in modo che si sviluppi in maniera incrementale ho $\mu_t = \mu_{t-1} + B$ e se prendiamo per esempio $\mu_2 = \mu_1 + B = \mu_2 = \mu_0 + 2 * B$ quindi vedo che $\mu_t = \mu_0 + B * t$, se prendo la forma incrementale ci aggiungo il rumore $\mu_t = \mu_{t-1} + B + \eta_t$ quindi noto che è un random walk con drift, dove la parte casuale è η_t , B è il drift che dirige lo sviluppo di questa retta, inoltre possiamo vederla come retta ed ottengo $\mu_t = \mu_0 + B * t + \sum_i^t \eta_t$, noto quindi che posso allontanarmi della retta ma comunque in media è una retta.

Nella realtà il coefficiente B cambia con il tempo. Per farlo variare costruisco $\mu_t = \mu_{t-1} + B_{t-1} + \eta_t$ dove $B_t = B_{t-1} + \zeta_t$, in questo modo abbiamo ottenuto un local linear trend. Notiamo quindi che sia gli shock η_t che ζ_t influenzano la retta.

Se analizziamo i processi vediamo che B_t è un processo integrato di ordine 1, invece μ_t è un processo integrato di ordine 2, quindi la sua previsione sarà a sua volta un processo non stazionario.

Notiamo che se imponiamo:

- Le varianze dei WN uguali a 0 diventa di nuovo un modello deterministico.
- $VAR(\zeta_t) = 0$ riotteniamo il modello di prima, oppure posso mettere anche $B_0 = 0$ e quindi eliminiamo anche il drift, dove μ_t è un processo integrato di ordine 1.
- $VAR(\eta_t) = 0$ ottengo che la retta ha come componente stocastica solo il coefficiente angolare, così creo un processo chiamato random walk integrato o trend liscio, μ_t è un processo integrato di ordine 2.

La previsione di un local linear trend è una retta che parte dall'ultimo valore e che incrementa in maniera costante a secondo dell'ultimo valore osservato di B , la previsione sarà quindi $\hat{\mu}_{n+k|n} = \mu_n + B_n * k$ ed è corretta nel breve periodo, nel lungo l'errore fatto è elevato.

3.2 Cicli stocastici

La sua formula è $\begin{bmatrix} \psi_t \\ \psi_t^* \end{bmatrix} = \rho * \begin{bmatrix} \cos(\lambda) & \sin(\lambda) \\ -\sin(\lambda) & \cos(\lambda) \end{bmatrix} * \begin{bmatrix} \psi_{t-1} \\ \psi_{t-1}^* \end{bmatrix} + \begin{bmatrix} K_t \\ K_t^* \end{bmatrix}$ dove $\rho \in (0, 1)$, i K_t sono dei $WN(0, I_2 * \sigma_k^2)$, la matrice dei seni e dei coseni è detto matrice di rotazione oraria.

Per stimare il ciclo si usano delle sinusoidi, sia R la sua ampiezza, λ la sua frequenza, e fase ϕ che fa spostare l'inizio a destra o sinistra del valore della sinusoide, infine ho il periodo cioè ogni quanto si ripeta la sinusoide e so che $\lambda = \frac{2\pi}{periodo}$.

Per derivare tale formula dobbiamo specificare la sinusoide $f(t) = R * \cos(\phi + \lambda * t)$. La sinusoide in questo modo però è difficile tra trattare con i minimi quadrati allora la andiamo a trasformare come $R * \cos(\phi + \lambda * t) = R \cos(\phi) \cos(\lambda * t) - R \sin(\phi) \sin(\lambda * t)$ che è riscrivibile come $A * \cos(\lambda * t) + B * \sin(\lambda * t)$ cioè una forma stimabile in maniera lineare, con questa forma possiamo stimare $R = \sqrt{A^2 + B^2}$, posso stimare $\phi = \arccos(A/R)$ se $B \geq 0$ se no $\phi = -\arccos(A/R)$, su R possiamo risolverlo attraverso "atan2(-B,A)".

Ora dobbiamo inserire la formula incrementale come per il trend, per farlo estendiamo la dimensionalità a due usando le coordinate cartesiane, sia fissato un vettore $\psi_0 = \begin{bmatrix} \psi_0 \\ \psi_0^* \end{bmatrix}$ la sua distanza dal centro

è $r = \sqrt{\psi_0^2 + \psi_0^{*2}}$, ora se applichiamo la matrice di rotazione $R(\lambda) = \begin{bmatrix} \cos(\lambda) & \sin(\lambda) \\ -\sin(\lambda) & \cos(\lambda) \end{bmatrix}$, ottengo che $\psi_1 = R(\lambda) * \psi_0$ è un punto sulla stessa circonferenza ma ruotato in senso orario di angolo λ , questa operazione si può ripetere tutte le volte che vogliamo.

Se ci focalizziamo sul primo elemento del vettore ruotato otteniamo il punto proiettato sull'asse delle X quindi si comporta esattamente come un coseno, quindi la matrice di rotazione lo sposta.

Quindi la sinusoide potrò scriverla come $\begin{bmatrix} \psi_t \\ \psi_t^* \end{bmatrix} = R(\lambda) * \begin{bmatrix} \psi_{t-1} \\ \psi_{t-1}^* \end{bmatrix} + \begin{bmatrix} K_t \\ K_t^* \end{bmatrix}$, cioè la riscrittura iniziale, inoltre con l'aggiunta di rumore l'ampiezza dell'angolo e la lunghezza di R non è detto che sia rispettata infatti ho un ciclo stocastico.

Notiamo che manca il ρ detto fattore di smorzamento, infatti se notiamo che ψ è un processo del tipo $VARIMA(2, 2, 1)$, ed essendo non stazionario uno shock è in grado di turbare perennemente la ψ , ma questo nella realtà è fastidioso e quindi inseriamo il $0 < \rho < 1$ per renderlo stazionario.

So che esiste $E(\psi_t) = 0$ e inoltre $E(\psi_t * \psi_t') = \frac{\sigma_k^2}{1-\rho^2} * I_2$ che è identica alla varianza di un $AR(1)$ questo conferma la sua stazionarietà.

Un ciclo in media dura 4 anni se ho per esempio un osservazione ogni trimestre avrò che il ciclo si ripete dopo 16 osservazioni in media

3.2.1 Cicli stocastici superiori

E' un metodo per generare cicli fortemente smooth, così da riuscire a datare un punto di minimo e un punto di massimo in maniera agevole. Ricordiamo l'equazione $\begin{bmatrix} \psi_t \\ \psi_t^* \end{bmatrix} = \rho * \begin{bmatrix} \cos(\lambda) & \sin(\lambda) \\ -\sin(\lambda) & \cos(\lambda) \end{bmatrix} * \begin{bmatrix} \psi_{t-1} \\ \psi_{t-1}^* \end{bmatrix} + \begin{bmatrix} K_t \\ K_t^* \end{bmatrix}$, è detto anche ciclo di ordine 1, allora per renderlo di ordine superiore detto ciclo di ordine 2 faccio $\psi_t^{(2)} = \rho * R(\lambda) \psi_{t-1}^{(2)} + \psi_t^{(1)}$, è molto utile l'ordine 2 spingersi oltre è rindondante.

3.3 Stagionalità

Si può trattare con 2 metodi, attraverso le dummy come abbiamo già visto in precedenza oppure attraverso delle sinusoidi.

Sia γ_t la stagionalità dove $t \in 1...n$ numero intero, dove $s \in 1...n$ valore della periodicità, se ho stagionalità vale che $\gamma_t = \gamma_{t-s}$ se lo vedo sotto forma di dummy, vedo che la somma della stagionalità se deterministica è nulla cioè $\sum_{i=1}^{s-1} \gamma_{t-i} + \gamma_t = 0$ anche scritta come $\sum_{i=0}^{s-1} \gamma_{t-i} = 0$, perciò vale che $\gamma_t = -\sum_{i=1}^{s-1} \gamma_{t-i}$, per renderla stocastica diciamo che $\gamma_t = -\sum_{i=1}^{s-1} \gamma_{t-i} + \omega_t$ dove ω_t è un white noise, quindi avremmo che in media vale così ma non sempre.

L'altra forma per rappresentare la stagionalità è usando le sinusoidi, dirò in particolare che la stagionalità ha forma $\gamma_t = \sum_{j=1}^{[S/2]} a_j * \cos(\frac{2\pi*j}{S} * t) + b_j * \sin(\frac{2\pi*j}{S} * t)$, dove la frequenza cioè quanti giri completi compie in un periodo S è dettato dal termine $F = \frac{2\pi}{S}$, più la frequenza è alta più il percorso è meno smooth. Posso ricavarci S avendo la frequenza come $S = 2\pi / Freq$.

Se lo vedessimo come un vettore fatto in precedenza quindi attraverso i cicli avrei

$$\begin{bmatrix} \gamma_t^j \\ \gamma_t^{j*} \end{bmatrix} = \begin{bmatrix} \cos(\frac{2\pi}{S} * j) & \sin(\frac{2\pi}{S} * j) \\ -\sin(\frac{2\pi}{S} * j) & \cos(\frac{2\pi}{S} * j) \end{bmatrix} * \begin{bmatrix} \gamma_{t-1}^j \\ \gamma_{t-1}^{j*} \end{bmatrix} + \begin{bmatrix} w_t \\ w_t^* \end{bmatrix} \text{ e poi usiamo } \gamma_t = \sum_j^{[S/2]} \gamma_t^j.$$

I coefficienti a_j e b_j ci dicono l'ampiezza della sinusoide, notiamo anche che tale è un modello lineare solamente con delle trasformazioni al vettore delle X , ciò ci permette di stimarlo con una regressione, questo da un vantaggio, infatti essendo X matrice a rango pieno quindi ogni \sin è indipendente dal \cos se sottraggo colonne di dati la stima per gli altri coefficienti rimane invariata.

Noto che devo stimare sempre $S - 1$ parametri, nel caso dispari è ovvio infatti la parte intera della divisione moltiplicata per due sarà sempre $S-1$, invece se S è pari mi trovo con l'ultimo seno scritto come $\sin(\frac{2\pi}{S} * \frac{S}{2}t) = 0$ questo vale sempre per ogni t , quindi il b corrispondente è inutile da stimare.

La stagionalità viene usata per modellare dei cicli molto frequenti che avvengono in maniera smooth, invece le dummy vengono usate per modellare le stagionalità più improvvise e che non si verificano spesso.

Operativamente è molto semplice da creare la griglia di dati da inserire nel seno e nel coseno infatti dati il vettore di tempi t , il valore del vettore j , uso "outer(t, j) * $2\pi/S$ ".

Per trattare al meglio una serie storica possiamo usare dummy e sinusoidi insieme.

Posso costruire un modello del tipo trend stagionalità $Y_t = B_0 + B_1 * t + B_2 * T^2 + \gamma_t$, e per renderlo stocastico basta aggiungere gli errori sia per la stima di B_t sia per γ_t . Inoltre un numero alto di sinusoidi va a creare la stessa funzione che avremmo tracciando la linea di una media mobile.

4 Regressori

Se prendiamo la generica serie storica $Y_t = \mu_t + \psi_t + \gamma_t + \delta * X_t + \epsilon_t$ dove $\delta * X_t$ è la parte di regressione, a differenza della classica relazione che obbliga X_t ad essere cointegrato con Y_t va bene anche se non lo sono perchè ci sono altri elementi integrati.

Rispetto ai soliti modelli ARMA se X_t è stazionario ma gli altri elementi non lo sono è comunque permessa e sensata la regressione perchè andiamo a stimare a parte la parte integrata.

Abbiamo anche visto un trend stocastico come $\mu_t = \mu_{t-1} + \delta * X_t + \eta_t$ vediamo quindi che i regressori possono essere aggiunti anche in questo caso.

Questi tipi di regressori $\delta * X_t$ sono **detti statici**, i classici regressori che si usano nelle serie storiche. Le variabili dummy possono essere dei regressori, come abbiamo visto in precedenza, per esempio:

- Variabili additive per movimenti della durata di un singolo periodo e per movimenti repentini, la dummy sarà formata da se $T = T_0$ $D = 1$ se no $D = 0$.
- Temporary change cioè un cambiamento repentino in un particolare range di date, la dummy sarà formata da $D = 1$ se $T_0 < T < T_1$ se no $D = 0$.
- Lever shift cioè un cambiamento di livello repentino che permane, la dummy sarà formata da $D = 1$ se $T > T_1$ se no $D = 0$.

- Slope shift cioè un cambio di pendenza che crea discontinuità nella derivata, la dummy sarà formata da $D = T - T_0 + 1$ se $T > T_0$ se no $D = 0$.

I **regressori dinamici** invece effettuano il loro cambiamento in una particolare tempo ma hanno effetto anche sul futuro.

Data una dummy D_t classica ed ho una serie storica $Y_t = \mu_t + \dots$ dove il trend è $\mu_t = \mu_{t-1} + B_t + \delta * D_t + \eta_t$ vedo che ho un Lever shift solamente che in un particolare tempo, anche se in Y_t permane sempre, e $B_t = B_{t-1} + \delta * D_t + \zeta_t$ in questo caso come prima ho uno Slope shift in un particolare momento che si ripercuote su tutti i tempi futuri.

Ciò è dovuto al tipo di integrazione che hanno le variabili che permette il permanere dello shock.

Si possono inserire anche nella parte stagionale così da modificare la stagionalità, inserendo la dummy nella somma delle stagionalità così da modificare i coefficienti oppure mettendo la dummy nella varianza del white noise della stagionalità così da modificare repentinamente gli sbalzi diversi.

Le **funzioni di trasferimento** sono le regressioni effettuate su regressori dinamici, in particolare ho che $Y_t = \dots + \delta * X_t + \dots + \delta * X_{t-k} + \dots$, però se K è alto o addirittura infinito, ho molti parametri da stimare, quindi i gradi di libertà scendono dunque non ho abbastanza dati per avere stime precise. Ricordando le stime ARIMA $Y_t = \frac{\Theta(B)}{\phi(B)} * \epsilon_t = \frac{1 + \theta_1 * B + \dots}{1 - \pi * B - \dots}$ so che gli effetti di lungo periodo vengono stimati da $\phi(B)$ infatti ci dice se la serie è stazionario o no, questo ci aiuta a stimare i nostri coefficienti in particolare avrò $w_t = \frac{W(B)}{\delta(B)} * X_t = \frac{w_0 + w_1 * B + \dots}{1 - \delta * B - \dots}$ ho quindi $W_t = \delta_1 * W_{t-1} + \dots + \delta_r * W_{t-r} + w_0 * X_t + w_1 * X_{t-1} + \dots + w_s * X_{t-s}$, dove w_0 è il moltiplicatore di impatto immediato, notiamo che con questo modello gli effetti del passato anche lunghi hanno effetto nel futuro, e come un modello stazionario classico se $\delta < 1$ allora l'effetto decadrà esponenzialmente e avrò stazionarietà.

Essendo un pseudo modello ARMA possono esistere anche le sue semplificazione AR ed MA.

Inoltre notiamo anche che il valore W_0 è la funzione di impulso su un cambio unitario di X_t (guarda definizione di serie storiche).

Prendiamo in esempio il modello $W_t = \delta_1 * W_{t-1} + \delta_2 * W_{t-2} + w_0 * X_t$ so che per essere stazionaria deve valere $\delta_1 * X + \delta_2 * X^2 = 1$ se non valesse allora ho un movimento oscillatorio che non converge verso lo 0.

Il regressore dinamico entra nella nostra equazione come un regressore classico quindi $Y_t = \dots + w_t + \dots$ dove W_t è come definito in precedenza, inoltre posso avere un mix anche tra regressori dinamici e statici.

5 Ucm in forma State Space

5.1 Local linear trend

Sappiamo che è composto da 2 equazioni dove ho due elementi μ_t e B_t , quindi ho $\mu_{t+1} = \mu_t + B_t + \eta_t$ e $B_{t+1} = B_t + \zeta_t$, quindi la sua forma state space ha come equazione di transizione: $\begin{bmatrix} \mu_{t+1} \\ B_{t+1} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} * \begin{bmatrix} \mu_t \\ B_t \end{bmatrix} + \begin{bmatrix} \eta_t \\ \zeta_t \end{bmatrix}$ dove η e ζ sono scorrelati e la loro matrice di varianza e covarianza è Q , e come equazione di osservazione ho $Y_t = \begin{bmatrix} 1 & 0 \end{bmatrix} * \begin{bmatrix} \mu_t \\ B_t \end{bmatrix} + \epsilon_t$.

Si può semplificare in vari modi:

- Local level model, in questo caso $\mu_{t+1} = \mu_t + \eta_t$ di conseguenza $Y_t = \mu_t + \eta_t$.
- Random walk con drift in questo caso basta che metto $\zeta_t = 0$ sempre.
- Random walk liscio cioè ponendo $\eta_t = 0$.
- Imporre che tutte le varianze degli errori dell'equazione di transizione siano 0 ottengo così un trend lineare base.

5.1.1 Condizioni iniziali

Non abbiamo ancora messo nessuna condizione iniziale che è obbligatoria per impostare la forma state space. Si nota che se una componente è non stazionaria, come nel caso del trend, posso prendere valori diffusi cioè scegliere valori approssimativi o addirittura causali cioè $a_{1|0} = 0$ e $P_{1|0} = \infty$.

5.2 Ciclo stocastico

La sua forma è $\begin{bmatrix} \psi_{t+1} \\ \psi_{t+1}^* \end{bmatrix} = \rho * \begin{bmatrix} \cos(\lambda) & \sin(\lambda) \\ -\sin(\lambda) & \cos(\lambda) \end{bmatrix} * \begin{bmatrix} \psi_t \\ \psi_t^* \end{bmatrix} + \begin{bmatrix} K_t \\ K_t^* \end{bmatrix}$ ed ho matrice $Q = \begin{bmatrix} \sigma_k^2 & 0 \\ 0 & \sigma_k^2 \end{bmatrix}$, infine come matrice di transizione ho $Y_t = \begin{bmatrix} 1 & 0 \end{bmatrix} * \begin{bmatrix} \psi_t \\ \psi_t^* \end{bmatrix} + \epsilon_t$.

Possiamo anche unire linear trend con il ciclo e scriverlo in forma state space sapendo che $T = \begin{bmatrix} T_{lt} & 0 \\ 0 & T_c \end{bmatrix}$

e $Q = \begin{bmatrix} Q_{lt} & 0 \\ 0 & Q_c \end{bmatrix}$, invece $Z = \begin{bmatrix} Z_{lt} & z_c \end{bmatrix}$ unendo ottengo quindi

$$\begin{bmatrix} \mu_{t+1} \\ B_{t+1} \\ \psi_{t+1} \\ \psi_{t+1}^* \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \rho * \cos(\lambda) & \rho * \sin(\lambda) \\ 0 & 0 & -\rho * \sin(\lambda) & \rho * \cos(\lambda) \end{bmatrix} * \begin{bmatrix} \mu_t \\ B_t \\ \psi_t \\ \psi_t^* \end{bmatrix} + \text{errori} \text{ e come equazione di misurazione } Y_t = \begin{bmatrix} 1 & 0 & 1 & 0 \end{bmatrix} * \begin{bmatrix} \mu_t \\ B_t \\ \psi_t \\ \psi_t^* \end{bmatrix} + \epsilon_t.$$

5.2.1 Condizioni iniziali

Se $\rho < 1$ allora il ciclo è stazionario e le condizioni iniziali hanno senso e sono $a_{1|0} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ e $P_{1|0} = \begin{bmatrix} \sigma_k^2/(1-\rho^2) & 0 \\ 0 & \sigma_k^2/(1-\rho^2) \end{bmatrix}$.

5.3 Stagionalità in forma state space

5.3.1 Dummy stocastiche

So che $\gamma_t = -\gamma_{t-1} - \gamma_{t-2} - \gamma_{t-3} + \omega_t$ dove omega $WN(0, \sigma_w^2)$. Per vederlo in questo caso come state space avrò equazione di transizione: $\begin{bmatrix} \gamma_{t+1}^1 \\ \gamma_{t+1}^2 \\ \gamma_{t+1}^3 \end{bmatrix} = \begin{bmatrix} -1 & -1 & -1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} * \begin{bmatrix} \gamma_t^1 \\ \gamma_t^2 \\ \gamma_t^3 \end{bmatrix} + \begin{bmatrix} \omega_t \\ 0 \\ 0 \end{bmatrix}$, il periodo di stagionalità n viene modellata da $n-1$. Come nel modello AR in forma state space ottengo $\gamma_t^2 = \gamma_{t-1}^1$ e che $\gamma_t^3 = \gamma_{t-1}^2 = \gamma_{t-2}^1$. Ed ha come matrice di misurazione $Y_t = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} * \begin{bmatrix} \gamma_{t+1}^1 \\ \gamma_{t+1}^2 \\ \gamma_{t+1}^3 \end{bmatrix} + \epsilon_t$ dove Q è una matrice di 0 e nel elemento 1,1 ha σ_w^2 .

5.3.2 Stagionalità trigonometriche stocastiche

sia la frequenza $\lambda = \frac{2\pi}{S} * j$ allora so che $\gamma_t = \sum_{j=1}^{[S/2]} \gamma_t^{(j)}$ e la sua forma matriciale è $\begin{bmatrix} \gamma_{t+1}^j \\ \gamma_{t+1}^{j*} \\ \gamma_{t+1}^j \end{bmatrix} =$

$$\begin{bmatrix} \cos(\lambda_j) & \sin(\lambda_j) \\ -\sin(\lambda_j) & \cos(\lambda_j) \end{bmatrix} * \begin{bmatrix} \gamma_t^j \\ \gamma_t^{j*} \end{bmatrix} + \begin{bmatrix} w_t \\ w_t \end{bmatrix} \text{ ha sempre S-1 coppie possibili. Ho come equazione di transizione } \begin{bmatrix} \gamma_{t+1}^1 \\ \gamma_{t+1}^2 \\ \gamma_{t+1}^3 \end{bmatrix} = \begin{bmatrix} \cos(\lambda_1) & \sin(\lambda_1) & 0 \\ -\sin(\lambda_1) & \cos(\lambda_1) & 0 \\ 0 & 0 & \cos(\lambda_2) \end{bmatrix} * \begin{bmatrix} \gamma_t^1 \\ \gamma_t^2 \\ \gamma_t^3 \end{bmatrix} + \begin{bmatrix} \omega_t^1 \\ \omega_t^2 \\ \omega_t^3 \end{bmatrix}, \text{ in questo caso ho } Q \text{ matrice diagonale con}$$

diagonale uguale a σ_w^2 , ho come euqazione di transizione $Y_t = \begin{bmatrix} 1 & 0 & 1 \end{bmatrix} * \begin{bmatrix} \gamma_{t+1}^1 \\ \gamma_{t+1}^2 \\ \gamma_{t+1}^3 \end{bmatrix} + \epsilon_t$

5.3.3 Condizioni iniziali

So che ho non stazionarietà quindi come in precedenza impongo delle condizioni diffuse cioè $a_{1|0} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$

$$\text{e } P_{1|0} = \begin{bmatrix} \infty & 0 & 0 \\ 0 & \infty & 0 \\ 0 & 0 & \infty \end{bmatrix}.$$

5.4 Esempio

Proviamo a mettere sotto forma state space tutte le caratteristiche sopra elencate ho come equazione

$$\text{di transizione } \begin{bmatrix} \mu_{t+1} \\ B_{t+1} \\ \psi_{t+1} \\ \psi_{t+1}^* \\ \gamma_{t+1}^1 \\ \gamma_{t+1}^2 \\ \gamma_{t+1}^3 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \rho * \cos(\lambda_1) & \rho * \sin(\lambda_1) & 0 & 0 & 0 \\ 0 & 0 & -\rho * \sin(\lambda_1) & \rho * \cos(\lambda_1) & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & -1 & -1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix} * \begin{bmatrix} \mu_t \\ B_t \\ \psi_t \\ \psi_t^* \\ \gamma_t^1 \\ \gamma_t^2 \\ \gamma_t^3 \end{bmatrix} + \text{errori la}$$

$$Q = \begin{bmatrix} \sigma_\eta^2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma_\zeta^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_K^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_K^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_w^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \text{ per } a_{1|0} = 0 \text{ per } P_{1|0} \text{ ha sia condizioni diffuse che non nelle loro}$$

rispettive posizioni, si possono capire dalla matrice di prima.

Come equazione di misurazione ho $Z = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 0 & 0 \end{bmatrix}$.

6 Regressori nelle forme state space

6.1 Regressione statica

Sia Y_t la variabile risposta e X_t variabili dei regressori, sia $Y_t = \mu_t + X_t' * \alpha_t + \epsilon_t$ vedo che i coefficienti variano nel tempo, quindi se li voglio in forma state space devo specificare $\mu_{t+1} = \mu_t$, $\alpha_{t+1} = \alpha_t$, $T = I_k$ e $Z = \begin{bmatrix} 1 & X' \end{bmatrix}$, $Q = 0_k$ cioè una matrice di 0, $H = \sigma_\epsilon^2$, ho come condizione iniziale le diffuse quindi $a_{1|0} = 0$ e $P_{1|0} = \infty$ questa è anche interpretabile come una prior gaussiana.

Sapendo l'equazione di Y_t in forma state space posso racchiudere in $C_t = B' * X_t$, e vengono trovati massimizzando la massima verosimiglianza, viene utilizzato poche volte esclusivamente se ho una forma UCM molto complessa.

Se invece voglio una regressione con coefficienti che evolvono come random walk, cioè $Y_t = X_t' * \alpha_t + \epsilon_t$ dove $\alpha_{t+1} = \alpha_t + \eta_t$ dove gli errori sono $WN(0, \sigma_\eta^2)$ (sigma è una matrice diagonale), tale modello è utile su serie storiche molto lunghe infatti l'eterogeneità nel tempo è presente in periodi lunghi.

Si può evolvere anche in maniera liscia usando il solito metodo dello smooth trend, ho quindi il solito Y_t però $\alpha_{t+1} = \alpha_t + B_t$ e $B_{t+1} = B_t + \zeta_t$ dove ζ_t è un white noise, quindi creo un percorso più smooth.

In questo caso vedo che $T = \begin{bmatrix} I_{k+1} & I_{k+1} \\ 0 & I_{k+1} \end{bmatrix}$, $Q = \begin{bmatrix} 0_k & 0_k \\ 0_k & \sigma_\zeta^2 \end{bmatrix}$, $Z = \begin{bmatrix} X' & 0' \end{bmatrix}$ (matrice riga di 0) e $H = \sigma_\epsilon^2$.

6.2 Regressione dinamica

La funzione di trasferimento è un filtro ARMA ad un regressore che possiamo osservare, so che $W_t = \frac{W_r(B)}{\delta_s(B)} X_t$, dove se sviluppiamo $w_r(B) = w_0 + w_1 * B + \dots + w_r * B^r$ (pseudo MA) esso modella il breve periodo, invece $\delta_s(B) = 1 - \delta_1 B - \dots - \delta_s * B^s$ ed esso descrive il lungo periodo ed è uno pseudo AR.

Prendiamo come esempio il caso che MA(0) e AR(1) ottengo quindi $W_t = w_0 * X_t + \delta_1 * W_{t-1}$ quindi l'effetto decade in maniera esponenziale se $\delta_1 < 1$, se no diventa degenere, è un caso molto usato forse

il più usato, ovviamente più il delta è vicino a 1 più il decadimento è lento. Inoltre le regressioni dinamiche e quelle statistiche possono essere combinate per creare misure ancora più accurate.

Proviamo a scrivere un modello $AR(1)$ in forma state space ho $\alpha_{1,t+1} = \alpha_{1,t}$ quindi una costante ed il nostro w_0 , e $\alpha_{2,t+1} = \alpha_{1,t} * X_{t+1} + \delta_1 * \alpha_{2,t}$ e come equazione di osservazione $Y_t = UCM + [0 \ 1] * \begin{bmatrix} \alpha_{1,t} \\ \alpha_{2,t} \end{bmatrix} + \epsilon_t$, ho $T = \begin{bmatrix} 1 & 0 \\ X_{t+1} & \delta_1 \end{bmatrix}$ e $Q = 0$ e $H = \sigma_\epsilon^2$.

Proviamo a ricostruire la forma state space di un $AR(2)$ $W_t = w_0 * X_t + W_t = w_0 * X_t + \delta_1 * W_{t-1} + \delta_2 * W_{t-2}$, ho come equazione di transizione $\alpha_{1,t+1} = \alpha_{1,t}$ $\alpha_{2,t+1} = \alpha_{1,t} * X_{t+1} + \delta_1 * \alpha_{2,t} + \delta_2 * \alpha_{3,t}$ e infine $\alpha_{3,t+1} = \alpha_{2,t}$, ho quindi $T = \begin{bmatrix} 1 & 0 & 0 \\ X_{t+1} & \delta_1 & \delta_2 \\ 0 & 1 & 0 \end{bmatrix}$ e $Z = [0 \ 1 \ 0]$ e $H = \sigma^2$.

Proviamo se è il numeratore di ordine 1 quindi $MA(1)$, cioè $W_t = \alpha_{1,t} * X_t + \alpha_{2,t} * X_{t-1}$ con coefficienti costanti ho una normale regressione.

Posso invece rappresentarla come $\alpha_{1,t+1} = \alpha_{1,t}$ e $\alpha_{2,t+1} = \alpha_{2,t}$ infine $\alpha_{3,t+1} = X_{t+1} * \alpha_{1,t} + X_t * \alpha_t$, quindi ho $T = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ X_{t+1} & X_t & 0 \end{bmatrix}$ e $Z = [0 \ 0 \ 1]$ e $H = \sigma_\epsilon^2$ e $Q = 0_n$.

Ora immaginiamo un $ARMA(1,1)$ ho $W_t = w_0 * X_t + w_1 * X_{t-1} + \delta_1 * w_{t-1}$, quindi in forma state space ho $\alpha_{1,t+1} = \alpha_{1,t}$, $\alpha_{2,t+1} = \alpha_{2,t}$, $\alpha_{3,t+1} = X_{t+1} * \alpha_{1,t} + X_t * \alpha_{2,t} + \delta_1 * \alpha_{3,t}$, dove $T = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ X_{t+1} & *X_t & \delta_1 \end{bmatrix}$ e $Z = [0 \ 0 \ 1]$ e $H = \sigma_\epsilon^2$.

Come ultimo ho un $ARMA(2,2)$ quindi $W_t = w_0 * X_t + w_1 * X_{t-1} + w_2 * X_{t-2} + \delta_1 * w_{t-1} + \delta_2 * w_{t-2}$, avrò come equazione di transizione $\alpha_{1,t+1} = \alpha_{1,t}$, $\alpha_{2,t+1} = \alpha_{2,t}$, $\alpha_{3,t+1} = \alpha_{3,t}$, $\alpha_{4,t+1} = X_{t+1} * \alpha_{1,t} + X_t * \alpha_{2,t} + X_{t-1} * \alpha_{3,t} + \delta_1 * \alpha_{4,t} + \delta_2 * \alpha_{5,t}$, infine $\alpha_{5,t+1} = \alpha_{4,t}$,

ho $T = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ X_{t+1} & X_t & X_{t-1} & \delta_1 & \delta_2 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$ e $Z = [0 \ 0 \ 0 \ 1 \ 0]$ e $H = \sigma_\epsilon^2$.

6.2.1 Regressori Dummy

Nella realtà però spesso i regressori dinamici sono delle dummy, chiamiamo D_t tale regressore, ho un $AR(1)$ allora ottengo $W_t = W_0 * D_t + \delta_1 * W_{t-1}$ possiamo esprimerlo come prima oppure in questa altra maniera, sia equazione di transizione $\alpha_{1,t+1} = \phi_t * \alpha_{1,t}$ dove $\phi_t = 1$ se $t < T$ quando diventa maggiore ho $\phi = \delta_1$, costruisco $Y_t = Z_t * \alpha_{1,t} + \epsilon_t$, dove $Z_t = 0$ se $t < T$ quando diventa maggiore ho $Z_t = 1$, noto che ho usato una sola variabile di stato e ciò è più efficiente.

Sempre in quest'ottica possiamo scrivere il modello $AR(2)$ sempre formato da dummy D_t dato il modello $W_t = W_0 * D_t + \delta_1 * W_{t-1} + \delta_2 * W_{t-2}$ è rappresentata come $\alpha_{1,t+1} = \phi_{1,t} * \alpha_{1,t} + \phi_{2,t} * \alpha_{2,t}$ dove $\alpha_{2,t+1} = \alpha_{1,t}$, dobbiamo definire $\phi_1 = 1$ se $t < T$ se maggiore $\phi_1 = \delta_1$, $\phi_2 = 0$ se $t < T$ se maggiore $\phi_2 = \delta_2$, ho quindi $Y_t = [\gamma_t \ 0] * \begin{bmatrix} \alpha_{1,t} \\ \alpha_{2,t} \end{bmatrix}$ dove $\gamma_t = 0$ se $t < T$ se maggiore $\gamma_t = 1$.

7 Filtro di Kalman

Sia $I_t = Y_t, \dots, Y_1$ informazione sulla serie storica, sia $a_{t|s} = P(\alpha_t | I_s)$ oppure se gaussiano $a_{t|s} = E(\alpha_t | I_s)$, sia matrice di varianza e covarianza $P_{t|s} = E((\alpha_t - a_{t|s})(\alpha_t - a_{t|s})')$. Si parla di previsore se $t > s$, si parla di filtro se $t = s$ e di smoother se $t < s$, avendo lo smoother più dati per fare le previsioni ha una varianza sempre minore e uguale alla categoria dei filtri, inoltre l'ultima stima sia per lo smoother e per il filtro sono uguali.

Il filtro di kalman che lavora in 2 fasi, la prima di previsione che ci permette da $a_{t|s}$, $P_{t|s}$ a $a_{t+1|s}$, $P_{t+1|s}$, e la fase di aggiornamento che ci permette di passare da $a_{t|t-1}$, $P_{t|t-1}$ a $a_{t|t}$, $P_{t|t}$ cioè ci fa ottenere il filtro.

Ricordiamoci che inizializziamo con $a_{1|0}$ e $P_{1|0}$ ottenuti dalla forma state space, con l'aggiornamento otteniamo $a_{1|1}$ e $P_{1|1}$ poi con la previsione previsione $a_{2|1}$ e $P_{2|1}$ ripeto questi passaggi fino ad ottenere $a_{n|n}$ e $P_{n|n}$.

Ricordiamoci le due equazioni della forma state space $Y_t = c_t + Z_t * \alpha_t + \epsilon_t$ e $\alpha_{t+1} = d_t + T_t \alpha_t + \nu_t$ imponiamo che ϵ_t sia indipendente da ν_s , imponiamo che sia un processo gaussiano quindi ϵ_t e ν_s sono $N(0, H_t)$ e $N(0, Q_t)$, quindi il nostro previsore lineare è anche ottimo, infine α_1 è indipendente da ν_t e ϵ_t .

Per dedurre il filtro di kalman iniziamo con l'analizzare lo step **predittivo** so che $a_{t+1|t} = E(\alpha_{t+1}|I_t) = d_t + T_t * E(\alpha_t|I_t) + E(\nu_t|I_t) = d_t + T_t * a_{t|t}$, visto che $E(\nu_t|I_t) = 0$ essendo indipendente dal passato e a media 0. Prevediamo $P_{t+1|t} = E((d_t + T_t * \alpha_t + \nu_t - d_t - T_t * a_{t|t})(identica)')$ ottengo $T_t * E(\alpha_t - a_{t|t})(/)'T_t' + E(\nu_t * \nu_t') = T_t * P_{t|t}T_t' + Q_t$ otteniamo quindi che $P_{t+1|t} = T_t * P_{t|t}T_t' + Q_t$.

L'errore di previsione in questo caso è chiamato innovazione cioè data una previsione $\hat{Y}_{t|t-1} = E(Y_t|I_{t-1}) = E(c_t + Z_t * \alpha_t + \epsilon_t|I_{t-1}) = c_t + z_t * a_{t|t-1}$, allora la sua innovazione è

$i_t = Y_t - \hat{Y}_{t|t-1} = c_t + Z_t * \alpha_t + \epsilon_t - c_t - z_t * a_{t|t-1} = z_t * (\alpha_t - a_{t|t-1}) + \epsilon_t$, da qua posso ricavarmi la matrice di varianza e covarianza di i_t cioè $F_t = E(i_t * i_t') = Z_t * E(\alpha_t - a_{t|t-1}) * (identica)' * Z_t' + E(\epsilon_t * \epsilon_t')$ da qua ne deduco che è uguale a $Z_t * P_{t|t-1} * Z_t' + H_t$, questa matrice serve per costruire le stime di massima verosimiglianza, inoltre l'innovazione è gaussiana e ortogonale al suo passato, infatti i_t è una sequenza di variabili IID di gaussiane.

Come step analizziamo quello di **aggiornamento** cioè passiamo da $a_{t|t-1}$ a $a_{t|t}$, in particolare so che $\begin{bmatrix} \alpha_t \\ Y_t \end{bmatrix} | I_{t-1}$ si dispone come una normale multivariata $N(\begin{bmatrix} a_{t|t-1} \\ \hat{Y}_{t|t-1} \end{bmatrix})$, $\begin{bmatrix} P_{t|t-1} & P_{t|t-1} * Z_t' \\ Z_t * P_{t|t-1} & F_t \end{bmatrix}$ in- oltre voglio ricavare le covarianze con $E((\alpha_t - a_{t|t-1}) * i_t) = E(\alpha_t - a_{t|t-1}) * (z_t * (\alpha_t - a_{t|t-1}) + \epsilon_t)'$ rimane $E((\alpha_t - a_{t|t-1})(\alpha_t - a_{t|t-1})' * Z_t' + E(\alpha_t - a_{t|t-1}) * \epsilon_t)$.

Ora voglio trovare $\alpha_t|Y_t, I_{t-1} = \alpha_t|I_t$ si dispone come una normale $N(a_{t|t}, P_{t|t})$ ora dobbiamo trovare momento primo e momento secondo della normale so che $a_{t|t} = a_{t|t-1} + P_{t|t-1} * Z_t' F_t^{-1} * i_t$ (formula previsore lineare) invece per trovare $P_{t|t} = P_{t|t-1} - P_{t|t-1} * Z_t' F_t^{-1} * Z_t * P_{t|t-1}$ (esempio 5.12 libro), può capitare che le matrici $P_{t|t}$ diventano costanti in esempi particolarmente semplici in particolare in un local level model (random walk con rumore).

7.1 Filtro di Hodrick Prescott

Spesso si usa il filtro di **Hodrick Prescott** per modellare il trend e ciclo, cioè si fitta un polinomio liscio ai dati e quella parte è detta di trend invece le oscillazioni sono modellate come ciclo o stagionalità.

Il polinomio è ottenuto minimizzando la funzione di perdita $\min(\sum_{t=1}^n (y_t - \tau_t)^2 + \lambda(\sum_{t=2}^{n-1} ((\tau_{t+1} - \tau_t) - (\tau_t - \tau_{t-1}))^2))$, cioè vado a minimizzare un mix tra minimi quadrati e la minimizzazione della derivata seconda deve essere piccola, cioè cerco una linea smooth che minimizzi però anche i minimi quadrati.

Se immaginiamo che $Y_t = \mu_t + \epsilon_t$ dove $\mu_t = \mu_{t-1} + B_{t-1}$ e $B_t = B_{t-1} + \zeta_t$, quindi la $VAR(\eta_t) = 0$, allora la miglior previsione lineare basata sull'informazione passata del nostro integrated random walk coincide con il filtro di Hodrick Prescott, infatti sia $P(\mu_t|Y_1...Y_n)$ se $\lambda = \frac{\sigma_\epsilon^2}{\sigma_\zeta^2}$, detto rapporto rumore segnale, ottengo il filtro.

8 Stime di massima verosimiglianza

Ricordiamoci delle osservazione classiche sono I.I.D indicate come $X_1 \dots X_n$ allora so che la funzione di verosimiglianza è $L(\theta) = f(X_1 \dots X_n) = \prod_{i=1}^n f(x_i)$ e quindi la sua logverosimiglianza è $\sum_{i=1}^n \log(f(x_i))$ però nel caso di serie storiche non ho indipendenza anzi le osservazioni sono correlate tra di loro.

Per risolvere ciò adoperiamo la funzione condizionata, cioè $f_{y|x}(y) = f(y|X) = \frac{f(y,x)}{f(x)}$ quindi so che $f(y,x) = f(x) * f(y|x)$. In questo modo posso modellare la verosimiglianza, ricordiamo che $f(y_1 \dots y_n) = \prod_{t=1}^n f(y_t|I_{t-1})$ allora se ipotizziamo che gli shock siano gaussiani anche $f(y_t|I_{t-1})$ lo è, dobbiamo quindi stimare momento primo e momento secondo.

Sapendo che $f(y_t|I_{t-1}) = N(\hat{Y}_{t|t-1}, F_t)$, ricordando che i_t è composta in maniera lineare da gaussiane, sia F_t la matrice di varianza e covarianza allora posso ricavarne le stime di massima verosimiglianza, allora la sua log verosimiglianza è $l(\theta) = \log(2\pi)(-\frac{nP}{2}) * -\frac{1}{2} \sum_t \log(Det(F_t)) - i_t' F_t^{-1} i_t$, ora posso massimizzare la log verosimiglianza ottenendo $\hat{\theta}$.

Essendo lo stimatore di massima verosimiglianza ne acquisisce le proprietà come $\sqrt{n}(\hat{\theta}_n - \theta_0)$ converge in distribuzione a $N(0, I^{-1})$ dove $I = -\lim_{n \rightarrow \infty} 1/n E(\sum_{t=1}^n \frac{\delta^2 \log(f(y_t|I_{t-1}))}{\delta^2 \theta})$ e invece l'informazione campionaria è $\hat{I}_n = -1/n \frac{\delta^2 \sum_{t=1}^n \log(f(y_t|I_{t-1}))}{\delta^2 \theta}$ e queste derivate vengono calcolate nel punto di massimo, quindi vogliamo l'Hessiana nel punto di massimo.

Da qua deduco che $\hat{\theta}_n$ si dispone come $N(\theta_0, (n * \hat{I}_n)^{-1})$ e da qua è semplice costruire dei test di ipotesi, ovviamente ciò vale solo se θ_0 non giace nei bordi parametrici cioè per esempio la varianza deve essere sempre maggiore di 0.

Spesso per gli ottimizzatori numerici cerchiamo la varianza rispetto a ψ^2 o comunque valori sempre positivi così da avere dei valori coerenti.

Metodo delta: Sia $\theta = g(\psi)$ che possono avere anche spazzi metrici diversi, dove la stima di massima verosimiglianza è ψ , sia G la matrice Jacobiana di $g(\psi)$, allora vale che $\sqrt{n} * (\hat{\theta}_n - \theta_0) \rightarrow N(0, G' * I^{-1} * G)$

8.1 Inizializzazione delle variabili di stato se ho stazionarietà

I valori stazionari cioè $a_{1|0} = E(\alpha_1)$ e $P_{1|0} = E(\alpha_1 - a_{1|0})(\alpha_1 - a_{1|0})'$ se ho un processo con elementi stazionari so che vale $\alpha_{t+1} = d^* + T^* \alpha_t^* + \nu_t^*$ allora il suo valore atteso esiste ed è $a_{1|0}^* = d^* + T^* a_{1|0}^*$ quindi $a_{1|0} = (I - T^*)^{-1} d^*$, stimo anche $P_{1|0}^* = E(d^* + T^* \alpha_t^* + \nu_t^* - a_{1|0}^*)(idem)'$ sapendo che $d^* = a_{1|0} * (I - T^*)$ si semplifica in $E((T^*(\alpha_t^* - a_{1|0}) + \nu_t^*)(idem)')$ avrò quindi essendo un processo stazionario $P_{1|0} = T P_{1|0} * t' + Q_t^*$. Sfruttando l'operatore vec, cioè trasforma la matrice in un vettore mettendo una sull'altra le colonne della matrice, e si sa che $vec(A * X * B) = (B' \otimes A) * vec(x)$. Quindi $vec(P_{1|0}) = vec(T P_{1|0} T') + vec(Q) = (T \otimes T) * vec(P_{1|0}) + vec(Q)$ infine ho $vec(P_{1|0}) = (I - T \otimes T)^{-1} * vec(Q)$. Ottengo quindi una formula chiusa per tutti i valori attesi e tutte le varianze

8.2 Smoothing

Data una forma state space classica $Y_t = c_t + Z_t * \alpha_t + \epsilon_t$, $\alpha_{t+1} = d_t + T_t * \alpha_t + \eta_t$, lo smoother calcola: $a_{t|n} = E(\alpha_t|I_n)$ e $P_{t|n} = E((\alpha_t - a_{t|n})(\alpha_t - a_{t|n})')$, detti parametri di state smoothing, dove $n > t$. Ci permette di predire il passato e prendi i valori predetti chiamati signal smoothing $\hat{Y}_{t|n} = c_t + Z_t * a_{t|n}$ ed è la previsione giusta al netto dei disturbi.

Il **Disturbance smoother** ci permette di trovare anomalie nei dati, in particolare calcola la proiezione lineare degli errori quindi $P(\eta_t|Y_1 \dots Y_n)$ o $P(\epsilon_t|Y_1 \dots Y_n)$, se data queste proiezioni $\hat{\eta}_t$ o $\hat{\epsilon}_t$ troviamo che $\frac{\hat{\eta}_t}{\sigma_{\epsilon t a_t}}$, cioè errore standardizzato (residuo ausiliario) per la sua varianza, supera 3 allora possiamo considerare quella un'osservazione anomala.

Vado a standardizzare perchè all'inizio e alla fine della storie storica ho più incertezza e mi porta ad avere varianze differenti nei punti della serie, la varianza è calcolata come $Var(\hat{\nu}_{t|n})$ sarà simile a 0.