Lezione 9 giugno 2023

Visione del file html # *Essentials of Time Series Analysis @ Metodi Probabilistici e Statistici per i Mercati Finanziari (MPSMF) 2021-2022*

Visione grafico "Now, we consider some real time series".

"As discussed above": noi vediamo la traiettoria passata nera, se ne verifica solamente una che ci porta ad oggi. A destra c'è il futuro, non so quale traiettoria si verificherà.

La proiezione migliore è una soluzione di compromesso, la *speranza condizionata del modello*(non dei dati), ovvero la retta rossa. Prima creo il modello, poi faccio la speranza condizionata. Le bande intorno sono le *bande di predizione*, cioè simili al concetto di intervalli di confidenza. Una predizione senza queste bande *non significa nulla*. Banda verde: 90%, banda blu: 95%, banda rossa 99%. Ovvero più è ampia la banda, più dati includo. Sarebbe più accurata una banda dell'80%, per avere più "precisione", anche se ricadono meno traiettorie. Come si costruiscono? E' simile alle bande di confidenza (dove lavoravo su media, non media condizionata). L'idea era:

- Se distribuzione è gaussiana e conosco la varianza, allora ci metto i quantili del 90/95% per la deviazione standard e ci costruisco intervallo.
- Senza conoscere la varianza, uso Student, prendo quantili Student, moltiplico per la varianza stimata e calcolavo intervalli, più ampi.

- Senza varianza nè distribuzione, dovevo sperare di avere tanti dati.

Qui è simile:

- Se conosco distrubuzione, e ho anche il rumore della traiettoria, allora prendo i quantili e faccio le bande di previsione.

- Se non riesco, posso simulare le traiettorie (1000,10000,...) e poi costruisco le bande di confidenza empiriche delle traiettorie simulate. Nell'esempio ce ne sono 3. Tiro giu verticalmente rispetto allo stesso tempo tutti i punti, ho un dataset sull'asse y, ci creo i quantili. Approccio di tipo *bootstrap*, sfruttando tante simulazioni. Devo avere il modello, perchè devo stimare la prima parte della traiettoria, dove avrò degli errori(residui=stima del rumore che colpisce la traiettoria). Invece di dire che distrubuzione abbiano i residui, li uso come rumore per calcolare la traiettoria futura. (E' come un sacchetto coi residui, estraggo residuo con reimbussolamento). Un buon modello *non fitta* perfettamente la traiettoria passata, perchè il residuo è 0, come faccio ad andare avanti? non posso. Il polinomio con cui 'fitto' la traiettoria passata non va bene per le future, perchè più è alto il grado, più è rigido nel futuro. Polinomio deterministico. Un buon modello non fitta perfettamente, ma ha *buoni residui*, cioè i residui sono *stazionari*, *omoschedastici*, *scorrelati*, se possibile *gaussianamente distribuiti* (difficile questo ultimo punto). Dato per buono che abbiamo questi residui, andando avanti, ad ogni step ci metto un residuo passato e lo rimetto dentro l'*urna*.

Procedendo nel file, subito dopo il grafico. Abbiamo serie storica, di cui vogliamo creare processo stocastico N-variato. Costruire modello da 0 è complicato, cerco di prendere modello noto (autoregressivi, media mobile, white noise) ed adattarlo per trasformarlo nel modello che noi vogliamo. N_t è l'errore nell'adattamento, ci deve essere sempre, sennò non posso fare predizioni futuri. E' un processo stocastico, ideale se è white-noise, ovvero pesca allo stesso modo, dalla stessa distribuzione, come un lancio della moneta. $F(t, X_t)$ è funzione deterministica. Y_t è una scrittura formale di un modello, come ad esempio la regressione lineare, che però è un caso specifico. X_t dipende dalle espressioni di Y ritardate, cioè non mi servono variabili esplicative, perchè ho solo dipendenza del passato. Spiego un processo rispetto alla sua storia ed altre variabili adattive.

Un processo predittivo **autoregressivo** dipende da un processo e da un disturbo. Nel processo predittivo **media mobile** dipende solo dagli errori passati.

La 3.2 (cioè Y_n) è una generalizzazione di una regressione. Per determinare F devo scomporla in parti deterministiche, e poi considerare il rumore. Il *trend/ciclo* è deterministica, ma varia nel tempo, la variabilità può essere periodica ma occasionale, ad esempio, se immetto un nuovo prodotto nel mercato, all'inizio la domanda salirà (prezzo sale), per poi scendere (prezzo scende), *non* è legato alla stagionalità, ma alla legge della domande e dell'offerta. La

componente stagionale è dovuta alle *stagioni*, ad esempio i gelati sono più venduti in estate che in inverno, non posso controllarlo. C'è anche nella vendita delle macchine: in estate, quando parto per le vacanze, vorrei avere una macchina in buono stato, oppure a gennaio con la tredicesima, a febbraio/marzo di meno. Scrivo allora come in formula 3.6.

As a consequence, in several cases we try to represent the process $\mathbf{Y}$ by an additive decomposition of the form

$$Y_t = m(t) + s(t) + g(X_t) + N_t, \tag{3.6}$$

for every $t \in \mathbb{T}$, where $g : \mathbb{R}^M \to \mathbb{R}^N$ is an appropriate function. In other cases, especially with reference to models for time series in financial markets, it might be more useful consider a multiplicative decomposition ot the form

$$Y_t = m(t) \times s(t) \times g(X_t) \times N_t, \tag{3.7}$$

for every $t \in \mathbb{T}$.

Un problema ricorrente con le serie storiche è che la *variabilità* varia a sua volta nel tempo. E' un bel problema. Si cerca di trasformare la serie storica tramite funzione di *Box-Cox* (il logaritmo ne è un esempio) per ricondurmi ad altra serie storica, possibilmente con variabilità meno variabile nel tempo, o ancora meglio omogenea. Posso sempre l'inversa.

Grafico "Gold Fixing Price"

Il trend lineare *verde* non va bene. Il trend esponenziale *rosso* (LOESS) va molto meglio, tuttavia la variabilità si assotiglia, quindi devo applicare *Box-Cox*, quale? il logaritmo. Differenza tra logaritmi: stima dei rendimenti. Scrivo modello $dS_t = \mu S_t dt + \sigma S_t dW_t$ dove $dS_t = S_{t+dt} - S_t$, differenza infinitesimale, mi dice che l'incremento che subisce il prezzo dello stock dal passaggio tra questi due tempi è dovuto dal valore attuale dello stock moltiplicata per la costante μ. Se non ci fosse $\sigma S_t dW_t$ sarebbe l'equazione l'esponenziale con soluzione $S_t = e^{\mu t}$. Essendoci la seconda componente, cosa stiamo trattando? $dW_t = W_{t+dt} - W_t$ è la differenza tra due v.a. gaussiane indipendenti, ovvero sempre una v.a. gaussiana. Cioè rappresenta un rumore gaussiano elementare. Questa ampiezza è moltiplicata per il valore del titolo S_t, più aumenta più cresce l'oscillazione.

Matematicamente potrei fare: $\frac{dS_t}{St} = \mu dt + \sigma dWt = dlog(S_\sigma)$

in pratica abbiamo ottenuto un modello più facile da studiare, perchè separo le due componenti. Con questo ultimo passaggio mi riconduco ad un *moto casuale* intorno ad una retta (*moto browniano con drift*), ovvero fisso una rete, e ci oscillo intorno.

> Penso ad una stanza con delle mosche, è un moto browniano. Se apro le finestre c'è uno stream di corrente che se le porta via, quindi c'è una corrente che le porta via, cioè un drift. L'unica zanzara buona è quella morta.

Dalla trasformata cerco di tirare fuori un trend (deterministico, come ad esempio la retta di regressione). Rimane la trasformata a cui ho tirato via il trend/ciclo, ovvero $\tilde{y_t}^0$. Tiro via anche la stagionalità e rimane $\tilde{y_t}^{0,*}$.

1. Explore the possibility to subject $\mathbf{y}$ to an invertible non linear transformation, so called *Box-Cox transformation*, often the *logarithm* or the *square root* transformation, to remove simple form of *heteroskedasticity*, which shows as a pronounced variation of the spread of the points in the time series graph around the regression line. This leads to the *homoskedastic* transformed time series $(\tilde{y}_t)_{t=1}^T \equiv \tilde{\mathbf{y}}$, such that

$$\tilde{y}_t = BC(y_t), \tag{3.8}$$

for every $t = 1, \dots, T$, where $BC : \mathbb{R}^N \to \mathbb{R}^N$ is the Box-Cox transformation considered.

2. Try to remove a mean component in the transformed time series $\tilde{\mathbf{y}}$ by time-regressions (e.g. a linear time-regression) or by smoothing procedures (e.g. a moving average). This leads to the *demeaned* homoskedastic transformed time series $\left(\tilde{y}_t^0\right)_{t=1}^T \equiv \tilde{\mathbf{y}}^0$ such that

$$\tilde{y}_t^0 = \tilde{y}_t - m(t), \tag{3.8}$$

for every $t = 1, \dots, T$.

3. Try to remove a seasonal component from the demeaned time series $\tilde{\mathbf{y}}^0$ by spectral decomposition (e.g. a linear combination of sinusoids) or by a deseasonalizing procedure (e.g. a seasonal average). This leads to the *deseasonalized* demeaned homoskedastic transformed time series $\left(\tilde{y}_t^{0,*}\right)_{t=1}^T \equiv \tilde{\mathbf{y}}^{0,*}$ such that

$$\tilde{y}_t^{0,*} = \tilde{y}_t^0 - s(t), \tag{3.9}$$

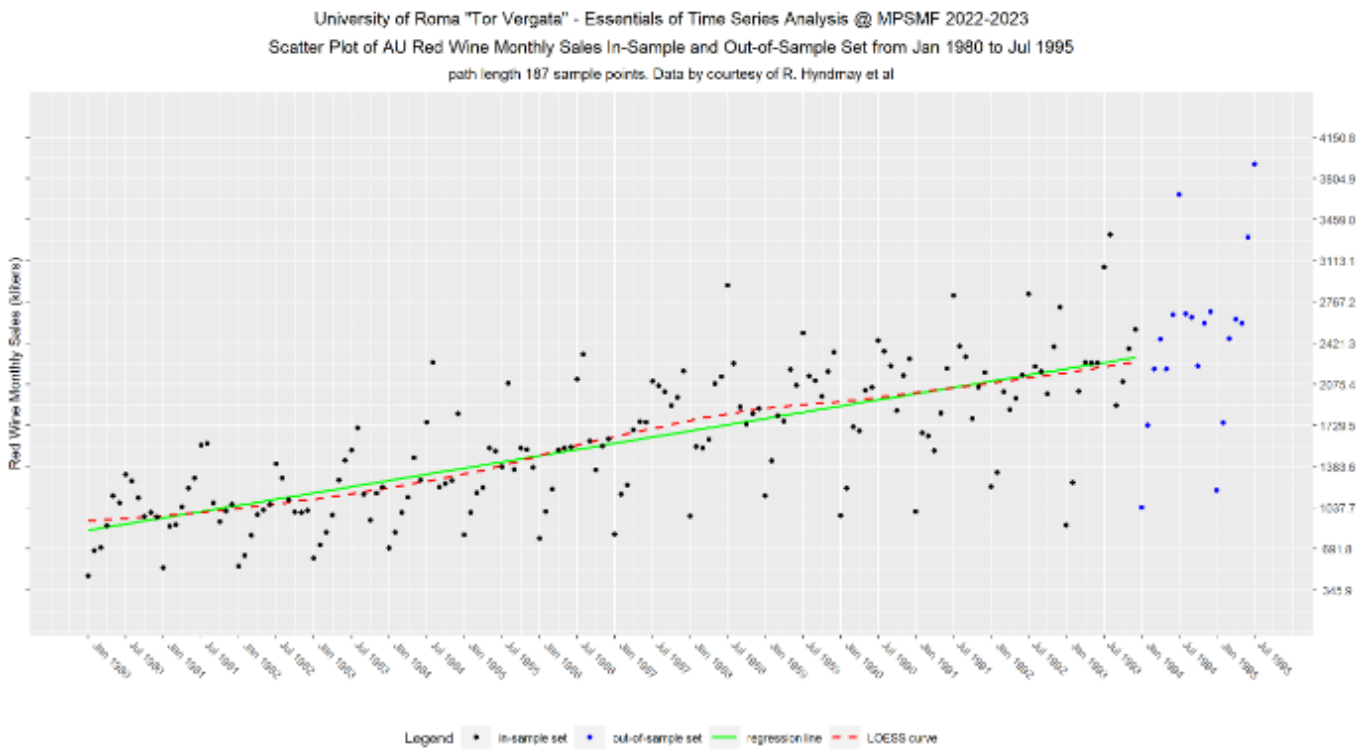
for every $t = 1, \dots, T$.

Questi sono modelli già studiati, come white-noise. Se residuo è esprimibile tramite *white-noise*, non mi rimane molto da fare.

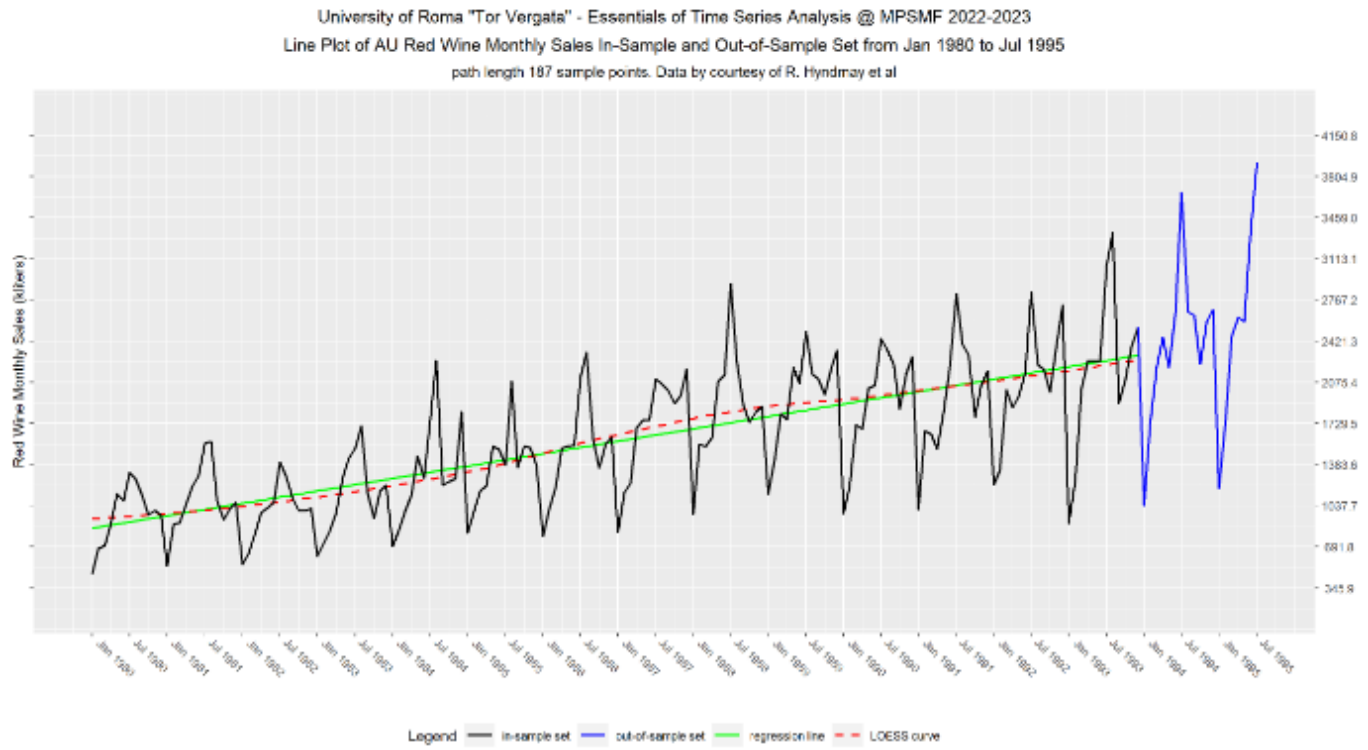
In Finanza, un fenomeno indipendente dalla stagionalità è ad esempio un titolo che cresce con una certa *dolcezza* grazie a delle notizie. Per via di altre notizie/mercato si inizia a vendere il titolo, il valore del prezzo del titolo inizia ad oscillare (*fibrillazione*), catturabile con modelli *GARCH*. Cioè variabilità volatile condizionata dalla variabilità precedente. E' dovuto dall'avversità al rischio, cioè rischio non viene gestito bene e si opera in modo irrazionale.

Analisi di serie storica "First, we split the RWS"

Abbiamo tabelle con indicie, anno, mese, ettoliti di litri venduti. Facciamo un plot con GGPlot.



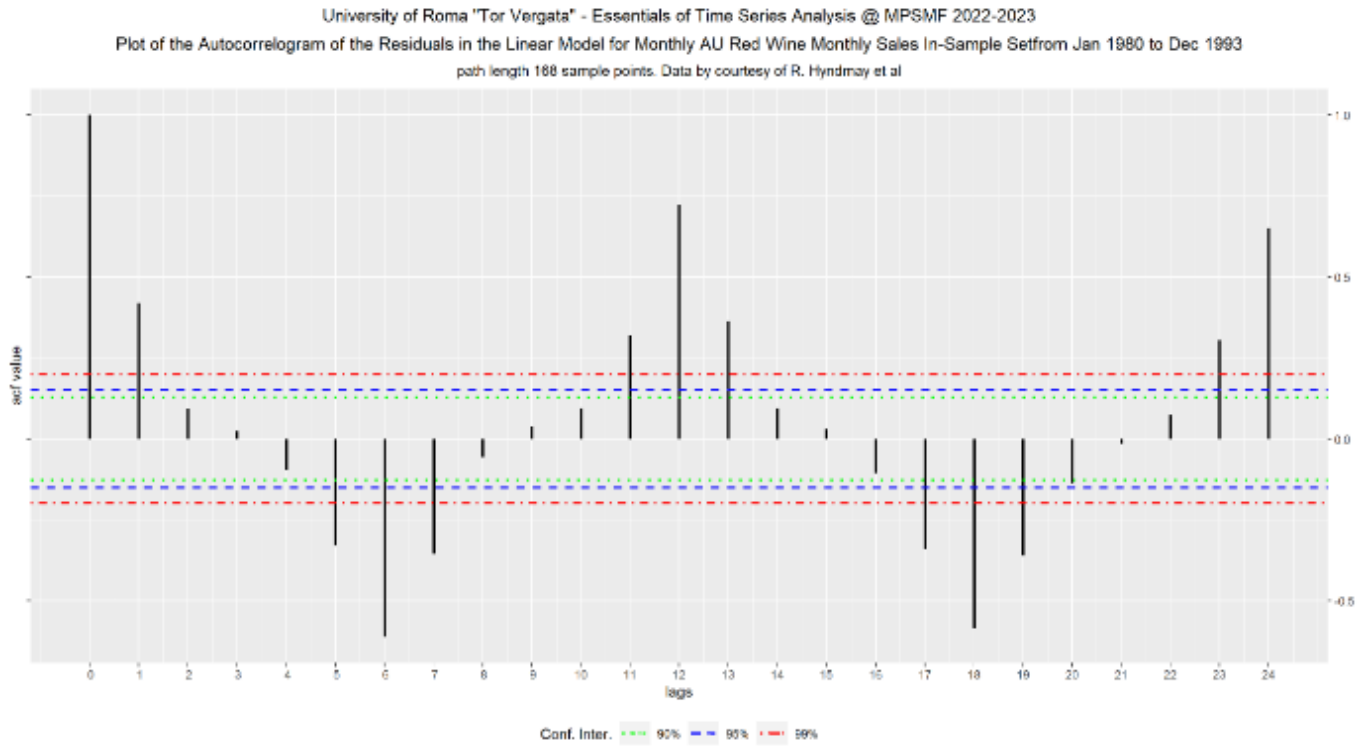
Nel secondo plot, line-plot, vediamo stagionalità, vediamo che nei vari *gennaio* ho picchi verso il basso, a luglio rialzo. Perchè? Perchè siamo in Australia, quindi le stagioni sono invertite, a gennaio sono al mare e ne consumano di più.



Prima del termine della parte nera vediamo retta simile alla LOESS. Quando voglio fare modello predittivo, tolgo la parte blu che sarà il mio test set, quella nera training set. La formula 3.10 $f(t, X_t) \doteq \alpha + \beta t, \quad \forall t \in T$ è predizione lineare, sarebbe stupendo se funzionasse, ma

purtroppo porta a brutti residui. Con una certa stagionalità non conviene mai provare la predizione lineare. Falliamo miseramente.

[.... skippato cose alla velocità della luce]



Residui.