FORECASTING

Il forecasting è l'analisi delle serie temporali (*asset finanziari o tassi di interesse*) allo scopo di creare modelli che predicono il futuro.

- Può riguardare la previsione di un singolo valore futuro o di molteplici valori futuri su una finestra temporale definita (horizon).
- Si applica a diversi scopi, come la previsione dei rendimenti o la previsione della volatilità.

Assunzione di Linearità

Si ipotizza che il **valore futuro** di una serie (ad esempio, il prezzo di domani) possa essere espresso come una **combinazione lineare** dei suoi **valori passati**.

$$L(X_{t+h}) = \sum_{j=0}^{p} \alpha_j X_{t-j}$$

- X_{t+h} è il valore futuro che vogliamo prevedere (al tempo t+h).
- Xt, Xt-1, ... sono i valori passati della serie. **Attenzione!** Questi valori non tengono conto di imprevisti (*come il covid*)
- α0, α1, ... sono dei coefficienti costanti che determinano l'importanza di ciascun valore passato.
- L indica che stiamo costruendo un *predittore lineare*.

L'assunzione di linearità è utile perché i modelli lineari sono molto semplici da analizzare (*specie se le variabili seguono una distribuzione gaussiana*).

Ad ogni modo, ogni serie (Xt) è composta da un componente deterministico (Ht) e una componente di rumore (Yt)

Xt = Ht + Yt

Componente Deterministica

È facilmente prevedibile. Può essere scomposta in 2 parti:

- **Trend:** La tendenza generale a lungo termine della serie a variare. È un movimento "lento" e "dolce". Offre una visione approssimativa della serie.
- **Seasonality:** Pattern che si ripetono sistematicamente a intervalli di tempo fissi (*es. ogni anno, ogni trimestre*)

Una serie può avere solo il trend, solo la stagionalità, entrambe, o nessuna delle due.

L'idea è eliminare dalla serie Xt le componenti Trend e Seasonality, isolando il rumore (Yt) su cui concentrare l'analisi modellistica più avanzata (ARMA).

Pannello "data": Mostra la serie temporale originale (Xt)

Pannello "remainder" (o Residuo): Mostra la componente aleatoria stimata (Yt = Xt - mt - st). Idealmente, questa serie dovrebbe apparire come rumore stazionario (senza trend o pattern stagionali evidenti), fluttuante attorno allo zero.

Si può vedere visivamente come la somma dei pannelli "seasonal", "trend" e "remainder" ricostruisca il pannello "data".

Componente di Rumore (o stocastica)

È imprevedibile e guidata da fattori casuali. Ma possiamo ulteriormente scomporla in 2 componenti

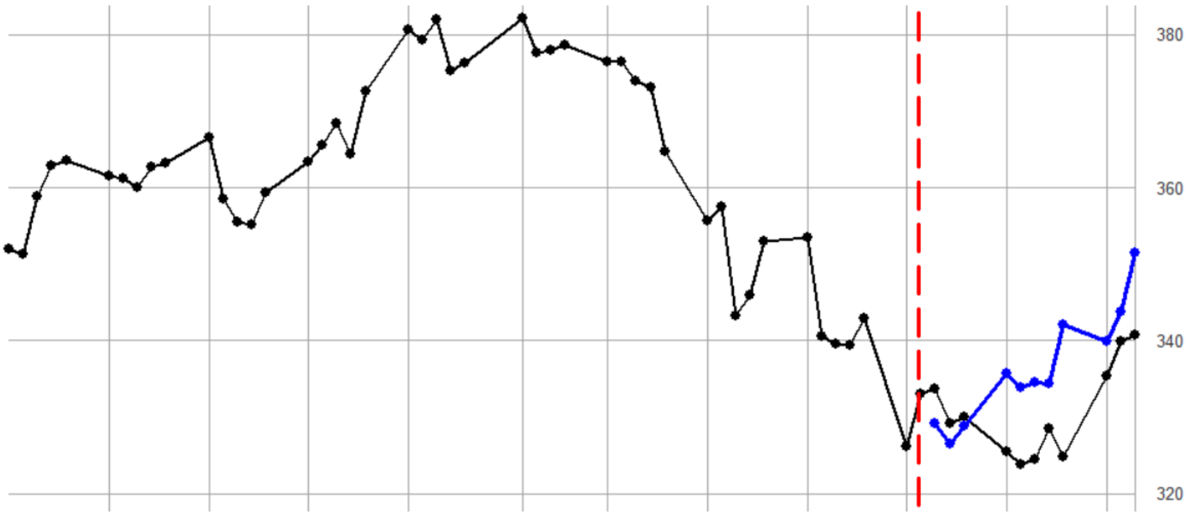
$$Y_t = E(Y_t \mid F_{t-1}) + a_t$$

- F_{t-1} -> tutti i valori passati della serie e altre informazioni.
- E(Yt | Ft-1) -> è la *migliore previsione* che possiamo fare per Yt basandosi solo sulle informazioni disponibili fino al momento precedente (t-1). Questa è la parte "prevedibile" (basata sul passato) della componente aleatoria.
- a_t -> ulteriore residuo del rumore

ARMA

L'assunzione di linearità funge da base per i modelli come ARMA.

ARMA esprime il valore corrente (o futuro) come combinazione lineare di valori passati e/o errori di previsione passati. Serve per analizzare la componente aleatoria (Yt) delle serie.



La linea blu rappresenta la previsione del modello rispetto alla realtà dei fatti. Possiamo vedere che ha previsto male

Stazionarietà

Una serie temporale si dice stazionaria se le sue **proprietà** statistiche fondamentali **non cambiano nel tempo** (media, varianza, covarianza rimangono costanti)

ARMA si basa sull'assunzione che la serie sia stazionaria, altrimenti funziona male. È molto utile infatti per calcolare le serie dei **rendimenti**, che sono tipicamente stazionari.

Peccato che tendenzialmente le serie dei **prezzi degli asset** finanziari siano NON stazionarie (salgono e scendono).

ARIMA

Si tratta di una **generalizzazione del modello ARMA**. Si aggiunge la componente **"Integrated" (I)**, che permette ad ARIMA di gestire serie temporali **non stazionarie**, cosa che ARMA non può fare direttamente.

Il modello ARIMA è definito da tre ordini:

1. **AR - Autoregressive:** parte del modello che assume che il valore corrente della serie Xt dipenda linearmente dai suoi p valori passati
2. **I - Integrated:** parte che si occupa della non stazionarietà. Indica quante **D** volte la serie temporale originale deve essere **differenziata** per renderla stazionaria. La differenziazione consiste nel calcolare la differenza tra un'osservazione e quella precedente (ΔXt = Xt - Xt-1). Se D=0, la serie è già stazionaria e il modello si riduce a un ARMA. Se D=1, si applica la differenziazione una volta, e così via.
3. **MA - Moving Average:** parte del modello che assume che il valore corrente della serie Xt dipenda linearmente dagli errori di previsione passati.

Come si trova l'ordine di D: Si usa il test di stazionarietà (come l'ADF test).

ADF (Augmented Dickey-Fuller)

È uno dei test più comuni per verificare la non stazionarietà.

- **Ipotesi Nulla (H0):** La serie è **non stazionaria**
- **Ipotesi Alternativa (H1):** La serie è **stazionaria**

Dobbiamo trovare prove *contro* l'ipotesi nulla per poterla rifiutare. Quindi ADF esegue una regressione lineare sulla serie Xt:

$$X_t = \alpha + \beta t + \varphi X_{t-1} + \theta_1 \Delta X_{t-1} + \theta_2 \Delta X_{t-2} + \cdots + \theta_{p-1} \Delta X_{t-p+1} + \varepsilon_t$$

Se (φ=0) non ci sono prove, la serie è considerata non stazionaria (H0). Dunque la serie NON tende a tornare verso una media o un trend (random walk = serie si muove per rumori)

Altrimenti c'è un ritorno alla media/trend (φ<0), la serie è considerata stazionaria (H1).

Altrimenti, c'è un modo alternativo per trovare la stazionarietà. Si guarda il **p-value** associato:

- **Se p-value < soglia di significatività (es. 0.05):** Si rifiuta l'ipotesi nulla (H0). Si conclude che ci sono prove sufficienti per dire che la serie è **stazionaria.**
- **Se p-value ≥ soglia di significatività (es. 0.05):** Non si rifiuta l'ipotesi nulla (H0). Si conclude che *non* ci sono prove sufficienti per rifiutare la non-stazionarietà; quindi, si considera la serie **non stazionaria**.

ARIMAX e SARIMAX

ARIMAX (ARIMA with eXogenous inputs):

Gli vengono aggiunte una o più *variabili esogene* (*notizie, utili aziendali o social media*).

Il problema è che usa la regressione lineare, dunque i dati perdono di significato. È necessario avere previsioni anche per le variabili esogene future se si vuole fare forecasting con l'ARIMAX.

SARIMAX (Seasonal ARIMAX):

Risolve i problemi di ARIMAX e aggiunge la capacità di modellare la *seasonality*.

Forecasting volatility

Volatilità = misura della variazione del prezzo di un asset nel tempo.

volatilità è diversa dalla liquidità

La volatilità dei mercati finanziari non è costante nel tempo. Ci sono periodi di alta turbolenza (alta volatilità) e periodi di relativa calma (bassa volatilità).

Il **problema** dei modelli ARIMA è che assumono che la **varianza degli errori sia costante nel tempo** (omoschedasticità). Poiché la volatilità finanziaria cambia (eteroschedasticità), questa assunzione viene violata. Ciò può rendere le stime e le previsioni ARIMA meno affidabili quando si applicano a serie finanziarie.

Limitazioni

È importante notare che l'assunzione di linearità è, appunto, un'*assunzione*. Molte serie temporali reali, specialmente quelle finanziarie, mostrano comportamenti **non lineari** che non possono essere catturati da modelli puramente lineari.

Dunque esistono anche altri tipi di modelli:

- **ARCH/GARCH ->** sono modelli creati per risolvere i problemi della famiglia ARIMA e sono in quindi in grado di catturare la varianza di una serie.
- **Reti Neurali Artificiali (ANN)**
- **k-Nearest Neighbors (kNN)**
- **Support Vector Machines (SVM)**