



Previsione dei mercati azionari: un'estensione dell'approccio Engle-Granger alle tecniche di machine learning ed ensemble learning

Relatore: Chiar.mo

Prof. Valerio Sullo

Laureando:

Alberto Sartini

Background generale

$$I_t = \ln \left(\frac{Y_t}{E_t/P_t} \right) = \ln Y_t - \ln E_t + \ln P_t,$$

dove:

- P_t : S&P 500
- E_t : Utili aziendali
- Y_t : Tasso d'interesse a 10 anni USA

Domanda di tesi

“L'approccio lineare di Engle-Granger è sufficiente per prevedere il rendimento a $t+1$ dell'indice S&P 500, o l'utilizzo di tecniche non lineari di machine learning può migliorare le previsioni?”

Approccio Engle-Granger

1° Step:

$$\ln P_t = c + \beta_Y \ln Y_t + \beta_E \ln E_t + \varepsilon_t$$

2° Step: Forma **ECM**

$$\Delta \ln P_t = \delta + \gamma \varepsilon_{t-1} + \sum_{i=1}^{p-1} \phi_i \Delta \ln E_{t-i} + \sum_{i=1}^{p-1} \vartheta_i \Delta \ln Y_{t-i} + \sum_{i=1}^{p-1} \xi_i \Delta \ln P_{t-i} + u_t$$

Algoritmi di machine learning

- Random Forest
- Support-vector regression (SVR)
- Reti neurali artificiali feedforward (ANN)

Dataset

- **Frequenza:** mensile
- **Variabili:** P, E, Y
- **Intervallo temporale:** 31/01/1945 – 31/12/2023
- **Fonte:** Robert J. Shiller

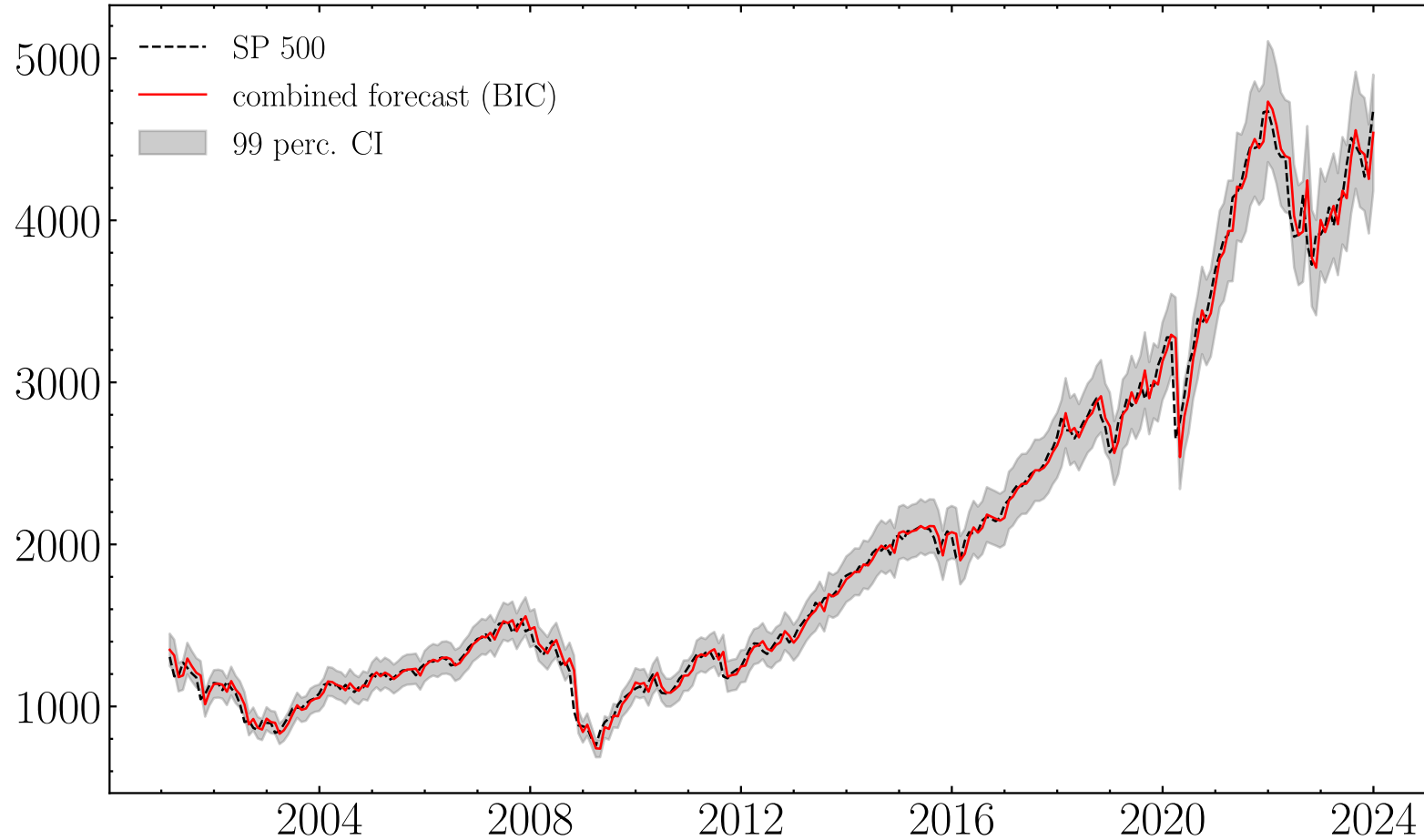
Ciclo di backtest

- 1: Sia $D.backtest = \{x_t | t = 1, 2, \dots, T\}$ il dataset di backtest di dimensione $(T \times k)$.
 - 2: Sia $\mathcal{M} = \{M_1, M_2, \dots, M_n\}$ un insieme finito di modelli, con $i = 1, 2, \dots, n$.
 - 3: **for** ogni istanza x_t del dataset $D.backtest$ **do**
 - 4: Caricamento dati **fino al periodo** t .
 - 5: Stima Approccio Engle-Granger.
 - 6: Memorizzazione forecast Engle-Granger
 - 7: **for** ogni modello $M_i \in \mathcal{M}$ **do**
 - 8: Stima i -esimo modello di Machine Learning.
 - 9: Memorizzazione forecast i -esimo modello di Machine Learning.
 - 10: **end for**
 - 11: Stima modelli di Ensemble Learning.
 - 12: Memorizzazione forecast modelli di Ensemble Learning.
 - 13: **end for**
-

Risultati

Modello	RMSE	Hit Ratio %
Engle-Granger	3.994	62.55
Random Forest	3.946	65.09
Media pesata (test RMSE)	3.938	64.73
Combined Forecast (BIC)	3.912	64.00

S&P 500: modello vs. realtà



Conclusioni e sviluppi futuri

- Punti di forza:
 - Approccio ibrido al problema
 - Accurata selezione delle *features*
- Selezione strategica degli iperparametri:
 - Ricerca bayesiana
 - Approcci metaeuristici

Grazie per l'attenzione.

Modello	RMSE	Hit Ratio %
Engle-Granger	3.994	62.55
Random Forest	3.946	65.09
SVR	4.906	65.82
ANN	4.008	62.91
Media Semplice	3.970	65.09
Media Pesata (training R2)	3.972	64.36
Media Pesata (training RMSE)	3.966	64.73
Media pesata (test RMSE)	3.938	64.73
Best Model (BIC)	3.969	64.73
Combined Forecast (BIC)	3.912	64.00
Best Model (AIC)	3.969	64.73
Combined Forecast (AIC)	3.932	64.00

RMSE

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

HIT RATIO

$$\textit{Hit Ratio}(\%) = \frac{\sum_{i=1}^n \mathbb{1}\left(\text{sign}(y_{\text{true},i}) = \text{sign}(y_{\text{pred},i})\right)}{n} \times 100$$