A Comparative Study of Financial Time Series Forecasting based on Ridge Regression and the Chronos Model

Wang Shuning Sabrina

1. Research Background and Objectives

With the rapid development of financial markets, accurately forecasting stock price movements is of great importance to investors and financial institutions. Traditional statistical methods and recently developed deep learning approaches each have their own advantages in the field of time series forecasting. The objectives of this study are:

- To compare the performance of the Ridge regression model and the Chronos pretrained time series model in predicting stock log returns;
- To investigate the differences between the two models in capturing nonlinear characteristics and temporal dependencies in financial time series;
- To provide an experimental framework for financial time series forecasting that integrates traditional statistical methods with deep learning approaches.

2. Data and Preprocessing

2.1 Data Source

The dataset used in this experiment is obtained from Kaggle and consists of daily closing prices of S&P 500 constituent stocks, covering a total of 501 stocks. To ensure consistency in time length across stocks, only the most recent 250 trading days of data for each stock are retained.

The original data are stored in a wide-table format. In this study, the data are transformed into a long-table format, where each row corresponds to the observation of a specific stock on a specific day, resulting in a standard panel data structure.

2.2 Data Processing

- **Data cleaning**: The data are converted into panel format and missing values are removed.
- **Return calculation**: Log returns are computed as

$$r_t = \log\left(\frac{P_t}{P_{t-1}}\right)$$

  This form has desirable numerical stability and time additivity and is therefore widely used in financial modeling. To reduce the impact of outliers, returns are clipped to the interval [−0.2,0.2].

- **Feature construction**: Features are constructed based on historical returns:

  - Lagged features: lag_1, lag_2, lag_3, lag_5, lag_10
  - Rolling statistical features: rolling mean (windows of 5, 10, 20) and rolling standard deviation (windows of 5, 10, 20)
    These features are designed to capture short-term momentum effects and changes in volatility, respectively.

- **Target variable**: Next-day log return r_{t+1}.

- **Dataset split**: For each stock, the data are split chronologically into a training set (first 80%) and a test set (last 20%, i.e., the final 50 days). This ensures that no future information is used during prediction.

3. Modeling Methods

- **Ridge Regression**

  Ridge regression is a linear regression model with an L2 regularization term. Its objective function is given by:

$$\min \ \| y - X\beta \|^2 + \alpha \| \beta \|^2$$

  This model mitigates multicollinearity and prevents overfitting in high-noise data, making it a widely used baseline model in financial forecasting tasks.

- **Chronos Time Series Model**

  Chronos is a pretrained time series model based on the Transformer architecture. It learns temporal structures by training on large-scale, general-purpose time series datasets. In this assignment, the most recent 30 days of returns for each stock are used as input context to predict the next-day return. The final prediction is obtained by averaging multiple sampled outputs.

- **Qwen Model**

  The Tongyi Qwen model is accessed via the DashScope API. Historical returns are provided as text input, and the model is asked to generate predictions along with explanations. Since the model does not have an explicit time series modeling structure, it is used only as a heuristic baseline for comparison.

4. Experimental Results and Analysis

4.1 Experimental Results

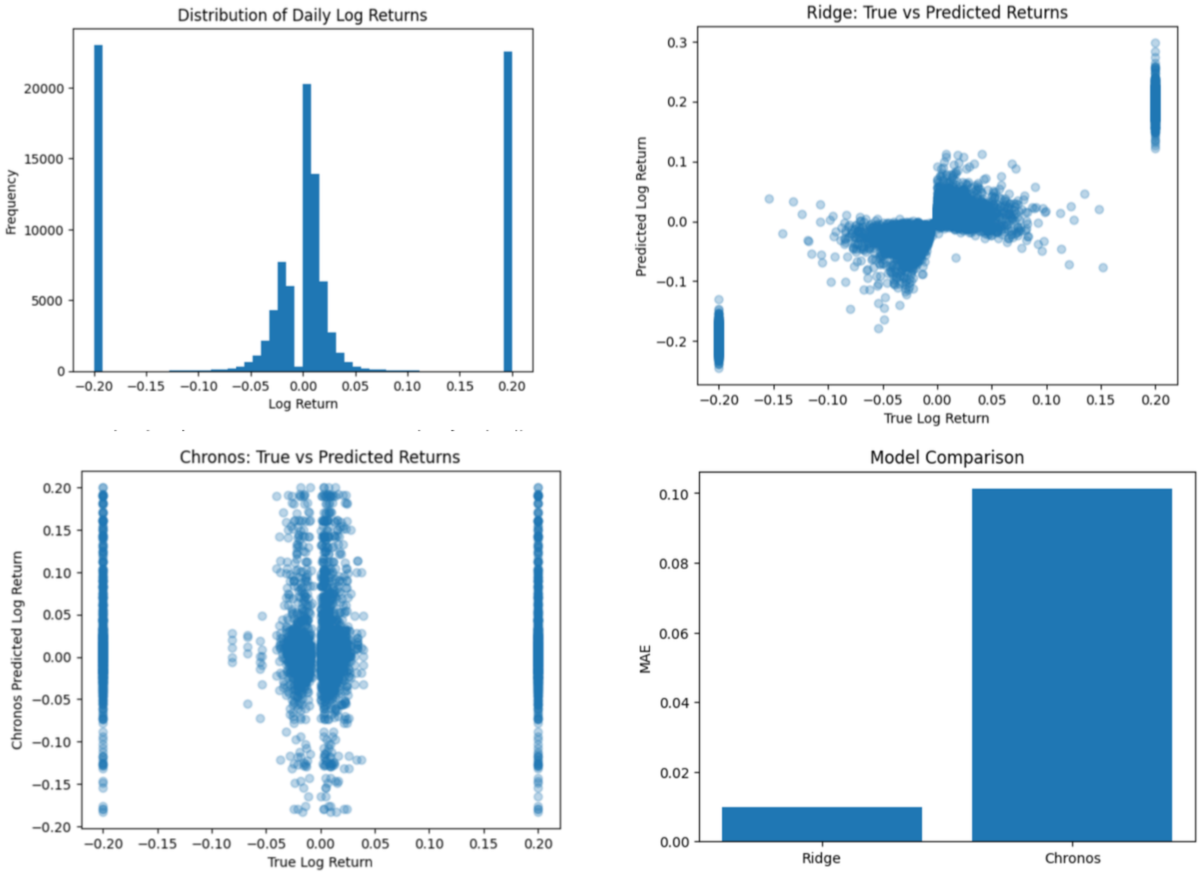
The evaluation metrics used are:

- **MAE (Mean Absolute Error)**: measures the average absolute prediction error;
- **RMSE (Root Mean Squared Error)**: measures the square root of the average squared prediction error.

Model	**MAE**	**RMSE**	**Number of test samples**	**Number of test stocks**
Ridge	0.009996	0.015016	25,050	501
Chronos	0.101196	0.138047	1,000	20
Qwen	No stable improvement	–	–	–

Ridge regression significantly outperforms Chronos in terms of error metrics, while Chronos performs poorly on this task.

4.2 Visualization Analysis



a. The distribution of daily returns shows that returns are highly concentrated around zero, indicating strong noise.
b. The comparison plot between Ridge predictions and true values shows a certain degree of correlation.
c. The comparison plot between Chronos predictions and true values exhibits weak correlation.
d. A bar chart comparing MAE across models further highlights the superior performance of Ridge regression.

5. Discussion

Based on the experimental results, Ridge regression demonstrates clear advantages. This can be attributed to its systematic modeling approach. Carefully designed lagged and rolling statistical features effectively capture short-term memory and volatility characteristics of financial time series, providing high-quality inputs to the model. The inclusion of L2 regularization plays a crucial role in preventing overfitting and improving generalization performance on unseen data. In addition, Ridge regression adopts a global training strategy by pooling data from all stocks, enabling it to learn common market patterns across stocks. Its high computational efficiency also makes it well suited for large-scale financial datasets.

In contrast, the Chronos model performs poorly in this task for several reasons. First, model capacity may be a limiting factor: the t5-small version used has a relatively small number of parameters, which may be insufficient to capture the complex nonlinear patterns and noise structures in financial time series. More fundamentally, there is a domain mismatch issue: the distribution of the pretraining data differs significantly from that of financial return series. Since the pretrained model is used directly without any fine-tuning or domain adaptation, it suffers from poor transferability. From an input perspective, the model relies solely on a single return series from the past 30 days, with no additional features, resulting in limited information. Moreover, predictions are generated via random sampling and averaged, a mechanism intended to reflect predictive uncertainty. However, in highly noisy and weakly structured tasks such as financial return prediction, this approach may introduce additional randomness and obscure already weak predictive signals.

This study also has several limitations. Due to computational and time constraints, the Chronos model is evaluated on only 20 stocks, which may limit the representativeness of the results and the generalizability of the conclusions. The success of Ridge regression largely depends on the quality of manual feature engineering, which requires domain expertise and introduces subjectivity. Furthermore, neither model accounts for broader market regimes (e.g., bull or bear markets), and ignoring such macro conditions may affect robustness under extreme market environments.

6. Conclusion and Future Work

The main conclusion of this study is clear: in the task of predicting next-day log returns for S&P 500 constituent stocks, Ridge regression with effective feature engineering significantly outperforms the untuned Chronos pretrained model. This result suggests that in financial time series forecasting, traditional statistical models combined with carefully designed features remain highly competitive and practical. At the same time, it highlights a key issue: general-purpose pretrained time series models often require domain-specific adaptation and fine-tuning to realize their full potential when applied directly to high-noise domains such as finance.

Future research can proceed along several directions. At the model level, larger Chronos variants (e.g., chronos-t5-large) can be explored, along with systematic fine-tuning on financial time series to bridge the domain gap. Ensemble learning frameworks could also be investigated to combine the stability of traditional models with the ability of deep learning models to capture complex patterns. From an information perspective, the feature set can be further expanded by incorporating classical technical indicators (such as RSI and MACD), market sentiment indicators derived from news or social media, macroeconomic variables, and event-related information. Methodologically, multimodal forecasting frameworks can be developed, and model robustness under different market volatility regimes can be studied in depth. Ultimately, predictive signals should be integrated into concrete investment strategies and rigorously backtested to assess their real economic value.

The practical significance of this assignment lies in providing an empirical reference for method selection in financial time series forecasting. In scenarios with limited computational resources and high requirements for model interpretability, traditional linear models such as Ridge regression remain reliable and efficient choices. In contrast, in data-rich environments where some degree of "black-box" behavior is acceptable and sufficient model-tuning capability is available, well-adapted deep learning models may overcome the limitations of traditional approaches and exhibit greater predictive potential. The choice and combination of these methods should be carefully determined based on specific business needs, data conditions, and resource constraints.