

A COMPARATIVE PREDICTION INTERVAL ANALYSIS ON CLASSICAL FORECASTING WITH BOOTSTRAPPING TO ENHANCE ACCURACY

HUỖNH BỬU KHANG^{1,2} - NGUYỄN THỦY DƯƠNG^{1,2*}

¹Industrial & System Engineering Major, Ho Chi Minh City University of Technology (HCMUT)

Chuyên ngành Kỹ thuật hệ thống công nghiệp, Trường Đại học Bách khoa (HCMUT)

²Vietnam National University Ho Chi Minh City (VNU-HCM)

Đại học Quốc gia Thành phố Hồ Chí Minh

* Email: duong.nt@hcmut.edu.vn

Submission Date: July 6, 2026

Revision Date: July 22, 2026

Publication Acceptance Date: August 5, 2026

Abstract:

Time-series forecasting is widely used in economics and plastics manufacturing, where prediction intervals are essential for quantifying forecast uncertainty and supporting informed decision-making. This study develops a distribution-free bootstrap procedure for constructing prediction intervals for classical time-series forecasting models. Unlike conventional approaches, which typically assume Gaussian errors and emphasize point forecasts, the proposed method generates empirical estimates of predictive uncertainty while enhancing forecast performance. Applied to a case study in plastics manufacturing, the procedure improves overall forecasting accuracy by 9.29%.

Keywords: *block bootstrap, enhanced forecasting, forecasting procedure, plastic manufacturing, empirical prediction interval*

Phân tích so sánh khoảng dự báo của các mô hình dự báo cổ điển bằng phương pháp Bootstrap nhằm nâng cao độ chính xác

Tóm tắt:

Dự báo chuỗi thời gian được ứng dụng rộng rãi trong kinh tế và sản xuất nhựa, trong đó khoảng dự báo đóng vai trò quan trọng trong việc định lượng độ bất định của dự báo và hỗ trợ ra quyết định. Nghiên cứu này tái xây dựng khoảng dự báo cho các mô hình dự báo chuỗi thời gian cổ điển mà không dựa trên các giả định về phân phối. Các phương pháp truyền thống thường dựa vào giả định Gaussian và chủ yếu tập trung vào dự báo điểm. Bằng cách tích hợp kỹ thuật bootstrap, các mô hình cổ điển có thể cải thiện hiệu suất dự báo đồng thời cung cấp các ước lượng thực nghiệm về độ bất định. Khi áp dụng vào trường hợp nghiên cứu trong ngành sản xuất nhựa, quy trình đề xuất cải thiện độ chính xác dự báo tổng thể 9,29%.

Từ khóa: *Bootstrap khối, nâng cao dự báo, quy trình dự báo, sản xuất nhựa, khoảng dự báo thực nghiệm.*

1. Introduction

In plastic manufacturing, accurate forecasts of production demand and material use are essential for reducing overproduction, waste, and environmental impacts. Manufacturing data often exhibit seasonality, irregular fluctuations, and non-stationary behavior, while practical datasets may contain missing values, outliers, irregular sampling, and measurement errors. Therefore, preprocessing procedures such as cleaning, resampling, smoothing, and transformation are important for improving data quality and model stability (Baumann et al., 2023). Classical approaches, including autoregressive integrated moving average (ARIMA), seasonal ARIMA (SARIMA), exponential smoothing, and structural time series models, remain widely used due to their interpretability, computational efficiency, and theoretical foundations.

Although they offer a transparent and interpretable structure, classical models commonly assume normally distributed errors and primarily produce point forecasts (Box & Jenkins, 1976; Harvey, 1989; Hipel & McLeod, 1994). When these assumptions are violated, uncertainty estimates may become unreliable, potentially affecting production planning and decision-making. Prediction intervals therefore complement point forecasts by quantifying the range of plausible future observations at a specified probability level. On the other hand, machine learning and deep learning approaches, including recurrent neural networks (RNNs), Long Short-Term Memory (LSTM), and DeepAR, can capture nonlinear patterns while generating probabilistic forecasts that account for predictive uncertainty (Salinas et al., 2020). Nevertheless, these approaches generally require larger datasets and greater computational resources and offer less interpretability, while classical models remain competitive and can outperform machine learning methods in some cases (Kumar et al., 2024).

Bootstrapping estimates uncertainty by resampling model residuals to approximate forecast distributions without imposing strict distributional assumptions (Efron & Tibshirani, 1994). Unlike parametric approaches, bootstrap-based prediction intervals can reflect observed variability without requiring Gaussian errors. Thus, bootstrap techniques can extend classical models to provide uncertainty estimates while retaining their interpretability and computational advantages.

Accordingly, this study applies bootstrap techniques to classical forecasting models to construct prediction intervals without Gaussian assumptions. The revised models are evaluated in terms of forecast accuracy and uncertainty representation to assess whether bootstrapping can improve predictive performance while providing reliable uncertainty estimates in plastic manufacturing. Section 2 presents the theoretical background, Section 3 describes the case study experimental setup, Section 4 discusses the results while Section 5 concludes the study.

2. Background

2.1. Decomposition-based models

2.1.1. MSTL/STL Decomposition

Seasonal-Trend decomposition using Loess (STL) separates a time series into trend and seasonal components through iterative locally weighted regression, allowing seasonal patterns to evolve gradually while remaining robust to outliers (Cleveland et al., 1990).

For series with multiple seasonal cycles, Multiple Seasonal-Trend decomposition using Loess (MSTL) extends STL by extracting several seasonal components simultaneously (Bandara et al., 2021). Given an observation X_t , Eq.(1) mathematically expresses the MSTL decomposition.

$$X_t = \hat{S}_{1,t} + \hat{S}_{2,t} + \dots + \hat{S}_{n,t} + \hat{T}_t + \hat{R}_t \quad (1)$$

where $\hat{S}_{n,t}$ represent multiple seasonal components while $\hat{T}_t$ covers the trend components and $\hat{R}_t$ is the remainders. This formulation allows the algorithm to capture different seasonal cycles simultaneously, such as daily and weekly patterns.

2.1.2. The SARIMA Model

Seasonal Autoregressive Integrated Moving Average (SARIMA), an extension of the classical ARIMA framework developed by Hipel and McLeod (1994), incorporates explicit seasonal terms in the form of Eq. (2) through backshift operators:

$$\Phi_p(B^s)\Phi_p(B)(1-B)^d(1-B^s)^D Y_t = \Theta_q(B^s)\theta_q(B)\varepsilon_t \quad (2)$$

The variable Y_t denotes the observed value of the series at time t , while ε_t represents the white-noise error component. The model also contains the polynomial terms $\phi_p(B)$ and $\theta_q(B)$, which correspond to the non-seasonal autoregressive (AR) and moving average (MA) components, respectively. Their seasonal equivalents are represented by $\Phi_p(B^s)$ and $\Theta_q(B^s)$, denoting the seasonal AR and MA polynomials. In addition, the operators $(1-B)^d$ and $(1-B^s)^D$ perform non-seasonal and seasonal differencing, respectively, to eliminate non-seasonal and seasonal trends from the series.

2.1.3. The UCM Model

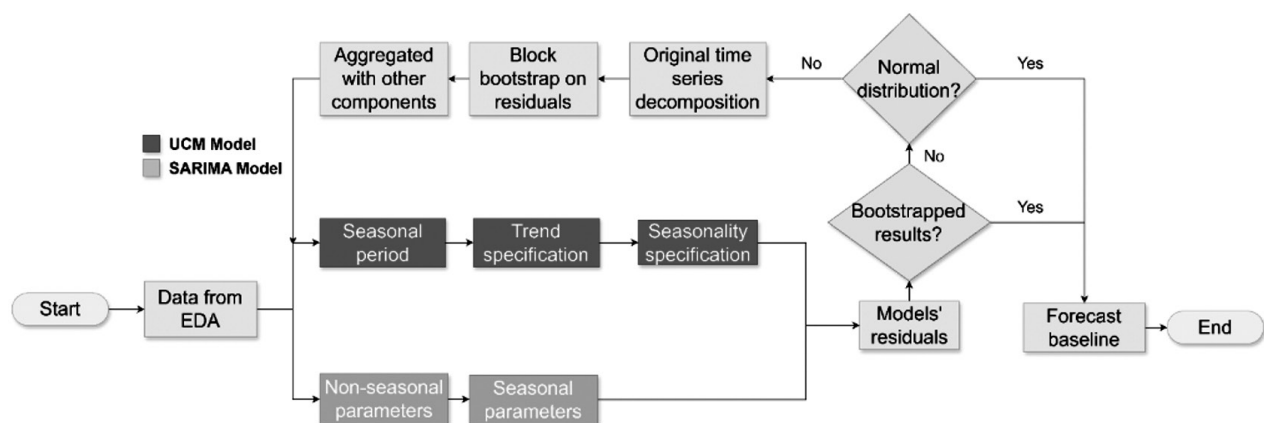
Unobserved Components Model (UCM), or Structural Time Series Model, built by Harvey (1989), decomposes time series into interpretable latent components. The model expresses the observed series in Eq. with μ_t denotes trend, γ_t as the seasonal component, ψ_t are cyclical or other stochastic effects and ε_t represents irregularity.

$$Y_t = \mu_t + \gamma_t + \psi_t + \varepsilon_t \quad (3)$$

Each component is modelled as a state-space process, allowing trends or seasonal patterns to evolve dynamically over time. Parameter estimation and filtering are performed through the Kalman filter (Harvey, 1989), providing optimal state estimates and full uncertainty quantification. Because UCM allows each component to be stochastic, it is particularly useful when underlying structures change gradually or unpredictably.

2.2. Bootstrapping procedure

Figure 1. Bootstrapped classical forecast procedure



Bootstrap resampling provides a flexible means of quantifying forecast uncertainty when the error distribution is unknown or analytically difficult to specify. Introduced by Efron and Tibshirani (1994), it approximates the empirical sampling distribution through repeated resampling, enabling confidence or prediction intervals without strong parametric assumptions. For time series, however, independently resampling observations can destroy serial dependence. The block bootstrap mitigates this issue by resampling consecutive observations in blocks, thereby preserving local temporal dependence and producing more realistic simulated series (Hyndman & Athanasopoulos, 2021).

As shown in Figure 1, the proposed framework combines Exploratory Data Analysis (EDA), MSTL decomposition, block bootstrap resampling, and forecasting to generate probabilistic forecasts. EDA first characterizes the series in terms of trend, multiple seasonal or cyclical patterns, and residual behavior,

after which MSTL separates the series into trend, seasonal, and residual components. Rather than resampling the original observations, overlapping blocks of predefined length are drawn with replacement from the residual component and concatenated to form bootstrapped residual sequences. Each sequence is subsequently recombined with the original trend and seasonal components to reconstruct a simulated time series.

3. Bootstrapped classical models case study

The dataset was collected and cleaned from a plastic manufacturing context, comprising daily observations for mold M_217 from January 3, 2016, to November 30, 2024. The data exhibited a weekly seasonal pattern; therefore, a block bootstrap with a block length of 7 was employed to preserve the underlying weekly dependence structure. The resulting data were subsequently aggregated to a monthly frequency for model development and forecasting. The analysis was conducted on an i7-11800H laptop equipped with an RTX 3050Ti mobile GPU, using the statsmodels and pmdarima Python packages. The models were trained using data through Q3 2024 to forecast the following quarter, while the remaining observations were reserved for testing and evaluation. Forecasting performance was compared using the metrics defined in Eq. (4) to Eq. (8).

$$MAE = \frac{1}{n} \sum_{t=1}^n |Y_t - \hat{Y}_t| \quad (4) \quad MPE = \frac{100\%}{n} \sum_{t=1}^n \left(\frac{Y_t - \hat{Y}_t}{Y_t} \right) \quad (8)$$

$$MAEP = \frac{100\%}{n} \sum_{t=1}^n \left| \frac{Y_t - \hat{Y}_t}{Y_t} \right| \quad (5) \quad \sum_{t=t_0+L}^{t_0+L+S} Z_{i,t} \quad (9)$$

$$WMAPE = \frac{100\%}{\sum_{t=1}^n Y_t} \sum_{t=1}^n |Y_t - \hat{Y}_t| \quad (6) \quad L_\rho(Z_i, Z_i^\rho) = 2(Z_i^\rho - Z_i) \rho I_{Z_i^\rho > Z_i} - (1 - \rho) I_{Z_i^\rho > Z_i} \quad (10)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (Y_t - \hat{Y}_t)^2} \quad (7) \quad \rho - risk = \frac{\sum_i L_\rho(Z_i, Z_i^\rho)}{\sum_i Z_i} \quad (11)$$

While point-forecasts are being validated with the above equations, the ρ -risk metric is utilized for comparing the models quantitatively, following Salinas et al. (2020). First, the window of comparison is considered as $(L, L+S)$ while the target value of item i is defined as shown in Eq.(9). For quantile level ρ , ρ -risk penalizes under-forecasting by ρ and over-forecasting by $1-\rho$, making it consistent with quantile regression. The formula for the ρ -quantile loss is noted in Eq.(10) while the formula for the ρ -risk, which is sum of all ρ -quantile normalized in each span, is formulated in Eq.(11).

4. Results

As reported in Table 1., the revised models were evaluated using mean forecasts, while Table 2. compares ρ -risk across rolling monthly windows (0,1), (1,2), and (0,2). The original, chosen by the company executive from Prophet, SARIMA, UCM, and ETS, produced the largest errors, with MAE of 505,743, RMSE of 521,749, MAPE of 24.92%, and WMAPE of 25.16%. Reconstruction substantially improved both models, with SARIMA achieving MAE of 279,372, RMSE of 279,412, MAPE of 13.94%, and WMAPE of 13.90%, while UCM performed better with MAE of 258,743, RMSE of 262,970, MAPE of 13.01%, and WMAPE of 12.87%. UCM therefore achieved the highest forecast accuracy at 87.13%, compared with 86.10% for SARIMA.

At the 0.9 quantile level, both models yielded low ρ -risk, with UCM achieving the lower average, 0.0138 compared with 0.0145 for SARIMA. SARIMA performed better in (0,1) and (0,2), at 0.0145 and 0.0165, respectively, while UCM was substantially better in (1,2), at 0.0066 compared with 0.0126. Next, at the 0.5 level, UCM again achieved the lower average risk, 0.1319, compared with 0.1396. SARIMA

Table 1. Original and revised procedure error metrics comparison

Error metrics	Original	Revised	
		SARIMA	UCM
MAE (EA)	505,743	279,372	258,743
RMSE (EA)	521,749	279,412	262,970
MAPE (%)	24.92	13.94	13.01
WMAPE (%)	25.16	13.90	12.87
MPE (%)	-5.24	13.94	13.01
% FA (1-WMAPE)	74.84	86.10	87.13

Table 2. Models ρ -risk metric comparison

ρ	Windows	Revised	
		SARIMA	UCM
$\rho = 0.9$	(0, 1)	0.0145	0.0175
	(1, 2)	0.0126	0.0066
	(0, 2)	0.0165	0.0172
	Average	0.0145	0.0138
$\rho = 0.5$	(0, 1)	0.1454	0.1602
	(1, 2)	0.1331	0.1050
	(0, 2)	0.1403	0.1306
	Average	0.1396	0.1319
$\rho = 0.1$	(0, 1)	0.4081	0.4251
	(1, 2)	0.3572	0.2983
	(0, 2)	0.3498	0.3022
	Average	0.3717	0.3419

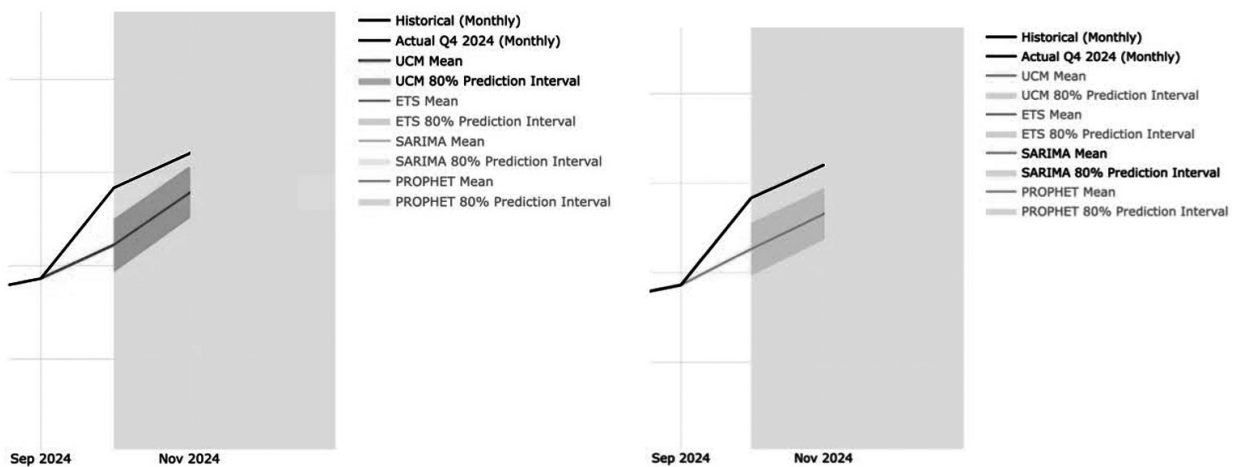
remained superior in (0,1), at 0.1454 compared with 0.1602, whereas UCM performed better in (1,2) and (0,2), at 0.1050 and 0.1306 against 0.1331 and 0.1403. Last with the 0.1 level, UCM likewise recorded the lower average, 0.3419 compared with 0.3717, outperforming SARIMA in (1,2) and (0,2), while SARIMA remained better in (0,1), at 0.4081 compared with 0.4251.

Overall, UCM consistently attained the lower average ρ -risk, although neither model dominated across all windows. SARIMA was strongest in (0,1), whereas UCM showed greater strength in (1,2) and at lower quantiles. Figure 2. shows comparable prediction-interval ranges, with the empirical intervals helping narrow the gap between the under-forecasted mean predictions and observed values, thereby providing useful information for supply-chain decision-making.

5. Conclusion and future works

This study extends the original forecasting framework by applying bootstrap methods to classical models, reducing reliance on Gaussian assumptions while enabling comparison of forecast performance and uncertainty. Consistent with Hyndman and Athanasopoulos (2021), the results demonstrate that classical models can remain effective when appropriately specified. Compared with the original procedure, UCM reduced MAE and RMSE by almost 50%, while reconstruction yielded the reported 9.29% improvement in overall forecasting

Figure 2. Experimental results of UCM (red) and SARIMA (yellow)



accuracy. These findings support bootstrapping as an effective enhancement to classical forecasting in plastic manufacturing. However, the test range remains limited and should be expanded to better assess predictive capability. Future work could investigate ensemble learning to integrate classical and modern methods, while probabilistic forecasting could strengthen comparisons of predictive performance and uncertainty ■

Acknowledgment:

We acknowledge the support of time and facilities from Ho Chi Minh City University of Technology (HCMUT), VNU-HCM for this study.

REFERENCES:

- Bandara, K., Hyndman, R. J., & Bergmeir, C. (2021). MSTL: A seasonal-trend decomposition algorithm for time series with multiple seasonal patterns. arXiv. <https://doi.org/10.48550/arXiv.2107.13462>
- Baumann, E., Hsu, C., Buba, H., & Cox, T. (2023). An introductory approach to time-series data preparation and analysis. Annual Conference of the PHM Society, 15(1). <https://doi.org/10.36001/phmconf.2023.v15i1.3561>
- Box, G. E. P., & Jenkins, G. M. (1976). Time series analysis: Forecasting and control (2nd ed.). Holden-Day.
- Cleveland, R. B., Cleveland, W. S., McRae, J. E., & Terpenning, I. (1990). STL: A seasonal-trend decomposition procedure based on Loess. Journal of Official Statistics, 6(1), 3–73.
- Efron, B., & Tibshirani, R. J. (1994). An introduction to the bootstrap. Chapman & Hall.
- Gneiting, T., & Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. Journal of the American Statistical Association, 102(477), 359–378.
- Harvey, A. C. (1989). Forecasting, structural time series models and the Kalman filter. Cambridge University Press.
- Hipel, K. W., & McLeod, A. I. (1994). Seasonal autoregressive integrated moving average models. In Developments in water science (Vol. 45, pp. 419–462). Elsevier.
- Hyndman, R. J., & Athanasopoulos, G. (2021). Forecasting: Principles and practice (3rd ed.). OTexts.
- Kumar, P., Verma, S., Varma, M., Yadav, K. K., & Singh, A. K. (2024). Comparative analysis of time series forecasting: Traditional statistical approaches vs. machine learning methods. International Journal of Agriculture Extension and Social Development, 7(9), 192–198. <https://doi.org/10.33545/26180723.2024.v7.i9c.1054>
- Montgomery, D. C., Jennings, C. L., & Kulahci, M. (2015). Introduction to time series analysis and forecasting (2nd ed.). Wiley.
- Salinas, D., Flunkert, V., Gasthaus, J., & Januschowski, T. (2020). DeepAR: Probabilistic forecasting with autoregressive recurrent networks. International Journal of Forecasting, 36(3), 1181–1191. <https://doi.org/10.1016/j.ijforecast.2019.07.001>