Forecasting India's Monthly GST Revenue under Structural Breaks: Evidence from SARIMAX, Exponential Smoothing, XGBoost, and an Ensemble

Ghanshyam Kumar Dhuware[1], Dharmendra Sahu[2]
[1]Research Scholar, School of Computer Science, SAM Global University, Bhopal, Madhya Pradesh, India
[2]Professor, School of Computer Science, SAM Global University, Bhopal, Madhya Pradesh, India

Accurate Goods and Services Tax (GST) revenue forecasts are essential for fiscal planning, cash management, and the timely identification of collection shortfalls. This study evaluates whether model complexity improves monthly gross GST revenue forecasting in India when the available public time series is short and contains a major structural break. An 80-observation monthly series covering August 2017-March 2024 was compiled from official Government of India releases. A leak-free chronological design was used: August 2017-March 2022 for initial estimation, April 2022-March 2023 for validation, and April 2023-March 2024 for final testing. Five approaches were compared: a seasonal-naive benchmark, Holt-Winters exponential smoothing, SARIMAX with deterministic trend and a COVID-19 intervention, XGBoost with lag and rolling features, and an inverse-validation-error weighted ensemble. On the held-out test period, SARIMAX achieved the lowest mean absolute error (INR 4,886 crore), root mean square error (INR 8,021 crore), mean absolute percentage error (2.77%), and symmetric mean absolute percentage error (2.87%). Its MAPE was 73.17% lower than the seasonal-naive benchmark. Holt-Winters and the weighted ensemble were competitive but less accurate, while XGBoost performed poorly, with a test MAPE of 12.17%. Residual autocorrelation was not significant at lag 12 (Ljung-Box p = 0.658). The findings show that, for a small public-revenue series with a pronounced intervention, a carefully specified statistical model can outperform a more flexible machine-learning model. The paper provides a reproducible benchmark for extending GST forecasting with macroeconomic and compliance indicators.

Keywords: GST revenue; tax forecasting; SARIMAX; exponential smoothing; XGBoost; structural break; India

JEL classification: C22; C53; H20; H68

How to cite this article: Ghanshyam Kumar Dhuware, Dharmendra Sahu. (2026). Forecasting India's Monthly GST Revenue under Structural Breaks: Evidence from SARIMAX, Exponential Smoothing, XGBoost, and an Ensemble, International Journal of Scientific Modern Research and Technology (IJSMRT), ISSN: 2582-8150, Volume-24, Issue-01, Number-04, July-2026, pp. 27-33, URL: https://www.ijsmrt.com/wp-content/uploads/2026/07/IJSMRT-26070104.pdf

IJSMRT-26070104

I. INTRODUCTION

India implemented the Goods and Services Tax on 1 July 2017 as a destination-based indirect tax that subsumed multiple central and state levies. Because GST is collected monthly and is closely linked to transactions, imports, compliance, and price levels, its revenue series is important for both budget management and high-frequency fiscal monitoring. Monthly GST collections exhibit a rising long-run level, recurring calendar effects, filing-related volatility, and unusual shocks. The collapse in collections during the first COVID-19 lockdown is especially important: models estimated without an explicit intervention can mistake a one-off disruption for a recurring seasonal feature. Conversely, highly flexible machine-learning models may overfit a series containing only a few dozen observations. Existing Indian GST forecasting research has compared exponential smoothing, TBATS, neural networks, and hybrid time-series methods. However, an open empirical question remains: does a nonlinear machine-learning model improve out-of-sample accuracy when the series is short and contains a large policy/economic intervention? This paper answers that question through a transparent comparison of statistical, machine-learning, and ensemble forecasts under a strictly chronological validation design.

The study contributes in four ways. First, it constructs a reproducible monthly gross GST series from official releases. Second, it separates model selection from final evaluation through validation and test windows. Third, it includes a deterministic intervention for the COVID-19 collection shock. Fourth, it reports a policy-relevant negative result: the carefully specified SARIMAX model materially outperforms XGBoost and an error-weighted ensemble.

II. RELATED LITERATURE AND RESEARCH GAP

Tax-revenue forecasting traditionally relies on regression, autoregressive integrated moving-average models, and exponential smoothing. These methods are attractive because they are parsimonious, interpretable, and effective when serial dependence and seasonality dominate the signal. Forecast-combination research suggests that ensembles may improve robustness when component models contain complementary information.

Recent research has introduced neural networks and tree-boosting methods to fiscal forecasting. Thayyib et al. [1] evaluated TBATS, ETS, artificial neural networks, NNAR, and hybrid approaches for Indian GST revenue, emphasizing nonlinear behavior and model combinations. XGBoost [2] offers regularized gradient-boosted trees and can model nonlinear interactions among lagged and engineered predictors. Yet a small monthly sample constrains effective model complexity and increases the risk that flexible models learn transient shocks rather than persistent relationships.

The present study addresses three gaps

- Limited evidence from an explicitly separated validation and test design;
- Limited treatment of the pandemic as an intervention rather than ordinary seasonality; and
- Insufficient reporting of cases in which statistical models outperform machine learning.

The central research question is *Which forecasting approach provides the most accurate one-year-ahead monthly GST revenue predictions under limited data and a major structural break?*

III. DATA AND STUDY DESIGN

The dependent variable is monthly gross GST revenue collected in India, measured in INR crore. The series was compiled from the GST Council revenue archive and Ministry of Finance/Press Information Bureau monthly releases [3-6]. The final dataset contains 80 observations from August 2017 through March 2024. Gross collections include CGST, SGST, IGST, and cess as reported in the monthly release. The use of gross rather than net revenue avoids changes introduced by refund timing, but it does not remove filing and settlement effects.

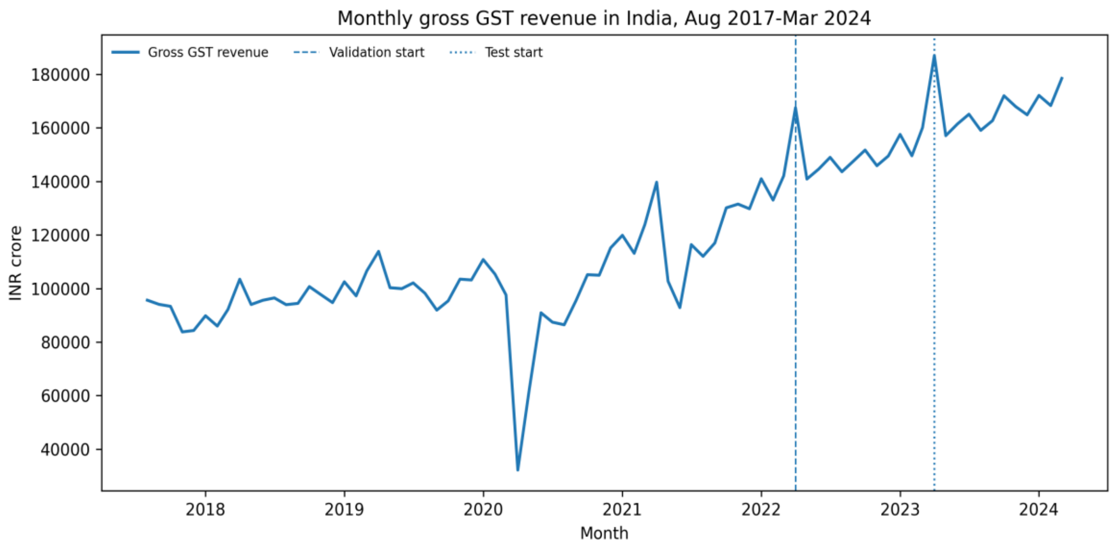

Figure 1. Monthly gross GST revenue and chronological data partitions.

The series rises markedly over time but contains a sharp trough in April-May 2020. By March 2024, the official release reported monthly gross revenue of approximately INR 1.78 lakh crore and FY 2023-24 gross revenue of INR 20.18 lakh crore [6]. Figure 1 also shows the validation and test boundaries used to avoid look-ahead bias.

Table 1. Chronological data partitions.

Partition	Period	n	Purpose
Initial training	Aug 2017-Mar 2022	56	Estimate candidate models and generate validation forecasts
Validation	Apr 2022-Mar 2023	12	Select specifications and ensemble weights
Full training	Aug 2017-Mar 2023	68	Re-estimate selected models before final testing
Test	Apr 2023-Mar 2024	12	Final out-of-sample comparison; untouched during selection

The observed range is INR 32,172 crore to INR 187,035 crore. The April 2020 minimum coincides with the first nationwide lockdown and delayed economic activity. The sample is therefore neither stationary in level nor well represented by a smooth deterministic trend alone. The modeling strategy combines differencing, trend representation, seasonal mechanisms, and an intervention indicator.

IV. METHODS

Seasonal-naive benchmark

The benchmark assigns each forecast the revenue observed in the same month one year earlier. It is difficult to beat when annual seasonality is stable and provides a policy-relevant reference against which added modeling complexity can be assessed.

Forecast (t) = Revenue (t - 12)

Holt-Winters exponential smoothing

An additive-error, additive-trend, additive-seasonality specification with a 12-month seasonal period was selected on validation performance. The method updates level, trend, and seasonal states recursively. Additive seasonality was preferred because the magnitude of within-year variation was more stable than a purely multiplicative pattern after the pandemic shock.

SARIMAX with intervention

Candidate autoregressive integrated moving-average specifications were compared on the validation period. The selected model was ARIMA(2,1,0) with deterministic trend and a COVID-19 intervention indicator. No seasonal ARIMA term was retained after validation, indicating that recent differences and intervention-adjusted dynamics captured the relevant dependence more effectively than an explicit seasonal polynomial.

Phi (B) (1 - B) y(t) = beta 0 + beta1 Trend(t) + beta2 COVID(t) + error(t)

Here, B is the backshift operator; COVID(t) equals 1 from April through September 2020 and 0 otherwise. Parameters were estimated by maximum likelihood. The residuals were assessed with the Ljung-Box statistic at lag 12.

XGBoost

XGBoost was trained as a recursive multi-step forecaster using lags 1, 2, 3, 6, and 12; three- and six-month rolling means; sine and cosine month encodings; a time trend; and the COVID intervention. The fixed hyperparameters were 150 trees, maximum depth 2, learning rate 0.03, subsample 0.90, column subsample 0.90, minimum child weight 2, L2 penalty 10, and L1 penalty 0.1. These conservative settings were chosen to reduce overfitting in the small sample.

Predicted revenue (i) = Sum [k = 1 to K] $f_k(x_i)$

Error-weighted ensemble

The ensemble combined Holt-Winters, SARIMAX, and XGBoost. Weights were proportional to the inverse validation RMSE, so a model with lower validation error received greater weight. The resulting weights were 0.278 for Holt-Winters, 0.553 for SARIMAX, and 0.169 for XGBoost.

w_i = (1 / RMSE_i) / Sum_j (1 / RMSE_j)

Ensemble forecast = Sum_i w_i Forecast_i

Accuracy Measures

Forecasts were compared using mean absolute error (MAE), root mean square error (RMSE), mean absolute percentage error (MAPE), symmetric MAPE (sMAPE), and the coefficient of determination (R^2). MAE and RMSE are reported in INR crore; percentage measures facilitate comparison across periods.

MAE = (1 / n) Sum |Actual(t) - Forecast(t)|

RMSE = Square_root [(1 / n) Sum (Actual(t) - Forecast(t))^2]

MAPE = (100 / n) Sum |(Actual(t) - Forecast(t)) / Actual(t)|

sMAPE = (100 / n) Sum [2|Actual(t) - Forecast(t)| / (|Actual(t)| + |Forecast(t)|)]

V. RESULTS

Validation performance

Table 2. Validation-period forecast accuracy, April 2022-March 2023.

Model	MAE	RMSE	MAPE (%)	sMAPE (%)	R^2
Seasonal Naïve	26,621	28,648	17.85	20.00	-14.341
Holt-Winters	12,753	15,028	8.34	8.83	-3.222
SARIMAX	4,840	7,558	3.06	3.17	-0.068
XGBoost	23,902	24,761	15.74	17.17	-10.461
Weighted Ensemble	9,599	11,672	6.22	6.50	-1.547

SARIMAX produced the lowest validation errors, with MAPE of 3.07% and RMSE of INR 7,558 crore. The ensemble improved substantially on the seasonal-naive and XGBoost models but remained less accurate than SARIMAX. The negative validation R^2 values reflect the limited variance in the twelve validation observations and should not be interpreted as a contradiction of the low percentage errors; for time-series forecast comparison, MAE, RMSE, and MAPE are the primary measures.

Final test performance

Table 3. Held-out test accuracy, April 2023-March 2024.

Model	MAE	RMSE	MAPE (%)	sMAPE (%)	R^2
Seasonal Naïve	17,383	17,531	10.34	10.91	-3.653
Holt-Winters	7,714	10,844	4.51	4.70	-0.780
SARIMAX	4,886	8,021	2.77	2.87	0.026
XGBoost	20,769	22,213	12.17	13.05	-6.471
Weighted Ensemble	7,233	10,144	4.15	4.32	-0.558

The final test results confirm the validation ranking. SARIMAX achieved MAE of INR 4,886 crore, RMSE of INR 8,021 crore, MAPE of 2.77%, and sMAPE of 2.87%. Its MAPE was 73.17% lower than the seasonal-naive benchmark. SARIMAX was also the only approach with a positive test R^2, although the value was modest (0.026).

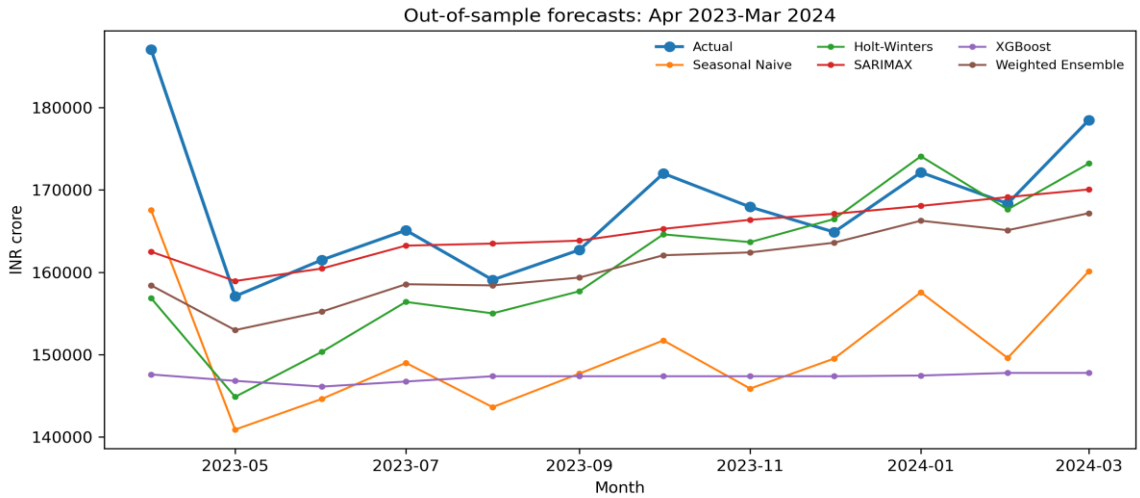

Figure 2. Actual and predicted monthly gross GST revenue during the held-out test year.

Figure 2 shows that SARIMAX follows the level and month-to-month movement of the test series more closely than the alternatives. Holt-Winters tracks the general trend but misses some peaks. The XGBoost recursive forecast is systematically low because tree models do not extrapolate a rising level easily when future observations exceed the training range. The ensemble inherits part of this downward bias.

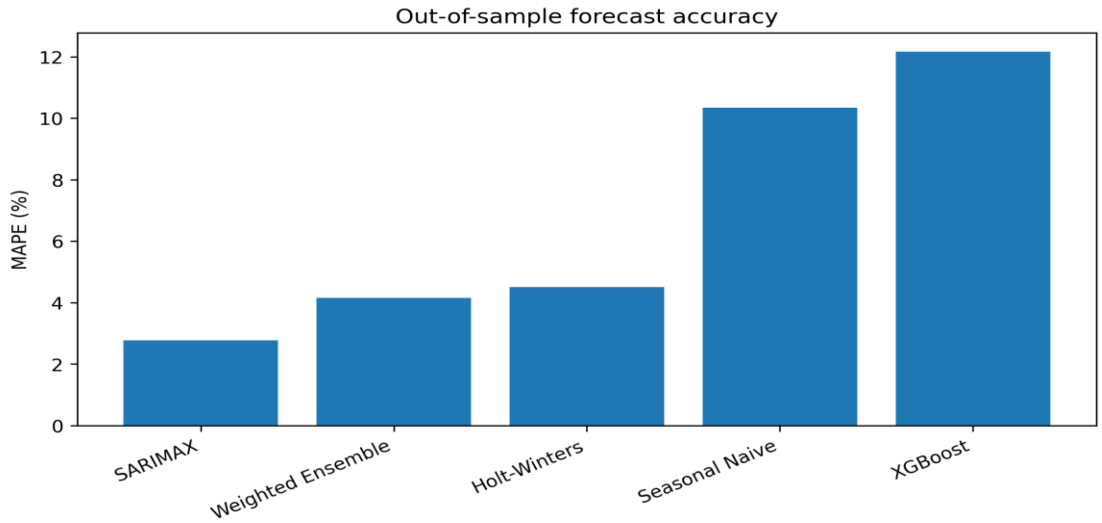

Figure 3. MAPE ranking on the held-out test period.

Residual diagnostics

The selected SARIMAX model had AIC 1399.23 and BIC 1410.10. The Ljung-Box statistic at lag 12 was 9.52 with $p = 0.658$; therefore, the null hypothesis of no residual autocorrelation was not rejected at conventional significance levels. This diagnostic supports the adequacy of the selected dependence structure, while the large residuals around the pandemic period remain visible.

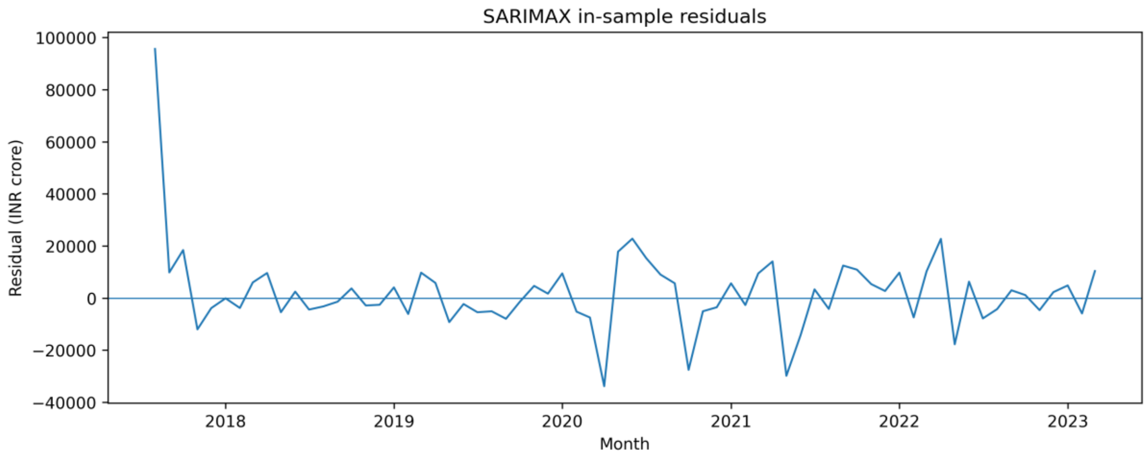

Figure 4. In-sample residuals from the selected SARIMAX model.

VI. DISCUSSION

The principal result is not that machine learning is generally unsuitable for tax forecasting, but that model flexibility must be matched to sample size, data-generating structure, and forecast horizon. The monthly GST series contains only 80 observations, one extraordinary collapse, and a strong upward level shift. Under these conditions, a parsimonious differenced autoregressive model with an explicit intervention provides an effective bias-variance trade-off.

XGBoost underperformance has three practical explanations. First, regression trees predict through partitions of the historical response and are weak extrapolators beyond the observed training range. Second, recursive forecasting compounds errors because each predicted value becomes a future lag. Third, the number of engineered predictors is large relative to the usable post-lag sample. A richer panel of state-level observations or contemporaneous macroeconomic covariates could change this result.

The ensemble did not improve on SARIMAX because forecast combination cannot guarantee gains when one component is materially biased. Inverse-error weighting assigned most weight to SARIMAX, yet the remaining XGBoost component still pulled forecasts below the observed series. This demonstrates why ensemble performance must be verified rather than assumed.

For fiscal managers, the 2.77% test MAPE suggests that a relatively transparent statistical model can provide a useful baseline for monthly monitoring. Forecast deviations larger than the model's typical error can be treated as signals for further investigation, but they should not be interpreted automatically as compliance failures because refunds, deadlines, rate changes, imports, and settlement timing also affect monthly collections.

VII. LIMITATIONS, AND FUTURE RESEARCH

The chronological validation design, untouched final test year, seasonal-naive benchmark, and residual diagnostic improve the credibility of the comparison. Nevertheless, several limitations remain:

- The national aggregate contains only 80 monthly observations, limiting the effective complexity of machine-learning models.
- The model uses revenue history, deterministic trend, seasonal encodings, and a pandemic intervention, but not contemporaneous IIP, inflation, imports, e-way bills, return filings, or policy-rate variables.
- Monthly official figures may be revised or affected by filing extensions, refunds, and administrative timing.
- The final test window contains only 12 observations; performance should be re-evaluated as additional months become available.
- The study reports point forecasts and does not evaluate calibrated probabilistic intervals.

Future research should construct state-level or tax-component panels, assess dynamic regression and Bayesian structural time-series models, incorporate e-way bills and macroeconomic indicators at their publication lags, and compare direct versus recursive multi-step machine-learning forecasts. Forecast reconciliation across CGST, SGST, IGST, cess, and national totals would also be valuable.

VIII. CONCLUSION

This paper compared seasonal-naive, Holt-Winters, SARIMAX, XGBoost, and weighted-ensemble forecasts for India's monthly gross GST revenue. Using a strictly chronological validation and test design, SARIMAX with trend and a COVID-19 intervention delivered the best out-of-sample performance. Its test MAPE of 2.77% was 73.17% lower than the seasonal-naive benchmark, while XGBoost and the ensemble were less accurate. The results show that greater algorithmic complexity does not necessarily improve fiscal forecasts when the available series is short and affected by a major structural break. The provided data and code establish a reproducible benchmark for future multivariate and state-level GST forecasting studies.

REFERENCES

[1] Thayyib, P. V., Thorakkattle, M. N., Usmani, F., Yahya, A. T., & Farhan, N. H. S. (2023). Forecasting Indian Goods and Services Tax revenue using TBATS, ETS, neural networks, and hybrid time series models. Cogent Economics & Finance, 11(2), 2285649. DOI: 10.1080/23322039.2023.2285649

[2] Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785-794. DOI: 10.1145/2939672.2939785

[3] Goods and Services Tax Council. (2024). GST Revenue archive. Ministry of Finance, Government of India. Available at: https://gstcouncil.gov.in/gst-revenue

[4] Government of India, Press Information Bureau. (2024, April 1). Second highest monthly gross GST revenue collection in March at INR 1.78 lakh crore; records 11.5% year-on-year growth. Available at: https://www.pib.gov.in/PressReleasePage.aspx?PRID=2016802

[5] Government of India, Goods and Services Tax Council. (2019). GST Council newsletters and monthly revenue releases, 2017-2021. Available at: https://gstcouncil.gov.in/

[6] Government of India, Ministry of Finance. (2021-2024). Monthly GST revenue collection press releases. Press Information Bureau. Available at: https://www.pib.gov.in/

[7] Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). Time Series Analysis: Forecasting and Control (5th ed.). Wiley.

[8] Hyndman, R. J., & Athanasopoulos, G. (2021). Forecasting: Principles and Practice (3rd ed.). OTexts. Available at: https://otexts.com/fpp3/

[9] Hyndman, R. J., & Khandakar, Y. (2008). Automatic time series forecasting: The forecast package for R. Journal of Statistical Software, 27(3), 1-22. DOI: 10.18637/jss.v027.i03

[10] Gardner, E. S. (1985). Exponential smoothing: The state of the art. Journal of Forecasting, 4(1), 1-28. DOI: 10.1002/for.3980040103

[11] Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2020). The M4 Competition: 100,000 time series and 61 forecasting methods. International Journal of Forecasting, 36(1), 54-74. DOI: 10.1016/j.ijforecast.2019.04.014

[112] Zhang, G. P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing, 50, 159-175. DOI: 10.1016/S0925-2312(01)00702-0