# Tractor Sales Forecasting and Marketing Impact Analysis

## Abstract

This paper presents a comprehensive analysis and forecasting model for tractor sales at PowerHorse, a tractor and farm equipment manufacturer. Using twelve years of monthly sales data, the project applies advanced time series analysis techniques, including decomposition, ARIMA, and SARIMAX modeling, to develop a robust forecasting system. Furthermore, the study investigates the impact of PowerHorse’s marketing programs on tractor sales using regression with ARIMA errors. The results provide actionable insights for production planning, inventory management, and evaluation of marketing effectiveness.

## Introduction

PowerHorse has demonstrated consistent growth since its establishment post-World War II. However, the company faces challenges in controlling inventory and production costs due to sales variability. Effective forecasting of tractor demand is essential for optimizing production planning, reducing costs, and ensuring smooth supply chain operations.

The South-East Asian unit, established 15 years ago, is under particular pressure to optimize planning due to limited supplier bargaining power. Forecasting sales and quantifying the marketing program’s contribution to demand are crucial steps toward ensuring sustainable margins and strategic planning.

## Problem Definition

* **Objective 1:** Develop a robust ARIMA-based forecasting model to predict tractor sales for the next three years.
* **Objective 2:** Evaluate the impact of PowerHorse’s marketing expenditure on sales using ARIMAX modeling.
* **Business Goal:** Enable proactive inventory and production planning, reduce costs, and critically assess the effectiveness of marketing efforts.

## Data Description

* **Source:** Monthly tractor sales figures provided by PowerHorse MIS team.
* **Period Covered:** 12 years of sales data.
* **Additional Dataset:** Monthly marketing expenditures (last 4 years) for impact analysis.
* **Preprocessing:**
  + Extracted month and year from provided data.
  + Created datetime index assuming sales were reported on the first of each month.
  + Cleaned and restructured data for time series analysis.

## Exploratory Analysis

1. **Visualization of Sales Data:**
   * Time series plot revealed a clear upward trend.
   * Seasonality visible with peaks recurring at regular intervals.
   * No significant evidence of long-term cyclic patterns (business cycles typically >7 years).
2. **Decomposition:**
   * Trend: Increasing sales over time.
   * Seasonality: Annual seasonality, peaking around consistent months.
   * Residual: Random fluctuations without long cycles.

## Methodology

### Step 1: Stationarity Analysis

* **Differencing (d=1):** Applied first-order differencing to remove trend.
* **Log Transformation:** Reduced variance instability in the series.
* **Differenced Log Series:** Achieved stationarity in both mean and variance.
* **Conclusion:** Differencing order (d=1) was necessary for ARIMA modeling.

### Step 2: ACF and PACF Analysis

* Plotted autocorrelation (ACF) and partial autocorrelation (PACF).
* Identified presence of AR and MA components in residuals.
* Seasonal spikes at lag 12 confirmed monthly seasonality in tractor sales.

### Step 3: ARIMA Model Selection

* Conducted grid search for parameters (p, d, q)(P, D, Q)s.
* Used SARIMAX implementation from statsmodels.
* Evaluated models using **AIC** and **BIC**.
* **Best Fit Model:** ARIMA(0,1,1)(1,0,1)[12].

## Model Implementation

### In-Sample Prediction

* Fitted best ARIMA model to training data.
* Computed dynamic predictions starting from 2012.
* **Mean Squared Error (MSE):** Low, indicating strong in-sample accuracy.

### Out-of-Sample Forecasting

* Forecasted tractor sales for 36 months (2015–2017).
* Constructed **95% and 99% confidence intervals**.
* Visualized forecast along with uncertainty bands.
* **Insight:** Long-term forecasts carry higher uncertainty; periodic re-training recommended.

### Residual Diagnostics

* Residuals displayed characteristics of white noise.
* **Normality:** Confirmed by KDE and QQ plots.
* **No Autocorrelation:** ACF and PACF confirmed randomness.
* **Conclusion:** Model adequately captured patterns in the data.

## Marketing Impact Analysis

### Correlation Analysis

* Initial scatterplot showed high correlation (>0.8) between marketing spend and sales.
* Differencing revealed that the correlation was spurious due to overall growth.
* Adjusted analysis reduced correlation to ~0.41, indicating weaker true relationship.

### Lagged Effect Investigation

* Tested correlation with lagged marketing expenditures (1–3 months).
* Very weak correlations observed (0.17 for previous month, negligible for quarter).

### ARIMAX Modeling

* Built SARIMAX models with marketing expenditure as exogenous variables.
* Compared AIC values of models with and without regressors.
* **Findings:** ARIMA without marketing expenditure had lower AIC, suggesting that marketing spend was not significantly improving predictive accuracy.

## Results

* **Forecasting:** ARIMA(0,1,1)(1,0,1)[12] model achieved high accuracy and reliable forecasts.
* **Marketing Impact:** Marketing expenditures showed weak correlation with sales and added limited predictive power.
* **Business Insight:** The unit should re-examine marketing strategies, as the impact appears limited in driving sales.

## Discussion

* **Strengths:**
  + Comprehensive forecasting model capturing seasonality and trend.
  + Rigorous statistical validation of residuals.
  + Combined pure ARIMA and ARIMAX approaches for holistic evaluation.
* **Limitations:**
  + Long-term forecasts are uncertain; frequent model updates required.
  + ARIMAX analysis cannot fully substitute structured program evaluation for marketing.
* **Future Work:**
  + Extend forecasting with additional external regressors (e.g., economic indicators).
  + Incorporate advanced models (Prophet, LSTM) for comparison.
  + Implement automated monitoring for forecast updates.

## Conclusion

This project successfully developed a reliable forecasting model for tractor sales and provided insights into marketing impact. The ARIMA model offers practical utility in inventory and production planning. However, the limited influence of marketing spend suggests the need for PowerHorse management to reassess their promotional strategies. Forecasting, when integrated with broader business intelligence, can enable better decision-making and sustainable growth.

## References

* Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time Series Analysis: Forecasting and Control*.
* Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and Practice*.
* Statsmodels Documentation: https://www.statsmodels.org