DATA SCIENCE & ANALYTICS TECHNICAL GUIDE

Understanding SARIMAX Outputs

A step-by-step diagnostic guide for junior analysts on evaluating, interpreting, and refining statsmodels time-series reports.

As a data analyst, fitting a time-series model in Python using `statsmodels` takes just a few lines of code. However, the real engineering begins *after* you run `.summary()`. A model summary is far more than a wall of numbers—it is a diagnostic health report. It tells you whether your model is ready for production deployment, over-fitting the training set, or missing critical underlying signals.

1. High-Level Summary & Model Identity

When you print a model summary, the top section lays out the administrative details, dataset scope, and the exact mathematical structure applied to your target variable.

```
SARIMAX Results
==========================================================================================
Dep. Variable:                  log_sales   No. Observations:                 1628
Model:     SARIMAX(0, 1, 1)x(0, 1, 1, 7)   Log Likelihood                 512.391
Date:                    Wed, 05 Aug 2026   AIC                          -1018.781
Time:                            09:00:40   BIC                          -1002.627
Sample:                        01-01-2013   HQIC                         -1012.785
                             - 06-16-2017
Covariance Type:                      opg
==========================================================================================
```

Key Structural Concepts

- **Target Variable (`log_sales`):** The model forecasts sales transformed via natural logarithms ($\ln$). Taking logs stabilizes variance over time and allows coefficients and errors to be interpreted in percentage terms.
- **Sample & Observations ($N = 1{,}628$):** Spans daily entries over roughly 4.5 years (Jan 1, 2013 to Jun 16, 2017).
- **Model Order $(p, d, q) \times (P, D, Q)_s$:**
  - **Non-Seasonal Component $(0, 1, 1)$:** $p=0$ (no direct autoregressive terms), $d=1$ (first-differencing applied to eliminate trend), and $q=1$ (1 non-seasonal moving average term).
  - **Seasonal Component $(0, 1, 1)_7$:** $s=7$ establishes a **7-day weekly seasonality**. $D=1$ subtracts last week's same-day log sales, and $Q=1$ includes a 7-day seasonal moving average term.

Model Comparison Metrics (AIC / BIC)

Akaike Information Criterion (AIC: -1018.781) & Bayesian Information Criterion (BIC: -1002.627): These metrics score model quality by balancing goodness-of-fit with parameter complexity (penalizing over-parameterization). Absolute values do not matter—when comparing multiple models on the exact same dataset, **the model with the lowest (most negative) AIC/BIC is preferred**.

2. Coefficient Evaluation & Parameter Significance

The middle table shows the estimated weights assigned to each model component along with their standard errors and statistical confidence intervals.

```
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ma.L1         -0.4042      0.016    -24.904      0.000      -0.436      -0.372
ma.S.L7       -1.0149      0.007   -154.566      0.000      -1.028      -1.002
sigma2         0.0297      0.001     40.771      0.000       0.028       0.031
==============================================================================
```

Interpreting the Estimated Parameters

- **Statistical Significance ($P > |z|$):** All terms have $p = 0.000$ (well below the 0.05 threshold), confirming that both the short-term shock parameter (`ma.L1`) and seasonal shock parameter (`ma.S.L7`) are statistically significant.
- **Residual Variance (`sigma2` = 0.0297):** Represents the variance of the white-noise error term. Taking the square root ($\sqrt{0.0297} \approx 0.1723$) reveals that typical daily prediction errors are approximately **17.2%.**

> ⚠️ **Junior Analyst Warning: The Over-Differencing Hazard**
>
> Notice that the seasonal MA coefficient `ma.S.L7` is **-1.0149**. In time-series analysis, when an MA parameter reaches or exceeds **-1.0**, it indicates that the series has been **over-differenced** ($D=1$ was too strong). Over-differencing introduces artificial negative autocorrelation into the series and leads to inaccurate long-term prediction intervals.

3. Residual Diagnostics: Validating White Noise Assumptions

For a time-series forecast to be valid, residual errors must behave as uncorrelated random noise with zero mean and constant variance (White Noise).

```
===================================================================================
Ljung-Box (L1) (Q):                   7.98   Jarque-Bera (JB):              1813.93
Prob(Q):                              0.00   Prob(JB):                         0.00
Heteroskedasticity (H):               1.31   Skew:                             1.00
Prob(H) (two-sided):                  0.00   Kurtosis:                         7.80
===================================================================================
```

Diagnostic Test	Reported Values	Status	Analytical Interpretation
Ljung-Box (Q) Autocorrelation Test	Q = 7.98 **Prob(Q) = 0.00**	FAILED	Rejects the null hypothesis of independence. Leftover temporal patterns remain uncaptured in the residuals at lag 1.
Jarque-Bera (JB) Normality Test	JB = 1813.93 **Prob(JB) = 0.00** Skew = 1.00, Kurtosis = 7.80	FAILED	Residuals are heavily non-normal. Kurtosis of 7.80 (normal = 3.0) and positive skew indicate extreme unmodeled demand spikes.
Heteroskedasticity (H) Variance Stability	H = 1.31 **Prob(H) = 0.00**	FAILED	Error variance grows over time (31% increase from start to end of dataset), violating constant variance assumptions.

4. Step-by-Step Analyst Refinement Workflow

Step 1: Address Over-Differencing (Adjust D parameter)

Because `ma.S.L7` ≈ -1.0, test candidate models setting $D = 0$ with a seasonal AR parameter instead, e.g., `SARIMAX(0, 1, 1)x(1, 0, 1, 7)`. Compare AIC scores.

Step 2: Incorporate Exogenous Features (X)

The high kurtosis (7.80) and failed Ljung-Box test indicate unmodeled external shocks. Add holiday flags, promotional events, and calendar effects as `exog` inputs to absorb outliers.

Step 3: Re-Evaluate Residual Diagnostics

Run standard plot diagnostics (`model_fit.plot_diagnostics()`) to verify that residual ACF plots fall within confidence bounds and Q-Q plots align with the normal line.

Step 4: Forecast & Apply Bias-Corrected Back-Transformation

To convert log predictions back to actual monetary units without underestimating mean expected values, apply half-variance log-normal correction:

$$\hat{y}_{actual} = exp(\hat{y}_{log} + \sigma^2 / 2)$$

💡 Summary & Takeaway

While this SARIMAX(0,1,1)x(0,1,1)$_7$ baseline captures fundamental 7-day seasonality, it suffers from over-differencing and fails residual white-noise tests. Following the 4-step workflow above will produce a robust, production-ready forecasting model.