Forecasting for Fashion Retail Using PyBATS: Incorporating Seasonality, Holidays, and Exogenous Features

Sourish Dey

August 2, 2023

1 Introduction

Accurate demand forecasting is essential for large online retail chains to manage inventory, pricing, and promotions efficiently. In this paper, we explore the **PyBATS** (Bayesian time series analysis) framework for forecasting in a fashion retail setting, which handles seasonality, holidays, and exogenous features such as markdowns, advertising, and discounts. The model used is the **Dynamic Generalized Linear Model** (DGLM), which is effective for time series data with non-Gaussian distributions like retail sales.

2 The PyBATS Forecasting Algorithm

PyBATS specializes in **Dynamic Generalized Linear Models (DGLMs)**, which allow for flexible modeling of time series with Poisson, Bernoulli, or Normal distributions. The algorithm can handle seasonality through harmonic components, account for holidays as dummy variables, and incorporate exogenous features like marketing activities.

2.1 Model Components

The DGLM model is defined as:

$$y_t = F_t' \theta_t + \epsilon_t$$

Where:

- y_t is the observed sales at time t.
- F_t represents the predictor variables, including seasonal components and exogenous features.
- θ_t is the state vector of model coefficients.
- ϵ_t is the error term.

The forecasting steps in PyBATS include:

1. **Model Initialization**: The model is initialized with prior beliefs about the coefficients, incorporating seasonality (e.g., weekly, monthly patterns).
2. **Sequential Updating**: As new data arrives, the model updates its parameters in a Bayesian framework.
3. **Forecasting**: The model generates forecasts by drawing samples from the predictive distribution and calculating point forecasts (mean, median).

3 Seasonality and Exogenous Variables

Retail sales are influenced by **seasonality** (e.g., higher demand during weekends or holidays), as well as external factors like **advertising** and **markdowns**. PyBATS allows for the inclusion of such variables, making it highly effective for complex retail environments.

3.1 Handling Seasonality

Seasonal effects are modeled using harmonic components. For example, weekly seasonality with a 7-day period can be represented by sine and cosine functions:

$$\text{Seasonal Effect} = \beta_1 \sin\left(\frac{2\pi t}{7}\right) + \beta_2 \cos\left(\frac{2\pi t}{7}\right)$$

This allows the model to capture repeating patterns effectively.

3.2 Incorporating Holidays and Special Events

Holidays often cause sharp changes in demand, requiring special handling in the model. PyBATS incorporates holidays as **indicator variables**, which help prevent the model from overreacting to these anomalies

4 Assumptions of PyBATS

The PyBATS model operates under several assumptions:

- **Linearity**: The relationship between predictors and sales is assumed to be linear.
- **Bayesian Framework**: The model operates in a fully Bayesian framework, updating beliefs as new data arrives.
- **Exogeneity**: External variables like promotions and markdowns are assumed to be independent of the response variable, except through their inclusion in the model.

5 Forecasting Process

For a retail chain with thousands of SKUs and hundreds of stores, the forecasting process with PyBATS can be broken down into the following steps:

1. **Data Preprocessing**: Historical sales data, along with exogenous features such as markdowns, advertising, and holidays, is prepared. 2. **Model Fitting**: A **Poisson DGLM** is chosen for count data (sales) and fitted using the **PyBATS analysis function**. 3. **Incorporating Seasonality**: Weekly seasonality and holiday effects are added to improve forecast accuracy. 4. **Forecast Generation**: The model generates forecasts over the desired horizon (e.g., 30 days), and draws samples from the forecast distribution.

6 Evaluation Metrics

The performance of the PyBATS model is evaluated using several accuracy metrics:

- **Weighted Mean Absolute Percentage Error (WMAPE)**:

$$WMAPE = \frac{\sum_{t=1}^{n} |y_t - \hat{y}_t|}{\sum_{t=1}^{n} y_t}$$

- **Symmetric Mean Absolute Percentage Error (SMAPE)**:

$$SMAPE = \frac{1}{n} \sum_{t=1}^{n} \frac{|y_t - \hat{y}_t|}{(y_t + \hat{y}_t)/2}$$

- **Weighted Absolute Percentage Error (WAPE)**:

$$WAPE = \frac{\sum_{t=1}^{n} |y_t - \hat{y}_t|}{\sum_{t=1}^{n} y_t}$$

7 Results

In our application of PyBATS to a fashion retail chain, the model demonstrated strong predictive performance, as shown in Table 1.

Table 1: Forecasting Results

Metric	**Value**
WMAPE	0.33
SMAPE	0.28
WAPE	0.29

The model successfully captured seasonality, holiday effects, and the impact of exogenous variables like advertising and markdowns, providing accurate demand forecasts for the retail chain.

8 Conclusion

The PyBATS framework is well-suited for forecasting in complex retail environments with seasonal and exogenous influences. Its Bayesian approach allows for continuous updating of forecasts, making it highly flexible and accurate for fast-paced industries like fashion retail.