Treatment Effect of Renovation on Retail Store Using Time-Series Regression

Samuel Atkinson

**Intro:**

Soccer Village(SV) Blue Ash location underwent renovations to a new retail store concept completed in June 2020([See store here](https://www.google.com/maps/@39.273266,-84.358258,3a,75y,319.95h,90t/data=!3m5!1e1!3m3!1sjvfqs9RTCDEAAAQXID2NxA!2e0!3e2)). It is the only store in the company that has operated under both the old and new design.

**Objective:**

The goal is to determine if there is a positive average treatment effect (ATE) that can be observed in the 12 months following renovations.

**Data**

Weekly sales data broken down by product group code (PGC) was collected Via MACH POS software. This process was done manually by transferring reports to an Excel Spreadsheet and then running a Python script to reformat the data that could be saved as a .csv file to ultimately be read into a Pandas Dataframe. This process was necessary as historical sales data sampled at the required time intervals cannot be queried with MACH software. The training data is comprised of 206 points with 8 attributes starting in May 2018 and ending in May 2022. Calculating a Treatment Effect requires that the dataset is divided into pre-treatment and post-treatment, or in this case pre-renovation and post-renovation. The pre-treatment data is what will be used for modeling, forecasting, to compare against the actual post-treatment data.

**Tools:**

Initial Data collection was done with Microsoft Excel and initial visualizations with Tableau Public. The main libraries used in Python were: NumPy, Pandas, Matplotlib, StatsModels and Scikit Learn.

**Cleaning and First Visualizations:**

Up until this point I was hoping to model the trend of replica and apparel in the store. The rationale being that these items are considered non-necessity or “add on sales” and would therefore be a good measure of consumer response to renovations. The difficultly with this which became clear in the first visualizations, is that replica sales trends span four years due to the spacing of major professional international tournaments and therefor two years of data would be insufficient for modeling.

Chart, line chart

Description automatically generatedChart, line chart

Description automatically generated

*Figure 1 Caption*

Figure 2 shows footwear (units sold), which accounted for half of the store’s gross sales in 2021, to be more reliably seasonal over a 12-month period.

The effects of COVID-19 were mitigated by eliminating the period where the store was closed as well as the two weeks after reopening from the dataset. Additionally, to account for the drop in sales in the weeks preceding the shutdown I chose a model that most accurately fit the training data’s shape, but which overshot it slightly. Because the effects of COVID-19 took place at the end of the Pre-Renovation portion of the dataset, overfitting the model in this case results an unrealistic negative trend when forecasted beyond the training data.

**Time Series Forecasting with Seasonal Autoregressive Integrated Moving Average (SARIMA)**

ACF and PACF plots are used to specify values for SARIMA hyperparameters.

Chart

Description automatically generated

Chart

Description automatically generated with medium confidence

Graphical user interface

Description automatically generated

An 80/20 training/test split was used for fitting and evaluating the model using NRMSE as the measure.

Graphical user interface, text

Description automatically generatedChart, histogram

Description automatically generated

This Model achieves an NRMSE of .2125. The same hyperparameters were used to forecast from 05/31/2020 to 05/03/21. The only difference being that the entirety of the pre-renovation data set was used as the training data making a 65/33 training split.

**Calculating Average Treatment Effect (ATE)**

Treatment effect for a single discrete event can be written as the following:

Where Dt is a binary variable indicating the treatment status. If untreated Dt = 0 and if treated Dt = 1. ATE for a single discrete event is represented by:

Where yt(1) is the observed footwear unit sales and y(t)(0) is the forecasted footwear unit sales. If we let T = the total number of observations, and t indicate an individual observation. ATE is calculated by;

Graphical user interface, text

Description automatically generated

Chart, line chart, histogram

Description automatically generated

The ATE calculated for the 12-month period following the completion of renovations is 17.074 units/week. The average sale at the Blue Ash store during this time period was $97.07 therefor the total Treatment Effect in sales is $86,186.34