Forecast on the Durables Market in Hungary

Project Summary

Nadezhda Gesheva

20 July 2018

The aim of this joint collaborative project is to create a 12 month forecast on the durables market. Retailers and manufacturers are constantly striving for maximum profitability and customer satisfaction. In order to achieve this, business units rely not only on historical data but also on reliable forecasts so that they would have the intuition of where the markets are heading.

The dataset for the analysis contains data for sales and quantity sold of smartphones on the Hungarian market for 6 years (2012-2017) on a monthly basis. These are also available with certain segmentations. In addition, there are 25 macro variables that will be investigated to see whether there is a correlation between sales and a certain macro parameter.

This overall analysis has 3 main focus areas:

1. Forecasts for total sales through various models – benchmark models; models relying on autocorrelation such as ARIMA and dynamic forecasting, taking into account exogenous(independent) features.

Exploratory Data Analysis on **total sales** data. Model creation for forecast of sales relying only on autocorrelation – benchmark models such as exponential smoothing and state space models, as well as more advanced ones such as ARIMA. The forecast models' performance was compared using cross validation and the best performing one in terms of low cross validation errors were the Error, Trend, Seasonality model (ETS) and ARIMA. Each time series model consists of a function that describes the observed data, and some state equations that describe how the unobserved components or states (level, trend, seasonality) behave over time. Hence, these are referred to as state space models. This is a very fast and reliable method for forecast creation.

The runner up model was seasonal ARIMA with a drift. I extend the ARIMA model for total sales in order to allow other information and not only the sales series to be included in the model. Furthermore, correlation analysis between total sales and a set of independent macro predictors provided by the partner company GfK. This list of 19 features was further expanded by my independent search for additional macroeconomic parameters that influence the sale of smartphones. The total set of predictors reached 25 features. As an additional predictive model, I employed dynamic forecasting by implementing a time series linear regression model with ARIMA errors. The dependent variables were predicted with the help of an Error, Trend, Seasonality (ETS) model, aka state space model. In the dynamic forecasting model I allow the errors from a time series linear regression to contain autocorrelation and hence I create a time series linear regression with ARIMA errors.

2. Sales segmentation for forecasting purposes – three breakdown approaches.

Market segmentation for forecasting purposes

Similarly to the analysis for total sales series, I conducted an identical analysis for the 3 types of sales breakdown:

- internet vs traditional sales;
- low priced products(up to 70 000 HUF) vs higher range priced products(equal to and higher than 70 000 HUF);

- main channel(CSS/OER/TCR) of product purchase vs alternative(Consumer Electronic Stores and Mass Merchandisers/DIYSs).

The goal of this analysis is to see whether a particular sales segmentation would create a better forecast than the forecast on total sales. This conclusion was reached with the help of cross validation with a rolling window for 12 month period. Only the low priced products versus the upper price range products produced forecasts which when summed up had lower error rate than the total sales. The internet and traditional sales' independent forecasts had poorer performance than the total sales since internet sales experienced a huge increase in 2017(the last year in our data sample). This spike was hard to be forecasted accurately by the predictive models hence the poor performance of this sales breakdown. Moreover, the main channel vs secondary channel of product purchase turned out to have similar cross validation errors to the total sales errors. Since we aim for simplicity and accuracy, my recommendation would be to use the total sales series forecast instead of this third way of sales breakdown.

To summarize this section of the analysis, it seems that only 1 market segmentation would bring value added when forecasting sales output with a 12 month horizon – this is the lower priced products vs higher price ranged products.

3. Additional analysis – demand forecasting – benchmark models and ones relying on auto-correlation such as ARIMA provide insight to where to expect demand of smartphones in Hungary in the next 12 months. No dynamic forecasting was attempted due to data length limitation which would have produced unreliable output.

Final Remarks:

The overall analysis was based on the 'KISS' principle - Keep it short and simple. Since, the dataset is available only for 6 years on a monthly basis, we have only 72 observations. The predictive models that I built take this into consideration, hence they aim for simplicity and good performance both on train set and through cross validation errors for a 12 month period.

Overfitting and producing poor forecasts are relevant issues when the data span is limited to a few years. It was important to produce reliable forecasts and suggestions for sales segmentations based on the analysis without falling into the trap of overfitting. In addition, the partner company was able to provide only macroeconomic data on monthly basis but no customer specific which could have been more relevant for determining causality and hence producing better forecasts. These might include but are not limited to:

- Were there certain discount policies at specific time periods (for e-purchases)?
- Was there a large number of new openings of physical locations of retailers?
- Any information about the customers – gender/age/nationality.

Therefore, the analysis should be taken with a grain of salt. Nevertheless, it provides valuable insights to what tomorrow has to offer for the smartphone market in Hungary and is a valuable starting point for further research and data analytics.

∞