Project 1: The Newsvendor Problem

Authors: Brandon Pover (bnp669), Vishal Gupta (vg22846),

Samuel Laplatney (sjl2953), Safiuddin Mohammed (sm76833)

Objective

In this analysis of the newsvendor model, we perform a comparison of different methods of using sample average approximation, a method to solve stochastic programming problems, to approximate future demand and profit. The standard newsvendor model, one that optimizes over profit only to find quantity using past demand data, is not an accurate approximation of reality; thus, extensions to this model will be analyzed for potential benefits.

The first extension will explore the assumption that if the quantity of newspapers printed is less than demand, a rush order can be placed at a larger cost to fill the remaining demand; otherwise, any newspapers printed over demand will be assumed to have a disposal fee. The second extension assumes that price impacts demand linearly with error which allows the model to jointly solve for an optimal price and quantity while adding randomness to past demand data. The third extension is to analyze how sensitive the optimal price and quantity is to the given demand data by using bootstrapped samples of the original data and refitting the linear relationship between price and demand many times.

The objective of this analysis and its extensions of the standard newsvendor model is to make a better approximation of reality and provide a better business plan for larger profits based on simulated randomness and average expectations.

Regression for Demand & Price

In order to estimate demand for our newspaper at different price points, we will assume that price impacts demand linearly with error. To stimulate demand, our first step was to regress demand on price using the dataset we were given. The figure below is a scatterplot of the given demand data and corresponding price data which illustrates the negative linear relationship between the two.

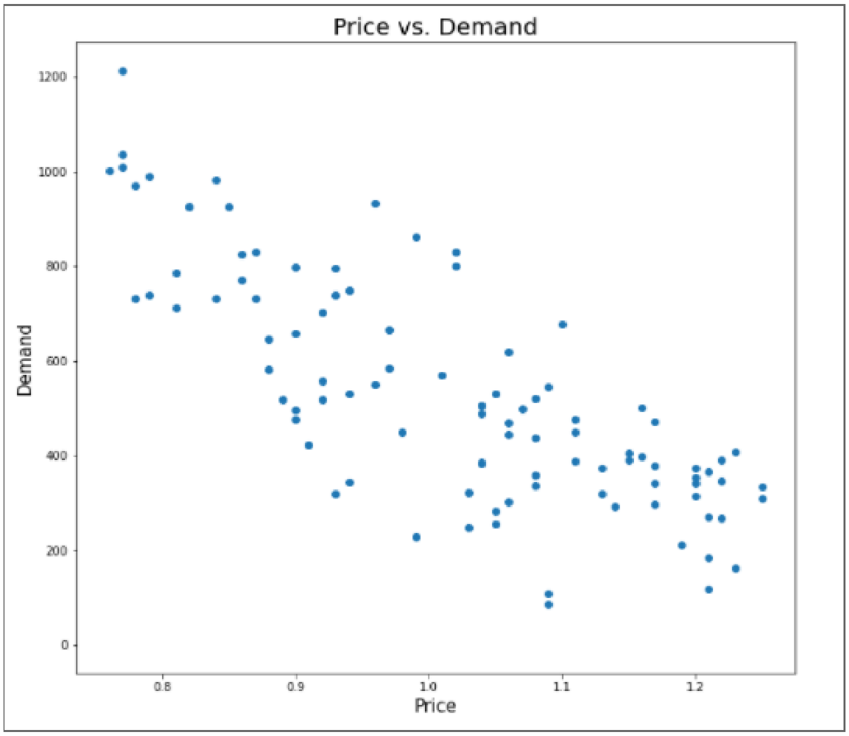

Figure 1

A linear regression was fit on the above data providing the following regression equation. Figure 2 provides a scatterplot of actual demand vs. predicted demand.

*Predicted demand = -1367.71*Price + 1924.72*

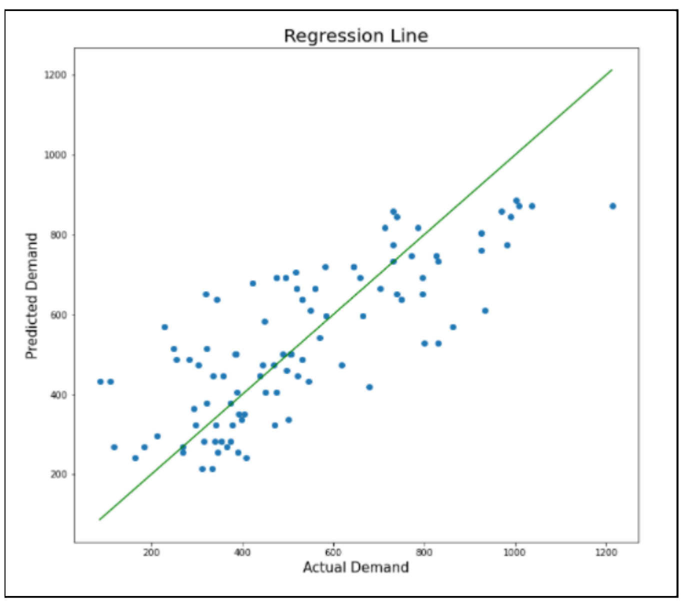

Figure 2

The residuals illustrated on the plot above were then used to simulate demand at different price points by plugging in a specific price and adding on the residual value to each of the data points. A price of 1 for each newspaper was then chosen to simulate demand data. Figure 3 provides descriptive statistics on the simulated demand data at a set price of 1. The simulated demand ranged from 210 units to 898 units with an average of 557 units and a standard deviation of 150 units.

	residuals	price	demand
count	9.900000e+01	99.0	99.000000
mean	1.860330e-13	1.0	557.005019
std	1.502125e+02	0.0	150.212535
min	-3.469109e+02	1.0	210.094127
25%	-1.045475e+02	1.0	452.457495
50%	1.289174e+01	1.0	569.896755
75%	1.002192e+02	1.0	657.224189
max	3.414211e+02	1.0	898.426119

Figure 3

Fixed Price Model

We formulated the described above into a Linear Program in order to maximize profit (h) for the coming day, given a price (p) and quantity printed (q) for the newspaper. In this scenario, the cost per newspaper (c) was 50 cents, the cost per rushed newspaper (g) was 75 cents, and the cost incurred from overprinting newspapers (t) was 15 cents per paper. To formulate the problem we create a constraint matrix of dimensions 198X100. The 198 rows in the matrix correspond to 2 constraints for each of the 99 days, together constraining that days' profit (h). The first constraint corresponded to profit in the case of rush orders:

$$h < pd - cq + g(q-d)$$

The second constraint for each day corresponded to the overprinting of newspapers:

$$h < pd - cq + t(d-q)$$

The elements of the objective function are the quantity printed tomorrow and the profit for each of the 99 days divided by 99. The lower bounds for all of the profit elements in the objective were set at negative infinity and 0 for the quantity printed element. The upper bound for quantity printed was set at infinity and the upper bounds for the profit elements were set at the maximum amount of revenue for each day.

The problem was then fed into gurobi and the optimal quantity to print was determined to be 471.87 and the expected profit was 231.48. Figure 4 is a plot of the simulated profit for each of the 99 days as the blue line, and the Expected profit for all days as the green line.

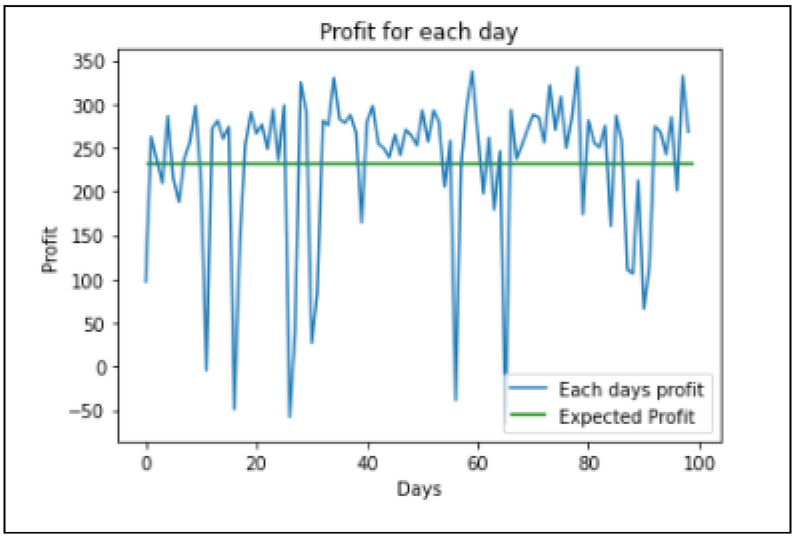

Figure 4

Dynamic Price Model

An extension to our model is added, by making the demand, a function of price. The objective function now becomes a quadratic expression in 'p'.

$$\max_q \frac{1}{n}\sum_{i=1}^{n}(pD_i - qc - g(D_i - q)^+ - t(q - D_i)^+)$$

$$(x)^+ = \max(x, 0)$$

D_i = Demand on Day i

q = Quantity of initial produce

p = Sale Price

c = Cost Price

g = Cost of Expedited process

t = Cost of excess Disposal

We are Given that the Demand on each day can be formulated as :

$Di = \beta o + p\beta_1 + \epsilon_1$ The linear regression equation

Replacing the Demand Linear Equation in the Objective Function we get the following equation:

$$Objective\ Function \Rightarrow p(D_i) - qc - g(D_i - q)^+ - t(q - D_i))^+$$

$$p(\beta o + p\beta_1 + \epsilon_1) - qc - g((\beta o + p\beta_1 + \epsilon_1) - q)^+ - t(q - \beta o + p\beta_1 + \epsilon_1))^+$$

$$p(\beta o + \epsilon_1) - [qc - g((\beta o + p\beta_1 + \epsilon_1) - q)^+ + t(q - \beta o + p\beta_1 + \epsilon_1))^+] + p\beta_1 p$$

$$p(\beta o + \epsilon_1) - h_i + p\beta_1 p$$

Where,

$$h_i = [qc - g((\beta o + p\beta_1 + \epsilon_1) - q)^+ + t(q - \beta o + p\beta_1 + \epsilon_1))^+]$$

Conditions:

$$h_i + (g - c)q - p\beta_i g > g(\beta_0 + \varepsilon_i)$$

$$h_i + (t + c)q - p\beta_i t > -t(\beta_0 + \varepsilon_i)$$

$$h > -\infty$$

$$q > 0$$

As we can see, the quadratic part of the equation is $p\beta_1 p$

Results:

```
In [13]: dynamic_price_nv.x[:2]
Out[13]: [535.291001278871, 0.9536264966232612]

In [14]: dynamic_price_nv.objVal
Out[14]: 234.42493487831734
```

- ➔ Optimal Price: 0.953 $
- ➔ Optimal Quantity : 535 newspapers
- ➔ Maximized Profit : 234.42$

This addition to the newsvendor model optimizes over both quantity and price which provides an alternative to the standard model where the price is fixed. Even though intuition is valuable from a business standpoint, there must be a balance between it and quantitative insights. The addition explores the relationship between the optimal price and quantity printed, which likely is significant over a large business like a publishing company. This relationship can have a huge impact on future revenue if not accounted for, so using a fixed price model is suboptimal at the least. Although this model has a higher expected revenue, the data given can also have a large impact on the results so this sensitivity must also be explored.

Sensitivity Analysis using Bootstrapping

The third extension of this problem is to use bootstrap sampling to gain an understanding of how sensitive the model is to a particular dataset. For this analysis, the basic idea of this method is that statistics like the mean or standard deviation of a population, all possible demand values, can be estimated by resampling the sample data, with replacement, and executing the extended newsvendor model on the resampled data many times.

More specifically, a sample of 99 observations was randomly selected from the original dataset with each observation having a small chance of being picked more than once. Thereafter, the bootstrap sample is used in a linear regression like *Figure 2* to find the new relationship between demand and price. The coefficients and residuals of this new regression are used in the dynamic price model which will output the optimal price and quantity based on this particular sample. This entire process is repeated 100,000 times to gain a realistic approximation of market fluctuations.

The results of this sensitivity analysis provide a useful insight into the bootstrap distribution of optimal quantities and prices. The mean and standard deviation of quantity is 535.91 and 32.72, respectively. Additionally, the mean and standard deviation of price is 0.95 and 0.01, respectively. A deeper dive into these metrics can be seen in the following:

	profit	price	quantity
mean	235.091219	0.954524	535.909305
std	8.956172	0.013572	32.719314
min	198.501774	0.899173	400.701230
25%	229.088467	0.945245	512.886425
50%	235.087181	0.953928	532.153331
75%	241.058219	0.963131	558.862595
max	278.872765	1.042542	685.389379

Figure 5

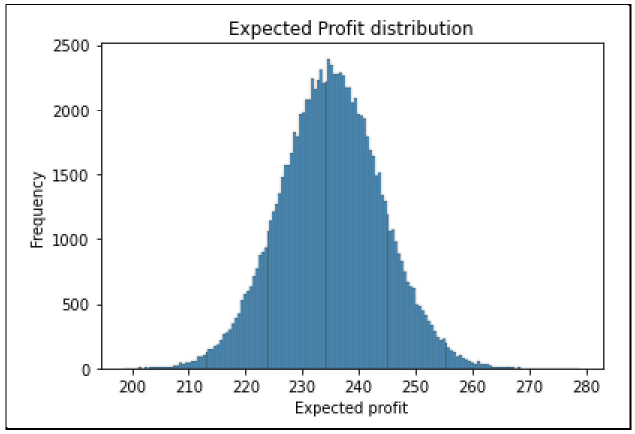

Figure 6

As shown, there is a tight range of optimal prices and quantities within the interquartile range, between the 25th and 75th percentiles. This adds assurance to the results dynamic newsvendor model, which had the result of 535.29 newspapers printed at a price of $0.95 per newspaper. These results are extremely close to the mean of the 100,000 bootstrapped samples. Furthermore, the histogram of expected profits shows a normal distribution at a mean of $235.09 with a standard deviation of $8.96. This allows the publishing company to assume with 99% confidence that profits will be between $261.96 and $208.22, controlling for all outside factors and using the mean optimal price and quantity. This gives the publishing company a useful range of estimates which they can use to forecast financials, investory, and work schedules. Additionally, the plot of all optimal quantities against optimal prices shows a familiar relationship that offers further support as seen below:

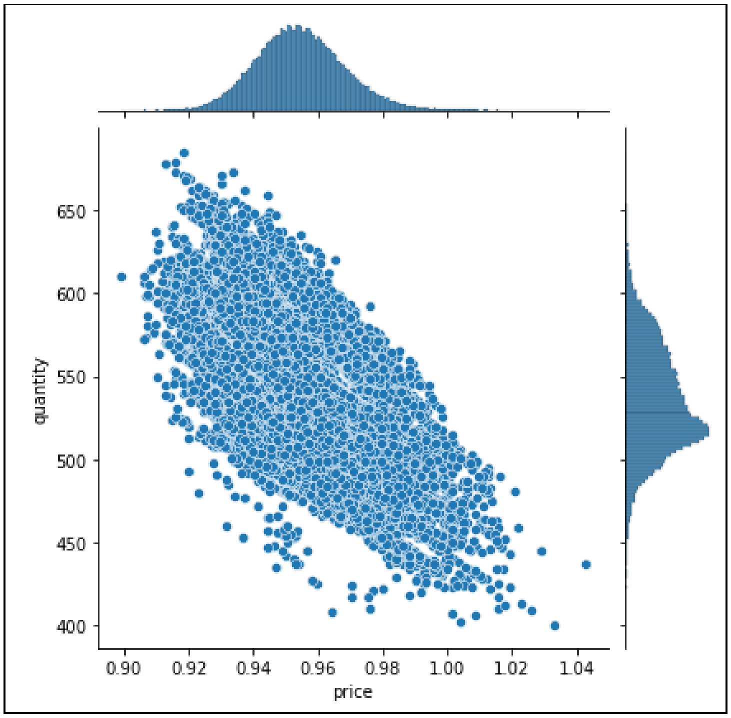

Figure 7

This figure maps out a relatively negative linear relationship between the optimal price and optimal quantity. As seen in *Figure 1*, there is also a negative relationship between demand and price, which can be related to the fact that the publishing company only wants to print enough newspapers as the demand requires. The similarity offers support for the assumption that the extensions of the newsvendor model provide a better approximation of market reactions to price.

Even though this model is seemingly well-grounded, it still relies heavily on foundational assumptions such as demand being linearly related to price. It is possible that consumers may not react in the same relationship as assumed. For example, if a potential customer sees a newspaper for a penny less than a competitor, will that realistically justify the publishing company printing 20 more newspapers to account for that supposed relationship. Consequently, each assumption should be carefully analyzed in the future when current demand data is available which allows the firm to measure the elasticity of demand and update the relationship between demand and price if necessary.

Conclusion

The existing model used within the organization to estimate the future demand and the optimal quantity of newspapers to produce uses a fixed price optimization approach. This approach limits the search space by fixing the selling price and produces a suboptimal solution that can lead to reduced overall profit. Since price is a significant determinant of the demand, which in turn affects the optimal quantity to produce in order to maximize profit, it is essential to consider price effects on demand. For different selling prices, the demand can vary and this varying demand can make it difficult to obtain the right quantity to produce in order to maximize the profits.

To shed more light on the difference between the two approaches we will use the confidence interval obtained from bootstrapping the demand for dynamic price approach for the key decision variables - price, quantity, and optimal profit. The comparison of the two approaches are outlined in Figure 8 below:

	Standard NV	**Dynamic Price**				
		Mean	**5%**	**95%**	**2.50%**	**97.50%**
Price	1	0.95	0.933	0.978	0.929	0.983
Quantity	490	535.90	487.85	591.26	477.41	601.83
Expected Profit	219.28	235.09	220.36	249.89	217.55	252.79

Figure 8

If taken a p-value of 5%, only price is outside the confidence interval that is expected as in the standard newsvendor model the price is set as fixed. But oftentimes a p-value of 5% is too stringent. Hence we will try to identify the confidence level with a p-value of 10%. Based on that we can make the following conclusions -

- Standard NV model gives statistically low profit
- Fail to assess any significant difference in quantity between Standard NV and Dynamic price model
- Standard NV model sell at high price statistically

With 10% p-value we are not able to reject the null hypothesis that there isn't any significant statistical difference between the optimal quantity obtained using the Standard

Newsvendor and Dynamic Pricing model, we are able to infer two things **- 1. The optimal price is lower using the Dynamic Pricing model and 2. The expected profit is higher.**

A clear advantage of the dynamic pricing model is that it adds disposal fees and rushed production costs, which helps model the problem more realistically as these costs can have a significant effect on the profits. Also, the dynamic pricing approach works best for both the consumer, allowing them to buy our products at lower prices, and the organization, as it ensures higher profit. Despite performing better than the standard newsvendor model, it has some limitations due to the assumption made about the linear relationship between demand and price that is used in the model formulation to optimize profit to get optimal price and quantity. If there are other variables involved in estimating demand or the relationship between the price and demand is non-linear, the formulation for the dynamic pricing model can become very difficult to solve. Additionally, if the past data becomes unrepresentative of future demand, then both models will be at risk of producing inaccurate results, which furthers the need to analyze more out-of-sample performance. Thus, the publishing company should rely on both the dynamic pricing model and business intuition about the market, its consumers, and the relationships explored in this analysis.