Project 1 - Stochastic Programming

Group #1: Yanqi Liang (yl29382), Brandt Green (bwg537), Prashanti Kodi (pk9759), Lucas Fernandez (lf23234)

Introduction

Companies often face the task of determining the appropriate price they should sell their product for. In addition, they must then determine the quantity of goods necessary to fulfill their expected demand. A company is typically best served by optimizing both decisions jointly- picking the optimal price & quantity together so that they maximize their expected profit. In this report we examine the results for a newspaper company if it uses the standard model for maximizing expected profit vs the extended model that jointly optimizes price and quantity to maximize expected profit.

Dataset

We have historical data containing price and demand numbers for the newspaper company. A brief snippet of the data and summary statistics below:

	price	demand
0	1.05	283
1	0.86	771
2	1.21	185
3	0.94	531
4	0.76	1002

	price	demand
count	99.000000	99.000000
mean	1.017677	532.828283
std	0.140726	244.150382
min	0.760000	87.000000
25%	0.900000	346.000000
50%	1.040000	490.000000
75%	1.135000	731.000000
max	1.250000	1213.000000

Fitting Linear Regression Model

First, we would like to estimate a demand function of the following linear form to derive an expected demand, given a price:

$$Demand_i = \beta_0 + \beta_1 * price_i + e_i$$

After fitting an linear regression using the given data, the intercept and coefficients are estimated to form the following equation:

$$Demand_i = 1924.72 + -1367.71 * price_i + e_i$$

Below is the plot of our data with the best fit line:

To determine optimal production and pricing decisions we will utilize following parameters:

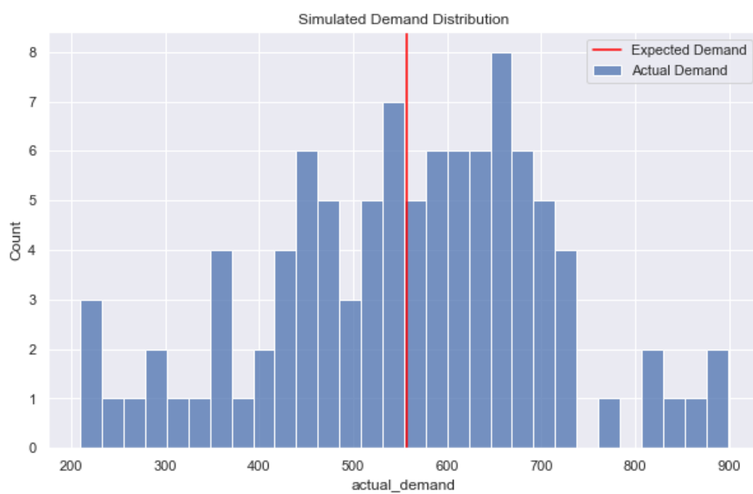
Parameter	**Value**
Rush Order Cost (g)	$.75
Disposal Cost (t)	$.15
Production Cost (c)	$.50
Price (p)	$1.00

Generate Demand Data: p=1

First, we want to determine what the predicted demand would be when we set the price of a newspaper equal to $1.00. We calculated the expected demand using the regression formula (with the coefficients calculated earlier) with price set to $1.00 and then added the residuals to the expected demand to generate the new data.

	price	predicted_demand	residuals	actual_demand
0	1	557.005019	-205.619393	351.385626
1	1	557.005019	22.515227	579.520247
2	1	557.005019	-84.785389	472.219630
3	1	557.005019	-108.067771	448.937249
4	1	557.005019	116.743975	673.748994
...	...	...	...	...
94	1	557.005019	-58.202391	498.802628
95	1	557.005019	112.098225	669.103245
96	1	557.005019	-115.296518	441.708501
97	1	557.005019	301.349231	858.354250
98	1	557.005019	44.443358	601.448378

To better understand how much demand can vary from our expectation, we plotted a histogram of the actual demand. There looks to be a lot of variability from our expectation!



Linear Programming: p=1

We want to find the quantity that achieves the below objective function where $(x)^{+}$= max $(x,0)$:

$$\max_{q} \frac{1}{n} \sum_{i=1}^{n} (p * D_i - q * c - g(D_i - q)^{+} - t(q - D_i)^{+})$$

This will give us the quantity to produce that provides the highest expected profit. We can actually simplify this problem by ignoring the first term, which will result in us minimizing our expected cost.

To do that, we created a variable, h, to represent negative cost and set the following constraints for h:

$$h_i <= -qc - gD_i + qp$$
$$h_i <= -qc - tq + t * D_i$$

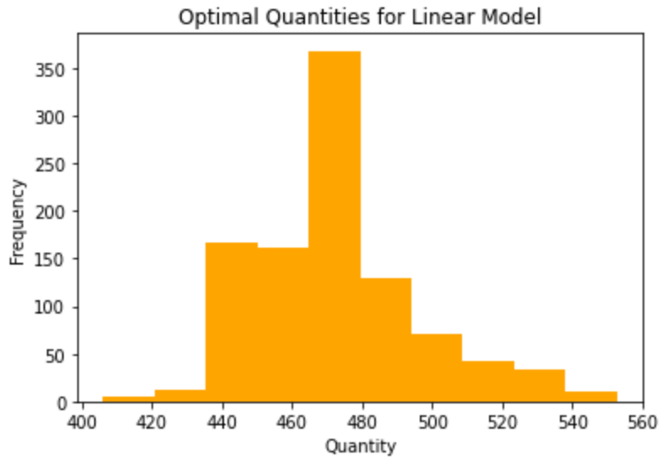
The code for the implementation is as follows:

```
c = 0.5
g = 0.75
t = 0.15
p = 1
nd = demand.shape[0]

# decision variables are (q, h1, h2, ..., h99) coeff of q would be 0, h1,h2... would have coeff of 1/99
obj = np.zeros(nd+1)
obj[1:] = 1.0/nd
ub = np.zeros(nd+1)
ub[0] = np.inf # the h variable can't be positive because it will always be negaitve (cost is always positive thus negative cost is negative)
lb = np.zeros(nd+1)
lb[1:] = -np.inf
rhs = np.zeros(2*nd)

direction = np.array(['<']*(2*nd))#hi needs to be less than the -cost (think of h under the two line of cost and we want to maximize hi so that its as close to 0 as possible)
```

```
A = np.zeros((2*nd,nd+1))#99*2 rows + 100 cols
for r in range(nd):
  #the constraint when P >= Di
  A[2*r,[0,r+1]] = [c-g,1]
  rhs[2*r] = -g*new_d.demand[r]
  #constraint when P <= Di
  A[2*r+1,[0,r+1]] = [c+t,1]
  rhs[2*r+1] = t*new_d.demand[r]
```

From the above code, we obtained the value of our optimal quantity of initially producing 472 newspapers. The expected profit is around $231.48 after calculation.

A quantities plot with 1000 sampling of the data using the linear NV model is shown below:



Quadratic Programming: price affects demand

In the above optimization, we assumed the price was given at $1.00. But now, we would like to consider the more realistic scenario where the newspaper company jointly chooses price and quantity together. To do so, we must reformulate our optimization objective and constraints so that everywhere we use our precalculated demand, we will replace the precalculated demand with our demand equation as a function of price (as estimated in the above regression). We now incorporate revenue as part of the objective function, which contains the quadratic part of the model, pDi. To represent the quadratic part in our objective function, we set up a Q matrix to contain beta_1 derived from our regression equation, which follows the below form:

$$\begin{pmatrix} x_1 & x_2 & x_3 \end{pmatrix} \begin{pmatrix} \beta & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

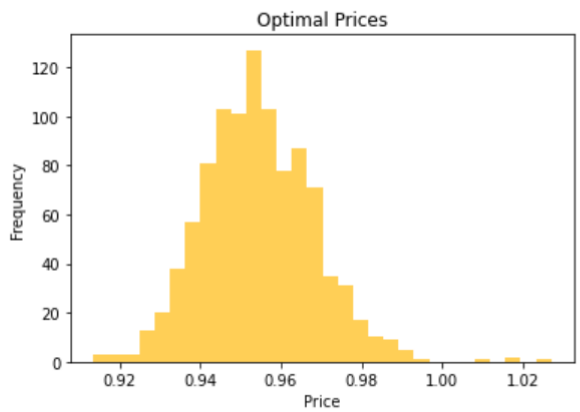
Beta_0 and the residuals will remain in the linear part of the objective function as coefficients for the price decision variable.

The constraints for determining costs changes slightly from the previous linear programming problem as the demand variable is replaced with the regression equation coefficients. As a result costs now become partially dependent on price.

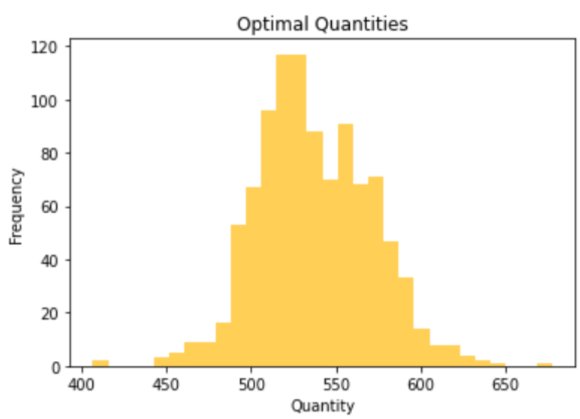
After the quadratic optimization ran, we discover that the optimal price is $0.95 and the optimal quantity to initially produce is 535 newspapers.

Bootstrap sampling: Extended NV Model

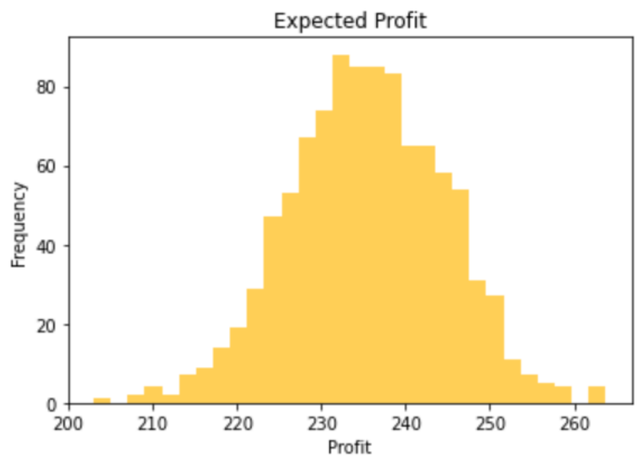
Now we want to see how sensitive our results are to the data set. To understand this, we created 1,000 bootstrapped samples of 99 observations from our original data. With this new dataset, we followed the exact same steps as listed above under Quadratic Programming to estimate the optimal quantity and price. The below plots shows the range of our price quantity and profit:



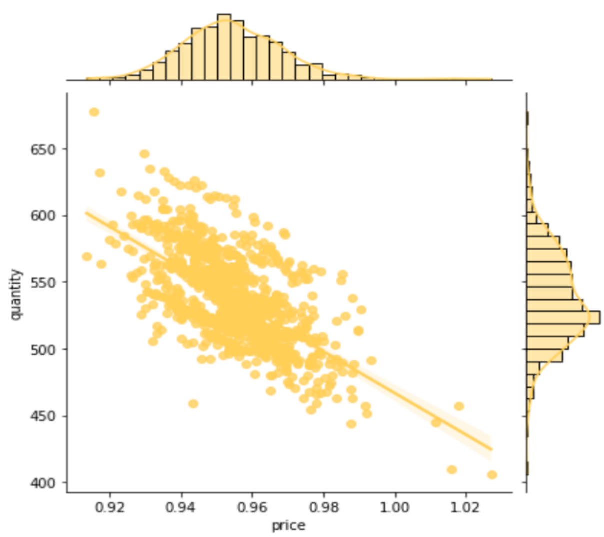
(i) Histogram of the optimal price from the simulations



(ii) Histogram of the optimal quantity from the process



(iii) Histogram of the expected profit



(iv) Scatterplot produced by taking the optimal quantity on the y-axis and optimal price on x-axis

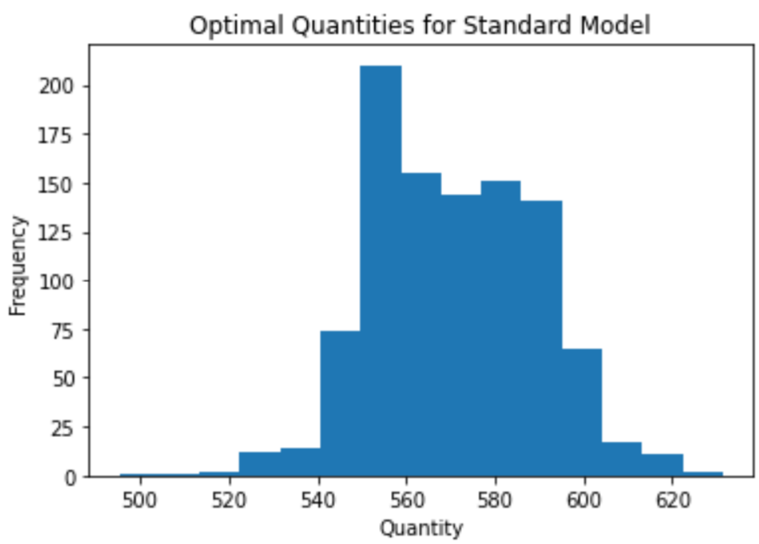
For price, we see that our optimal number ranges from around $.92 to $1. For quantity, the amount we should produce ranges from around 450 to 650. And last, our expected profit ranges from around 210 to 260. It's important to understand this variability, especially if the newspaper company may be making further decisions based on the project profits from this optimization.

Standard NV Model vs Extended NV Model

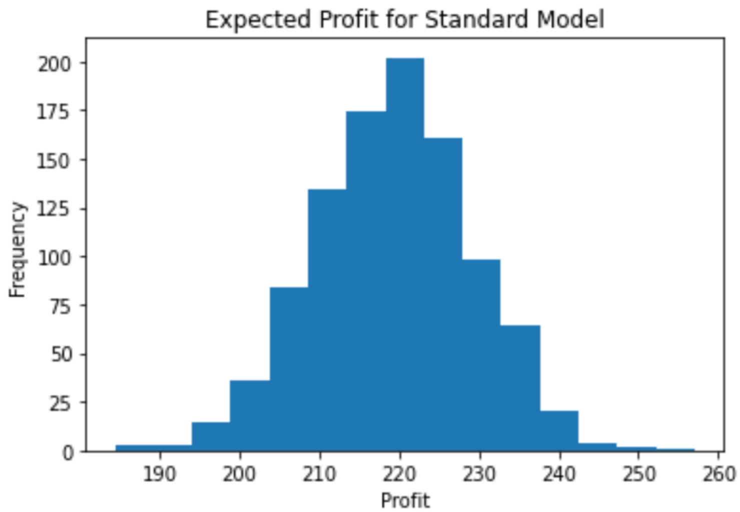
A standard NV model is also created to compare and contrast its result with the extended NV Models. The objective function of this model is to maximize the profit, however, it does not include the cost of rush ordering and disposal fees. Here, we treat the h variable as profit for each day and the constraints are as follow:

$$h \leq pD_i - cq$$
$$h \leq pq - cq$$

The Di of the constraint is generated for p = 1 using the regression equation as similar to the linear NV model. Next, we performed the same bootstrapping process on this model to see its range for optimal quantities and expected profit:



Looking at the optimal quantities for the standard model, we observe a range of higher quantities than in the optimal quantities plot from the other two extended NV models. The highest occurring optimal quantity for the standard model seems to be around 550, while for the quadratic NV model it's around 540 and around 470 for the linear NV model.



Next, looking at the expected profit for the standard model, we observe that it has a slightly lower range when compared to the quadratic NV model. The mode of profit seems to be around $220, which is lower than the quadratic model plot which showed $230 as most frequent. It is also lower than the expected profit calculated from the linear NV model which is around $231.

Below is a summary based on comparisons between the plots of the three models:

- Standard NV model shows the highest optimal quantity and the lowest expected profit.
- Quadratic NV model shows the highest expected profit.
- Linear NV model shows the lowest optimal quantity and the second highest expected profit.

As we can see from the summary and the plots, the standard model shows some of the worst results out of the three models. Although it has the highest quantity produced, its profit is significantly lower than the expected profit returned by the other two extended NV models.

To better understand our best choice of model, we have listed the advantages and disadvantages below for the models:

Standard NV Model

- Advantage: Lowered computational power and time consumption. Simple implementation.
- Disadvantage: Ignore relevant relationships in the real world, such as price effect on demand, the cost of rush ordering, disposal fees.

Linear NV Model (Extended NV Model)

- Advantages: Takes into account realistic scenarios such as rush ordering and disposal fees. Easier implementation than the quadratic model.
- Disadvantages: Did not take into account the relationship between price and demand.

Quadratic NV Model (Extended NV Model)

- Advantages: Takes care of all relevant relationships and scenarios such as price's effect on demand, rush ordering and disposal fee, which gives a more realistic picture of demand and the real world.
- Disadvantages: Highest computational power, more complexity and harder to implement.

Based on the summary of advantages and disadvantages of the standard model and extended models, it is best not to switch to the standard model as it will not increase your bottom line since it yields one of the lowest expected profit. To have the most accurate representation of the real world demand, the quadratic model is recommended, however, if you wish to set a constant price and save computational power, the linear model is also useful as its expected profit is only slightly lower than the quadratic model.