Question 1

These were the values of p, q, M and N(30) for all the 4 parts.

PART	P	Q	M	N(30)
1	0.0027924	0.2140974	34.4047289	0.9738414
2	0.0010607	0.1936843	Given	-
3	-	-	-	4.8904395
4	0.0006770	0.1948740	Given	4.4945176

Question 2

Note: Average P value = ***0.03****. Average Q value =* ***0.38****.* Source.

1. p = q = 0.05.

   p is slightly above average. q is **much lower** than average. As a result, the **innovation coefficient dominates**. Lack of imitation leads to lower adoptions over time.

2. p = 0.04, q = 0.06

   p is slightly above average, but lower than in 1. q is still much lower than average, but higher than in 1. Therefore, the **innovation**

coefficient still dominates, but a slight curve due to increase in the imitation coefficient can be seen for New Customers.

3. p = 0.06, q = 0.04

   p is above average, and higher than in 1 & 2. q is still much lower than average, and even lower than in 1 & 2. That is why the **innovation coefficient dominates**, even more, suppressing any curve that existed in 1 & 2 (New Customers).
4. p = 0.1, q = 0.1

   p is much higher than average. q is still on the lower side of the average. The **innovation coefficient continues to dominate**, but due to increase in q, the drop in New Customers is not as drastic as in 3.

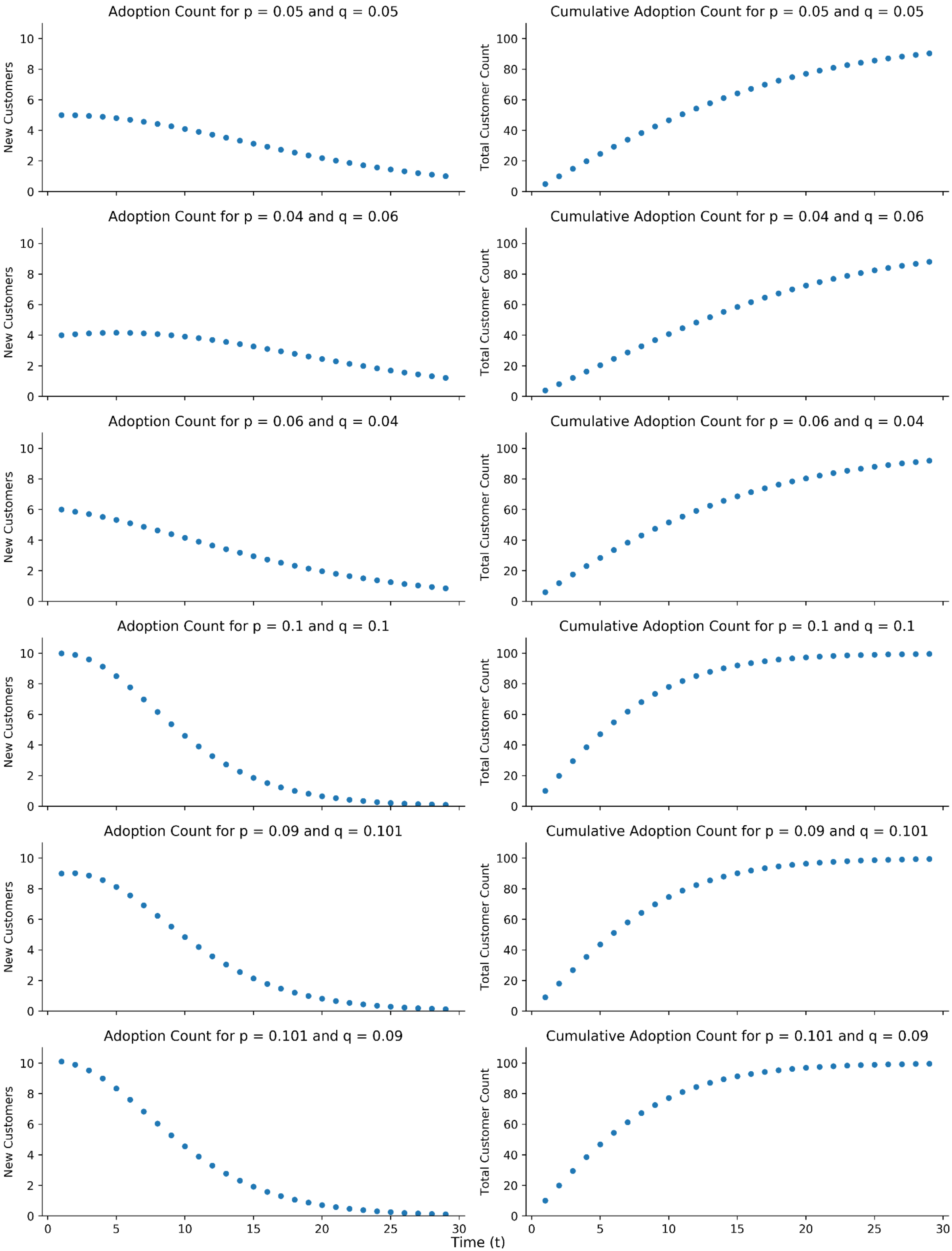
5. p = 0.09, q = 1.01

   p is still much higher than average, but lower than in 4. q is still on the lower side of the average, but a bit higher than in 4. The **imitation coefficient dominates** compared to 4, but in general, the **innovation coefficient** is still very much the driving factor.

6. p = 1.01, q = 0.09

   p is the highest among all the values. q is the lowest compared to 4 & 5. **The innovation coefficient is the clear dominator** here. The imitation coefficient barely has an effect, as indicated by the steep drop in New Customers over time.

The graph is available in the next page. To view the graph, open **images/Q2.png**.


Adoption Count for p = 0.05 and q = 0.05
Cumulative Adoption Count for p = 0.05 and q = 0.05
Adoption Count for p = 0.04 and q = 0.06
Cumulative Adoption Count for p = 0.04 and q = 0.06
Adoption Count for p = 0.06 and q = 0.04
Cumulative Adoption Count for p = 0.06 and q = 0.04
Adoption Count for p = 0.1 and q = 0.1
Cumulative Adoption Count for p = 0.1 and q = 0.1
Adoption Count for p = 0.09 and q = 0.101
Cumulative Adoption Count for p = 0.09 and q = 0.101
Adoption Count for p = 0.101 and q = 0.09
Cumulative Adoption Count for p = 0.101 and q = 0.09
New Customers
Total Customer Count
Time (t)

Question 3

Estimation error of M is larger in the slide 13/14 than in the slide 15/16, and consequently the forecasted curve beyond period 15 is much further from the true curve in the first case than in the second case. What differences in the 2 series or the nature of the noise (up to period 15) lead the difference described in the previous sentence?

Adoption 1: $p = 0.001$, $q = 0.2$, $M = 100$

Adoption 2: $p = 0.001$, $q = 0.5$, $M = 100$

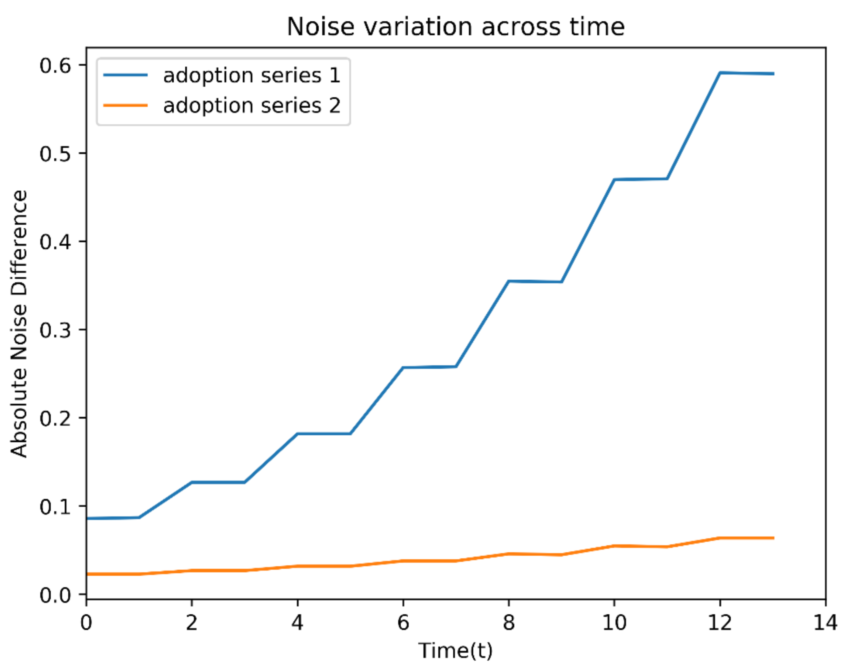
The possible causes for adoption 2 to be better estimated, compared to adoption 1 can be:

1. Noise Levels
2. p
3. q

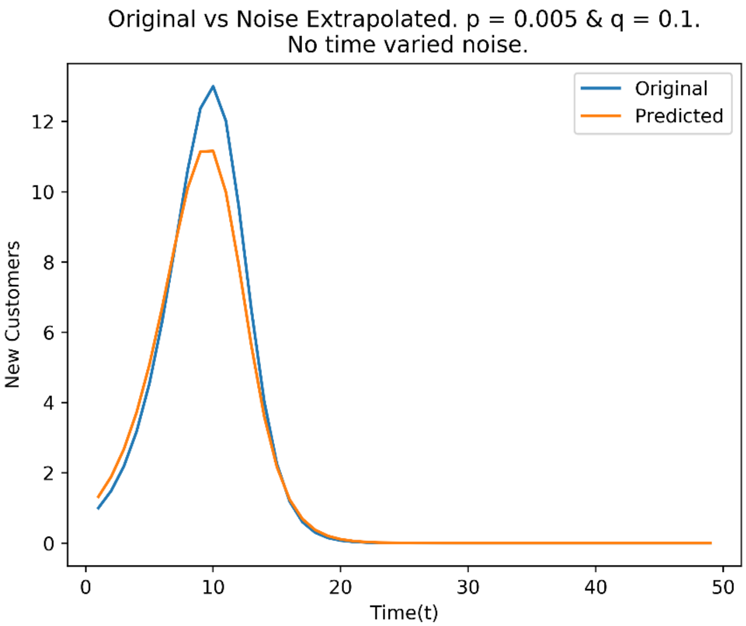
Noise Levels

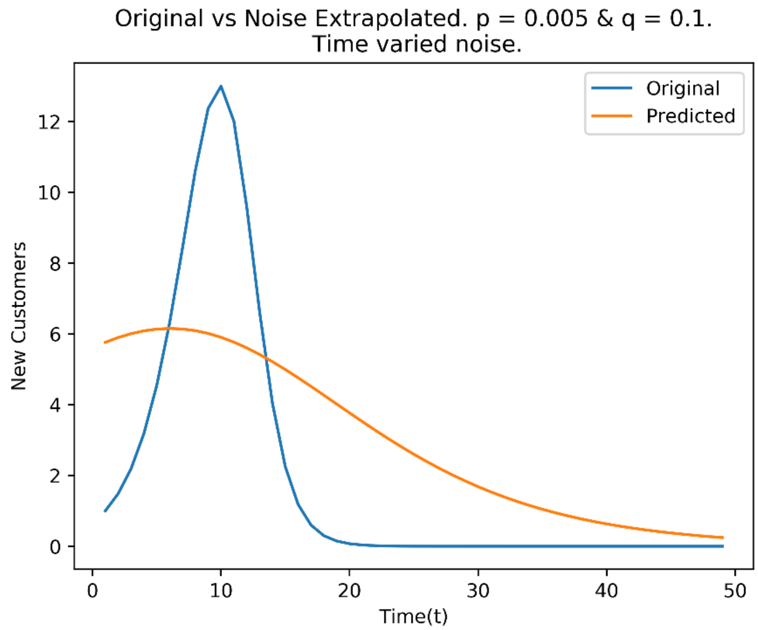
Noise can be added in various ways. The graph shows the absolute difference in adoption values for both the series, over time.

As we can see, the noise level increases over time for adoption 1. The noise is **time varied**. For adoption 2, the noise level remains the same over time.



To verify this, ***compare_noise()***, is defined in helper.py to help visualize the effects of various parameters.





On the left, there is still noise, but it's not time varied. The estimation is much closer.

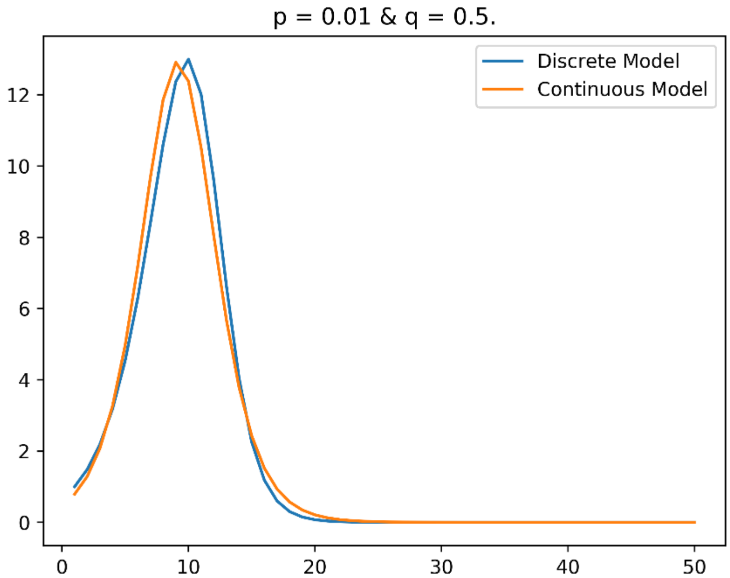
Both the graphs represent similar values of p & q. The difference is that on the graph to the right, the noise is **time varied.** The noise depends upon the time period. The graph looks nowhere close to the original.

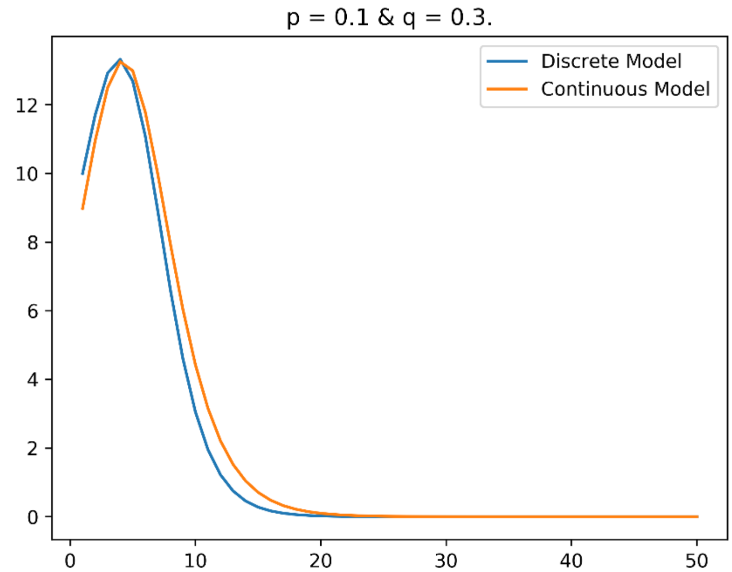
Notice that changing p & q values did not make the estimation worse in the non-time varied plot.

Time varied noise is the major culprit.

Question 4

In most cases, the observation is that: **continuous values are larger than discrete counterparts.**





As we can see, in both the cases, for different values of p & q, the continuous model is higher. The general trend noticed is that the **higher the p is**, the more likely it is that this relation holds.

Therefore, in Image 1, even though N(30) of continuous is greater, there is no defined relationship at all points. In Image 2, p is higher, and there is a more defined relationship (i.e., after t = ~7, continuous model dominates).

N(30) in 1.3: 4.8904395 (Discrete Model), p = 0.00106
N(30) in 1.4: 4.4945176 (Continuous Model), p = 0.000677

Clearly, the Discrete Model is higher. The p value in both the models are extremely low. This is the reason the that trend we see generally is not observed here.

Question 5

In order to determine the start-values, **multiple-start approach** used. It turns out that after performing ***curve_fit()***, all the returned values are identical. So, the following steps were taken:

1. Create two vectors containing p & q values, of a size.
2. Apply *curve_fit()* to the data, with each of these p & q values as the initial start point (p0).
3. Recreate the bass model from the obtained p & q values from curve fit.
4. Compute sum of squares between this model's N, and the original data.
5. Choose p, q corresponding to the minimum sum of squares.

After choosing 100 random values, the best p, q parameters were:

P = 0.00106072

Q = 0.19368439

Code in Q5.py.