Joe Standerfer, Spriha Gupta, Jasmine Kaur, Daniel Lesser

Marketing Analytics 95-832

Assignment 3: Demand Analysis using Linear Regression

In a familiar tale, the traditional marketing approaches of retailers have been upended in the age of computers and big data. For many decades, the dominant marketing strategy for these companies had been “Hi-Lo Pricing” - the use of periodic discounts to drive traffic while keeping prices high on average. However, the introduction of computers enabled a new, seemingly more effective, strategy to be born. Often referred to “Everyday Low Pricing”, large retailers such as Walmart began to use consumer data to set product prices. By correlating price and to sale quantity along with several other predictive factors, these retailers were able to set the optimal profit maximizing price for each product. In this case study we compared both approaches using orange juice sales data from a Chicago based grocery chain, Dominick’s Finer Foods. Specifically, we analyzed Dominick's sales of 64oz Tropicana Premium Orange juice and developed a series of regression models for predicting sales quantity given a price point.

The first regression model we created, estimated a sales quantity for the Tropicana orange juice using 3 factors: the log of the product’s sales price, the presence of in-store advertisements/ shelf-tags, and the existence of out-of-store advertisements. What we found was that the product demand is very elastic, or in other words the sales quantity responds strongly to changes in price. The coefficient we predicted for the log price was around -2.02. At that rate we can estimate the demand will increase 2% for each single percent decrease in price. The other two factors, in-store and out-of-store advertisements showed some effect on demand, but of a much lower magnitude than price. Their coefficients were 0.535 and 0.077 respectively. This first model, however, was far from perfect. When we tested the model estimates on a hold-out test dataset it achieved a mean-error of around 0.52. Charts of the model predictions as well as summary statistics can be seen in the Figures 1 through 4 in the Appendix.

A large portion of the prediction error can be traced back to a base assumption made by the model. By utilizing a single log price coefficient, the model assumes that all of Dominick’s stores have similar demand curves. However, this is likely not the case. Price sensitivity for a store’s product can be affected by a variety of factors unique to that store: surrounding demographics, closest competition, etc. To acknowledge that fact and thus lower the prediction error, we developed a second model – the ‘Store model’, that estimates distinct regression intercept and log price coefficient for each store. As expected, the prediction error for this “store model” is much lower than those of the original “pooled model”, at 0.36. Figures 5 through 7 in the appendix provide details on this model’s performance.

We tried to include the demographics of the stores – age, income, ethnicity and education to our model. However, we found that these variables were highly correlated to the ‘store’ attribute and hence, did not improve upon the model. It can be argued that ‘store’ already incorporates the demographic effects.

To further improve the regression model, we decided to consider another significant factor in product sales – competition. We created an improved model by including prices of competing orange juice products at the week level for every store. MinuteMaid, FloridaGold, HH Orange Juice were a few of the products from this category that we added for comparison. The mean error of this model came down to 0.33. Figures 8 through 10 in the appendix provide details on this model’s performance. This makes sense as, brands consider the prices of other competing products while setting their prices as well. If they set a price higher than the market price, they would lose market share. And if they set a price lower than the market price, they may gain a large proportion of market share.

Figure 11 shows the histogram of the residuals of the Pooled model. The large deviations from actual values are expected as the model has a high mean error of 0.52.

Figure 12 shows the histogram of the residuals of the Store model. The deviations of the predicted values range from -0.5 to 1 and most of the deviations are in the range [-0.5, 0.5]. Given that this model includes store effects and has a mean error of about 0.36, we would expect the predictions to deviate less as compared to the Pooled model.

Figure 13 shows the histogram of the residuals of the third improved model. Most of the predicted values differ by 0.2 in absolute value from their actual values. Amongst the three models, the improved model has the least deviations from the actual quantities. Also, the maximum deviation is 0.6.

Figures 14 through 16 show the time series plots of the errors across weeks for store 2. The pooled model shows high variations in as compared to the store model. The improved model has errors which are mostly centered about 0, amongst the three models. Both the error plots align with the low mean error of the improved model and strengthen our claim about the model. The implied optimal price for the log-linear demand model is given by:

where c = cost of the product

The optimal price for:

The pooled model: $4.35

Figures 17 and 18 show the computed optimal prices for each store for the Store model and the Improved model respectively.

The Mean difference between the store model predictions and the actual values = -2.44. For the Improved model, the difference = -1.24. According to our analysis, the improvised model gives the best results – as can be seen from the mean errors on the validation set. As Dominick’s Pricing Managers, we would implement the Improvised model for this case.

**Appendix:**

**Model 1 Figures**

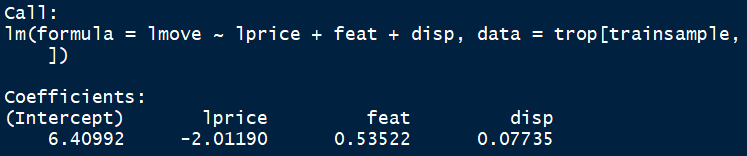


Figure 1: Pooled Model Regression Coefficients

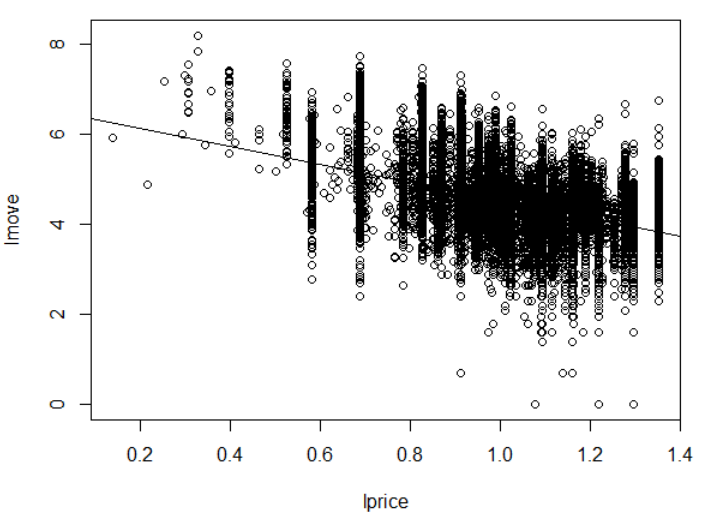


Figure 2: Pooled Model Predictions: All stores

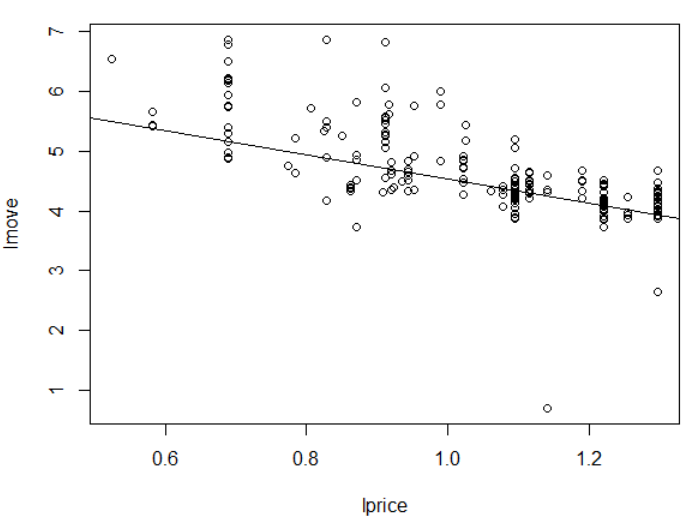


Figure 3: Pooled Model Predictions: Store 5

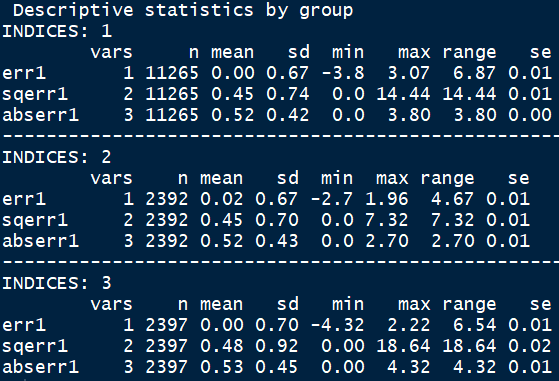


Figure 4: Pooled Model #1 Test Error

**Model 2 Figures**



Figure 5: Per-store Model #2 Regression Coefficients

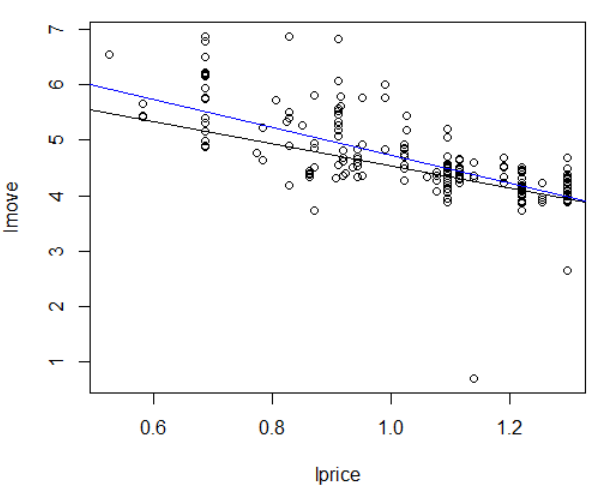


Figure 6: Per-Store Model #2 Predictions: Store 5 move predictions; price coefficent (black) vs all coefficients (blue)

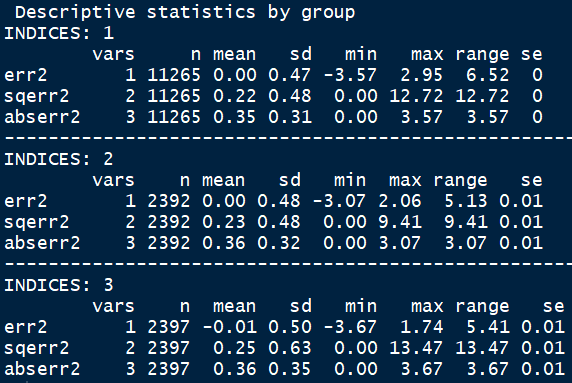


Figure 7: Per-Store Model #2 Test Error

**Model 3 Figures**



Figure 8: Improved Per-store Model #3 Regression Coefficients

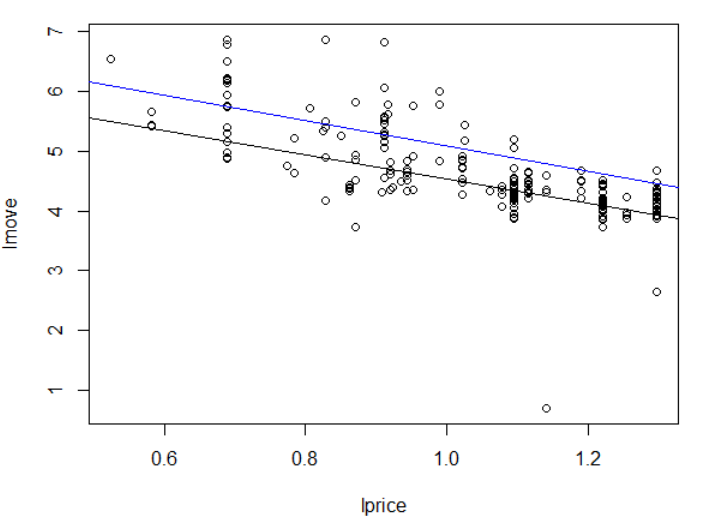


Figure 9: Model #3 Predictions - Store 5 move prediction; price coefficient (black) vs all coefficients (blue)

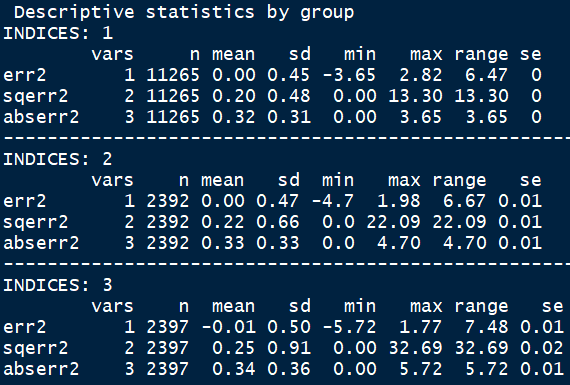


Figure 10: Improved Per-Store Model #3 Test Error

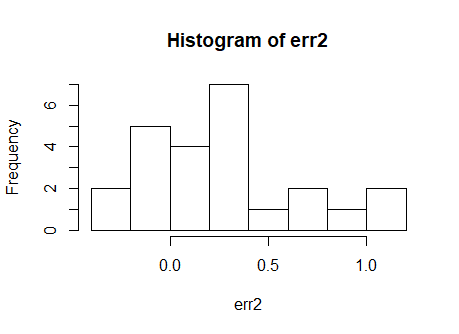


Figure 11: Histogram of errors of the Pooled model

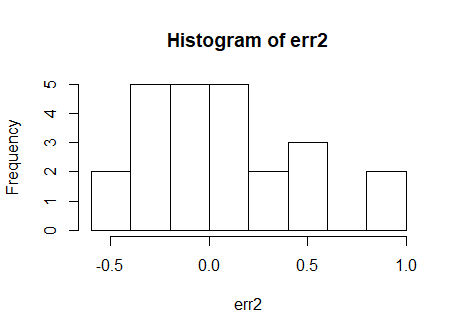


Figure 12: Histogram of errors of the Store model

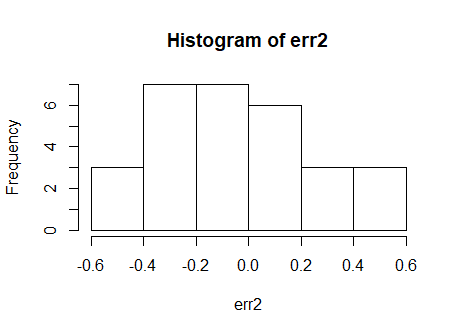


Figure 13: Histogram of errors of improved model

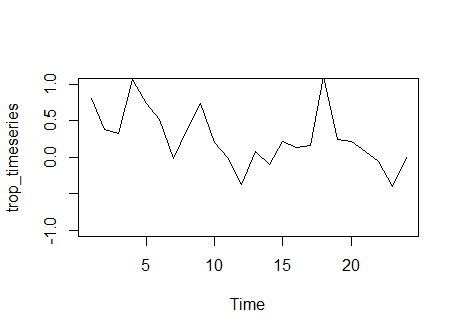


Figure 14: Pooled Model: Time Series plot of Store 2 price prediction errors

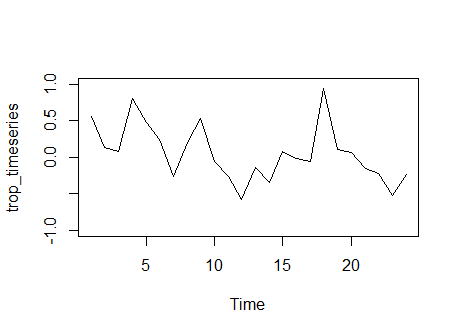


Figure 15: Store Model: Time series plot of store 2 price prediction errors

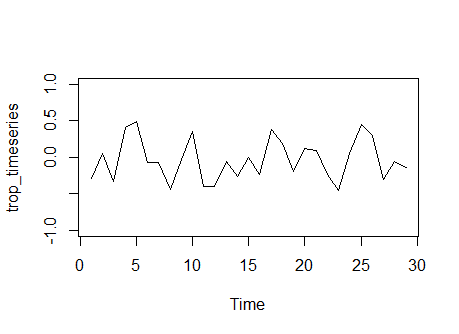


Figure 16: Improved Model: Time series plot of store 2 price prediction errors

|  |  |  |  |
| --- | --- | --- | --- |
| **Store** | **Actual Price** | **Predicted Price** | **Difference-Actual-Predicted** |
| store5 | 2.11992 | 4.180978142 | -1.990978142 |
| store8 | 2.11992 | 3.659411393 | -1.469411393 |
| store9 | 2.11992 | 4.272851606 | -2.082851606 |
| store12 | 2.11992 | 3.373024243 | -1.183024243 |
| store14 | 2.11992 | 5.585289357 | -3.395289357 |
| store18 | 2.11992 | 3.776096061 | -1.586096061 |
| store21 | 2.11992 | 3.550308613 | -1.360308613 |
| store28 | 2.11992 | 4.516822653 | -2.326822653 |
| store32 | 2.11992 | 7.006521049 | -4.816521049 |
| store33 | 2.11992 | 5.148811936 | -2.958811936 |
| store40 | 2.11992 | 3.316754834 | -1.126754834 |
| store44 | 2.11992 | 3.70793471 | -1.51793471 |
| store45 | 2.11992 | 6.295506139 | -4.105506139 |
| store47 | 2.11992 | 3.978118156 | -1.788118156 |
| store48 | 2.11992 | 5.038102382 | -2.848102382 |
| store49 | 2.11992 | 4.44169204 | -2.25169204 |
| store50 | 2.11992 | 4.119070921 | -1.929070921 |
| store51 | 2.11992 | 4.100710176 | -1.910710176 |
| store52 | 2.11992 | 4.405787488 | -2.215787488 |
| store53 | 2.11992 | 4.176320718 | -1.986320718 |
| store54 | 2.11992 | 4.374117249 | -2.184117249 |
| store56 | 2.11992 | 21.03175375 | -18.84175375 |
| store59 | 2.11992 | 3.419463285 | -1.229463285 |
| store62 | 2.11992 | 8.731137561 | -6.541137561 |
| store64 | 2.11992 | 3.7518465 | -1.5618465 |
| store67 | 2.11992 | 4.503264727 | -2.313264727 |
| store68 | 2.11992 | 4.245556453 | -2.055556453 |
| store70 | 2.11992 | 3.728460016 | -1.538460016 |
| store71 | 2.11992 | 3.606781919 | -1.416781919 |
| store72 | 2.11992 | 5.180219485 | -2.990219485 |
| store73 | 2.11992 | 4.690551836 | -2.500551836 |
| store74 | 2.11992 | 4.282168516 | -2.092168516 |
| store75 | 2.11992 | 4.248295529 | -2.058295529 |
| store76 | 2.11992 | 3.237451785 | -1.047451785 |
| store77 | 2.11992 | 3.995762645 | -1.805762645 |
| store78 | 2.11992 | 3.47151161 | -1.28151161 |
| store80 | 2.11992 | -17.70434825 | 19.89434825 |
| store81 | 2.11992 | 4.085320543 | -1.895320543 |
| store83 | 2.11992 | 3.348219063 | -1.158219063 |
| store84 | 2.11992 | 4.369698223 | -2.179698223 |
| store86 | 2.11992 | 4.012054634 | -1.822054634 |
| store88 | 2.11992 | 21.91998392 | -19.72998392 |
| store89 | 2.11992 | 3.622561016 | -1.432561016 |
| store90 | 2.11992 | 3.229477582 | -1.039477582 |
| store91 | 2.11992 | 3.868933177 | -1.678933177 |
| store92 | 2.11992 | 5.41550027 | -3.22550027 |
| store93 | 2.11992 | 4.833480864 | -2.643480864 |
| store94 | 2.11992 | 15.59136286 | -13.40136286 |
| store95 | 2.11992 | 3.350356711 | -1.160356711 |
| store97 | 2.11992 | 3.808257233 | -1.618257233 |
| store98 | 2.11992 | 3.779826985 | -1.419826985 |
| store100 | 2.11992 | 4.341124832 | -1.961124832 |
| store101 | 2.11992 | 7.926407513 | -5.536407513 |
| store102 | 2.11992 | 4.573543879 | -2.063543879 |
| store103 | 2.11992 | 3.863353698 | -1.433353698 |
| store104 | 2.11992 | 3.634282656 | -1.444282656 |
| store105 | 2.11992 | 3.254826554 | -0.924826554 |
| store106 | 2.11992 | 3.864649763 | -1.674649763 |
| store107 | 2.11992 | 4.631140411 | -2.441140411 |
| store109 | 2.11992 | 4.529841272 | -2.339841272 |
| store110 | 2.11992 | 5.340914867 | -3.150914867 |
| store111 | 2.11992 | 3.18516789 | -0.99516789 |
| store112 | 2.11992 | 10.24445992 | -7.874459916 |
| store113 | 2.11992 | 4.290589297 | -2.100589297 |
| store114 | 2.11992 | 3.561948808 | -1.171948808 |
| store115 | 2.11992 | 3.718350708 | -1.328350708 |
| store116 | 2.11992 | 4.984088201 | -2.794088201 |
| store117 | 2.11992 | 4.465134232 | -2.275134232 |
| store118 | 2.11992 | 4.224210907 | -2.034210907 |
| store119 | 2.11992 | 6.942615705 | -4.752615705 |
| store121 | 2.11992 | 3.671292428 | -1.321292428 |
| store122 | 2.11992 | 7.634461599 | -5.324461599 |
| store123 | 2.11992 | 3.566660402 | -1.376660402 |
| store124 | 2.11992 | 4.006538238 | -1.816538238 |
| store126 | 2.11992 | 3.189548109 | -0.849548109 |
| store128 | 2.11992 | 3.839386172 | -1.459386172 |
| store129 | 2.11992 | 4.075567616 | -1.655567616 |
| store130 | 2.11992 | 3.102838731 | -0.912838731 |
| store131 | 2.11992 | 4.531936048 | -2.111936048 |
| store132 | 2.11992 | 3.736389186 | -1.346389186 |
| store133 | 2.11992 | 2.966846525 | -2.966846525 |
| store134 | 2.11992 | 3.676263537 | -1.306263537 |
| store136 | 2.11992 | 3.864885486 | -3.864885486 |
| store137 | 2.11992 | 3.472399267 | -1.282399267 |
| store139 | 2.11992 | 2.993742143 | -0.803742143 |
|  | **Average** | 4.618641729 | -2.441347611 |

Figure 17: Computed Optimal prices for Store model

|  |  |  |  |
| --- | --- | --- | --- |
| **Store** | **Actual Price** | **Optimal Price** | **Difference** |
| 5 | 2.11992 | 4.002727394 | -1.81272739 |
| 8 | 2.11992 | 3.962440084 | -1.77244008 |
| 9 | 2.11992 | 6.150255484 | -3.96025548 |
| 12 | 2.11992 | 4.056764781 | -1.86676478 |
| 14 | 2.11992 | 11.61728064 | -9.42728064 |
| 18 | 2.11992 | 6.425445963 | -4.23544596 |
| 21 | 2.11992 | 3.523069975 | -1.33306998 |
| 28 | 2.11992 | 8.694515709 | -6.50451571 |
| 32 | 2.11992 | -35.05473678 | 37.24473678 |
| 33 | 2.11992 | 9.174422081 | -6.98442208 |
| 40 | 2.11992 | 3.478965082 | -1.28896508 |
| 44 | 2.11992 | 5.012989511 | -2.82298951 |
| 45 | 2.11992 | 8.017809482 | -5.82780948 |
| 47 | 2.11992 | 5.217918684 | -3.02791868 |
| 48 | 2.11992 | 19.55132655 | -17.3613265 |
| 49 | 2.11992 | 5.947080454 | -3.75708045 |
| 50 | 2.11992 | 5.180188521 | -2.99018852 |
| 51 | 2.11992 | 9.186282499 | -6.9962825 |
| 52 | 2.11992 | 7.413952736 | -5.22395274 |
| 53 | 2.11992 | 6.008964083 | -3.81896408 |
| 54 | 2.11992 | 4.530816348 | -2.34081635 |
| 56 | 2.11992 | -4.988131952 | 7.178131952 |
| 59 | 2.11992 | 3.867303686 | -1.67730369 |
| 62 | 2.11992 | -46.6393962 | 48.8293962 |
| 64 | 2.11992 | 4.62733575 | -2.43733575 |
| 67 | 2.11992 | 6.717499211 | -4.52749921 |
| 68 | 2.11992 | 4.710001771 | -2.52000177 |
| 70 | 2.11992 | 4.822893737 | -2.63289374 |
| 71 | 2.11992 | 4.381847317 | -2.19184732 |
| 72 | 2.11992 | 6.319668648 | -4.12966865 |
| 73 | 2.11992 | 5.357219386 | -3.16721939 |
| 74 | 2.11992 | 4.973085239 | -2.78308524 |
| 75 | 2.11992 | 5.650529875 | -3.46052987 |
| 76 | 2.11992 | 4.059838821 | -1.86983882 |
| 77 | 2.11992 | 7.538250017 | -5.34825002 |
| 78 | 2.11992 | 3.617245086 | -1.42724509 |
| 80 | 2.11992 | -8.906076653 | 11.09607665 |
| 81 | 2.11992 | 5.047036576 | -2.85703658 |
| 83 | 2.11992 | 4.168529846 | -1.97852985 |
| 84 | 2.11992 | 6.280178802 | -4.0901788 |
| 86 | 2.11992 | 3.998543674 | -1.80854367 |
| 88 | 2.11992 | -16.32626342 | 18.51626342 |
| 89 | 2.11992 | 4.001599566 | -1.81159957 |
| 90 | 2.11992 | 3.451830247 | -1.26183025 |
| 91 | 2.11992 | 4.25932431 | -2.06932431 |
| 92 | 2.11992 | 6.351234827 | -4.16123483 |
| 93 | 2.11992 | 6.757334075 | -4.56733407 |
| 94 | 2.11992 | -91.89869594 | 94.08869594 |
| 95 | 2.11992 | 4.189318721 | -1.99931872 |
| 97 | 2.11992 | 3.88623802 | -1.69623802 |
| 98 | 2.11992 | 4.826276268 | -2.46627627 |
| 100 | 2.11992 | 6.807349081 | -4.42734908 |
| 101 | 2.11992 | 11.86375965 | -9.47375965 |
| 102 | 2.11992 | 4.087434695 | -1.57743469 |
| 103 | 2.11992 | 4.337335661 | -1.90733566 |
| 104 | 2.11992 | 3.931844504 | -1.7418445 |
| 105 | 2.11992 | 3.85098673 | -1.52098673 |
| 106 | 2.11992 | 4.323382905 | -2.1333829 |
| 107 | 2.11992 | 4.902142593 | -2.71214259 |
| 109 | 2.11992 | 8.465462963 | -6.27546296 |
| 110 | 2.11992 | 19.17111445 | -16.9811144 |
| 111 | 2.11992 | 3.385172158 | -1.19517216 |
| 112 | 2.11992 | 16.69014403 | -14.320144 |
| 113 | 2.11992 | 5.281515066 | -3.09151507 |
| 114 | 2.11992 | 3.956282815 | -1.56628282 |
| 115 | 2.11992 | 4.035692726 | -1.64569273 |
| 116 | 2.11992 | 18.74111614 | -16.5511161 |
| 117 | 2.11992 | 5.116253265 | -2.92625326 |
| 118 | 2.11992 | 4.87271382 | -2.68271382 |
| 119 | 2.11992 | 15.37145698 | -13.181457 |
| 121 | 2.11992 | 4.846226878 | -2.49622688 |
| 122 | 2.11992 | 13.89452125 | -11.5845212 |
| 123 | 2.11992 | 3.545344037 | -1.35534404 |
| 124 | 2.11992 | 4.426540738 | -2.23654074 |
| 126 | 2.11992 | 3.421705523 | -1.08170552 |
| 128 | 2.11992 | 4.926961316 | -2.54696132 |
| 129 | 2.11992 | 4.596209105 | -2.1762091 |
| 130 | 2.11992 | 3.500092672 | -1.31009267 |
| 131 | 2.11992 | 7.041723162 | -4.62172316 |
| 132 | 2.11992 | 3.766856479 | -1.37685648 |
| 133 | 2.11992 | 4.106829213 | -4.10682921 |
| 134 | 2.11992 | 4.867557396 | -2.4975574 |
| 136 | 2.11992 | 13.54254215 | -13.5425421 |
| 137 | 2.11992 | 4.002648668 | -1.81264867 |
| 139 | 2.11992 | 3.170719262 | -0.98071926 |
|  | **Average** | 3.412326078 | -1.23503196 |

Figure 18: Computed optimal prices for Improved model