# Interm mid-term report

Julie Pecukonis and Matthew Raison

July 9, 2015

Professor Rachel Maitra

Our project goal is to optimize the location of grocery stores, specifically in the city of Boston. For our project to be successful, we would want to be able to use our model to place certain stores in distinct locations in order to maximize customer satisfaction and profit. This would not only be beneficial for the stores to gain revenue, but also for its customers, since the purpose is to optimize convenience and affordability for different demographics. We have been basing most of our work on the paper Store Choice in Spatially Differentiated Markets by Lucrezio Figurelli. He created a model that does exactly what we are trying to accomplish, but we would like to alter it to be more general. So we took the model that he created and are trying to work off of it. Just in the short time that we have been working on it, we have change variables multiple times. We wanted to work off of a basic model, and change it by adding and manipulating the variables to get the final model. This process has been mostly trial and error, but we are slowly making some progress. Our current model:

$$u_{hsbt} = v_{hsbt} + \alpha_h \cdot e(b, p_{hst}) + \gamma_h(d_{hs}) + X_{hs} \cdot \beta + \xi_{hs} + \epsilon_{hst}$$

| | | |
|---|---|---|
| $u_{hsbt}$ | = | utility for household $h$ buying bundle $b$ at store $s$ given the time $t$ since they've last gotten groceries |
| $v_{hsbt}$ | = | utility of bundle $b$ for household $h$ at store $s$ given the time $t$ since they've last gotten groceries |
| $\alpha \cdot e(b, p_{hst})$ | = | expenditure or price as related to the household $h$ store $s$ and time $t$ since the last time shopping, of bur |
| $X_{hs} \cdot \beta + \xi_{hs}$ | = | observed store component and error |
| $\epsilon_{hst}$ | = | error given household $h$ and store $s$ |

When initially approaching this equation we had to consider what data we would have access to. We

first solidified a unit of measurement for the household $h$, supposing that we should identify households by their average income. However when we went to find our data we were only able to get the average household income per zip-code, not per house. Continuing with our initial supposition by simply expanding the $h$ variable to encapsolate an entire zip-code area, we found that we could further define the store $s$ variable by zip-code as well using a similar method relating the store to the average household income of the neighborhood within which it existed.

To then further simplify the model above, we assumed that the bundle $b$ and subsequently time $t$ remain constant. That is to say we will assume for our initial model that each household is getting a standard bundle of groceries that they will exhaust within the same time as all other households, thus requiring them to revisit the grocery store on regular, unwavering basis.
additionally considering our variable $v$ can represent the a households predisposition to a store given that we are supposing $b$ and $t$ to be fixed.

Now considering all of our assumptions we simplify our equation to be:

$$u_{hs} = v_{hs} + \alpha_h \cdot e(p_{hst}) + \gamma_h(d_{hs}) + X_{hs} \cdot \beta + \xi_{hs} + \epsilon_{hst}$$

Under observance of the paper previously mention we will assume there are two main types of grocery stores, supermarkets and convenient stores. Additionally however we will consider subcategories within these types. Under supermarkets are fresh stores, limited assortment stores, and superstores, which include mass merchandisers and wholesale clubs. For our simplified model lets assume people are mostly shopping at supermarkets, and that utility for these stores will be based upon some ranking of the store types. Then we can consider the variable $v$ to be a ranking in quality for the stores. In this way we will be able to isolate the type of store to consider for our equation.

For function $\alpha \cdot e(b, p_{hst})$, requires a little more ingenuity to properly quantify for our model. We will start by assuming that the change in price of a bundle over the households income is proportional to the change in utility for that bundle. Intuitively this should make sense since to a high income household an even moderate price change probably won't change that households desire for that bundle. Now elaborating

on this assumed relation we get:

$$\begin{aligned}
\text{let } I_h = &\quad \text{income of household } h \\
\Delta\frac{p_{hs}}{I_h} \propto &\quad \Delta u_{hs} \\
\frac{\Delta u_{hs}}{\Delta p_{hs}} \propto &\quad \frac{1}{I_h} \\
\frac{\delta u_{hs}}{\delta p_{hs}} = &\quad \frac{c}{I_h} \\
\alpha_h \cdot e = u_{hs} = &\quad \frac{c \cdot p_{hs}}{I_h}
\end{aligned}$$

Consider $c$ to be less than 0 since the utility should decrease as price increases

Suppose $c = -1$ for our model

$$\alpha_h \cdot e = \quad \frac{-p_{hst}}{I_h}$$

We should then find the function $\gamma_h(d_{hs})$ intuitively as well. Given that we have obtained data for the distances between every zip-code in Boston, most of the work for this variable has already been done. We can then just let $d_s = D(Z_s, Z_h)$ where $D(Z_s, Z_h)$ is a function that returns the distance between the zip-code of the store, $Z_s$, and the zip-code of the household $Z_h$, then reason through the following:

$$\begin{aligned}
u_\gamma \propto &\quad k \cdot d_s \\
u_\gamma = -d_s &
\end{aligned}$$

Our current equation then looks like this:

$$u_{hs} = v_{hs} + \frac{-p_{hst}}{I_h} + -d_s + X_{hs} \cdot \beta + \xi_{hs} + \epsilon_{hst}$$

Immediate steps that we will need to take to implement fully our simplified model is to find a reasonable equation for the $X_{hs} \cdot \beta$ variable. We assume this variable will be proportional to the household income and the income of the neighborhood the store is in since this is how we are quantifying position. We should discuss whether this will necessarily make for the best model but given the data at our disposal this approach will probably be best. Considering the two errors negligible we should be able to impose our model on our data using ampl. Since we are trying to optimize position of a grocery store We will want to look for a

maximum value for our equation given the store type. Then we must also consider if the found maximum is greater than the utility of not going to a different grocery store or no grocery store at all. Given our assumptions however we may be able to ignore the utility of not going to any grocery store.

$$X_{hs} \cdot \beta = -k \cdot \frac{Z_h}{|Z_h - Z_s|}$$

Since we assume that the bundle size is fixed and that the time between shopping trips is also fixed we should be able to reason that the utility of not going to any grocery store will be 0. Recall from the paper the utility of not shopping function:

$$\begin{aligned}
\sum_j f_j(b_{t-j}) + \epsilon_{h0t} \quad &= \\
&= \sum_j f_j(C) + \epsilon_{h0t} && \text{where } C \text{ is a constant.} \\
&= f_j(C) \cdot j && \text{assume error is negligible} \\
&= \hat{C} \\
&= 0 && \text{considering we are ultimately looking for a maximization.}
\end{aligned}$$

This only works however because maximum can be scaled, making constants somewhat inconsequential. As we elaborate on our equation and add more variables, like bundle size, this equation may no longer hold.

As far as field testing our model goes we will need to find real stores for reference. We will need to research what stores have either failed in what areas (zip-codes) or what stores have succeeded. Depending on how strong our model is at predicting whether a type of supermarket is in a certain zip-code we may find little need in this research. It is still some what unclear what some of our next steps are until we start using our data to test. Thus the most pressing goal for us is to solidify the model by determining that last component, $X_{hs} \cdot \beta$. Additional we should revisit the $\gamma$ function to see what Figurelli suggests as a function. Our initial thinking is that a direct relation like we have may be too simple to properly model someone's willingness to drive farther and farther distances.

$$u_{hs} = v_{hs} + \frac{-p_{hst}}{I_h} + -d_s + -k \cdot \frac{Z_h}{|Z_h - Z_s|}$$

$$u_{hs} = v_{hs} + \frac{-p_{hst}}{I_h} + -k \cdot \frac{Z_h}{|Z_h - Z_s|}$$