**Introduction**

Cleanly summarizes brainstorming in [this](https://docs.google.com/document/d/1H-qqbAOXen2Vnzcd4F6FLtCeTU8EBRH9eXskM3waR1Q/edit?usp=sharing) doc. We want to model a demand system based on the gasoline market in California, **with access only to market-level data**. We have no consumer-level data, instead our data looks like the following:

* Gas station-level, **daily** retail prices (OPIS)
* Gas station-level, **annual** quantities (PIIRA A15)

This memo summarizes possible approaches to estimating this demand system, the advantages and disadvantages of these different types of models, before concluding with a discussion of how we can address the gasoline market-specific problem of spatial competition/consumption.

I think we have the following outstanding questions which can help us determine the appropriate model/approach to the spatial component of the gasoline market:

* Do we want a non-parametric or parametric model?
* What kind of data can we get access to?
  + Is it possible to develop some algorithm for coding what services each of the gas stations in our dataset offer? Car wash, mechanic service, food mart, etc.
  + Can we get any data on commuting patterns on a zip-code level?
  + Can we get some kind of measure of the density of businesses/residences on a zip-code level?

**Demand System Estimation Techniques**

For a modern review of demand estimation, see this [*NBER* article](https://www.nber.org/system/files/working_papers/w29305/w29305.pdf) by Berry and Haile.

1. **Linear Demand Model**

In the linear demand model, we can write consumer *i*’s utility from good *j* takes the following form:

Here, utility is only a function of price; this is obviously not borne out in our data, where there is the significant and growing gap in branded and unbranded prices.

1. **Multinomial Logit Model**

Let consumer *i*’s utility from good *j* take the following form:

where:

* \delta\_{jt} is the mean utility for good *j* in market *t*
* \mu\_{ijt} captures heterogeneity in individual’s preferences
* \epsilon\_{ijt} captures idiosyncratic heterogeneity and accommodates estimation

Assuming \epsilon\_{ijt}’s distribution is i.i.d. Type I extreme value, we can derive the market shares:

Where J\_t represents the set of all products in market *t*.

*Pure Logit Model*

We can simplify this model by assuming away systematic heterogeneity in individuals’ preferences. This limits the richness of our model but makes estimation very straightforward as we can write:

And thus

Where x\_{jt} is a vector of non-price characteristics of product *j* in market *t* and \xi\_{jt} are unobserved product characteristics.

Then, each market share is:

Then we can write and estimate:

Where s\_{0t} is the market share of the outside good. This is clearly fully-identified; we have observed shares for each good (yearly quantity data), prices (daily price data), and we can easily develop some gas station-specific characteristics (mainly a measure of its proximity to commuter routes, an issue I address in the subsequent section). We would also incorporate brand into this vector of characteristics.

p\_{jt} is endogenous, as it is correlated with unobserved product characteristics \xi\_{jt}. Thus, we would have to instrument it–something we can discuss later and which I mention in the BLP/Random Coefficients section.

The advantage of this approach is that it’s easy/straightforward and very computationally cheap. We could execute this immediately, no more modeling/data needed. The disadvantage is that it’s not rich; it allows for no systematic heterogeneity in individuals’ preferences, restricting elasticity estimation. This runs counter to our story of branded market power, which is predicated on there being different ‘types’ of consumers preferring different brands.

Materials for “Pure” Logit Model:

* Section 5.1.1, pgs. 34-35 in [Berry and Haile (2021)](https://www.nber.org/system/files/working_papers/w29305/w29305.pdf)
* Gortmaker’s BLP Workshop Pure Logit [slides](https://github.com/Mixtape-Sessions/Demand-Estimation/tree/main/Slides)

1. **BLP/Random Coefficients/Mixed Logit**

Pure Logit is hobbled by a clear shortcoming–it allows no systematic heterogeneity in consumer preferences. So, let’s reintroduce a non-zero consumer systematic heterogeneity. Now, once again:

And then we can write

Where x’\_{jt} represents *all* observed characteristics, including price.

Then, if we want to capture systematic heterogeneity, we need to estimate a new coefficient, \beta\_{it} which variables at the product-consumer level. In a parametric model, we need to draw this from some distribution, most specified as:

Where \pi shifts preferences according to observed demographics drawn from a data source like the census (the array of which is referred to as y\_{it}). Then v\_{it} refers to unobserved preferences among the population of consumers.

This allows to decompose \beta\_{it} into two parts: 1) the homogenous component (denoted by \beta) and 2) the heterogeneous component (denoted by \Sigma v\_{it} + \pi y\_{it}) such that:

Then, market shares are determined according to the following equation:

Which we identify using the BLP estimator/GMM.

Advantages of BLP:

* Allows for *systematic* heterogeneity in consumer preferences

Disadvantages of BLP:

* Parameters are unknown; estimating them requires an iterative non-linear optimization problem, meaning there is no mathematical guarantee of a solution.
* This computational difficulty has helped to cause a lack of a standardized implementation of the model, resulting in numerical inconsistencies.

Materials for BLP Modelling/Estimation:

* [Nevo (2000)](https://www.jstor.org/stable/pdf/2600994.pdf?refreqid=fastly-default%3A701666696a190ab7808d5b076e188ea8&ab_segments=&initiator=&acceptTC=1) for good introduction/application of model
* Section 4 (starting on pg. 21) in [Berry and Haile (2021)](https://www.nber.org/system/files/working_papers/w29305/w29305.pdf)
* [Conlon and Gortmaker (2020)](https://pyblp.readthedocs.io/en/stable/background.html#supply) for estimation of BLP (w/PyBLP)
* Gortmaker’s BLP Workshop Mixed Logit [slides](https://github.com/Mixtape-Sessions/Demand-Estimation/tree/main/Slides)
* [Knittel and Metaxoglou (2014)](https://www.jstor.org/stable/pdf/43554912.pdf) for a discussion of flaws in numerically estimating BLP

1. **Micro BLP**

Micro BLP approach connects microdata on the consumer level to the BLP estimator; given that we don’t have consumer-level data, this approach is likely not applicable/appropriate for our model.

Materials for Micro BLP:

* [Conlon and Gortmaker (2024)](https://chrisconlon.github.io/site/micro_pyblp.pdf)

**Potential Solutions to Spatial Aspect of Gas Station Competition**