THE UNIVERSITY OF CHICAGO

THE COMPETITIVE EFFECTS OF MULTI-STORE SHOPPING: UNCOVERING COMPLEMENTARITIES ACROSS BUNDLES

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE SOCIAL SCIENCES
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

KENNETH C. GRIFFIN DEPARTMENT OF ECONOMICS

BY
NADIA LUCAS

CHICAGO, ILLINOIS
JUNE 2024

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

DISCLAIMER

Researcher's own analyses calculated (or derived) based in part on data from Nielsen Consumer LLC and marketing databases provided through the NielsenIQ Datasets at the Kilts Center for Marketing Data Center at The University of Chicago Booth School of Business. The conclusions drawn from the NielsenIQ data are those of the researcher and do not reflect the views of NielsenIQ. NielsenIQ is not responsible for, had no role in, and was not involved in analyzing and preparing the results reported herein.

ACKNOWLEDGMENTS

I'd like to thank Ali Hortaçsu, Michael Dinerstein and Jean-Pierre Dubé for excellent comments and advising. I would also like to thank Ignacia Cuevas, Arjun Gopinath, Josh Higbee, Gabriel Jardanovski, and Camilla Schneier for helpful comments and discussions. I'd like to thank all the participants in the UChicago IO Lunch and IO Reading Group run by Eric Richert for providing a forum for comments and feedback on this work. I would also like to thank the Kilts Center at Booth for data availability. None of the findings in this paper reflect the views of NielsenIQ or the Kilts Center. Email: nrlucas@uchicago.edu.

ABSTRACT

The nature of competition between entire retail storefronts has always been largely debated in antitrust. Looking at product offerings gives suggestive evidence to whether stores directly compete. I propose an estimator that uses consumer behavior to evaluate how much consumers treat entire storefronts as complements or substitutes. I develop the estimation strategy for the level of complementarity between two storefronts, taking into account repeated purchase behavior over a long panel. I first set up the challenges of using a classical complementarity framework to estimate this. Then I propose the estimation strategy which has two major steps. The first step is a nested logit discrete choice framework to come up with preference parameters for consumers across different items. This step consists of estimating both an inner and outer nest. The next step of the estimation strategy is to use those preference parameters to construct a complementarity term across each pair of retailers in the sample, in line with our intuition about complementarity across bundles. Finally, I report these estimates as correlations between observable characteristics across the two stores and show that generally consumers that shop at multiple stores tend to value one store having high brand variety and might use this store as a focal store in their shopping trip that they might rely on for the coverage of variety.

1 Introduction

Competition between different types of stores is a hotly debated topic in the U.S. Federal Trade Commission (FTC). Mergers are often allowed or denied depending on the definition of a certain market and who the players in the market are. A number of examples of how the FTC treats competition between different storefronts applies to food/grocery mergers alone. There is often disagreement about whether different types of stores compete with one another but with little empirical evidence directly from how consumers behave when shopping at different types of storefronts.

Often, mergers shed light on how tricky but important it is to formally define a market of competitors. A historical example is a merger between two natural food companies. The merger originally went through because "a district court concluded that PNOS (premium, natural, and organic supermarkets) was not a distinct market and that [Company 1] and [Company 2] compete within the broader market of grocery stores and supermarkets." [1] However, it was then overturned when the FTC contended that these two retailers are the two largest operators of what it called PNOS. The nature of competition would depend on whether consumers treat the stores as complements or substitutes both between the two stores themselves and within the broader market.

If consumers treat the stores as complementary to one another, then we might feel the merger would have a lower anti-competitive effect on consumer welfare. Complementary treatment could mean, for example, they treat one retailer as having better fresh produce options but the other has a larger selection of pre-packaged snack foods. The lower anti-competitive effect would then be due to the ability of each store to change prices. One store changing the prices of any single good would not drive shopping behavior as heavily towards the other, so there would be less incentive to change prices at the individual retailer level.

1. The FTC 2010 press release cited in the bibliography details the court proceedings and the changing rulings of the US FTC over time throughout the merger process

However, if found that consumers treat the stores more as substitutes and will often shop for deals on the same goods at one store or the other by going to both, the merger would have a larger anti-competitive effect on consumer welfare as post-merger price increases would benefit the retailers more than in the pre-merger period.

There is evidence documented in Bhattacharya et al. [2023] that mergers raise prices around 1.5% but with substantial heterogeneity. Around a quarter of mergers even end up decreasing prices at least around 5.1%. This motivates how important it is to ex-ante understand the nature of competition between retailers and which case we are in.

Furthermore, it is important to think about not only the level of complementarity across the two stores but to understand how complementary each store is with other grocery formats. Shoppers might treat the two natural foods retailers as substitutes but also treat them as similar levels of substitutes with other traditional grocery retailers in the same area. This would then lead to lower anti-competitive effects on consumer welfare as a result of a merger as the broader market the two natural foods retailers competes in is larger.

Lately there has been speculation about whether small grocery store formats and private-label only store formats compete with traditional supermarket grocery store formats. This is especially relevant to a current merger between two large traditional grocery store retailers. The nature of competition could perhaps flip the opinion of a regulator if small grocery formats were deemed to compete with traditional grocery formats.

I propose a novel estimation strategy that ranks how consumers might treat entire stores as either complements or substitutes. The level of complementarity or substitutability between stores is a good indicator of how much they could be perceived as being competitors. By the classical definition of complementarity, the utility a consumer gains from consuming a set of complementary goods is larger than the additive utility of consuming each good separately. A classical example of this is hamburger patties and hamburger buns. Substitutes are the opposite of this. They are often goods that would yield nearly identical utility no matter which of the goods is consumed, we can think of different brands of napkins as being substitutes with one another.

Complementarity and substitutability is often well-defined across individual products in a consumption bundle. However, we can extend this thought process to behavior at the store level. One might think of combining two stores on the same trip where the value of leaving the house is higher if I can stop at two stores to complete my tasks. Another way of thinking about it is some might prefer certain stores for certain items and rotate between them to fill out your pantry each month. My model will capture the latter but could have extensions to the former with more data.

I estimate a model in my empirical setting of shopper behavior in Cook County, IL from 2006-2018. I use NielsenIQ Scanner Data from the stores in this geographic region to construct the product availability and subsequent prices. This is a weekly-level dataset that tracks the quantity and prices of universal product codes (UPCs) sold in each store for each week of the sample. I also employ the NielsenIQ Homescan Panel which follows a panel of shoppers over the same time period and reports which trips they take and what they bought on each trip.

I model the overall consumer treatment of stores visited by consumers to determine which pairs of stores exhibit shopping behavior that is more or less complementary. The first step in doing this is I employ a standard, nested logit demand model where the nests are at the store-choice/category level. I define store-choice as a store visited as part of a bundle of stores visited in each month of the shopper panel. Category is defined as a product category in line with the NielsenIQ scanner data. I employ this method on a subset of product categories. Traditionally, as Thomassen et al. [2017b] demonstrated, multi-store shopping with product categories presents an N-squared dimensionality problem in both the dimension of the pair of store choice and in the product-category level. I get around the first issue by restricting store choice to a small geographic market (approximately 5 digit zip code level). Thus there are few combinations of pairs of stores. I get over the second issue of dimensionality by employing a nested factorization technique that employs machine learning created by Donnelly et al. [2021]. This method scales computationally linearly in the number of product categories so if I define each nest at the store-choice/category level, I don't have to worry about cross-category decision-making blowing up the computational complexity of the demand estimation. In this way it also fits into the broader machine learning literature in Marketing which include papers such as Ruiz et al. [2020] and Smith et al. [2023] that employ machine learning techniques in both demand estimation and optimal pricing policies.

This setting is also a good use of the nested factorization technique more generally and it demonstrates how, given improved computational methodology, we can get around some of the dimensionality issues that plague multi-store multi-category demand estimation if we are careful about defining the choice sets for each consumer.

The second step of the estimation strategy is to construct counterfactual estimates of single and multi-store utility. I take the structure of the indirect utility function quite seriously and construct counterfactual store-level utilities based on what would be available to each consumer if they were to visit both stores or only one or another. I am careful not to double-count overlapping offerings between the two stores. I then subtract the bundled store-level utility from the sum of the utilities constructed at each individual store level. The difference is an estimate of the complementarity between the two stores.

Once I have this estimation of the level of complementarity between two stores, I correlate this measure with various characteristics of the stores. I find that, partially by construction, stores with a larger differences in number of brand offerings are often most complementary to one another. I additionally find that, controlling for number of products available, most consumers prefer a high variety store in the bundle. I also find some evidence that consumers don't necessarily prefer to put a private-label heavy store in their bundle.

The finding about the high-variety store suggests that consumers might perceive high-variety stores as places they might rely on in their shopping bundle as the value of the low-variety stores increase when paired with the high-variety store. This could have implications with defining competition that the FTC has particularly struggled with. They have been switching the decision to include specialty stores or low-variety stores in the market for traditional grocery stores that are generally high variety.

The paper builds on many techniques for the demand estimation procedure. It is fundamentally using micro data of individual consumer purchase behavior rooted in the works outlined in Dubé and Rossi [2019]. I build off of classical discrete choice demand estimation techniques pioneered by Berry et al. [1995], Berry et al. [2004], and Nevo [2001] that treat utility as a linear combination of product and user characteristics. I additionally build off of the empirical strategies outlined in Smith [2004] which combine estimates of store choice with the consumer behavior at that store in an empirical setting of smaller retailers in the UK.

This paper fits into a broader literature on the estimates of complementarity between goods. A canonical example of how to evaluate complementarity comes from Gentzkow [2006] which estimates whether online and print newspapers are complementary goods in an empirical setting. The paper employs a demand estimation using the introduction of the new good, online newspapers. Then with a measure of complementarity, one can make a broader estimate of how demand for the other good would change with price changes in the potentially complementary good. This paper outlines an effort to take seriously the consumer behavior across goods that could potentially have complementary effects when a new type of good is introduced in a market. There is also an effort to incorporate this into a BLP demand estimation framework with complementarity as Wang demonstrates

The paper also fits the broader literature of analyzing multi-stop shopping behavior and how category-level competition impacts how stores interact with one another. Thomassen et al. [2017b] discusses how one-stop shoppers have larger pro-competitive impacts by employing a large multi-store, multi-category discrete choice model to model consumer behavior when deciding what to purchase at potentially multiple stores. Vroegrijk et al. [2013] specifically study how consumer's treat private-label only discounters in their consumption bundle as often stops to these stores are bundled with stops to other stores. Both of these papers deal heavily with complementarity between stores: the first is agnostic about the nature of complementarity and uncovers its effects by making assumption to reduce dimensionality in the discrete choice framework. The second parameterizes complementarity between stores and includes an estimate of complementarity that stems from some metric of distance between the product characteristics.

There is relatedly much work on retail assortment and how that can play into shopper behavior such as work by Briesch et al. [2009], as well as location amenity of retail choice which is the natural application of a complementarity term in the broader application of understanding overall market structure and competition. This is related to work on retail locations as local amenities as shown in Almagro and Dominguez-Iino [2022] and Ramos-Menchelli and Sverdlin-Lisker [2022] and trip level costs of shopping and demand estimates with trips and travel times documented in Athey and Schmidt [2018], Bell et al. [1998] and Jindal et al. [2020]. There is also evidence that store location might not be that restrictive to consumer behavior as evidenced in Allcott et al. [2019] which report that food deserts are largely driven by demand instead of supply.

My technique of estimating complementarity structurally can provide another angle to an often difficult-to-measure but important component of estimates of multi-store competition. Overall, this paper proposes a useful metric in using shopper behavior to suggest how complementary or substitutable stores are. It creates a framework for thinking about complementarity across bundles of goods, which can be difficult to define. It also suggests a more general framework where other characteristics such as distance traveled or size of the store could be used in the same model framework to add robustness and provide a more clear understanding of the boundaries of any given market. This metric can be used to evaluate how shoppers actually treat stores within their consumption bundle and better evaluate what could happen as a result of a merger in an application to antitrust litigation.

The next section describes the sources of data used. Section 3 presents summary statistics and background. Section 4 introduces a model with some examples that motivate the empirical strategy. Section 5 describes the structural estimation strategy. Section 6 presents all results, and section 7 concludes the paper.

2 Empirical Setting and Data

My empirical setting will be shopping behavior using both the NielsenIQ Household Survey (HMS) and the NielsenIQ Scanner data[2], in Cook County, IL. I additionally include some summary statistics on store composition from the publicly available SNAP retailer dataset from the US Department of Agriculture (USDA). As per NielsenIQ regulations, I do not combine the information from these two datasets as I cannot identify the retailers in the NielsenIQ data. I only use the latter dataset to understand trends in store-level composition to motivate further the empirical exercise.

2.1 NielsenIQ Consumer Panel Data

The NielsenIQ Consumer Panel Data (HMS) is a longitudinal survey of approximately 60,000-70,000 households per year on shopping trips and purchase behavior. The HMS panel includes household characteristics such as household income, demographics, and the makeup of the household, e.g. presence of children and how many.

The HMS panel includes every trip taken by each household and the anonymized store code of the place visited. It also includes the Universal Product Code (UPC) of all items

2. NielsenIQ data is available through a partnership with the Kilts Center for Marketing at the University of Chicago Booth School of Business

purchased along with all purchase quantities, and the total dollar amount spent.

From this panel, we can construct overall shopping decisions regarding which stores consumers frequent and whether we can classify them as a single- or multi- stop shopper.

2.2 NielsenIQ Retailer Scanner Data

The NielsenIQ Retailer Scanner Data is a weekly dataset of all prices and quantities purchased at participating retailers across the US. I use the data from the years 2006-2018 as well to construct the choice sets each consumer sees in each product category. The quantities and prices sold at each store are reported at the Universal Product Code (UPC) level which delineates the brand each product is sold at. Importantly, the prices are normalized for products of different sizes so they can be compared one to one.

Product Category Subsets

The product categories I ultimately chose were bread, detergent, mushrooms (frozen and fresh), berries (frozen and fresh strawberries), potatoes (frozen and fresh) and carrots (frozen and fresh).

2.3 USDA Historical SNAP Retailer Locator Data

This is a dataset from the US Department of Agriculture (USDA) which reports information on the stores that accept aid from the government's Supplemental Nutrition Assistance Program (SNAP). I use this data just to look at the general composition of store types in the US over time to report general trends in the next section. Since all grocery stores accept SNAP, the universe of stores in this dataset is larger than that in NielsenIQ and we can use this data to understand general trends in store type composition.

3 Background

I would like to uncover store-level complementarities in a revealed preference framework. In a setting where retailers decide both where to locate and what products to offer on the shelves, this type of demand estimation could prove quite useful in understanding the competitive landscape between retailers that might poise themselves to be complementary storefronts and not a one-stop shop. This is increasingly the case in the UK and is becoming more commonplace in urban areas in the US as the number of unconventionally small grocery stores in the US has been growing over the last 20 years.

I present a few stylized facts about the composition of retailers in the US over the last few decades, focusing specifically on the supermarket and grocery industry. In many industries, not just the grocery industry, there has been a rising trend in specialized stores. This is documented in recent literature from Ekerdt and Wu [2024] and Rossi-Hansberg et al. [2021b] at the broader firm level. At the grocery-specific level, I document a few stylized facts. These facts are all from the dataset of SNAP-accepting retailers which has information on the stores, the exactly locations of the stores, and the date of opening or closing. I restrict a store's market to the zip code level. I motivate complementary storefronts with a small case study about some private-label heavy storefronts, documenting their rise and the implications for neighboring stores. There is a trend in increasing specialty stores/small grocery store formats in the US. This is coupled with a well-documented upward trend in superstores as well as displayed in Leung and Li [2022] and Holmes [2011]. And, interestingly, we see that the traditional grocery store format has fallen in a way that mirrors the rise in super stores, suggesting a replacement effect.

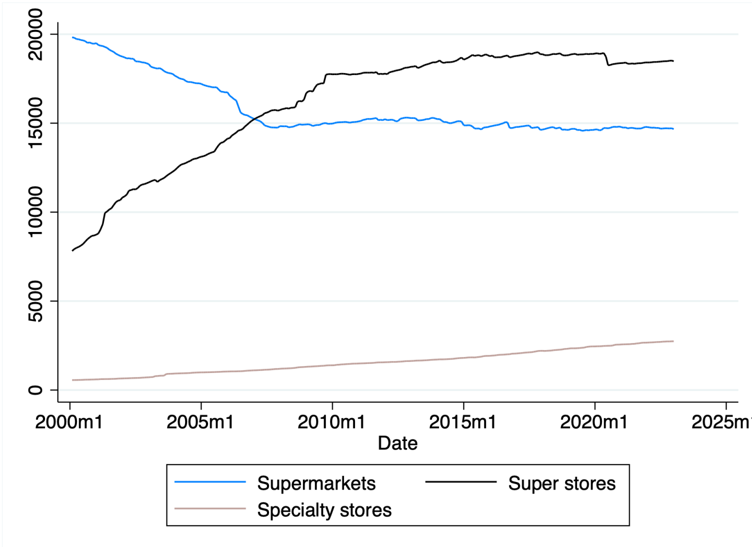

Figure 1: Composition of stores by type

If we zoom in on just the count of private-label specialty stores, we see a dramatic rise in these storefronts over time. They are often touted as being able to exist more easily in urban settings due to their small square-footage. This means they do not need to take over real estate that is traditionally meant for grocery stores and can more easily exist in urban real-estate settings.

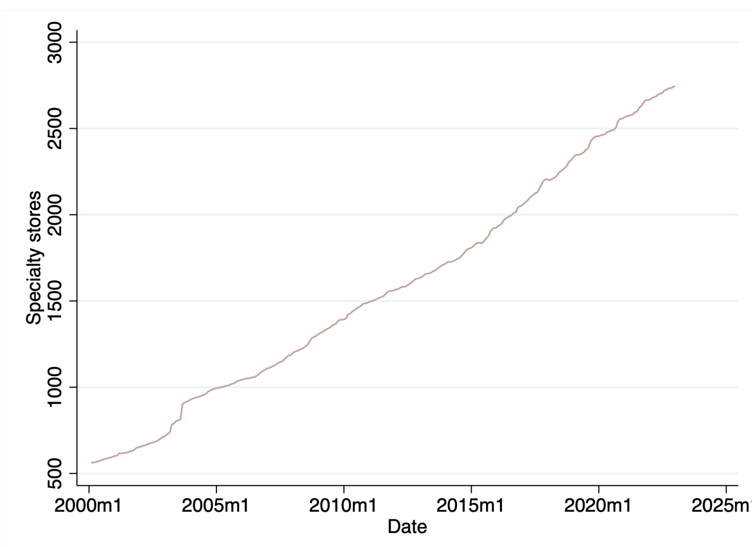

Figure 2: Trend of private label specialty stores

This phenomenon documents the rise of the private-label only specialty store. These stores reportedly mimic store formats of small groceries in Europe which have long since seen store formatting similar to the square footage of these smaller specialty stores. It has been reported in The Grocer magazine these stores are often used as one stop of many in the broader shopper bundle. This could be due to the fact that they are both conveniently located but small enough that they generally cannot support the needs of an entire grocery bundle.

4 Model Illustration

4.1 Single Good Case

In order to measure the level with which consumers treat stores as complements, it is nice to think of a store more generally as a bundle of offerings. We model complementarities across bundles within a general complementarities framework. Recall the Gentzkow [2006] complementarities definition: Take any two goods, i and j, and denote the bundle of the two goods as r. The utility of consuming bundle r is:

$$U^r = U^i + U^j + \Gamma^r \tag{1}$$

The intuition here is that there is some additional level of utility that is gained from both goods being consumed together. This is on top of the utility of both goods being consumed separately and summed together. A positive Γ^r denotes that there are synergies to consuming i and j together as a bundle. The canonical example of this is hot dogs and and hot dog buns. One could think that the utility of consuming both together is larger than the utility of consuming just the hot dogs plus the utility of consuming just the buns.

4.2 Extension to Bundles

This exact logic breaks down when being applied to two *bundles* of goods. The level of complementarity between two goods is generally captured in this residual utility term. However, when we think about two storefronts being complementary, the intuition behind this residual utility term breaks down. The reason why is because of *reallocation* among goods in the bundle.

Illustrative Counterexample

We want to understand why *reallocation* of purchase behavior breaks down the classic complementarity definition when applying it to two bundles of different varieties. Take the following illustrative example with two stores. We can think of one of the stores as being high variety and one of them as being low-variety:

Store 1	Store 2
A	A
B	B
C	
D	

Figure 3: Store pair example

We care about utilities net of price of a consumer that prefers to buy goods A, B, C, and D. They are indifferent about where they buy good A, but goods B and D are complements, and they are even more complementary if bought at different stores. Additionally, they do not gain any additional utility from re-purchasing a single item. We can liken this to a weekly shopping trip with perishable goods where the agent only cares about buying one of each possible item.

We can illustrate the utilities as follows under the notation U_A^1 denoting the utility of buying good A from store 1. The following is the underlying utilities for the consumer for all goods bought individually:

$$U_A^1 = U_A^2 = 1 \tag{2}$$

$$U_B^1 = U_B^2 = -1 \tag{3}$$

$$U_C^1 = 1 \tag{4}$$

$$U_D^1 = -1 \tag{5}$$

Now we are in a world where B and D are complements and even more so if they are bought at different stores:

$$U_B^1 + U_D^1 = 3 \tag{6}$$

$$U_B^2 + U_D^1 = 4 \tag{7}$$

Assume travel costs are zero. The total utility from going to just store 1 is 5. The agent would receive:

$$U_A^1 + U_B^1 + U_C^1 + U_D^1 = 5 \tag{8}$$

The total utility from going to just store 2 is 1. The agent would receive:

$$U_A^2 = 1$$

as the only good purchased from store 2 alone would be good A.

However, if we went to both stores together, the utility from going to both stores would be:

$$U^{(1,2)} = U_A^2 + U_B^2 + U_C^1 + U_D^1 = 6 \tag{9}$$

Now let's put this back in our original Gentzkow framework.

$$U^{(1,2)} = U^1 + U^2 + \Gamma^{(1,2)} \tag{10}$$

$$6 = 5 + 1 + \Gamma^{(1,2)} \tag{11}$$

$$\implies \Gamma^{(1,2)} = 0 \tag{12}$$

The definition of complements in the original sense would classify these two stores, or bundles, as independents as there is no additional utility gained from the utility of going to both stores. However, this yields a contradiction as by revealed preference, an agent would be prefer to go to both stores than to any one store on it's own.

$$U^{(1,2)} > U^1 > U^2 \tag{13}$$

We've also set up the example to show that there are complementarities to be had from visiting both stores, however, it does not work to use demand estimates to estimate store-level utility and back out the complementarity term. This is due to consumer reallocation within bundle.

Reallocation means that consumers will reoptimize within their bundle. This leads to a double-counting of certain products, in this case product A that will lead to extra utility being counted.

When comparing the competitiveness between two different stores, it is novel to use consumer behavior to estimate how shoppers might treat each store as substitutes or complements. If the utility of multi-stop shopping and going to both stores is much higher than the utility of going to one or the other, we can think of them as complements. This will come out if we employ a demand estimation strategy that captures utility will be higher if purchases are generally more frequent at stores if consumers also go to the other. This is in line with the monotonicity axiom outlined in Iaria and Wang [2021b]. However, due to the above counterexample, we cannot blindly estimate the utility of going to one store or the other and then add them to get an additive complementarity term.

4.3 Modeling Complementarities across Bundles

I now define a simple way to use the estimates of consumer demand to construct some type of complementarity term across bundles. The intuition is to construct utility estimates of each product category both at each store individually, and when the bundle of stores is visited. The complementarity term is defined as:

$$U^{(1,2)} = \widehat{U}^1 + \widehat{U}^2 + \Gamma^{(1,2)} \tag{14}$$

This implies that for any product category, c

$$\widehat{U}^1_c + \widehat{U}^2_c = \frac{U^1_c + U^2_c}{\sum_{v=1,2}\{\mathbb{1} \text{ if } c \text{ would be purchased individually from } v\}} \tag{15}$$

$\widehat{U}^1$ and $\widehat{U}^2$ are constructed such that there will be no double-counting of goods in any individual product category.

Application to Example

Let's revisit our above example with the same framework, and we make the assumption that on each trip, the consumer purchases at most one item in each of the product categories A, B, C, or D.

Store 1	Store 2
A	A
B	B
C	
D	

Figure 4: Store pair example

In this scenario, we could average the utility of the goods A, B from each of the stores 1 and 2, if they are respectively purchased at those stores, such that:

$$\widehat{U}_A^1 + \widehat{U}_A^2 := \frac{1}{2}U_A^1 + \frac{1}{2}U_A^2 \tag{16}$$

However, since product B would not be individually purchased if a consumer only went to store 2, no such re-scaling would occur

$$\widehat{U}_B^1 + \widehat{U}_B^2 := U_B^1 \tag{17}$$

In our example,

$$\widehat{U}^1 + \widehat{U}^2 = \widehat{U}_A^1 + \widehat{U}_A^2 + \widehat{U}_B^1 + \widehat{U}_B^2 + \widehat{U}_B^1 + \widehat{U}_B^2 + \widehat{U}_C^1 + \widehat{U}_C^2 + \widehat{U}_D^1 + \widehat{U}_D^2 \tag{18}$$

$$= \frac{1}{2}U_A^1 + \frac{1}{2}U_A^2 + U_B^1 + U_C^1 + U_D^1 \tag{19}$$

$$= 5 \tag{20}$$

Thus

$$\Gamma^{(1,2)} = U^{(1,2)} - \widehat{U}^1 - \widehat{U}^2 \tag{21}$$

$$= 1 \tag{22}$$

Note, that we can interpret this technique in two ways. The first is that there could be a re-scaling of any double-counted product categories between the bundle of stores. In this way, we do not allow for reallocation to effectively contaminate the estimate of a complementarity term.

The other way we can interpret this technique is that $\widehat{U}^1 + \widehat{U}^2$ are an averaging of two different scenarios. The consumer then is allowed to go to store 1 first and treat it like a single-stop shop. Then we allow them to go to store 2 next to fill the holes of everything they were not able to buy at store 1. Then we allow them to switch the ordering of these stores in the same thought experiment. The $\widehat{U}^1 + \widehat{U}^2$ would then be an averaging of these two scenarios.

This mathematically captures the same idea as above where if you were going to store 1 first, you would purchase all four goods at store 1 and nothing at store 2. If you go to store 2 first, you purchase A at store 2 and then B, C, and D at store 1. Both give a utility of 5, the average of which would be 5.

This is another way in which we could imagine $\Gamma^{(1,2)}$ really does capture some level of complementarity between the two storefronts. The caveat for measuring complementarity in this way is that it is constructed from indirect utility functions and thus can only be interpreted in an ordinal fashion.

In the next section, I propose a method for empirically estimating $\Gamma^{(1,2)}$ in two steps. The first step is setting up a demand estimation procedure that stems from an indirect utility function which captures cross-store complementarity by carefully defining the consumer's choice sets. I then counterfactually estimate $\widehat{U}_c^1 + \widehat{U}_c^2$ for each product category $c \in \mathcal{C}$ the product categories available across both storefronts. Armed with the product category-level counterfactuals defined for each bundle, I come up with $\Gamma^{(h,k)}$ which is the complementarity term for each pair of stores in the sample (h, k).

Finally I will present my results which correlate $\Gamma^{(h,k)}$ with parameters of interest that could help to define which characteristics of paired stores would lead to higher or lower complementarity. This can eventually inform estimates of consumer welfare in merger analysis.

5 Empirical Strategy

The structural estimation to uncover complementarities across storefronts is two-fold in my setting. The first is to run a very careful demand estimation that allows for decisions by shoppers to purchase goods at either store and treats each of these products as different, as motivated in the stylized example. The second is to construct the complementarity term making sure not to double-count any given product by using the utilities each consumer gets by going to only one store or another and constructing a counterfactual utility based on user and item observables, and item availability.

The motivation behind the careful demand estimation that is run is that we want to uncover and construct a counterfactual utility that a consumer will get by going to only one store or another inside of their bundle. This will help to understand how much additional utility they might get by combining and going to both stores. I take seriously the coefficients from the demand estimates so we need to set up the demand estimation in a clever way.

Once we have the demand coefficients, I construct the $\Gamma^6 r$ terms by taking seriously each demand coefficient. I would like to construct each of the Γ^r terms by building up the complementarity term. I first describe in detail the final estimation strategy, and each latent parameter that is estimated. Then I use each estimated latent parameter to build back up the complementarity term, Γ^r for each pair of stores, $h, k \in r$.

It is important to note that I have an ordinal, not a cardinal measure at the end of the day. As utilities do not have units, all I can do is order the pairs of stores by more or less complementary given my estimation strategy. After presenting the estimation strategy and summarizing the final results, I will correlate the complementarity terms with various covariates the suggest they do in fact capture ordinal levels of complementarity.

My estimation strategy extends on a multinomial logit discrete choice demand estimation, and employs a nested factorization technique to overcome issues of dimensionality. I take the notation from Donnelly et al. [2021] to first introduce the nested factorization technique, then I expand upon it to fit a model that can uncover complementarities empirically.

There are two key ingredients to model well in a demand estimation framework for complementarities among stores. The first is product category level demand. The second is to be careful with including both store choice and product choice.

5.1 Product Category Demand in a Single Store Framework

I first discuss product category level demand. In order not to totally explode the dimensionality of the problem, we want to estimate how the store-category level utility plays a role in each consumer's buying decisions. We can use a nested logit framework for intuition about this. This piece of the estimation strategy is taken from Donnelly et al. [2021] for a method of demand estimation that will scale linearly with the number of product categories. General nested frameworks are often used to deal with issues of large dimensionality concerns[3]. This becomes increasingly important as in order to measure complementarities among stores, we need to take into account many product categories to capture shopping behavior illustrated in the example above.

In order to fix ideas to build up the estimation strategy, consider the one-store case with multiple product categories.

For each product category $c \in \mathcal{C}^h$ where $\mathcal{C}^h$ is the set of product categories offered at store h, I allow the consumer to purchase only one product per product category which we will define as the UPC $j \in \mathcal{J}_c^h$ where $\mathcal{J}_c^h$ is the set of all UPCs for a product category c offered in store s. For each product category, we can define the indirect utility of consumer i shopping for category c at time t to be

3. also see Iaria and Wang [2021a] for more discussion on how generalized nested logits can aid in reducing dimensionality

$$U_{ict} = \vartheta_i\beta_c + \rho_c y_{it} + \psi_i x_c' - \phi_i\lambda_c IV_{ict} + \epsilon_{ict} \tag{23}$$

where y_{it} are consumer observables, x_c are category level observables, and $\{\vartheta_i, \beta_c, \rho_c, \psi_i, \phi_i, \lambda_c\}$ are latent choice parameters to be estimated. Each individual product category has an entire inclusive value term that is additionally estimated in the first step of the procedure. Recall the inclusive value is built up from characteristics at the UPC level, $v \in \mathcal{V}_c^s$

$$IV_{ict} = E[\max_{v=1...V_c^s} u_{ivt}] = \log \sum_{v=1}^{V_c} a_{vt} \exp u_{ivt} \tag{24}$$

and at the product level we have mean product level utility and price

$$u_{ijt} = \theta_i\beta_j + \rho_j Y_{it} + \sigma_i X_j - \gamma_i\lambda_j P_{jt} + \epsilon_{ijt} \tag{25}$$

where P_{jt} is the price of product j in time t, Y_{it} is the vector of all consumer-level observables for individual i at time t, and X_j are the product-level observables. Due to data limitations, these are fixed over time. The latent choice parameters to be estimated in this step are $\{\theta_i, \beta_j, \rho_j, \sigma_i, \gamma_i, \lambda_j\}$.

This is exactly the case of the empirical strategy in Donnelly et al. [2021] and the estimation strategy works at the single store level. I build on this intuition to now allow for multi-store estimates where the utility of any single good purchased at any store is different depending on the other stores they go to. In order to do this, I introduce a new index that I use to redefine the space of the consumer choice sets. This is $v \in V_c$ where v is the specific UPC or product which is store specific. The broader choice set it comes from V_c includes all the product categories in the store visited also indexed by the bundle of stores visited. Thus the items in each product category are now specific to the bundle of stores visited which blows up the number of choices in each product category. And each product in each category can be purchased many different ways in each market. [4]

5.2 Demand Estimation with Multi-Stop Shopping

Step One: Inner Nests in the Multi-Store Case

Once we understand the one-store case, we can add a bit more notation to this framework to encompass the multi-store case. We must make the following assumptions in order to limit issues of dimensionality:

1. Assume each individual only has the option to shop at stores in their zipcode of residence, also known as the market, m.

2. We only consider the shopping behavior of the individual at the monthly level, which is our time period, t.

3. Consumers do not go to more than two stores in any given time period t, these two stores are chosen first by frequency of trips, then tiebreakers on dollars spent per store

4. Consumers are considered single-stop shoppers for the month if more than 80% of their trips are to the same store.

5. Assume that if they go to a certain retailer, that retailer is the one associated with their zipcode

6. All retailers in each zipcode are treated as identical and their product offerings are combined and weighted based on the total sold across all stores

7. Consumers only buy at most one of each item in each product category

4. An alternative demand specification which expands along the dimension of the number of nests instead of within nest is outlined in Appendix 1

Note that we do not assume that identical products purchased at different stores are identical. They are included as different products in the demand estimation. This requires some of the above assumptions to be made in order to not drastically blow up the dimensionality of the problem as each product is assigned to a specific nest which is defined as the store/choice/category level. This means that the private label organic mushrooms at store h when a shopper goes to both stores h and k are treated as a different product than the private label organic mushrooms at store h when a shopper only goes to store h in the shopping period.

As the nested factorization method assigns each latent parameter at the user and product level[5], we have many coefficients that are estimated and for each product category.

Overall I have to construct each of the nests in a careful way in order to capture the appropriate consumer behavior. The size of each individual nest can blow up the computational complexity of the problem which is the main constraint in the model estimation. Therefore, in order to capture the complexity of the multi-store behavior, I make coarser the individual store variety.

I classify the varieties of each product category at each store as private label, organic, and I have some product specific bins as well. On average, there are about 8-16 varieties of each product at each specific store. For example, the product category of bread, I also break down into white and wheat.

For example, in the product category of bread, the list of all products in the product category would be:

1. private label, non-organic, white

2. private label, organic, white

3. private label, non-organic, wheat

5. see Appendix 2 for more details on the estimation strategy

4. private label, organic, wheat
5. branded, non-organic, white
6. branded, organic, white
7. branded, non-organic, wheat
8. branded, organic, wheat

I detail all of these covariates in the Results section of the paper. Note that some of these covariates are common across all product categories but some of them are not. I make the call to separately estimate the coefficient of each of these covariates within nest and therefore do not allow shoppers to have the same propensity for private label goods across the board. Rather they have a different affinity for private label goods depending on the product category.

The indirect utility from the inner nests is constructed at the product level so for

$$u_{ivt} = \theta_i^T \beta_v + W_i^T \rho_v + \sigma_i^T X_v - \gamma_i^T \lambda_v price_{vt} + \epsilon_{ivt} \tag{26}$$

Step one in the estimation is to sort each item into different product categories and understand that they are purchased conditionally if the product category is chosen. The latent parameters in this step of the estimation are $\{\theta_i, \beta_v, \rho_v, \sigma_i, \gamma_i, \lambda_v\}$.

In this step, we estimate the conditional purchase probability of purchasing product v in product category c given that a purchase is made in category c. This is due to the assumption of the error term ϵ_{ivt} being distributed as Type 1 Extreme Value.

$$P\left(y_{ivt} = 1 \mid \sum_{v=1}^{V_c} y_{ivt} = 1\right) = \frac{\exp(u_{ivt})}{\sum_{k=1}^{V_c} \exp(u_{ikt})} \tag{27}$$

There ends up being a lot of nests given this framework and within each nests, we also need to include each possible combination of stores as the same product purchased at the same store but with a different pair of stores visited in a time period is considered a different product. That's why reducing the variation in the actual products offered instead of including all UPCs helps to make this model tractable.

Step Two: Outer Nests

We can rewrite our category-level demand as now also a function of both the store, h, and the choice made by the consumer of which bundle of stores (either 1 or 2) to shop at, r, indexed by the market of stores in each consumer's zipcode, m:

$$U_{ict} = \vartheta_i \beta_c + \rho_c y_{it} + \psi_i x'_c - \phi_i \lambda_c IV_{ict} + \epsilon_{ict} \tag{28}$$

The product-category level utility is still the same, however, under the hood in the inclusive value term we have

$$IV_{ict} = E[\max_{v=1...V_c^{hr}} u_{ivt}^{hr}] = \log \sum_{v=1}^{V_c^{hr}} a_{vt}^{hr} \exp u_{ivt}^{hr} \tag{29}$$

where we expand the set of feasible products in each product category. These would now include $\mathcal{V}_c^{hr}$ which is all products purchased in at store h in product category c where the bundle of stores visited in that period is r.

The outer nest is then defined in a way that would replace 'price' in a standard demand framework with the inclusive value term: the outer nest is estimated by:

$$U_{ict} = \theta_i^T \beta_c + W_i^T \rho_c + \psi_i^T X_c - \phi_i^T \lambda_c IV_{ict} + \mu_c \delta_t + w_{ct} + \epsilon_{ict} \tag{30}$$

the outer nest is estimated at the product category/store/choice level, where c is specific to store h and bundle of stores chosen, r. We end up getting an observation per user/category for each coefficient in the outer nest estimation.

The outer nest assumes the error term ϵ_{ict} is distributed Type 1 Extreme Value and thus estimates the probability of purchasing from the nest at all

$$P(y_{ict} = 1) = \frac{\exp(u_{ict})}{1 + \exp(u_{ict})} \tag{31}$$

This captures complementary behavior within store, where the utility of a certain product in a specific store depends on the other stores visited, expanding your choice set. This would capture the case of consumers buying more of certain goods at store 1 if they pair them with other goods at store 2.

In order to make the final part of the estimation strategy work, we have to be careful with how to estimate the utility of the outside option. This would be the utility the consumer gets by going to store h if they are single-stop shopping or if they are multi-stop shopping, stores h, k and then not buying anything from product category c. This is important in constructing the estimates of $\widehat{U}^h + \widehat{U}^k$ which go into the final complementarity estimates. [6]

The demand estimation runs into issues of power with the space of the choice set. Thus, I made some calls where I drop all combinations of stores that are not often paired together both across time and geography in my sample.

Step Three: Constructing Complementarity

The complementarity estimation strategy uses demand estimates from the inner and outer nests.

The inner nests are used to correct for the accounting issue of double-counting each

6. Given the treatment of the outside option, this makes the interpretation of the coefficients that are estimated from the outer nest a bit more difficult to interpret. As the end goal is the complementarity measures, I do not think it gets in the way of the final estimates.

product category when calculating store-level utility and the outer nests provide a store-level utility term that we difference at the end.

We first construct the inclusive value terms for all the categories offered by each pair of retailers we want to measure the complementarity term for. For each individual i, we only assume they purchase the items at the retailer that gives them the highest utility for that item, so the inclusive value term for each product category c

$$\widehat{IV}_{ict}^{(h,k)} = E\left[\max_{v=1\ldots V_c} U_{ivt}\right] = \log \sum_{v=1}^{V_c} \exp u_{ivt} \tag{32}$$

for V_c being the set of products that would yield the consumer the maximum utility net of price the consumer gets of each item purchased in pair of retailers, h, k. We are therefore not allowing for products to be double-counted in the inclusive value term.

We then construct the inclusive value terms for the pair of retailers as if we were forced to go to each retailer separately. I do this in two ways by changing V_c

Method 1: The set of goods we can choose from in each product category is randomly plucked from each store from the single-stop utilities and overlap is not allowed. If a good is not available in the store, the outside option utility of going to the store and not buying anything is used.

Method 2: Similar to method 1 but without randomness. We construct the utility by taking the mean of having to buy your entire bundle at each store separately, taking into account the outside option utility of going to the store and not buying anything if the product is not available. These yield similar results.

We can explicitly write out

$$\widehat{IV}_{ict}^{h} + \widehat{IV}_{ict}^{k} = \log \sum_{v=1}^{V_c} \exp \frac{u_{ivt}}{\{\mathbb{1} \text{ if } v \text{ would be purchased at the shop on its own}\}} \tag{33}$$

After this is all taken into account, we can use these new inclusive value estimates for each store, $\widehat{IV}_{iht}$ and we construct the total difference in utilities by taking each store-level utility where each product-category is at the product-category-store level. The stores visited are h, k and I denote the bundle of stores as r. We can see that the utility of the bundle of stores would be

$$\widehat{U}_{irt} = \sum_{c^r \in C_r} \theta_i^T \beta_{c^r} + W_i^T \rho_{c^r} + \psi_i^T X_{c^r} - \phi_i^T \lambda_{c^r} \widehat{IV}_{ic^r t} + \mu_{c^r} \delta_t + w_{c^r t} + \epsilon_{ic^r t} \tag{34}$$

And we construct each store-level utility as follows:

$$\widehat{U}_{iht} = \sum_{c^h \in C_h} \theta_i^T \beta_{c^h} + W_i^T \rho_{c^h} + \psi_i^T X_{c^h} - \phi_i^T \lambda_{c^r} \widehat{IV}_{ic^h t} + \mu_{c^r} \delta_t + w_{c^h t} + \epsilon_{ic^h t} \tag{35}$$

Now constructing the estimates of Γ_r we see

$$\Gamma^{(h,k)} = \widehat{U}_{it}^{(h,k)} - \widehat{U}_{it}^{h} - \widehat{U}_{it}^{k} \tag{36}$$

These estimates of Γ^r are completely ordinal. By definition, utility is ordinal so they only make sense when comparing the constructed measure with characteristics of each store or set of stores. Thus, I present what this complementarity measure yields when looking at various characteristics at the store-level.

For notation, $\Gamma^r := \Gamma^{(h,k)}$ for $(h, k) \in r$ to simplify some definitions of store-level characteristics going forward.

6 Results

In this section I outline the main results from the paper. I recap the empirical strategy as I present the results from both the inner and then the outer nests. I detail specifically what design choices went into constructing the consumer-level decision-making. I then present the results of the complementarity estimates by correlating the estimates with various characteristics of the stores to then interpret which store-level characteristics are most likely associated with complementary shopping behavior.

For scale, the subset of data I am using is Cook County, IL. This subset longitudinally follows 855 households, which is proportional to the fraction of US population in Cook County. I collapse the data to the month-level for a total of 144 months for a total of 123,120 shopping periods. Within each period, I allow the shopper to go to at most two retailers. I restrict the sample to the six most frequented retailers in the dataset.

6.1 Inner Nests

Recall that the inner nests are estimated at the product level and are estimated each additively by:

$$u_{ivt} = \theta_i^T \beta_v + W_i^T \rho_v + \sigma_i^T X_v - \gamma_i^T \lambda_v price_{=vt} \tag{37}$$

As we get many latent parameters out of the demand estimation, and then linearly construct each of the complementarity terms. I first report the coefficient estimates for σ_i and $\gamma_i^T * \lambda_v$ as a sanity check on the direction of all the coefficients on product observables and prices. I report the mean and standard deviation of the σ_i terms which are the coefficients on product observables. I have different product observables for each product category in the estimation.

In general, there seems to be a premium on organic products, especially for mushrooms, berries, and bread. Consumers seem to be relatively indifferent about whether the product is private label or not, with only detergent exhibiting large brand preferences. Frozen food actually has a premium on mushrooms, potatoes, and berries. Potatoes make sense as most of the frozen potatoes in the sample are processed to include french fries and tater tots. Frozen berries are also convenient. As carrots keep well, it does make sense that there is little premium on frozen carrots.

With some of the observables I allow to differ among the products, I see effects that make sense with certain characteristics of mushrooms. Sliced mushrooms have a premium. I also see a large premium in what I've classified as a 'fancy' mushroom. This is an indicator for whether the mushroom is not white, baby bella, or portabello. So this would include mushrooms such as shiitake, oyster, wood ear, morel, or chanterelle. There is a large premium on these premium mushrooms.

Similarly, we see the same patterns in carrots where consumers prefer either shredded or baby carrots. There also seems to be premiums on both russet and sweet potatoes. The type of frozen berries do not seem to matter much, mixed is defined as frozen berries mixed with other fruits and medley is a berry medley. White and wheat bread do not have a premium. With respect to detergent, almost the entire effect is in the brand, and even liquid versus solid does not have much of a premium.

Table 1: Demand estimates summary - Item level observables by product category

Attribute	Mushroom	Carrot	Potato	Berry	Bread
private label	0.00302	1.29	-0.00138	0.00146	0.000506
	(0.0959)	(0.0514)	(0.0964)	(0.0964)	(0.0959)
organic	1.414	-0.000916	-0.0408	1.860	1.695
	(0.048)	(0.0955)	(0.0392)	(0.0315)	(0.0229)
frozen	1.840	-0.000428	1.702	1.815	
	(0.059)	(0.096)	(0.023)	(0.0293)	
sliced	1.537				
	(0.028)				
fancy	3.155				
	(0.0564)				
shredded		3.167			
		(0.0586)			
baby		3.156			
		(0.041)			
russet			3.243		
			(0.0362)		
sweet			3.094		
			(0.0513)		
mixed				0.595	
				(0.0919)	
medley				0.00283	
				(0.0961)	
white					0.00553
					(0.0959)
wheat					0.00174
					(0.0958)[7]

7. Recall that this is a summary of the mean and standard deviation of all the coefficient estimates in the latent parameter set that can be interpreted as the coefficient of item observables, aggregated to the item category level. While the standard deviation gives a sense of the dispersion of the estimate, it cannot be interpreted directly as a standard error. The mean coefficients are reported with the standard deviations in parentheses.

Due to formatting, I cannot include the last column, the estimates for detergent, in the same table. The only attributes relevant to detergent are whether or not it is liquid and whether or not it is private label.

Table 2: Demand estimates summary - Detergent only

Attribute	Detergent
private label	-0.149
	(0.0257)
liquid	0.00268
	(0.096)

I also report price sensitivity which is calculated at the individual and item level. I combine the individual/product level price sensitivities and give the following decomposition among individuals and items purchased. Overall, these coefficients make sense and are negative in nearly all cases. The magnitudes are, however, quite low compared to the observable level coefficient estimates, meaning that variety offered plays more of a role than prices. This is actually more suggestive then that stores can be treated more as complements depending on the distance between their varieties offered. There was also not much change in prices throughout the sample period over time which is in line with the findings in DellaVigna and Gentzkow [2019].

Table 3: $\gamma_i * \lambda_v$ summary

	price sensitivity
count	1813007
mean	-0.010849
std	0.005284
min	-0.042187
25%	-0.014247
50%	-0.010772
75%	-0.007373
max	0.009173

The mean utilities for the inner and outer nests are reported in Appendix 3. By now in the estimation strategy, the magnitudes of the coefficients no longer become easy to interpret as we begin to use a number of scale parameters to construct the inclusive value terms. What now matters is relative differences between mean utilities.

Given our setup of the inner nest, the mean utility terms reported correspond to: $\theta_i^T \beta_v + W_i^T \rho_v + \sigma_i^T X_v$.

In the inner nest, we want to summarize what are the mean utilities of the products at each store net of prices. Before getting all the inclusive value terms, these are the average utilities for each product as we divide products into where they are purchased and the store-level choice sets. There should be no ranking of these model parameters as they summarize how good the store offerings are. It is nice to see that within retailers where the product is purchased, the mean utility term is relatively stable with a few dips for some retailers that are rarely paired with others such as 916. This is a bit of a drawback in doing the estimation strategy separating products based also on the pair of shops visited.

Recall the outer nest is estimated by:

$$U_{ict} = \theta_i^T \beta_c + W_i^T \rho_c + \psi_i^T X_c - \phi_i^T \lambda_c IV_{ict} + \mu_c \delta_t + w_{ct} + \epsilon_{ict}$$

The reported mean utilities for the outer nest are $\theta_i^T \beta_c + W_i^T \rho_c + \psi_i^T X_c + \mu_c \delta_t + w_{ct}$ The outer nests do not capture a lot of variation, most likely because the nests were too similar but the nearly identical coefficients leads to the majority of the estimation coming from the rearrangement and estimation of the inclusive value terms. Thus the subtraction between each of these mean utility terms when we linearly combine $\widehat{U}^r - (\widehat{U}^h - \widehat{U}^k)$ is quite negligible as a component of the overall complementarity term.

At the store-level, for store h, this utility can just be summed up over the entire product category levels

$$U_{iht} = \sum_{c^h \in C_h} \theta_i^T \beta_{c^h} + W_i^T \rho_{c^h} + \psi_i^T X_{c^h} - \phi_i^T \lambda_{c^r} IV_{ic^h t} + \mu_{c^r} \delta_t + w_{c^h t} + \epsilon_{ic^h t} \tag{38}$$

6.2 Model Fit and Precision

I report the fit and precision of the model. The model employs a machine learning technique and works to maximize a log-likelihood function detailed in Appendix 2. The dataset is split into a training, test, and validation set. I assign 67% of the data to the training set, 22% to the test set, and 11% to the validation set. Due to the very high dimensionality of the choice set, the model precision is quite low. Model precision means that the model predicts correctly exactly which product a consumer would buy in every product category while also predicting correctly which stores they went to each month and which store they ended up buying the product from. For future work, one could dive deeper into how the model precision responds to different definitions of choice sets and delineations of product categories.

However, we can compare model precision as compared to a baseline prior of all the coefficients being 0, which is our original prior. When estimating the latent parameters, we start with each random coefficient being distributed independently $\mathcal{N}(0, 1)$.

Inner Nest Precision

I have about 4.8 million observation in the training data, 1.4 million observations in the test data, and about 700,000 observations in the validation data. We look to see how the precision improves in the validation data over the baseline prior of each random coefficient being distributed $\mathcal{N}(0, 1)$. The very first iteration of the validation set being tested, before the first gradient descent, yielded a precision of 0.0019%, so barely better than zero. Within only 100 iterations of the gradient descent we already get up to 0.049% which is a 25-fold increase in precision over the naive prior. By the maximum number of iterations, we only improve to 0.05% precision which means that the model being so highly dimensional is not able to get past a pretty baseline improvement from the first 100 iterations of the nested factorization algorithm.

Outer Nest Precision

We consolidate nearly all of the product-level variation when moving to the second step of the estimation strategy going from the inner to the outer nest. As evidenced by the results, nearly all of the explanatory power comes from the estimates of the inner nests. In the outer nest, the validation set begins at about a 0.042% accuracy and increases to only about 0.047% by the maximum iteration, meaning that there is not much in the outer nest that can help with model fit.

6.3 Estimation of Γ^r and Correlations with Store-Level Measures

Brand Variations: A count of the difference in brand variation between the two shops (high variety and low variety would yield a higher brand variation measure). This measure will be denoted for each store h and product category $c \in C$ as

$$UPC_h^{avg} = \frac{\sum_{c \in C} UPC_c^{count}}{\sum_{c \in C} 1}$$

Which is just an average of brand variation among the product categories in our sample. This is a good statistic to sanity check the model because mechanically, the brand variation over time should be a big reason why stores are deemed more or less complementary. The availability of brands in each product category in my definition of Γ_{hk} is what will lead to higher or lower complementarity. This is because consumers will get more utility from more variety in each bundle.

Table 4: Brand Variation Summary

	Average UPCs/product
count	23030
mean	121.121803
std	34.436802
min	40.833333
25%	96.000000
50%	110.833333
75%	147.166667
max	225.333333

This table actually denotes some of the issues I had in originally constructing the demand estimates which is that there are a very high number of UPCs sold in each product category. This includes UPCs that could denote seasonal products, and often brands will have many UPCs for very similar products. In my estimates, I narrow the products down and include many UPCs for each single product I define as being sold in each store. For example, a product that is the private-label organic white bread sold in any store will count as only one variant in the actual estimation. In this way, this measure is not exactly mechanically related to the estimates of the complementarity term, Γ_r. However, the measure is still a nice summary of the type of store and whether or not there is high or low variety of products offered.

Brand Variation Results The regression for the complementarity term between any pair of stores in market m and time t uses the absolute value of the difference between the two levels of brand variation: $UPC_{mt}^{avg,hk} =\mid UPC_{mt}^{avg,h} - UPC_{mt}^{avg,k} \mid$

$$\Gamma_{mt}^{hk} = \alpha_{mt}^{hk} + \beta_{mt}^{hk} UPC_{mt}^{avg,hk} + \gamma_t + \rho_m + \epsilon_{mt}^{hk}$$

And the estimates for the result of correlation is as follows:

Table 5: Correlation with overall brand variation

	coef	std err	[0.025	0.975]
const	0.2877***	0.014	0.261	0.314
UPC_{mt}^{hk}	0.0039***	3.58e-05	0.004	0.004
N	41436			
R-squared	0.299818			

In this first case, the UPC count is clearly very correlated with the complementarity measure and this is by construction. This sanity check means that the estimates of complementarity, which depend heavily on product availability, are performing in the way we would like.

Private Label Measure: A count of private label offerings at each store i

$$PL_i = \frac{UPC_{pl}^{count}}{UPC^{count}}$$

Reported as the fraction of all UPCs that are private label, we see on average stores having nearly a quarter of all UPCs offered as being private label.

Table 6: Private Label Summary

	Private Label Percentage
count	23030
mean	0.230779
std	0.083751
min	0.003791
25%	0.181795
50%	0.215539
75%	0.299143
max	0.392437

We need to come up with a pairwise measure of private label goods. One way of doing this is to use the difference between any two stores as private label or not. I also come up with a measure that is average private label to denote whether they are both heavily private label or not. For now we denote the private label difference between any two stores, h and

k in market m and time t as the absolute value of the difference between the two levels of private label offerings:

$$PL_{mt}^{hk} = \mid PL_{mt}^{h} - PL_{mt}^{k} \mid \tag{39}$$

Then we perform a similar regression

$$\Gamma_{mt}^{hk} = \alpha_{mt}^{hk} + \beta_{mt}^{hk} PL_{mt}^{hk} + \gamma_t + \rho_m + \epsilon_{mt}^{hk} \tag{40}$$

The results are as follows:

Table 7: Correlation with difference in private label offerings

	coef	std err	[0.025	0.975]
const	0.4012***	0.015	0.371	0.431
PL_{mt}^{hk}	0.3067***	0.017	0.274	0.339
N	41436			
R-squared	0.103201			

Now we also see a very high correlation between private label and complementarity, however, this is likely due to how correlated private label composition is with overall brand variation. When we control for total brand variation and simply look at whether the coefficient for private label is still relevant, it actually goes the opposite direction. This means that consumers are less likely to pair store that are heavily private label with those that are not, conditioning on overall brand variation.

Table 8: Private label correlation controlling for brand variation

	coef	std err	[0.025	0.975]
const	0.2916***	0.014	0.265	0.318
PL_{mt}^{hk}	-0.0488***	0.015	-0.078	-0.019
UPC_{mt}^{hk}	0.0039***	3.67e-05	0.004	0.004
N	41436			
R-squared	0.299995			

This result warrants diving a bit deeper into how exactly to measure the difference between two shops in terms of their private label offerings.

Brand Coverage Measure: A measure of overlapping brand coverage between each pair of stores. The measure is as follows for any given retailers h and k.

$$Coverage_{hk} = \left(1 - \frac{UPC_h}{UPC_{total}}\right) * \left(1 - \frac{UPC_k}{UPC_{total}}\right)$$

To fix ideas here, if the measure is 0, that means that at least one of the stores offers nearly all the UPCs that the combination of both stores offer. That means that the product offerings are not so complementary. The measure can go up to 0.25 which would be an even split of all the UPCs offered between both stores. On average, this measure is about .12 for each pair of stores in each market in the sample.

Table 9: Brand Coverage Summary

	Coverage Measure
count	41454
mean	0.118830
std	0.026442
min	0.002975
25%	0.114522
50%	0.121378
75%	0.130413
max	0.200983

Note the store count is larger in this result because coverage is measured at the pairwise retailer level so there is an expansion of these measures within each market for this measure

The coverage term is already defined as pairwise so the regression is:

$$\Gamma^{hk}_{mt} = \alpha^{hk}_{mt} + \beta^{hk}_{mt} Coverage^{hk}_{mt} + \gamma_t + \rho_m + \epsilon^{hk}_{mt}$$

Table 10: Coverage measure correlation

	coef	std err	[0.025	0.975]
const	0.4254***	0.016	0.394	0.457
$Coverage^{hk}_{mt}$	0.0542	0.046	-0.037	0.145
N	41436			
R-squared	0.0958661			

We see that this measure on it's own is not statistically significantly correlated withe complementarity term. This makes sense as we can think about two ways in which people use storefronts as complementary. One in which they go to two different stores that each sell different sets of goods, or another way in which they go to a high and a low variety store. The low variety store could be convenient for many reasons and the high variety store serves as a complement. The brand coverage variable might be indicative of which direction people usually act on.

To fix this idea more, we can try also controlling for the total brand variation term which we know is highly correlated with the complementarity terms by construction. We see in this case the sign for coverage becomes negative.

Table 11: Brand coverage controlling for Brand count

	coef	std err	[0.025	0.975]
const	0.3706***	0.014	0.343	0.399
UPC^{hk}_{mt}	0.004***	3.62e-05	0.004	0.004
$Coverage^{hk}_{mt}$	-0.764***	0.041	-0.845	-0.683
N	41436			
R-squared	0.3055967			

This is highly suggestive that consumers prefer a high variety store in their pairing. By the construction of the Γ_r term, I would have expected it to actually be going in the other direction because there is an implicit penalty in the $\widehat{U_{iht}}$ for low availability in each single store. This implies that Γ_r would be larger if brand coverage were high and each store split their offerings more evenly. However, since most consumers most likely preferred pairing a high variety store in each shopping period, the utility terms for bundles with any high variety terms is driving this result.

Finally, just to sanity check that there are no odd relationships between the correlations, we run all three together.

Table 12: All correlations

	coef	std err	[0.025	0.975]
const	0.3709***	0.014	0.343	0.399
PL_{mt}^{hk}	-0.0096	0.015	-0.039	0.02
$Coverage_{mt}^{hk}$	-0.7603***	0.042	-0.842	-0.679
UPC_{mt}^{hk}	0.004***	3.69e-05	0.004	0.004
N	41436			
R-squared	0.305604			

We see there is not much change in the magnitudes although the private label measure does seem the least correlated with the complementarity terms estimated.

7 Conclusion

This paper seeks to understand the nature of competition between storefronts by modeling them as pairs of bundles in a shopping period. I first introduce a model of complementarity among goods and then modify it in order to capture complementarity across bundles. The modified framework lends itself nicely to a demand estimation strategy where we build up counterfactual utilities of each single store. These utilities will be able to account for the challenge of reallocation within bundles.

I employ a nested factorization demand estimation technique to estimate preferences in a discrete choice framework. I extend the demand estimation to include decisions both at the store and the product level. Due to the nested structure, I do not have to worry about the estimation strategy blowing up with the number of product categories. The estimation can therefore be extended to more product categories to capture more consumer behavior and scale linearly with this.

I use my demand estimation framework to then reconstruct the utility that each consumer would get by going to one store and then another, forcing the number of products purchased to be the same as if they went to both stores. This will get rid of the accounting issue of double counting products from each individual bundle.

From these estimates, I piece together the residual term, or the complementarity between two stores. I then correlate the complementarity with various store-level attributes. Then we can gain an understanding of which attributes at the store level lead to higher or lower levels of store-level complementarity.

Given that the data I am using is anonymous at the store and retailer level, I am only able to construct my own store-level covariates. With more information about particular stores, there could be more covariates in understanding what these complementarity measures are correlated with to help inform which aspects of stores are relevant for competition in merger analysis.

A limitation of my model is that the magnitudes of the complementarity term are less interpretable unless compared directly with one another. Thus, it gives a sense of the ordering of complementarity in storefronts. This does mean that the model needs to be run on many pairs of storefronts to understand the ordering of how shoppers treat each pair of retailers.

Further steps that could be taken with this model is to include travel costs. One of the main limitations of the model is that it assumes traveling is costless. Travel costs could be put in the outer nest to account for the additional utility a store would need to overcome for a shopper to choose to go there in any given period.

Another interesting angle is to look at the level to which online grocery or online retail complements or substitutes storefront retail. I did not have data on online retailers either but this method could be applied to understand how consumers treat online versus storefront retailers in their overall consumption bundle.

Overall I find that complementarity among storefronts is highly correlated with the difference in the number of brand offerings. This mechanically makes sense in the model. When controlling for the number of brand offerings, it seems not to matter whether the storefronts are very different in terms of their private label brand offerings with the sign of private label difference being slightly negative.

Additionally, I find that distance in store coverage is also negatively correlated with the complementarity measure. This suggests that consumers prefer a high-variety store in their shopping bundle, regardless of if they use a low variety store in their bundle. This suggests that traditional and low-variety private label stores could act more as complements than direct competitors to one another.

REFERENCES

Aboutaleb, Y. M. , Ben-Akiva, M. , and Jaillet, P. . Learning structure in nested logit models. *arXiv preprint arXiv:2008.08048.*, 2020.

Allcott, Hunt , Diamond, Rebecca , Dubé, Jean-Pierre , Handbury, Jessie , Rahkovsky, Ilya , and Schnell, Molly . Food Deserts and the Causes of Nutritional Inequality. *The Quarterly Journal of Economics*, 134(4):1793–1844, 2019.

Almagro, Milena and Dominguez-Iino, Tomas . Location Sorting and Endogenous Amenities: Evidence from Amsterdam. *Working Paper*, 2022.

Athey, Susan, Blei-David Donnelly Robert Ruiz Francisco and Schmidt, Tobias . Estimating Heterogeneous Consumer Preferences for Restaurants and Travel Time Using Mobile Location Data. *AEA Papers and Proceedings*, 108:64–67, May 2018.

Bell, David R. , Ho, Teck-Hua , and Tang, Christopher S. . Determining Where to Shop: Fixed and Variable Costs of Shopping. *Journal of Marketing Research*, 35(3):352–369, 1998. ISSN 00222437.

Berry, Steven , Levinsohn, James , and Pakes, Ariel . Automobile Prices in Market Equilibrium: Part I and II. *Econometrica*, 63(4):841–890, July 1995.

Berry, Steven , Levinsohn, James , and Pakes, Ariel . Differentiated Products Demand Systems from a Combination of Micro and Macro Data: The New Car Market. *Journal of Political Economy*, 112(1):68–105, 2004.

Bhattacharya, Vivek , Illanes, Gastón , and Stillerman, David . Merger Effects and Antitrust Enforcement: Evidence from US Consumer Packaged Goods. *Working Paper*, (31123), April 2023.

Briesch, Richard A. , Chintagunta, Pradeep K. , and Fox, Edward J. . How Does Assortment Affect Grocery Store Choice? *Journal of Marketing Research*, 46(2):176–189, 2009.

Brynjolfsson, Erik and Milgrom, Paul . Complementarity in Organizations. 2012.

Compiani, Giovanni and Smith, Adam N. . Boundaries of Differentiated Product Markets and Retailer Pricing. *Working Paper*, 2020.

DellaVigna, Stefano and Gentzkow, Matthew . Uniform Pricing in U.S. Retail Chains. *The Quarterly Journal of Economics*, 134:2011–2084, November 2019.

Donnelly, Robert. , Ruiz, Francisco J.R. , Blei, David , and Athey, Susan . Counterfactual inference for consumer choice across many product categories. *Quantitative Marketing and Economics*, 19:369–407, 2021.

Dube, Jean-Pierre H. , Hitsch, Guenter J. , and Rossi, Peter E. . Income and Wealth Effects on Private-Label Demand: Evidence from the Great Recession. 2016.

Dubé, Jean-Pierre . Multiple Discreteness and Product Differentiation: Demand for Carbonated Soft Drinks. (23), September 2004.

Dubé, Jean-Pierre H and Rossi, Peter . Microeconometric Models of Consumer Demand. *Elsevier*, 2019.

Ekerdt, Lorenz K. and Wu, Kai-Jie . The Rise of Specialized Firms. *Census Working Papers*, (CES-24-06), 2024.

Erdem, Tülin and Keane, Michael P. . Decision-Making under Uncertainty: Capturing Dynamic Brand Choice Processes in Turbulent Consumer Goods Markets. *Marketing Science*, 15(1):1–20, 1996.

Ershov, Daniel , P. Laliberté, Jean-William , Marcoux, Mathieu , and Orr, Scott . Estimating Complementarity With Large Choice Sets: An Application to Mergers. March 2021.

Ershov, Daniel , P. Laliberté, Jean-William , Marcoux, Mathieu , and Orr, Scott . Estimating Complementarity With Large Choice Sets: An Application to Mergers. September 2023.

FTC. *FTC Matter/File Number: 0710114*, Docket Number: 9324.

Gentzkow, Matthew . Valuing New Goods in a Model with Complementarities: Online Newspapers. (12562), October 2006.

Hausman, Jerry and Leibtag, Ephraim . Consumer Benefits from Increased Competition in Shopping Outlets: Measuring the Effect of Wal-Mart. *Journal of Applied Econometrics*, 22(7):1157–1177, 2007. ISSN 08837252, 10991255.

Hendel, Igal . Estimating Multiple-Discrete Choice Models: An Application to Computerizzation Returns. (168), October 1994.

Holmes, T. J. . The Diffusion of Walmart and Economies of Density,. *Econometrica*, 79(1): 253–302, 2011.

Iaria, Alessandro and Wang, Ao . An Empirical Model of Quantity Discounts with Large Choice Sets. *CEPR Discussion Paper*, No. DP16666, 2021a.

Iaria, Alessandro and Wang, Ao . A note on stochastic complementarity for the applied researcher. (199), February 2021b.

Jindal, Pranav , Zhu, Ting , Chintagunta, Pradeep K. , and Dhar, Sanjay K. . Marketing Mix Response Across Retail Formats - The Role of Shopping Trip Types. *Journal of Marketing*, 84(2):114–132, 2020.

Leung, J. and Li, Z. . Big-box store expansion and consumer welfare. *Working Paper*, 2022.

Manzini, Paola , Mariotti, Marco , and Ulku, Levent . Stochastic Complementarity. *The Economic Journal*, n/a(n/a), 2018.

Miyauchi, Y. , Nakajima, K. , and Redding, S. J. . Consumption access and the spatial concentration of economic activity: Evidence from smartphone data. *NBER Working Paper*, (w28497):253–302, 2021.

Nevo, Aviv . Measuring Market Power in the Ready-to-Eat Cereal Industry. *Econometrica*, 69(2):307–342, 2001.

Ramos-Menchelli, Daniel and Sverdlin-Lisker, Diana . Markets and the Proliferation of Small Firms: Evidence from Mom-and-Pop Shops in Mexico. *Working Paper*, 2022.

Rossi-Hansberg, E. , Sarte, P. D. , and Trachter, N. . Diverging trends in national and local concentration. *NBER Macroeconomics Annual*, 35(1):115–150, 2021a.

Rossi-Hansberg, Esteban , Sarte, Pierre-Daniel , and Trachter, Nicholas . Diverging Trends in National and Local Concentration. *NBER Macroeconomics Annual*, 35:115–150, 2021b.

Ruiz, Francisco J. R. , Athey, Susan , and Blei, David M. . SHOPPER: A probabilistic model of consumer choice with substitutes and complements. *The Annals of Applied Statistics*, 14(1):1 – 27, 2020.

Smith, Adam N. , Seiler, Stephan , and Aggarwal, Ishant . Optimal Price Targeting. *Marketing Science*, 42(3):476–499, 2023.

Smith, Howard . Supermarket Choice and Supermarket Competition in Market Equilibrium. *The Review of Economic Studies*, 71(1):235–263, January 2004.

Thomassen, O. , Smith, H. , Seiler, S. , and Schiraldi, P. . Multi-category competition and market power: a model of supermarket pricing. *American Economic Review*, 107(8): 2308–2351, 2017a.

Thomassen, Øyvind , Smith, Howard , Seiler, Stephan , and Schiraldi, Pasquale . Multi-category Competition and Market Power: A Model of Supermarket Pricing. *American Economic Review*, 107(8):2308–51, August 2017b.

Vroegrijk, M. , Gijsbrechts, E. , and Campo, K. . Close Encounter with the Hard Discounter: A Multiple-Store Shopping Perspective on the Impact of Local Hard-Discounter Entry. *Journal of Marketing Research*, 50(5):606–626, 2016.

Vroegrijk, Mark , Gijsbrechts, Els , and Campo, Katia . Close Encounter with the Hard Discounter: A Multiple-Store Shopping Perspective on the Impact of Local Hard-Discounter Entry. *Journal of Marketing Research*, 50(5):606–626, 2013. ISSN 00222437.

Wang, Ao . A BLP Demand Model of Product-Level Market Shares with Complementarity. *Working Paper*.

Wang, Ao . A BLP Demand Model of Product-Level Market Shares with Complementarity. January 2021.

1 Appendix: Alternative Demand Specification

We can rewrite our category-level demand as now also a function of both the store, s, and the choice made by the consumer of which bundle of stores (either 1 or 2) to shop at, r, indexed by the market of stores in each consumer's zipcode, m:

$$U_{ict}^{sr} = \vartheta_i \beta_c^{sr} + \rho_c^{sr} y_{it} + \psi_i (x')_c^{sr} - \phi_i \lambda_c^{sr} IV_{ict}^{sr} + \epsilon_{ict} \tag{41}$$

The product-category level utility is still the same, however, under the hood in the inclusive value term we have

$$IV_{ict}^{sr} = E[\max_{v=1 \ldots V_c^{sr}} u_{ivt}^{sr}] = \log \sum_{v=1}^{V_c} a_{vt} \exp u_{ivt}^{sr} \tag{42}$$

This blows up the dimensions on the category-level space which makes it much more tractable given that the indirect utility term can be estimated while scaling linearly in the number of product categories. Now we are treating each category as a separate product category given the store it was purchased in and the combination of the stores chosen by the consumer.

The difficulty with modeling this is there will have to be restrictions made on all the "unvisited" product categories so most product categories in this case would get zero purchases even though they were technically not even a part of the consumer's choice set to begin with. We are then only allowing consumers to choose within certain product categories mechanically

This estimation strategy will thus rely heavily on the outer nests picking up the store choice correctly in contrast with the current estimation strategy where most of the choice variation is picked up in the inner nests. Our current estimation strategy also yields a simpler construction of the counterfactuals of $\widehat{IV}_{mt}^{(h)} + \widehat{IV}_{mt}^{(h)}$ which is why it was ultimately chosen.

2 Appendix: Nested Factorization Details

In more detail, I present the mechanics of the demand estimation used which is the nested factorization technique outlined in Donnelly et al. [2021]. As this estimation strategy is mechanically the same for both the outer and inner nests, I detail exactly how each step of the estimation strategy works just on the inner nest. Recall the utility function for the conditional purchase probability of each product $v \in V_c$

$$u_{ivt} = \theta_i^T \beta_v + W_i^T \rho_v + \sigma_i^T X_v - \gamma_i^T \lambda_v price_{vt} + \epsilon_{ivt} \tag{43}$$

The latent parameters in this step of the estimation are $\updownarrow = \{\theta_i, \beta_v, \rho_v, \sigma_i, \gamma_i, \lambda_v\}$. The observables are y and X. y is the vector of all product purchase outcomes, 0 or 1 given whether we observe the purchase in every market at every month. X is the matrix of item observables in every market at every month. We want to use the observables to estimate the posterior joint distribution of the latent parameters.

$$p(l \mid y, X) = \frac{p(l)\Pi_t p(y_{ijt} \mid l, x_{ijt})}{p(y \mid X} \tag{44}$$

The posterior estimation is not computationally intractable so instead the authors employ variational inference. The family of distributions of l is chosen to be a mean-field family where all latent variables are jointly distributed independently normal. Note, this tracks with our original prior on all latent variables being distributed $\mathcal{N}(0, 1)$. This is specified as a flexible parameterized family of distributions $q(l; v)$. Variational parameters v are chosen to minimize the Kullback-Leibler divergence. Kullback-Leibler divergence is an approximation of a distance function. Minimizing the distance is equivalent to maximizing the following

$$\mathcal{L}(v) = E_{q(l;v)[\log(p(y|x,l) - \log q(l;v)]} \tag{45}$$

See Appendix 2 in Donnelly et al. [2021] for a longer exposition on how this is maximized in a computationally tractable way using gradient descent. Overall this shows that this setting, while having unusually many product categories for a multinomial discrete choice method to converge, is a nice application for a machine learning technique with gradient descent.

3 Appendix: Demand Estimates from the Inner and Outer Nests

Table 13: Demand Estimates Summary - inner nest

Purchased retailer	Pair of retailers	mean utility
6	6	0.118711
6	6 151	0.111944
6	6 61	0.111858
6	6 858	0.110627
6	6 916	0.110451
61	61	0.109835
61	61 151	0.105921
61	61 858	0.105091
61	61 916	0.071995
61	6 61	0.112447
151	151	0.072500
151	151 858	0.116531
151	151 916	0.074104
151	6 151	0.121675
858	151 858	0.104719
858	61 858	0.107634
858	6 858	0.109983
858	858	0.106084
858	858 916	0.084027
916	151 916	0.062587
916	61 916	0.072799
916	6 916	0.113233
916	858 916	0.090485
916	916	0.082697

Now to look at the demand estimates for the outer nest

Table 14: Demand Estimates Summary - outer nest

Purchased retailer	Pair of retailers	mean utility
6	6	0.129271
6	6 151	0.129312
6	6 61	0.129302
6	6 858	0.129235
6	6 916	0.129284
61	61	0.129291
61	61 151	0.129307
61	61 858	0.129301
61	61 916	0.129306
61	6 61	0.129315
151	151	0.129269
151	151 858	0.129284
151	151 916	0.129257
151	6 151	0.129322
858	151 858	0.129297
858	61 858	0.129306
858	6 858	0.129252
858	858	0.129259
858	858 916	0.129282
916	151 916	0.129352
916	61 916	0.129301
916	6 916	0.129256
916	858 916	0.129277
916	916	0.129311