

Goods Prices and Availability in Cities

Jessie Handbury

David E. Weinstein*

University of Pennsylvania

Columbia University and NBER

November 19, 2013

Abstract

This paper uses detailed barcode data on purchase transactions by households in 49 U.S. cities to overcome a large number of problems that have plagued spatial price index measurement. We identify two important sources of bias. Heterogeneity bias arises from comparing different goods in different locations, and variety bias arises from not correcting for the fact that some goods are unavailable in some locations. Eliminating heterogeneity bias causes 97 percent of the variance in the price level of food products across cities to disappear relative to a conventional index. Eliminating both biases reverses the common finding that prices tend to be higher in larger cities. Instead, we find that price level for food products falls with city size.

*We wish to thank Paul Carrillo, Donald Davis, Jonathan Dingel, Gilles Duranton, Zheli He, Joan Monras, Mine Senses, and Jonathan Vogel for excellent comments. Molly Schnell and Prottoy Aman Akbar provided us with outstanding research assistance. David Weinstein would like to thank the NSF (Award 1127493) for generous financial support. Jessie Handbury would like to thank the Research Sponsors' Program of the Zell-Lurie Real Estate Center at Wharton for financial support.

1 Introduction

The variation in prices and price indexes across locations is as central to economic geography and international economics as inflation is to macroeconomics. However, the methods used to construct prominent spatial price indexes are significantly cruder than those used to construct inflation rates and other inter-temporal price indexes. For example, while the U.S. Consumer Price Index (CPI) compares the relative prices over time of identical goods sold in the same store, regional price indexes compare different (but similar) goods purchased in different stores.¹ Moreover, the U.S. CPI accounts for product entry and exit. Evidence suggests that product availability varies across locations as well as over time, yet even the latest spatial price indexes do not account for these differences.²

This paper uses detailed barcode data documenting purchase transactions by households in 49 U.S. cities to overcome these obstacles in spatial price index measurement. In order to give some sense of the magnitude of the heterogeneity and variety biases in standard indexes we focus on two phenomena: the spatial variation in price indexes, which is itself the subject of the purchasing power parity (PPP) debate, and the correlation of price indexes with population, which yields a common agglomerating force across many New Economic Geography (NEG) models. Our use of better data enables us to replicate prior results from these areas and demonstrate a number of novel findings.

First, we precisely measure prices of *identical* goods sold in comparable stores across 49 U.S. cities to properly estimate spatial price differences. While standard price indexes show a positive correlation between average prices and city sizes, this correlation almost entirely disappears when we compare transaction prices of identical products purchased in the same stores. If we define purchasing power parity (PPP) deviations as differences in the average price of traded goods, we find that 97 percent of the variance in PPP deviations for groceries across U.S. cities can be attributed to heterogeneity biases in the construction of price indexes.

Second, while average product prices do not vary much across space, we find dramatic differences in product availability. The detail of our transaction-level data allows us to quantify these differences. We estimate the number of varieties of products available in each city and find that a doubling in city size is associated with a 20 percent increase in the number of available products.

¹The ACCRA (American Chamber of Commerce Researchers Association) index of U.S. urban prices, used in important papers such as Chevalier (1995), Parsley and Wei (1996), Albouy (2009), and Moretti (2013), is an example of such an index.

²The Bureau of Economic Analysis (BEA) recently released regional price (RPP) indexes for the U.S. The RPP methodology, outlined in Aten (2005) and Aten and Martin (2012), makes some headway towards adjusting for product and store heterogeneity. Product heterogeneity has also been partially addressed in the latest Penn World Table (PWT), which compares quality-adjusted prices across countries (Feenstra et al., 2012). Data limitations mean that neither the PWT or the BEA's RPP indexes compare identical goods in different markets (which is critical for the approach used in this paper), nor do they adjust for variety differences.

Finally, we use data on the purchase quantities, as well as transaction prices, to demonstrate that the differences in variety availability yield economically significant variation in the price level across cities.³ When we use the data to construct a theoretically-rigorous price index that corrects for product, purchaser, and retailer heterogeneity *and* accounts for variety differences across locations, we find that the price level is actually lower in larger cities. Consumers spend less, on average, to get the same amount of consumption utility in larger cities.

The association between city population and price levels plays an important role in many urban and NEG models. NEG models typically predict that price indexes over tradable goods are lower in larger cities (see, *e.g.*, Fujita (1988); Rivera-Batiz (1988); Krugman (1991); Helpman (1998); Ottaviano et al. (2002); Behrens and Robert-Nicoud (2011)). This prediction is at odds with empirical work demonstrating that prices are higher in larger cities (DuMond et al., 1999; Tabuchi, 2001). One reason that these studies have not been deemed fatal for the theory is that it is easy to modify NEG models to generate higher housing prices in cities (see, *e.g.*, Helpman (1998)). Our paper suggests data problems in the construction of urban price indexes are sufficiently large to explain the seemingly contradictory evidence: variety- and heterogeneity-adjusted price indexes are lower in larger cities. If we also extract non-traded land price components from goods prices to construct a producer price index, we find that the price of purely traded goods falls even more sharply. This result is consistent with Helpman (1998) and Suedekum (2006) whose work suggests that, while the price of purely traded goods should be lower in cities, the inclusion of non-traded goods prices in the index can produce an inconclusive result.

A key difference between this paper and earlier work is that we work with barcode data, so the prices we compare are for *identical* goods. Our dataset includes the prices for hundreds of thousands of goods purchased by 33,000 households in 49 cities in the U.S. Critically, the data lets us know the price of each good, where it was purchased, and information about the purchaser. Consistent with earlier analyses, if we aggregate our data and compare the prices of *categories of goods*, we find that the elasticity of the grocery price level with respect to population is 0.042. This implies that a New Yorkers pay 16 percent more for similar, but not necessarily identical, groceries than people in Des Moines (population 456,000). However, when we adjust this index, step-by-step, for the various biases we identify in the standard methodology, we end up with our final estimate for the correct elasticity: -0.011. In other words, when they are estimated properly, grocery price indexes do not rise, but rather fall, with population.

One of the most important classes of bias are “heterogeneity biases,” which arise from not being careful about which prices are being compared. For example, the price of an item like a “half-gallon of whole milk” can vary enormously depending on a number of sources of un-

³We use the word “price” to refer to the price of a particular good and the term “price level” to refer to a price index, or some weighted average of relative prices across goods.

derlying heterogeneity. “Product heterogeneity biases” arise because there are many varieties of whole milk that differ enormously in price, *e.g.*, name brand vs. store brand, organic vs. non-organic, etc.⁴ “Retailer heterogeneity biases” arise because high-amenity stores may systematically charge different prices to low-amenity stores for the same good. Finally, “purchaser heterogeneity biases” arise because shoppers who search intensely for the lowest price can often purchase the same good in the same store for less. Regional price indexes typically do not correct for these biases because without barcode data it is difficult to find the same good in the same store chain in two different locations.⁵ To get some sense of the magnitude of these biases among goods that are available in more than one location, we regress dis-aggregate log prices against log population with product, purchaser characteristic, and store controls. We find that controlling for these heterogeneity biases reduces the elasticity of price with respect to population from 0.042 to 0.006 (86 percent). This indicates that the large positive elasticity in the aggregate data is due to the fact that consumers in large cities tend to purchase higher quality varieties in nicer stores and shop less intensely (presumably because rich people have a higher opportunity cost of time). Although statistically different from zero, the elasticity that remains after controlling for heterogeneity is not economically meaningful; it implies prices of commonly-available goods are approximately equal in large and small cities. Indeed, between 95-97 percent of the variance in PPP deviations across cities disappears once we correct for these biases.

A second major source of bias is variety bias. Variety biases arise because consumers do not have access to the same set of products in all locations. These biases have been studied in the context of the CPI by Broda and Weinstein (2010), but there is reason to believe that they are much more important in the regional context. The difference in product availability between New York and Des Moines, for example, is likely to be much greater than the difference in product availability in the U.S. economy from one year to the next. In order to quantify this effect, we adapt some well-developed statistical procedures to the problem of estimating the number of varieties in cities. Our results indicate very large differences in variety availability. We estimate that there are approximately four times more types of grocery products available in New York than in Des Moines.

In order to quantify the variety bias we need to put more structure on the problem. We use a spatial variant of the Constant Elasticity of Substitution (CES) exact price index developed in the seminal work of Feenstra (1994). The CES structure is commonly employed in NEG models

⁴Just to give one simple case of this, in Westside Market in New York on August 18, 2013, a half gallon of Farmland whole milk sold for \$2.47 while a half gallon of Sky Top Farms whole milk sold for \$6.59.

⁵The food component of the BEA RPP is based on BLS data. The BLS is careful to keep products and stores constant over time, but uses random sampling to select the stores and products for which prices are collected in each location. ACCRA provides field agents with detailed instructions to collect prices for products and in stores meeting certain specifications. These instructions leave a large scope for product and store heterogeneity in prices.

and is well suited for our data.⁶ When calculated over varieties available in more than one city and using prices adjusted for product, purchaser, and store heterogeneity, the theoretically-rigorous CES index yields almost the same elasticity of price with respect to population as the price regression above. An advantage of the CES framework is that it enables us to make an additional adjustment for the fact that small cities offer consumers substantially fewer purchase options. Given the important difference in product availability across locations, we find that variety bias is extremely important economically. Correcting for the variety bias further lowers the elasticity to -0.011. In other words, when we correct for heterogeneity and variety biases, the standard result that prices rise with city size is reversed.

This paper complements large literatures studying international price and variety differences. Simonovska (2010) and Landry (2013), for example, use micro price data to document international price differences of identical products. Barcode price data has also been used extensively in the study of PPP convergence (see a recent survey by Burstein and Gopinath (2013)) and PPP convergence (see, *e.g.*, Broda et al. (2008); Burstein and Jaimovich (2009); Gopinath et al. (2011)). Hummels and Klenow (2005) document that larger countries both export more varieties of products; while Bernard et al. (2007) and Eaton et al. (2011) document that larger countries import more varieties of products.

There is less work on intranational price and variety differences. Parsley and Wei (1996) use the ACCRA data to examine convergence to the law of one price in the U.S. Crucini and Shintani (2008) use similar data from the Economist Intelligence Unit, to examine the persistence of law of one price deviations for nine U.S. cities. This work on deviations from the law of one price does not address the question of how much of the difference in observed prices across cities reflects unobserved heterogeneity in products or retailers. The only other paper, to our knowledge, to compare prices of identical goods within countries is Atkin and Donaldson (2012), who use spatial price differences as a proxy for intranational trade costs in developing countries.

A nascent literature has documented that larger and more dense areas in the U.S. have more varieties of restaurants (Schiff, 2012). Unfortunately, the lack of price data and the inability to control for quality differences across restaurants in different locations make it difficult to accurately measure the welfare implications of these variety differences. Recent work by Couture (2013) uses household travel patterns to estimate the substitution between restaurants but, without an additional price or quality measure, he cannot separately identify price from quality so must assume that these two factors are perfectly correlated.

In complementary work, Handbury (2012) uses the same data as the current paper to cal-

⁶Recent NEG models have also used the quadratic linear framework developed by Ottaviano et al. (2002). While quadratic linear framework is tractable for theoretical analysis, it is difficult to estimate and, therefore, not well-suited for price index measurement.

culate variety-adjusted city-specific price indexes for households at different income levels and finds that high-income households face relatively lower price indexes in cities with higher per capita incomes. Consistent with the PPP variance results here, Handbury (2012) finds that these intra-income differences are driven entirely by variety differences across cities. Both papers point towards the relevance of the extensive variety margin in explaining PPP deviations across cities.

The rest of the paper is structured as follows. Sections 3 and 4 explore how identical goods prices and goods availability vary across cities. In Section 5 we summarize these results using an urban price index that adjusts for the heterogeneity and variety biases in standard indexes. Finally, in Section 6, we control for retail rents in order to estimate how producer, or wholesale, prices vary across cities. Section 7 concludes.

2 Data

The primary dataset that we use is taken from the Nielsen HomeScan database. These data were collected by Nielsen from a demographically representative sample of approximately 33,000 households in 52 markets across the U.S. in 2005.⁷ Households were provided with Universal Product Code (UPC) scanners to scan in every purchase they made including online purchases and regardless of whether purchases were made in a store with scanner technology.⁸ Each observation in our data represents the purchase of an individual UPC (or barcode) in a particular store by a particular consumer on a particular day. We have the purchase records for grocery items, with information on the purchase quantity, pre-tax price, and date; the name or type of the store where the purchase is made; and demographic information on the household making the purchase.⁹

Figure 1 presents the basic structure of our data. A barcode, u , uniquely identifies a product. For example, “Horizon 1% Milk in a Half-Gallon Container” has a different barcode than “Horizon 2% Milk in a Half-Gallon Container.” Nielsen provides product characteristics for each barcode, including its brand, a detailed description of the type of good that Nielsen refers to as a “module,” and a more aggregate description of a good that Nielsen refers to as a product “group.” For example, “Horizon 1% Milk in a Half-Gallon Container” is sold under the “Hori-

⁷The Nielsen sample is demographically representative within each market.

⁸In cases where panelists shop at stores without scanner technology, they report the price paid manually. Since errors can be made in this reporting process, we discard any purchase records for which the price paid was greater than twice or less than half the median price paid for the same UPC, approximately 250,000 out of 16 million observations.

⁹Nielsen provides a store code for each transaction in the data. For all but 800,000 of 16 million transactions, the store code identifies a unique store name. For the remaining observations representing 4.4 percent of sales in the data, Nielsen’s store code refers to one of approximately 60 store categories, such as “Fish Market,” “Cheese Store,” “Drug Store,” etc.

zon” brand in the “Milk” module that fits within the “Dairy” product group. We group barcodes with the same brand and in the same module into “brand-modules.” For example, “Horizon Milk,” “Horizon Butter,” and “Breakstone Butter,” constitute three different brand-modules in the “Dairy” product group, the first of which is in the “Milk” module and the latter two are in the “Butter” module. The 2005 HomeScan sample we consider contains transaction records for almost 350,000 UPCs that are categorized into 597 modules, 27,853 brands, and 55,559 brand-module interactions and 63 product groups.¹⁰ Detailed descriptions of the Nielsen data and the sampling methods used can be found in Broda and Weinstein (2010).

Figure 1: Terminology

Universal Product Code (UPC) or Barcode $(u \in U_g)$	\subset	Brand-Module $(b \in B_g)$	\subset	Product Group $(g \in G)$
<i>e.g.</i> Horizon 1% Milk in a Half-Gallon Container		<i>e.g.</i> Horizon Milk		<i>e.g.</i> Dairy
$N=348,646$		$N = 55,559$		$N = 63$

Although the Nielsen dataset contains data for 52 markets, we classify cities at the level of Consolidated Metropolitan Statistical Area (CMSA) where available and the Metropolitan Statistical Area (MSA) otherwise. For example, where Nielsen classifies urban, suburban, and ex-urban New York separately, we group them all as New York-Northern New Jersey-Long Island CMSA. We use population, income distribution, and racial and birthplace diversity data from the 2000 U.S. Census and 2005 retail rents from REIS.^{11,12} The population and retail rents for the markets included in the analysis are listed in Table A.1, along with market IDs we will use to identify markets in the charts below. There are two cases in which Nielsen groups two MSAs into one market. In these cases, we count the two MSAs as one city, using the sum of the population and the population-weighted mean retail rents.

¹⁰This sample excludes the “random weight” product group. The quality of random weight items, such as fruit, vegetables, and deli meats, varies over time as the produce loses its freshness. We cannot control for this unobserved quality heterogeneity.

¹¹Specifically, we use the combined effective rents for community and neighborhood shopping centers. Effective rents adjust for lease concessions.

¹²We replicated the analysis below using total manufacturing output and food manufacturing output as alternative measures of city size and reached the same qualitative conclusions. This is not surprising, as the data for total manufacturing output and food manufacturing output from the 2007 U.S. Economic Census were highly correlated with population across the markets in our sample, with coefficients of 0.70 and 0.73, respectively.

3 Measuring Retail Prices in Cities

While our ultimate goal is to construct the first theoretically-founded urban price index, we begin by exploring the data. Variation in the price index across cities is driven by differences in the prices of identical goods and the variety availability. Our reduced-form analysis explores each of these factors. In this section, we focus only on the price data and then address goods availability and the construction of an urban price index in Sections 4 and 5, respectively.

3.1 Evidence From Categories of Goods

A common method to compare price levels across cities within countries relies on unit value indexes such as those published by the Council for Community and Economic Research (formerly the American Chamber of Commerce Research Association (ACCRA)). ACCRA collects prices in different cities across the U.S. for a “purposive” (*i.e.*, non-random) sample of items that selected to represent categories of goods. For each item, ACCRA’s price collectors are instructed to record the price of a product that meets certain narrow specifications, *e.g.*, “half-gallon whole milk,” “13-ounce can of Maxwell House, Hills Brothers, or Folgers ground coffee,” “64-ounce Tropicana or Florida Natural brand fresh orange juice,” etc. ACCRA takes the ratio of the average price collected for each item in each city and quarter relative to its national average in that quarter. The ACCRA COLI is a weighted average of these ratios, where item weights are based on data from the U.S. Bureau of Labor Statistics 2004 Consumer Expenditure Survey.¹³

While we will deal with a host of problems arising from comparing prices of similar, as opposed to identical, products later, our first task is to replicate the standard result that if one uses the standard ACCRA methodology, the price index for tradable goods rises with population in our data. In order to establish this stylized fact, we obtained the ACCRA COLI data for 2005 and measured the association between log population and four different indexes: ACCRA’s aggregate, or composite, cost-of-living index; their grocery index; and two food price indexes that we built using the ACCRA item-level price ratios and weights. We refer to these two constructed indexes as the ACCRA food index and Nielsen food index. The ACCRA food index replicates the ACCRA methodology, but only over food items. The Nielsen food index replicates the ACCRA food index using Nielsen price data. To build this index, we first identified the set of UPCs in the Nielsen data whose characteristics match the ACCRA specifications for each food item represented in the ACCRA index. We then calculated the average price observed in the Nielsen data for the set of UPCs matching each item in each city and the ratio of each of these city-specific item unit values to their national average. The Nielsen food index for each city is the weighted average of these Nielsen unit value ratios across items using ACCRA

¹³See <http://www.coli.org/Method.asp> for more details.

item weights.

Table 1: Category Price Indexes vs. Population

	ACCRA Price Indexes			Ln(Nielsen
	Ln(Composite Index _c)	Ln(Grocery Index _c)	Ln(Food Index _c)	Food Index _c)
	[1]	[2]	[3]	[4]
Ln(Population _c)	0.132*** [0.0209]	0.0727*** [0.0167]	0.0718*** [0.0168]	0.0423*** [0.00939]
Constant	2.706*** [0.306]	3.562*** [0.245]	3.541*** [0.245]	3.982*** [0.138]
Observations	47	47	47	47
R-Squared	0.47	0.30	0.29	0.31

Standard errors in brackets; *** p<0.01, ** p<0.05, * p<0.1.

We regressed the log of each price index for each city on the log of the city’s population and report the results in Table 1. As one can see from the table, there is a very strong positive association between each of these price indexes and population. Although the composite ACCRA index, which includes land prices, rises the steepest with population, we see a very similar pattern for the grocery and food price indexes. A one log-unit rise in city size is associated with a four percent increase in the food price index when we build it using Nielsen data and seven percent when we rely on the ACCRA data. While the magnitude of the slope coefficient varies across the indexes, none of these differences are statistically significant.¹⁴ In particular, all four indexes, using both the ACCRA and the Nielsen data, imply that there is strong positive association between city size and the price level. These coefficients are economically significant as well. Even the smallest coefficient, from the Nielsen price index regression, is economically significant, indicating that a consumer in New York pays 16 percent more for food items than a person in Des Moines.¹⁵

3.2 Controlling for Product, Buyer, and Retailer Heterogeneity

There are three types of heterogeneity biases that may generate the positive correlations observed above: product heterogeneity bias, retailer heterogeneity bias, and purchaser heterogeneity bias. If consumers in larger cities systematically purchase higher quality (*i.e.*, more expensive) varieties within a product category, then a higher average price level in a city might

¹⁴This partially reflects the fact that the correlation coefficient between the various price indexes ranges from 0.8 to 0.9.

¹⁵Other indexes show a similar pattern. When we regress the BEA RPP index for goods against log population using our sample of cities, the coefficient on log population is 0.026 and statistically significant at the 1 percent level. There are many potential reasons for the difference between this elasticity and those reported in Table 1, including that the RPP covers a different, broader set of goods than we have in the Nielsen data and partially adjusts for product and store heterogeneity.

just reflect the fact that consumers in that city buy more expensive varieties of that product category. Similarly, retailer heterogeneity bias can arise because consumer in large cities might purchase goods in stores that offer systematically higher amenities. For example, some grocery stores, like Whole Foods, offer nicer shopping experiences than mass-merchandisers. Finally, if there are a higher fraction of wealthy people in large cities, and rich people look for bargains less than poor people, purchaser heterogeneity might mean that purchase prices may reflect different shopping intensities of consumers.

As we mentioned earlier, our objective is obtain a standardized price measure that reflects the prices of identical goods purchased in different locations but at similar stores and by consumers with similar shopping intensities. Essentially, we are trying to do the spatial equivalent to the time-series methodology employed in the construction of the U.S. Consumer Price Index, which measures price changes for identical products, purchased in the same store, by field agents with common shopping instructions.

Our methodology for doing this is quite straightforward. Let $P_{ucr h}$ be the average price that a household h paid for UPC u in store r in city c .¹⁶ We refer $P_{ucr h}$ as the “unadjusted price” and define $p_{ucr h}$ as $\ln(P_{ucr h})$. We can then construct an adjusted price index by running the following regression:

$$p_{ucr h} = \alpha_u + \alpha_c + \alpha_r + \mathbf{Z}_h \boldsymbol{\beta} + \varepsilon_{ucr h} \quad (1)$$

where α_u , α_c , and α_r are UPC, city, and store fixed effects, respectively, and \mathbf{Z}_h denotes a vector of household characteristics and $\boldsymbol{\beta}$ are the corresponding coefficients. Household demographic dummies are included for household size, as well as the gender, age, marital status, and race of the head of household; in addition, we control for household income, which is correlated with shopping intensity. Our store fixed effects take a different value for each of the approximately 600 retail chains in our sample that serve at least 2 cities. For stores that we observe serving a single city, we restrict α_r to be the same for all stores of the same type, where type is defined in one of seven “channel-IDs”: grocery, drug, mass merchandiser, super-center, club, convenience, and other.¹⁷ The α_r are designed to capture store amenities, and the $\mathbf{Z}_h \boldsymbol{\beta}$ capture factors related to purchaser heterogeneity.

The city fixed effects, α_c , can be thought of as city price indexes that control for the types of products purchased, the store in which the purchase occurred, and the shopping intensity

¹⁶HomeScan panelists record purchases for each transaction they make while participating the survey and data records are identified using a calendar date. We aggregate the data to the annual frequency, summing purchase values and quantities across transactions in the 2005 sample. The average price paid is, therefore, the sum of the dollar amounts that a household h paid for UPC u in store r over all of the transactions where we observe the household purchasing that UPC in that store, divided by the sum of the number of units that the household purchased across the same set of transactions. We identify the “store” r that a transaction occurs in using Nielsen’s store code variable.

¹⁷We apply the same restriction to stores whose codes refer to store categories (such as “Fish Market,” “Cheese Store,” etc.) rather than store names.

of the buyer. We then can test whether standardized urban prices co-vary with population by regressing the city fixed effects on log population, *i.e.*,

$$\hat{\alpha}_c = \alpha + \gamma \ln(\text{Pop}_c) + \varepsilon_c, \quad (2)$$

where $\ln(\text{Pop}_c)$ is the log of population in city c . In this specification, γ tells us how prices vary with population after we control for the different bundles of products purchased in different cities. An advantage of this two-stage approach as opposed to simply including co-variates of interest in equation (1) is that our city price level estimates are not affected by what we think co-varies with urban prices. Thus, we separate the question of whether urban prices rise with population from the question of how to correctly measure urban prices. We will use this feature of the methodology in Section 5. However, as a robustness check, we will also show in Section 3.4.1 that our two-stage approach is not qualitatively important for our results.

3.3 Evidence from Barcode Prices

Recall that in Section 3.1 above, we showed that products from the same category were purchased for higher unit values in larger cities. The results in Table 1 indicated that a one log unit rise in city size is associated with a four percent rise in the unit value of groceries. We will now demonstrate that almost all of this effect can be explained by product, retailer, and purchaser heterogeneity biases. In other words, past studies have found that there are higher traded goods prices in larger cities because big cities have different (less price sensitive) consumers purchasing different (more expensive) varieties of products in different (more expensive) stores.

Table 2 presents results from estimating equations (1) and (2). The first key difference from Table 1 is that we are now gauging price differences between identical products, or UPCs, sold in different cities.¹⁸ In the first column of the table, we present the results from a specification that only adjusts for product heterogeneity. In other words, instead of running the regression specified in equation (1), we compute the city price index by only regressing prices on UPC dummies and city dummies. This method for computing the price index corrects for product heterogeneity, which are contained in the UPC fixed effects, but does not adjust for purchaser and retailer heterogeneity. In the second panel, we report the results from regressing the estimated city dummy coefficients on log population. We obtain a coefficient of 0.0139, which is only one third as large as the coefficient we obtained in Table 1 when we used the ACCRA methodology to generate a price index and regressed that on population. This result indicates that two-thirds of the positive relationship between prices and city size in the unit value index reflects the fact that people in larger cities purchase far more high-priced varieties of goods than

¹⁸In all regressions, we weight the data by the transaction value which gives more weight to goods that constitute higher expenditure shares.

residents of small cities.¹⁹

Table 2: Identical Product Price Indexes vs. Population

Panel A				
	P_{ucr}^1			
	[1]	[2]	[3]	[4]
Ln(Income _{<i>h</i>})	-	0.0114***	-	0.00805***
	-	[0.000961]	-	[0.000525]
UPC Fixed Effects	Yes	Yes	Yes	Yes
City Fixed Effects	Yes	Yes	Yes	Yes
Household Demographic Dummies ²	No	Yes	No	Yes
Store Dummies ³	No	No	Yes	Yes
Observations	15,570,529	15,570,529	15,570,529	15,570,529
Number of UPCs ⁴	348,645	348,645	348,645	348,645
R-Squared	0.948	0.948	0.953	0.953
Panel B				
	City Fixed Effect Coefficient from Panel A			
	[1]	[2]	[3]	[4]
Ln(Population _{<i>c</i>})	0.0139***	0.0130***	0.00603***	0.00568**
	[0.00400]	[0.00396]	[0.00215]	[0.00214]
Constant	-0.245***	-0.229***	-0.117***	-0.110***
	[0.0586]	[0.0581]	[0.0315]	[0.0314]
Observations	49	49	49	49
R-Squared	0.916	0.205	0.187	0.143

Standard errors in brackets; *** p<0.01, ** p<0.05, * p<0.1. Panel A standard errors are clustered by city.

Notes:

1. $p_{ucr} = \ln(P_{ucr})$ where P_{ucr} is the total expenditures by household h on UPC u in store r in city c in 2005 divided by the total quantity of UPC u purchased by household h in store r in city c during 2005. Observations in the Panel A regression are weighted by the total expenditure of household h on UPC u in store r .
2. Household demographic dummies are for household size, male and female head of household age, marital status, race, and hispanic.
3. Regressions with store dummies include one of seven channel-ID dummies is Nielsen does not provide a store name or the store identified only has sales in one city and a store name dummy otherwise.
4. Random weight UPCs have been dropped from the sample.

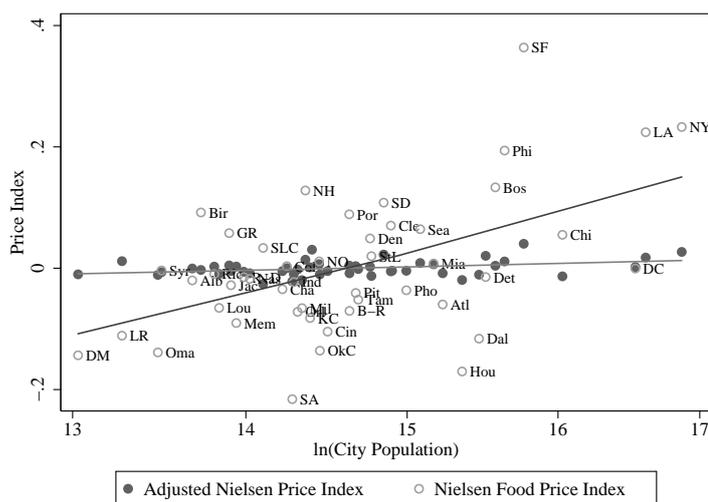
In Column 2 of Table 2, we adjust the urban price index for both product heterogeneity and purchaser heterogeneity. The positive coefficient on household income indicates that high income households systematically pay more for the same goods as poorer households. Some of this may be due to the fact that high-income households have a higher opportunity cost of time and therefore shop less intensively and also have a greater willingness to shop in high amenity stores. Alternatively, some of this positive association may be due to the fact that stores that

¹⁹One possible concern with these results is that shifts in the weighting of the data or some other factor associated with the shift from the price index methodology to the regression methodology is responsible for the drop. We investigate this possibility as a robustness check and show in Appendix Section C that this concern is not warranted.

cater to richer clientele are able to charge higher markups. While we will disentangle these two forces later, for the time being we simply note that controlling for purchaser heterogeneity causes the coefficient on log population to fall by another ten percent.

Interestingly, controlling for store fixed effects in Column 3 has a much more substantial impact on the elasticity of urban prices with respect to population than controlling for purchaser heterogeneity: more than halving the coefficient. The large impact of controlling for store heterogeneity implies that a second important reason why prices appear higher in larger cities is that residents of large cities disproportionately shop in stores that charge high prices *in all cities*. The most obvious source of this sort of heterogeneity is differences in amenities—rich households tend to purchase nicer varieties of goods and shop in nicer stores—but we will also examine the possibility that markup variations are explaining this in Section 3.4.2. Finally, if we control for both product, purchaser, and retailer heterogeneity in Column 4 the coefficient collapses to only 13 percent of its magnitude in Table 1. Most of this fall arises from adjusting for retailer heterogeneity, which reflects that shoppers in larger cities purchase more items in high-amenity stores. The coefficient on household income remains positive and significant, which means that richer households pay more for the same UPC *even in the same store*. We interpret this as evidence for the impact of purchaser heterogeneity on prices. Interestingly, the magnitude of the coefficient on income in Column 4 is about 70 percent as large as in Column 2, indicating that most of the reason why richer households pay more for the same UPC is due to their lower shopping intensity within stores and not to their choosing to shop in nicer stores.

Figure 2: Estimated Price Levels vs. Log City Population



Notes:

1. The market labels on the ACCRA price indexes reference the city represented, as listed in Table A.1.
2. City price indexes are normalized to be mean zero.

Figure 2 presents plots of price indexes computed using the ACCRA methodology in presented in the final column of Table 1 and the price indexes generated in Column 4 of Table 2. The hollow circles indicate the price indexes computed using the ACCRA methodology and the solid circles indicate those computed after correcting for the various forms of heterogeneity in the data. As one can see from the plot, there is a dramatic collapse in the relationship between urban prices and population once one controls for product, purchaser, and retailer heterogeneity. Indeed, the slight positive association that we identified in Table 2 is almost imperceptible, indicating that its economic significance is minor. Moreover, most of the dispersion in city price indexes disappears once we control for heterogeneity yielding very small deviations from the fitted line. After adjusting for the various forms of heterogeneity bias, the variance in urban price levels falls by 95 percent. In other words, purchasing price parity for tradables hardly varies across US cities.

3.4 Robustness

In this section we consider two robustness checks. First, we demonstrate that our results are similar when we estimate the association between city population and prices in a one-step estimation procedure. Second, we demonstrate that our results are not due to variable markups.

3.4.1 One-Step Estimation Procedure

Thus far we have been working with a two-step procedure for estimating the association between city population and prices. In Table 3 we repeat our exercise using a one-step procedure in which we replace the city fixed effects in equation (1) with the log of population in each city. The results are extremely similar to our earlier ones. The coefficients on log population are, in particular, very close to their counterparts in Panel B of Table 2, which tells us that our choice of a two-step procedure does not qualitatively affect the results.

3.4.2 Amenities vs. Mark-ups

One of the important adjustments that we make is for store amenities. Our methodology assumes that if consumers in a given city pay more for identical products when they buy them at one type of store relative to other stores within the same city, the higher price must reflect a difference in store amenities. An alternative explanation is that the higher price reflects a higher markup. If stores that are prevalent in larger cities charge higher markups, our results might be due to the fact that our method of eliminating retailer heterogeneity would be eliminating markup variation across cities and therefore might be understating the high prices in large cities. In other words, if the store effects capture amenities, consumers do not necessarily find

Table 3: Are prices higher in larger cities?

	P_{ucr}^1			
	[1]	[2]	[3]	[4]
Ln(Population _c)	0.0155*** [0.00267]	0.0150*** [0.00363]	0.00650*** [0.00223]	0.00604*** [0.00219]
Ln(Income _h)	- -	0.00704*** [0.00149]	- -	0.00554*** [0.000727]
UPC Fixed Effects	Yes	Yes	Yes	Yes
Household Demographic Dummies ²	No	Yes	No	Yes
Store Dummies ³	No	No	Yes	Yes
Observations	15,570,529	15,570,529	15,570,529	15,570,529
Number of UPCs ⁴	348,645	348,645	348,645	348,645
Number of Cities	49	49	49	49
R-squared	0.947	0.947	0.953	0.953

Robust standard errors in brackets; *** p<0.01, ** p<0.05, * p<0.1. Standard errors are clustered by city and storename/type in specifications where store dummies are included and city otherwise.

Notes:

1. $p_{ucr} = \ln(P_{ucr})$ where P_{ucr} is the total expenditures by household h on UPC u in store r in city c in 2005 divided by the total quantity of UPC u purchased by household h in store r in city c during 2005. Observations are weighted by the total expenditure of household h on UPC u in store r .
2. Household demographic dummies are for household size, male and female head of household age, marital status, race, and hispanic.
3. Regressions with store dummies include one of seven channel-ID dummies is Nielsen does not provide a store name or the store identified only has sales in one city and a store name dummy otherwise.
4. Random weight UPCs have been dropped from the sample.

big cities to be more expensive because they are getting a higher-quality shopping experience in return for paying a higher price. If the store effects instead reflect markup differences due to differences in market power across stores offering the same shopping experience, then consumers are not getting anything in return for the relatively high prices charged by stores in large cities and will, therefore, perceive these stores as more expensive.

Although we cannot measure markups directly, we can look at store market share information in an attempt to assess how markups might vary in our data, first across cities, and then across retailers and, in particular, across retailers that locate disproportionately in large, relative to small, cities. In many variable markup demand systems involving strategic substitutes, markups positively covary with market shares. For example, Feenstra and Weinstein (2010) show that for the translog system, markups will positively covary with the Herfindahl index in the market. We can compute retailer Herfindahl indexes for each city by aggregating the purchases of consumers in each store. Not surprisingly, Herfindahl indexes are negatively correlated with city size ($\rho = -0.3$) reflecting the fact that consumers in large cities not only have more choices of products, but also more choices of where to purchase those products. This circumstantial evidence suggests that, if anything, we are understating the amenity effect because

stores in large cities are likely to face more competition and charge lower markups (which is also consistent with models like Melitz and Ottaviano (2008)).

We also can try to strip out the market power effect from our estimates more directly. In order to do this we control for differences in markups across retail chains, or types, by including store market shares in the regression where we estimate the store effects (α_r).²⁰ Specifically, we add the market share of each store in each city to equation (1) and estimate

$$p_{ucr h} = \alpha_u + \alpha_c + \alpha_r + \mathbf{Z}_h \boldsymbol{\beta} + \gamma \text{Share}_{rc} + \varepsilon_{ucr h}, \quad (3)$$

where Share_{rc} is store r 's market share in city c and γ is a parameter to be estimated. We interpret $\hat{\alpha}_r$ in this specification as the component of the store's idiosyncratic price that cannot be explained by its market power. We then subtract these $\hat{\alpha}_r$ estimates from observed prices, adjusting prices for the component that is potentially related to differences in amenities, but not the component related to differences in markups via market power. We regress these adjusted prices against city fixed effects to estimate urban price indexes that control for differences in amenities across stores, but still allow for price differences resulting from differences in market power:

$$p_{ucr h} - \hat{\alpha}_r = \alpha_u + \tilde{\alpha}_c + \mathbf{Z}_h \boldsymbol{\beta} + \tilde{\varepsilon}_{ucr h}, \quad (4)$$

The dependent variable is the store amenity-adjusted price and $\tilde{\alpha}_c$ is an urban price index that reflects systematic differences in prices across stores with different market shares, but not those related to unobserved heterogeneity between stores.

Table 4 presents the results of this exercise. The estimates for equation (3), in Column 1, suggest no significant relationship between a store's market share in a market and the price it charges.²¹ The coefficient on market share is positive but not statistically significant, indicating that the capacity of a retailer to exercise market power is quite limited in most cities.²²

Column 2 presents the results of estimating the price indexes according to equation (4), and Column 3 presents the results of regressing the resulting adjusted city fixed effects, $\tilde{\alpha}_c$, against population. Not surprisingly, given the lack of a significant effect in Column 1, we do not find that adjusting for market share qualitatively affects our results. The city price indexes that have

²⁰Recall that r denotes the store code for each transaction. Most store codes uniquely identify retail chains or standalone stores; others refer to one of 60 store categories. If a store only has sales in one city or we do not have the store name, we restrict α_r to be equal across stores with the same "channel-ID," which can take one of seven values: grocery, drug, mass merchandiser, super-center, club, convenience, and other. We do not group stores in this manner when calculating market shares: Share_{rc} represents the sales share of store code r in city c .

²¹We also tried non-linear specifications linking store shares with market power by including quadratic and cubic terms without finding a significant link.

²²This result is consistent with the previous literature on retailer market power in the U.S. grocery sector which finds that stores do not exploit market power in their pricing decisions. For example, Ellickson and Misra (2008) demonstrate that "stores in a particular market do not use pricing strategy as a differentiation device but instead coordinate their actions" and find that "firm size is not the primary determinant of pricing strategy."

Table 4: Are prices higher in larger cities, controlling for market power?

	$p_{ucr h}^1$	$p_{ucr h} - \hat{\alpha}_r$	City Fixed Effect from Column 2 ($\hat{\alpha}_c$)
	[1]	[2]	[3]
Ln(Income _{<i>h</i>})	0.00876*** [0.000572]	0.00877*** [0.000560]	- -
Store Share _{<i>rc</i>}	0.00128 [0.00977]	- -	- -
Ln(Population _{<i>c</i>})	- -	- -	0.00604** [0.00238]
Constant	0.549*** [0.0151]	0.550*** [0.0113]	-0.112*** [0.0348]
UPC Fixed Effects	Yes	Yes	N/A
City Fixed Effects	Yes	Yes	N/A
Household Demographic Dummies ²	Yes	Yes	N/A
Store Dummies ³	Yes	No	N/A
Observations	15,570,529	15,570,529	49
Number of UPCs ⁴	348,645	348,645	N/A
Number of Cities	49	49	N/A
R-Squared	0.934	0.933	0.121

Standard errors in brackets; *** p<0.01, ** p<0.05, * p<0.1. Column 1 standard errors are clustered by city.

Notes:

1. $p_{ucr h} = \ln(P_{ucr h})$ where $P_{ucr h}$ is the total expenditures by household h on UPC u in store r in city c in 2005 divided by the total quantity of UPC u purchased by household h in store r in city c during 2005. Observations in Columns [1] and [2] are weighted by the total expenditure of household h on UPC u in store r .
2. Household demographic dummies are for household size, male and female head of household age, marital status, race, and hispanic.
3. Regressions with store dummies include one of seven channel-ID dummies is Nielsen does not provide a store name or the store identified only has sales in one city and a store name dummy otherwise.
4. Random weight UPCs have been dropped from the sample.

been purged of these market power effects have almost the same association with population as we saw in Table 2, which suggests that most of what is being captured in the store fixed effects reflects store amenities and not market power.

4 Do Larger Cities Have More Varieties?

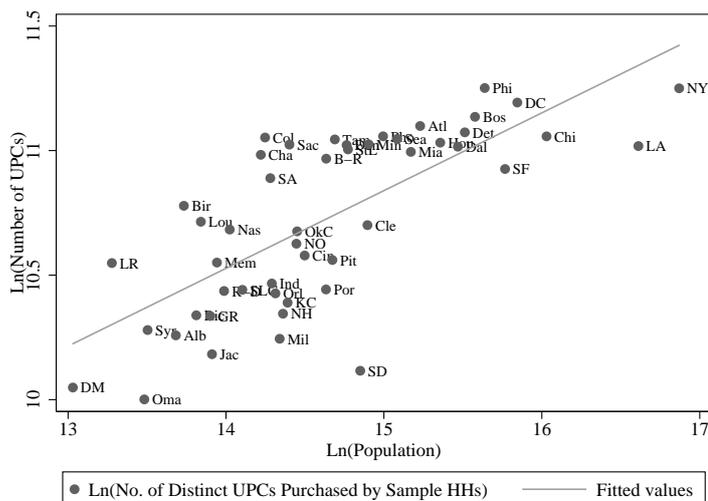
Many NEG models predict not only that the price level of tradable goods should be lower, but also that larger cities should offer consumers more varieties because more firms can enter in larger markets. This prediction also has implications for the price level in cities as varieties that are unavailable in locations are effectively priced at their reservation levels. In this sense, we can think of the variety bias as an extreme form of the well-known substitution bias that plagues fixed-weight price indexes—if prices are so high that goods are not consumed in small cities,

fixed-weight indexes will understate the true cost of living because high-priced goods that are not consumed will receive a weight of zero in the index. While we will deal with both the substitution and variety biases in Section 5, in this section we examine the underlying evidence on variety availability.

4.1 Data Overview

The simplest way to document that consumers in larger cities consume more varieties is to examine whether we observe more varieties being purchased in larger cities. Figure 3 shows the relationship between the log number of UPCs observed in the Nielsen sample for each city against log population. This relationship is upward sloping with a coefficient of 0.312 and standard error of 0.043. We cannot interpret this estimate as the elasticity of variety with respect to city size, however, because it is affected by sample bias: Nielsen tends to sample more households in larger cities.

Figure 3: Log Number of Distinct UPCs in Each City Sample vs. Log City Population



Notes:

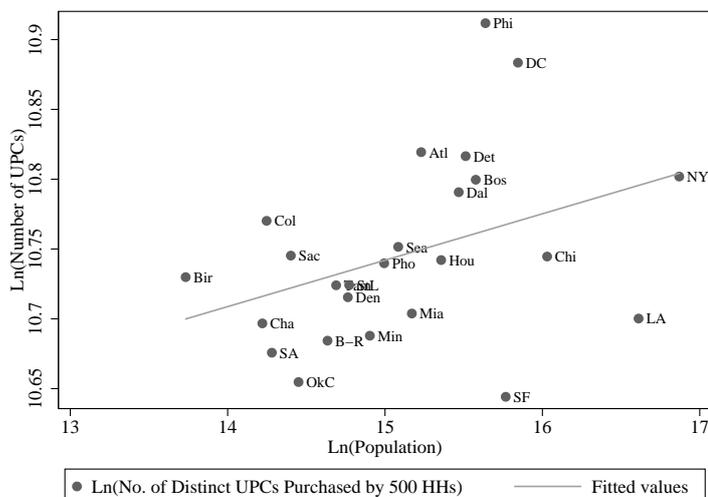
1. Numbers on plots reference the market ID of the city represented, as listed in Table A.1.

One way to deal with this bias is to instead examine whether the number of different varieties consumed by an equal number of households varies with city size. The basic idea is that any two households are less likely to purchase the same product in markets where there are more products to choose from. If there is less overlap in the varieties purchased by different households in larger cities, we expect to see equally-sized samples of households from these cities purchasing larger numbers of unique varieties.

We therefore restrict ourselves to only looking at 25 cities in which Nielsen sampled at least 500 households and compare the number of varieties purchased by a random sample of

500 households in each of these cities.²³ Figure 4 plots the aggregate number of *different* UPCs purchased by these randomly-selected households against the size of the city in which the households live. The results show a clear positive relationship between the variety of UPCs purchased by 500 households in a city and the population of the city. The slope of the linear regression fit is 0.033 with a standard error of 0.017. The large amount of noise in the 500-household variety counts indicates that this estimate may be subject to attenuation bias.²⁴

Figure 4: Log Number of Distinct UPCs Purchased by 500 Households in Each City vs. Log City Population



Notes:

1. Numbers on plots reference the market ID of the city represented, as listed in Table A.1.

These results are certainly suggestive of the notion that the number of varieties available

²³There is a trade-off between the number of households that we consider and the number of cities that can be included in the sample. As we decrease the number of households selected, we increase the number of cities in our sample (adding small cities disproportionately). However, as we work with smaller samples of households, we have a lot more noise because the number of barcodes purchased by a small sample of households can vary a lot depending on the households picked. This results in attenuation bias.

²⁴We have replicated this analysis looking at the purchases of different fixed numbers of households and, consistent with the attenuation bias hypothesis, we find that the estimated variety-city size relationship is increasing in the number of households under consideration. For example, the coefficient on city size is statistically zero when we consider the number of varieties purchased by samples of 116 households in all 49 cities, but increases to 0.05, statistically significant at the 5 percent level, when we look at the number of varieties purchased by 750 households, in the 23 cities where this is possible.

One reason why it is difficult to identify differences in the number of varieties available in a city in the purchases of small samples of households is due to the fact that many households purchase “popular” goods that other households in their city also purchase. To see the intuition here, suppose that all households purchase $N + 1$ products, N of which are the same across households in a city and one of which is drawn at random from the set of varieties available in the city. Regardless of how many varieties are available for purchase in a city, we can at most expect to see $N + 1$ unique varieties purchased by one household, $N + 2$ by two households in the same city, $N + 3$ by three, etc. The number of varieties purchased in a city will range from N to N plus the number of households sampled in a city.

in a location rises with number of inhabitants in that location, but neither provides a reliable estimate of the elasticity. In the next section, we take a more direct approach to estimating the variety-city size relationship: we use all of the information at hand to estimate the total number of varieties available in each location and then examine how these aggregate variety estimates vary with city size.

4.2 Estimating the Number of Varieties in Cities

The principle challenge that we face in measuring the number of varieties in a city is that our data is not a census of all varieties purchased in a city but rather a count of varieties based on a random sample of households. Fortunately, our problem is isomorphic to a well-studied problem in biostatistics: estimating the number different species in a general area based on the number of species identified in certain locations (see Mao et al. (2005, 2004)). Prior work in this area has solved the problem using parametric and structural approaches that yield very similar results in our data. Since the parametric approach is significantly simpler to explain, we focus on the parametric approach and relegate the the structural approach to Appendix A as a robustness check.

In order to obtain some intuition for this methodology, assume that the expected number of different products purchased by one household in city c is denoted by $S_c(1)$. The expected number of distinct products purchased in a sample of n households can be denoted by the “accumulation curve,” $S_c(n)$. Accumulation curves must be concave because every time the sample size rises by one household the probability of finding good that has not been purchased by any of the other households falls. Moreover, a critical feature of accumulation curves is that as the number of households surveyed rises, the number of observed varieties in a city must approach the true total number of varieties in a city. We can write this formally as $\lim_{n \rightarrow \infty} S_c(n) = S_c^T$, where S_c^T is the total number of distinct varieties available in the city.²⁵ In other words, the asymptote of the accumulation curve is the estimate for the total number of goods available in the city.

Estimation of S_c^T requires us to know the expected value of distinct varieties for each sample of households, *i.e.*, $(S_c(1), S_c(2), S_c(3), \dots)$, and also the functional form of $S_c(n)$. Estimating the expected number of distinct varieties purchased by a sample of n households, $S_c(n)$ is straightforward. The only econometric issue we face is that the number of distinct varieties we observe being purchased, $S_c(n)$, in a sample of n households is going to depend on exactly which households are in the sample. For example, our measure of $S_c(1)$, how many different goods one household purchases, depends on which household is chosen. In order to obtain an

²⁵This property is based on the assumption that all types of varieties have a positive probability of being purchased.

estimate of the expected number of goods purchased by a sample of n households, Colwell and Coddington (1994) propose randomizing the sample order I times and generating an accumulation curve for each random ordering indexed by i . The expected value of the number of varieties purchased by n households can then be set equal to the mean of the accumulation curves over I different randomizations, *i.e.*,

$$S_c(n) = \frac{1}{I} \sum_{i=1}^I S_{ci}(n).$$

We set $I = 50$.²⁶

Once we have our estimates for each $S_c(n)$, we can turn to estimating the asymptote, $S_c(H_c) = S_c^T$. Unfortunately, theory does not tell us what the functional form of $S_c(n)$ is, so we follow Jimenez-Valverde et al. (2006) by estimating the parameters of various plausible functional forms and use the Akaike Information Criterion (AIC) goodness-of-fit test to choose between a range of functional forms that pass through the origin and have a positive asymptote.

4.3 Results

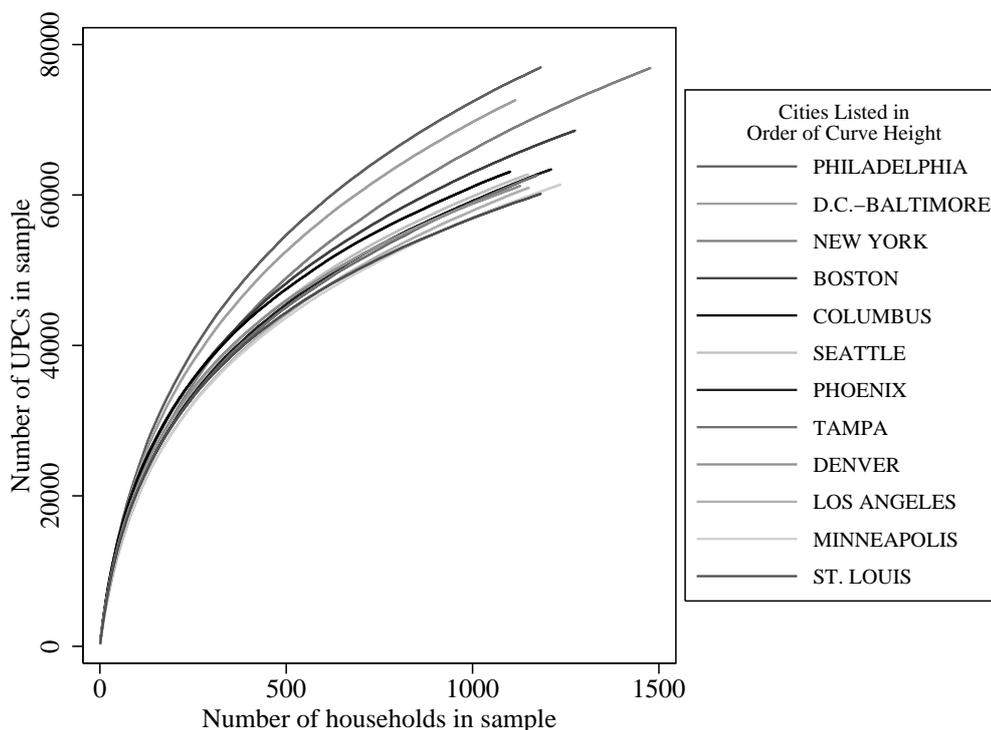
We can get a clear sense of how this methodology works by simply plotting the accumulation curves. Figure 5 presents a plot of accumulation curves for the twelve cities for which we have the largest samples. As one can see from the picture, the average sample of 1000 households in Philadelphia (population 6.2 million) purchased close to 70,000 different varieties of groceries. By contrast, the average sample of a 1000 households in Saint Louis (population 2.6 million) purchased closer to 50,000 different varieties. Moreover, these curves reveal that the four highest curves correspond to Philadelphia, D.C.-Baltimore, New York, and Boston, which are all among the five largest cities in our sample. In other words, this limited sample indicates that a given number of households tends to purchase a more diverse set of goods when that sample is drawn from a city with a larger population.

We can examine this more formally by estimating the asymptotes of the accumulation curves. Since we are not sure how to model the functional forms of these accumulation curves, we tried five different possible functional forms – Clench, Chapman-Richards, Morgan-Mercer-Flodin, Negative Exponential, and Weibull. We choose among these based on the Akaike Information Criterion (AIC). The Weibull was a strong favorite with the lowest AIC score in the majority of cities for which we modeled UPC count accumulation curves, and so we decided to focus on this functional form.

Once again, we can get intuition for how this methodology works by showing the fit for a sub-sample. Figure 6 plots the raw data and the estimated Weibull accumulation curve for our

²⁶The resulting estimates are less noisy, and their correlation with city size less subject to attenuation bias, than the 500-household variety counts studied in Section 4.1 above, each of which is just a single point on a single accumulation curve for each city.

Figure 5: UPC Accumulation Curves for Markets with 12 Largest Samples

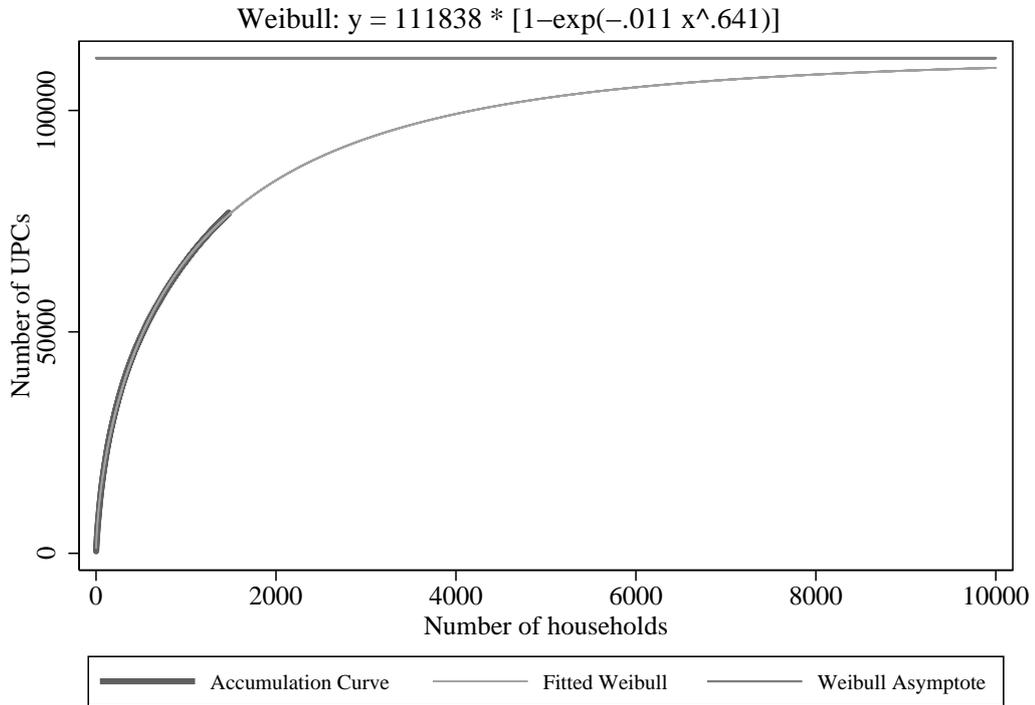


largest city, New York. A typical sample of 500 random households buys around 49,000 unique UPCs, and a sample of 1000 households typically purchases around 66,000 different goods. As one can see from the plot, the estimated Weibull distribution fits the data extremely well. The estimated asymptote is approximately 112,000 varieties, which is 35,000 more than we observe in our sample of 1500 New York households.²⁷

Figure 7 presents a plot of the log of the estimated Weibull asymptotes for each city against the log population in the city. As one can see, there is a clear positive relationship between the two variables—we estimate that households in larger cities have access to more varieties than households in smaller ones. It is interesting that the relationship between city size and the total number of varieties in a city is much stronger than the relationship between city size and the number of varieties purchased by a fixed sample of cities observed in Figure 4. This is consistent with the pattern observed in Figure 5: there is less dispersion in the number of unique UPCs purchased by a common set of household across cities increases as the number of household in that set increases. Overall, the data support relationship between the size of a city and the number of varieties available hypothesized by NEG models. Residents of New York have access just over 110,000 different varieties of groceries, while residents of small cities like

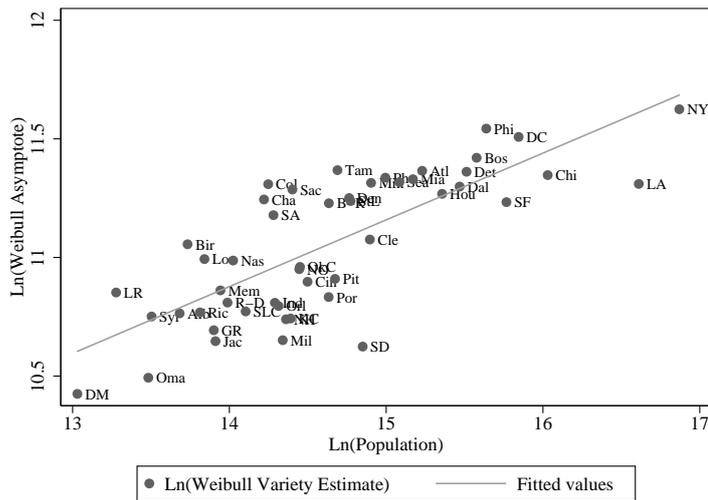
²⁷Since the number of households in a city is large, we obtain almost identical results regardless of whether set the number of varieties equal to $S_c(H_c)$ or $S_c(\infty)$.

Figure 6: Fitted UPC Accumulation Curve for New York



Omaha and Des Moines have access to fewer than 24,000.

Figure 7: Log Weibull Variety Estimate vs. Log City Population



Notes:

1. Acronyms on plots reference the city represented, as listed in Table A.1.

We test this relationship between city size and variety abundance formally in Table 5. Table 5 presents the results from regressing the log estimated number of varieties in a city on the log

of the population in the city. The first three columns of the table present regressions of the log sample counts of varieties in each city on the log of the city's population. The next three columns present regressions of the log estimate of number of varieties based on the Weibull asymptotes on city size. As one can see from comparing columns 1 and 4, the elasticity of variety with respect to population is slightly less using the Weibull estimate presumably because the Weibull corrects for the correlation between sample size and population in the Nielsen data. What is most striking, however, is that we observe a very strong and statistically significant relationship between the size of the city and the number of estimated varieties. Our estimates indicate that a city with twice the population as another one typically has 20 percent more varieties.

Table 5: Do larger cities have more UPC varieties?

	Ln(Sample Count _c)			Ln(Weibull Asymptote _c)		
	[1]	[2]	[3]	[4]	[5]	[6]
Ln(Population _c)	0.312*** [0.0432]	0.338*** [0.0678]	0.281*** [0.0971]	0.289*** [0.0373]	0.317*** [0.0582]	0.321*** [0.0841]
Ln(Per Capita Income _c)	-	-0.155 [0.341]	-0.043 [0.369]	-	-0.032 [0.293]	-0.038 [0.319]
Income Herfindahl Index	-	-0.952 [3.132]	-0.289 [3.246]	-	-1.302 [2.689]	-1.338 [2.809]
Race Herfindahl Index	-	0.064 [0.411]	0.115 [0.417]	-	0.147 [0.353]	0.145 [0.361]
Birthplace Herfindahl Index	-	0.006 [0.282]	0.029 [0.285]	-	0.068 [0.222]	0.067 [0.225]
Ln(Land Area _c)	-	-	0.087 [0.106]	-	-	-0.005 [0.0919]
Constant	6.158*** [0.632]	7.474** [3.391]	6.275* [3.704]	6.835*** [0.546]	6.790** [2.911]	6.856** [3.205]
Observations	49	49	49	49	49	49
R-squared	0.53	0.53	0.54	0.56	0.57	0.57

Standard errors in brackets; *** p<0.01, ** p<0.05, * p<0.1

One concern with these results is that they might be biased because larger cities have more diverse populations. In order to control for this we constructed a number of Herfindahl indexes based on the shares of MSA population with different income, race, and country of birth. These indexes will be rising in population homogeneity. In addition, we include the per capita income in each city. As one can see from columns 2 and 5 in Table 5, controlling for urban income and diversity does not alter the results.

Finally, we were concerned that our results might be due to a spurious correlation between city population and urban land area. If there are a constant number of unique varieties per unit area, then more populous cities might appear to have more diversity simply because they occupy more area. To make sure that this force was not driving our results, we include the log of urban land area in our regressions. The coefficient on land area is not significant in any of the

specifications, while the coefficient on population remains positive and very significant. These results indicate that controlling for land area and demographic characteristics does not qualitatively affect the strong relationship between city size and the number of available varieties. The R^2 of around 0.5 to 0.6 indicates that city size is an important determinant of variety availability. Thus, the number of tradable goods varies systematically with city size as hypothesized by the NEG literature.

5 The Price Level in Cities

5.1 Constructing an Exact Urban Price Index

In order to produce a theoretically-sound price index for a city, we need to take into account not only the heterogeneity measurement issues discussed in section 3 and the product availability differences discussed in Section 4 but also make adjustments for substitution biases. Progress can only be made by putting some more structure on the problem, and so we will assume that one can use a CES utility function to measure welfare in cities. The CES has the advantage of not only being tractable and estimatable, but also has been used in many economic geography models, such as Krugman (1991), so our results can be thought of as a reasonable structural estimate of what the price level is in this framework. Moreover, Feenstra and Weinstein (2010) show that this index yields very similar aggregate price levels to the translog (which is a second-order approximation of an arbitrary expenditure system) even when the number of varieties is varying. Thus, the CES assumption is unlikely to dramatically affect our results relative to other commonly used price indexes.²⁸

Feenstra (1994) developed the variety-adjusted price index for the CES utility function. Here, we will modify it so that it can be used with our data. In the original Feenstra paper, the price index was expressed in terms of price level in period t as compared to $t - 1$. Instead of working with two time periods, we modify the notation of the basic theory so that we can compare two locations. In particular, we will express the price level in each city as its level relative to the price level a consumer would face if the buyer faced the average price level in the U.S. and had access to all the varieties available in the U.S. This approach enables us to write the price level in every city relative to the the same U.S. national benchmark, which greatly

²⁸Given that the median number of UPCs purchased in a module by a single person household (conditional on purchasing anything in the module) is one, the data suggest that one should think of households as having heterogeneous ideal-type preferences, as opposed to the identical CES preferences that form the theoretical foundation for the variety-adjusted exact prices indexes used in this paper. This discrepancy, however, is not a problem for our analysis if we think of consumers as having a logit demand system. In particular, Anderson et al. (1987) have demonstrated that a CES demand system can arise from the aggregation of ideal-type logit consumers. We will therefore follow Anderson et al. (1987) and use the CES structure to evaluate aggregate welfare even though we know that the discrete choice model is a better depiction of reality at the household level.

simplifies the analysis.

5.1.1 Intuition

It is worth spending a little time on the intuition behind this index before plunging into the detail. Feenstra’s basic insight was that a variety-adjusted CES exact price index (EPI) can be written as a standard CES price index, which we term the “Conventional Exact Price Index” or CEPI, multiplied by a “Variety Adjustment” or VA term. The CEPI is a sales-weighted average of the prices of each good sold in the city where the weights adjust for conventional substitution effects. In our context, one can think of the CEPI in city c , $CEPI_c$, as the correct way of measuring the price level of the commonly available goods within each product group in city c relative to their national average price.

Since some goods are not available in each location, we need to adjust the $CEPI_c$ by the variety adjustment VA_c . The variety adjustment corrects the price index for the impact of missing varieties on consumer utility. It is based on two factors: the quality-weighted count of the goods unavailable in a location and the substitutability of these missing goods with other goods that are available in a location. The quality-weighted count of missing goods weights each good by how important it is in the utility function. Controlling for price, consumers will care more about not having access to high-quality varieties more than low-quality ones. In the CES setup, if two goods of have the same price, their relative market shares will be determined by their relative qualities. So, if consumers in a city do not have access to a good with a high *national* sales share, we know that they are missing a good with greater importance (or price-adjusted quality) than they would be missing if they could not buy a good with a low market share. As a result, the national market shares of goods not available in a location properly captures the quality-adjusted count of the unavailable goods.

The second adjustment comes from knowing the substitutability of the goods. Not having access to a good without close substitutes is worse than not having access to a goods with close substitutes. In order to account for this force empirically, we follow Broda and Weinstein (2010) by allowing consumers to value not having access to an entire brand-module differently than not having access to a good within a brand-module. For example, if Coke and Pepsi both sell two soft-drink barcodes under each of their brands (cans and two-liter bottles), it may be worse for consumers in a location not to have access to any Coke items than to not have access to just two-liter bottles. In terms of the formulas, this will require us to weight our quality-adjusted counts of missing products in a location by how similar those products are to products that are available in that location.

5.1.2 Notation

Before actually writing down the price index, we need to set forth some notation to make these concepts concrete. Our first task is to compute the quality-weighted counts of missing goods in each city that we can feed into our price index. Let $g \in \{1, \dots, G\}$ denote a product “group”, which we define in the same way as Nielsen to capture broadly similar grocery items. Let B_g and U_g respectively denote the set of all “brand-modules” and UPCs in a product group g , and U_b be the set of all UPCs in brand-module b .

Not all UPCs are sold in every city, so we now define the subsets of UPCs that we observe being purchased. Let v_{uc} denote the value of purchases of UPC u observed in the sample for city c .²⁹ We define $U_{bc} \equiv \{u \in U_b | v_{uc} > 0\}$ as the set of all UPCs in brand-module b that have positive observed sales in city c , and $U_{gc} \equiv \{u \in U_g | v_{uc} > 0\}$ as the set of all UPCs in product group g that have positive observed sales in city c . We similarly define $B_{gc} \equiv \{b \in B_g | \sum_{u \in b} v_{uc} \neq 0\}$ as the set of all brand-modules that have positive sales in city c in product group.

We next need to measure the share of available goods both within brand-modules and within product groups. The simplest of these to construct is the share of available UPCs within a brand-module. Let s_{bc} be share of *national* expenditures on brand-module b on UPCs U_{bc} that are sold in city c , *i.e.*,

$$s_{bc} \equiv \frac{\sum_{u \in U_{bc}} \sum_c v_{uc}}{\sum_{u \in U_b} \sum_c v_{uc}}. \quad (5)$$

s_{bc} tells us the expenditure share of UPCs within a brand-module that are available in a city using *national* weights. The numerator is the total amount spent *nationally* on all the goods available in city c on UPCs in brand-module b , while the denominator gives the total spent on brand-module b *nationally*. This ratio will be less than one whenever a UPC from brand-module b is unavailable in city c . Another way to think about this variable is to realize that s_{bc} is positively related to the number of available UPCs in a city and will be smaller if, holding fixed the number of unavailable UPCs, varieties with a high market share are unavailable. It is easiest to see what moves s_{bc} by considering an extreme case. If all varieties had the same price and quality and therefore the same market share, s_{bc} would equal the share of all varieties within a brand-module that are available in city c , and s_{bA} would equal s_{bB} only if cities A and B had the same number of UPCs available in brand-module b . In general, however, two cities with the same number of UPCs available in brand-module b will have equal values of s_{bc} if their unavailable varieties have the same aggregate importance in national consumption, or national expenditure share.

²⁹The national average price of a UPC is the total value of purchases of that UPC across all cities in the HomeScan sample divided by the total quantity that these purchases represent. In all of the analysis below, we work with nationally-representative values and quantities for each UPC, scaling the value and quantity of purchases in each city by the population in that city divided by the total number of household members represented in the Nielsen sample for that city (*i.e.*, the sum of the household sizes for the Nielsen sample households).

We define a quality-weighted count of the brand-modules available in a location analogously:

$$s_{gc} \equiv \frac{\sum_{b \in B_{gc}} \sum_c v_{bc}}{\sum_{b \in B_g} \sum_c v_{bc}}, \text{ where } v_{bc} = \sum_{u \in U_b} v_{uc}, \quad (6)$$

and s_{gc} is the expenditure share of brand-modules within product group g that are available in city c . The numerator of s_{gc} is the total amount spent *nationally* on the product group g brand-modules available in city c , and its denominator is the total spent on product group g *nationally*. While s_{bc} tells us about the availability of UPCs within brand-modules, s_{gc} tells us about the availability of brand-modules themselves.

Finally, it is useful to discuss the price data we use in the index. In our preferred specification, we will work with “adjusted prices” that correct for product, purchaser, and retailer heterogeneity biases. In the simplest case, where we only control for product heterogeneity biases, we set the adjusted price, $\tilde{P}_{ucr h}$, equal to the the actual price: $\tilde{P}_{ucr h} \equiv P_{ucr h} = \exp\{p_{ucr h}\}$. However, at other times we may want to correct for product and purchaser heterogeneity biases in the collection of price data that we documented in Section 3. In this case, we will set $\tilde{P}_{ucr h} \equiv \exp\{p_{ucr h} - \mathbf{Z}_h \hat{\beta}\}$, or to correct for product, purchase, and retailer heterogeneity biases we will set prices equal to $\tilde{P}_{ucr h} \equiv \exp\{p_{ucr h} - \hat{\alpha}_r - \mathbf{Z}_h \hat{\beta}\}$. Similarly, we can write the adjusted value and quantity of UPC u purchased in city c as $\tilde{v}_{uc} \equiv \sum_{h \in H_c} \sum_{r \in R_c} \tilde{P}_{ucr h} q_{ucr h}$ and $q_{uc} \equiv \sum_{h \in H_c} \sum_{r \in R_c} q_{ucr h}$, respectively, where $q_{ucr h}$ is the quantity that household h in city c purchases of UPC u in store r .

5.1.3 Definition

We define a variety-adjusted spatial price index is based on Broda and Weinstein (2010)’s extension of the Feenstra (1994) index. This extension is exact to a nested CES utility function and was developed in order to measure inflation over barcoded products, nested according to Nielsen categories defined above. We adapt Broda and Weinstein (2010)’s intertemporal price index to a spatial context, re-writing their Proposition 1 as follows:

Proposition 1. *If $B_{gc} \neq \emptyset$ for all $g \in G$, then the exact price index for the price of the set of goods G in city c relative to the nation as a whole that takes into account the differences in variety in the two locations is given by,*

$$EPI_c = \prod_{g \in G} [CEPI_{gc} VA_{gc}]^{w_{gc}}$$

where:

$$CEPI_{gc} \equiv \prod_{u \in U_{gc}} \left(\frac{\tilde{v}_{uc}/q_{uc}}{\sum_c \tilde{v}_{uc}/\sum_c q_{uc}} \right)^{w_{uc}}$$

$$VA_{gc} \equiv (s_{gc})^{\frac{1}{1-\sigma_g^a}} \prod_{b \in B_{gc}} (s_{bc})^{\frac{w_{bc}}{1-\sigma_g^w}}$$

w_{uc} , w_{bc} , and w_{gc} are log-ideal CES Sato (1976) and Vartia (1976) weights defined in Appendix B, σ_g^a is the elasticity of substitution across brand-modules in product group g , and σ_g^w is the elasticity of substitution among UPCs within a brand-module.

This index has been used extensively in the literature (see Feenstra (2010) for a survey of its use), so we will only briefly discuss its properties here. The index is exact in the sense that it answers the question of what is the unit cost of achieving a given utility level in various cities assuming CES preferences.

We can obtain some intuition for the formula by breaking it up into several components. The term in the square brackets is the exact price index for each product group, and EPI_c is just a weighted average of these product group price indexes where the Sato-Vartia weights take into account both the importance of each product group in demand and the ability of consumers to substitute away from expensive product groups. Each product-group price index is composed of two terms. $CEPI_{gc}$ is the conventional exact price index for product group g . It is a sales weighted average of the prices of each good sold in the city where the weights adjust for conventional substitution effects. One can think of $CEPI_{gc}$ as the correct way of measuring the price level of each product group in city c relative to its national average if no goods were unavailable in the city.

Since some goods are not available in each location, we need to adjust the $CEPI_{gc}$ by the variety adjustment VA_{gc} . The variety adjustment is composed of two terms. $(s_{gc})^{1/(1-\sigma_g^a)}$ corrects the index for the importance of missing brand-modules. As we argued above, s_{gc} provides a quality-adjusted count of the missing brand-modules in city c , and the exponent weights these counts by how substitutable they are. For example, if the Coke soft-drink brand were not available in a city, its importance for the price index would depend on its quality, that is how important Coke is nationally (s_{gc}), and how substitutable Coke is with other soft drinks (σ_g^a). If Coke is highly substitutable with other brand-modules, then σ_g^a will be large, and not having Coke available will not matter much, but if Coke is a poor substitute for other soft drinks, σ_g^a will be small and the quality-adjusted count, s_{gc} , will matter more. The second variety adjustment term is a weighted geometric average of variety adjustments for each brand-module available in city c , $\prod_{b \in B_{gc}} (s_{bc})^{\frac{w_{bc}}{1-\sigma_g^w}}$. Each brand-module variety adjustment, $(s_{bc})^{\frac{1}{1-\sigma_g^w}}$, corrects the index for the fact that even if brand-module b is available in a city, not all of the varieties within that brand-module may be available. Thus, if the two-liter bottle of Coke were

not available in a city, the impact on the price index would depend on how important two-liter Coke is (s_{bc}) and how similar two-liter Coke is to other Coke products (σ_g^w). We obtain the elasticities of substitution computed for UPCs within a brand-module and across brand-modules within a product group from Broda and Weinstein (2010).

5.2 Measuring the Share of Commonly Available Goods

Recall that we do not observe the full set of UPCs available in each city. Just as we estimated the raw count of UPCs available in each city, we now have to estimate the quality-adjusted counts (or national expenditure shares) of available brand-modules (s_{gc}) and UPCs (s_{bc}). In Section 4.2, we built an accumulation curve corresponding to the raw count of the UPCs represented in different-sized samples of households for each city. We now use the same method to build curves corresponding to the national market shares of the different brand-modules and UPCs represented in a sample. Our estimates of the s_{gc} and s_{bc} are simply the asymptotes of these shares accumulation curves in each market. If $s_{gc}(n)$ denotes the national expenditure share of brand-modules within a product group g purchased by n households in city c , our estimate of s_{gc} (the national market share of the brand-modules that are available in a city within a product group) is simply the value we expect $s_{gc}(n)$ to take when the number of households equals the number of households in the city, which we approximate using the asymptote of the accumulation curve, or $\lim_{n \rightarrow \infty} s_{gc}(n)$. We use a similar procedure to estimate the national market share of the UPCs that are available in a city within each brand-module that is available in the city, s_{bc} .

Unfortunately, we are unable to estimate a value of s_{bc} for every brand-module b in each product group g for each city c because household samples are sometimes too small to allow us to observe more than a few purchases of UPCs within many of the brand-modules available in a city. Since we do not want to estimate accumulation curves for brand-modules in which there are very few purchases, we pool the data across brand-modules in each city and estimate a common s_{bc} within each product group, which we denote \bar{s}_{gc} . In other words, \bar{s}_{gc} tells us the average national market share of UPCs available in a city within the set available brand-modules in a product group in that city. Since \bar{s}_{gc} does not vary at the brand-module level, this assumption simplifies the variety adjustment in Proposition 1 to:

$$VA_{gc} \equiv (s_{gc})^{\frac{1}{1-\sigma_g^d}} (\bar{s}_{gc})^{\frac{1}{1-\sigma_g^w}} \quad (7)$$

5.3 Price Level in Cities: Evidence

In this section we will present our variety-adjusted exact prices for each city. First, however, we illustrate how we estimate the brand-module and UPC share inputs into the variety-adjustment

component (s_{gc} and \bar{s}_{gc}) using the bread product group and show how these quality-adjusted counts vary across cities of different sizes.

Figures 8 and 9 plot the share accumulation curves for one of the largest product groups in terms of sales and the number of varieties available nationally (bread and baked goods) and using data from two cities with very different populations, New York and Des Moines. Figure 8 shows the s_{gc} accumulation curve, which tells us about the availability of brand-modules, while Figure 9 portrays the same curve for \bar{s}_{gc} (which tells us about the availability of UPCs within brand-modules). We estimate s_{gc} and \bar{s}_{gc} for each product group and city as the asymptotes of a Weibull function fitted to each of these curves. As one can see from the plots, the Weibull fits the share accumulation curves extremely well.³⁰ By examining the asymptote of the Weibull distribution in the first panel, we can see that New Yorkers have access to a set of bread brand-modules available that constitute 80 percent of national expenditures on bread. By contrast, residents of Des Moines, with a population less than a fortieth as large, have access to bread brand-modules that constitute only 62 percent of national expenditures. In other words, the quality-adjusted count of brand-module varieties is higher in Des Moines than in New York.

Figure 9 shows that even when brand-modules are available in a city, many of the UPCs that make up these brand-modules are often not available. For example, consumers in New York and Des Moines only had access to bread UPCs that accounted for on average 64 and 43 percent of their brands' national sales in a given module, respectively. This indicates that knowing that a firm sells a product in a city does not necessarily mean that all varieties of that product are available there.

In summary, the bread brand-modules available to New Yorkers account for 18 percent more of the national bread market than the brand-modules available in Des Moines and, conditional on a brand being available in their city, the varieties of these brand-modules available to New Yorkers account for 21 percent more of the national market for each brand-module than the varieties of these brand-modules available to Des Moines residents.

To demonstrate how these differences in the market shares of UPCs and brand-modules available in the two cities translate into a discount on the exact price index in New York relative to Des Moines, we can calculate the variety adjustments for bread in each city from equation (7). The elasticity of substitution between UPCs within brand-modules in the bread product group is 17.2, so the UPC variety adjustment for New York is $0.64^{1/(1-17.2)}$, or 1.028. The across-brand-module elasticity of substitution for the bread product group is 9.6, so the brand-module variety adjustment for New York is $0.80^{1/(1-9.6)}$, or 1.026. Assuming that all available goods are sold at the same price in all locations, our estimates imply that the fact that there are fewer bread varieties available in New York than nationally means that someone restricted to

³⁰We selected the Weibull since the AIC favored it for the raw count accumulation curves. We find that the AIC favors the Weibull distribution when estimating the share accumulation curves as well.

Figure 8: Brand-Module Share Accumulation Curves for the Bread Product Group

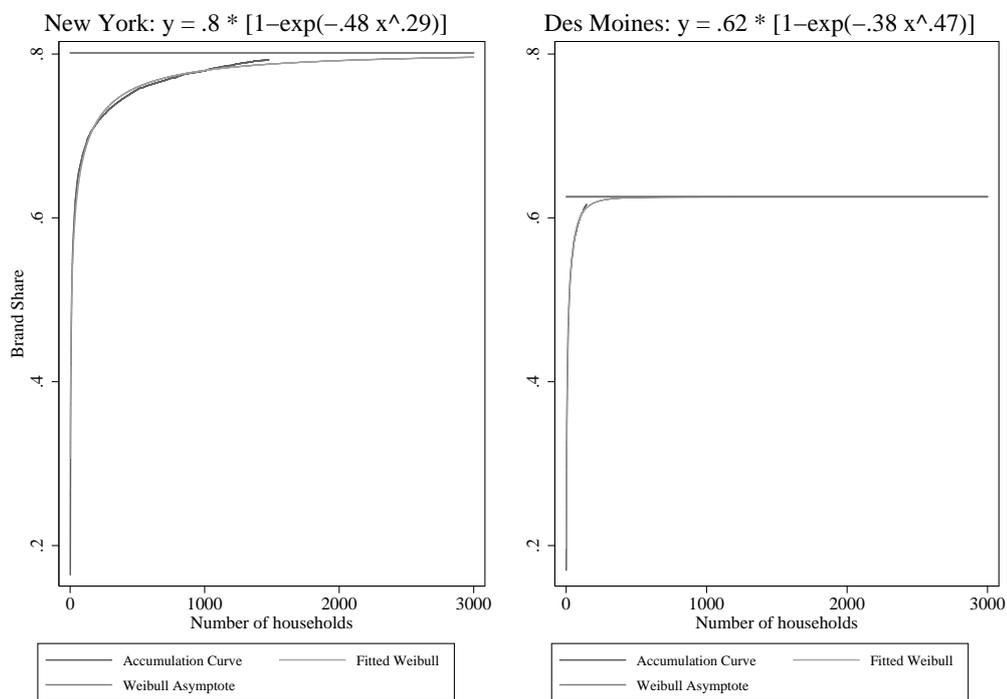
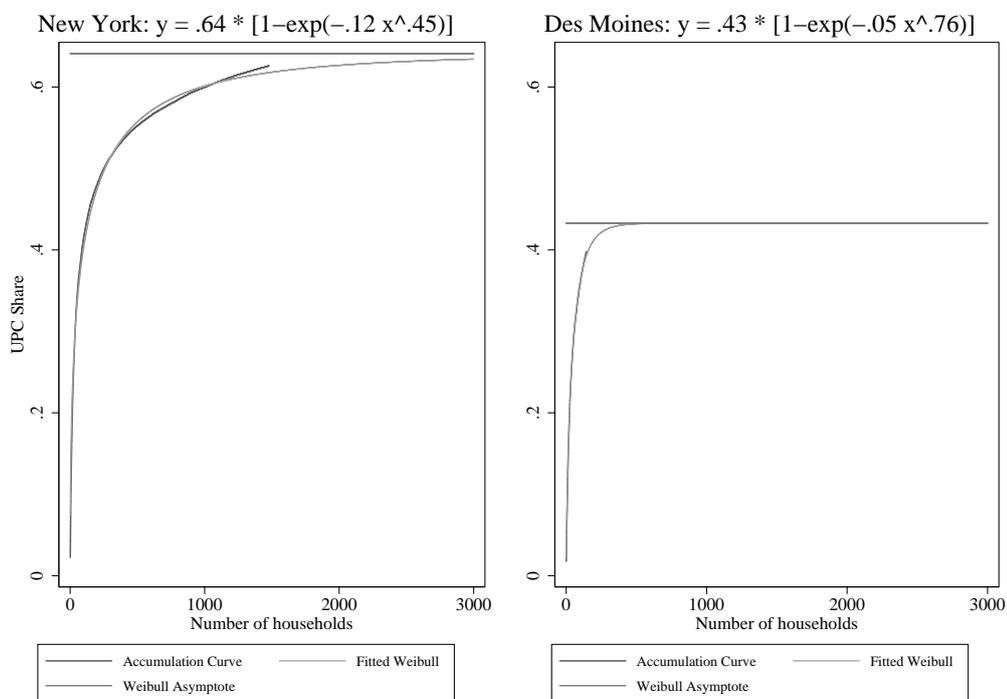


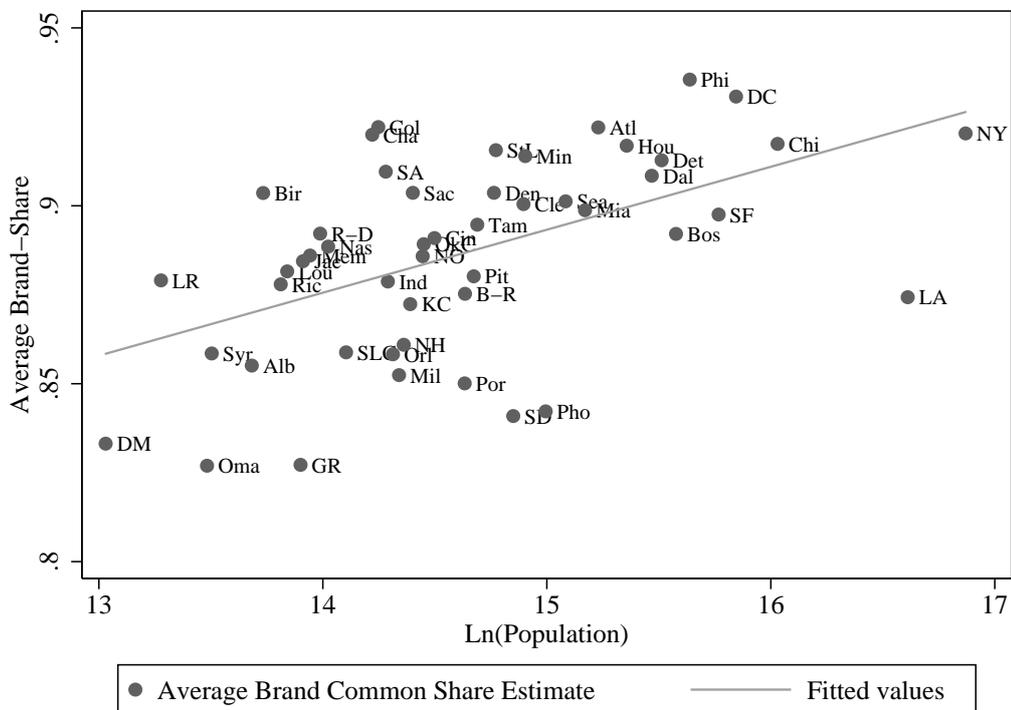
Figure 9: UPC Share Accumulation Curves for the Bread Product Group



only consume those varieties available in New York would face a price index that is 5.5 percent higher than someone with access to all national varieties at average national prices. A similar calculation shows that the variety adjustment for the bread product group in Des Moines is equal to 6.9 percent. The variety adjustment for New York relative to Des Moines is equal to $1.055/1.069$, or 0.987 , implying that if the prices of commonly available varieties are the same in the two cities, people living in New York face 1.3 percent lower costs for bread because they have access to more varieties.

Figures 10 and 11 plot the average asymptotes of the share accumulation curves (estimated using a Weibull distribution) in each city against the log of the population in each city to show that the results for bread in New York and Des Moines are representative of the sample as a whole. As one can see from the figure, there is a strong positive relationship between our estimates for the national market share covered by varieties available in each city and the city's population. Nationally, consumers spend 5 percent more on brand-modules and UPCs available in the largest cities than they do on those brand-modules and UPCs available in the smallest cities. Once again we see that people in larger cities have access to more varieties and/or more popular varieties, while those in smaller cities have more limited access.

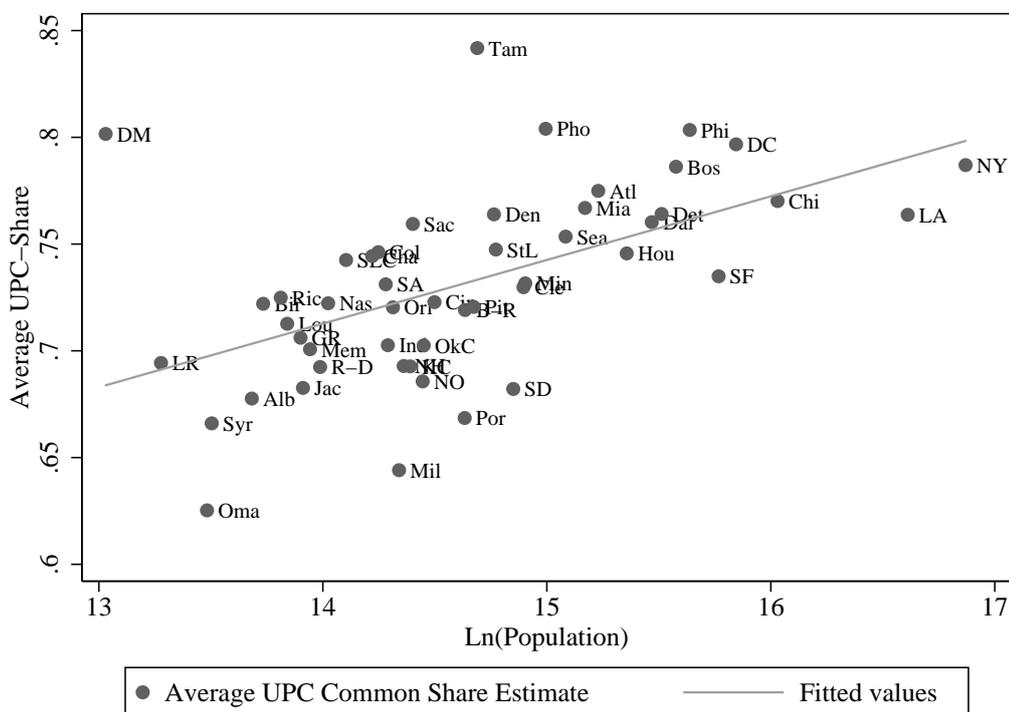
Figure 10: Average Brand-Module Share Estimate vs. Log City Population



Notes:

1. Common shares are averaged across product groups using national product group sales weights.
2. Acronyms on plots reference the city represented, as listed in Table A.1.

Figure 11: Average UPC Share Estimate vs. Log City Population



Notes:

1. Common shares are averaged across product groups using national product group sales weights.
2. Acronyms on plots reference the city represented, as listed in Table A.1.

If we compare the results in Figure 10 with those in Figure 7, we can see the importance of adjusting the data for the quality of the unavailable varieties. We saw in Section 4 that we estimated New York to have almost four times as many varieties as Des Moines. However, many of the brands-modules available in New York but not in Des Moines have extremely small market shares, implying that consumers do not care that much about them. Moreover, conditional on a brand-module being available, we can see from Figure 11 that Des Moines is an outlier in terms of having, on average, more UPCs available within that brand-module than New York. In general, both the shares of brand-modules available and of UPCs available in a brand-module, conditional on that brand-module being available, are increasing in city-size, but less so than the raw counts. Even though both raw and quality-adjusted counts of variety availability indicate that larger cities have more varieties, the differences in quality-adjusted counts are much smaller, presumably because the most popular brand-modules and UPCs tend to be available everywhere.

The asymptotes of the share accumulation curves, in conjunction with our price data and sales weights, enable us to estimate exact price indexes for each city. Proposition 1 defines the exact price index in terms of product group-specific conventional exact price indexes ($CEPI_{gc}$)

and variety adjustments (VA_{gc}). We aggregate across product groups to obtain values for each of these components for each city. Rearranging terms in the exact price index definition from Proposition 1, we can see that the exact price index is the product of each of these aggregate components:

$$EPI_c = \prod_{g \in G} [CEPI_{gc} VA_{gc}]^{w_{gc}} = \prod_{g \in G} [CEPI_{gc}]^{w_{gc}} \prod_{g \in G} [VA_{gc}]^{w_{gc}} = CEPI_c VA_c.$$

Table 6 presents our estimates for how the conventional exact price index ($CEPI_c$), the variety adjustment (VA_c), and their product, the exact price index (EPI_c), vary across cities. The first three columns in the table use unadjusted prices to compute the index. As we saw in Table 2, unadjusted prices are higher in larger cities, and these results translate into a common exact price index that rises with city size. These higher prices, however, are offset by an even stronger variety adjustment to prices arising from the greater availability of varieties in larger cities, leading to an exact price index that falls slightly, but not significantly with city size.

Table 6: Are price indexes higher in larger cities?

	Unadjusted Prices			Prices Adjusted for Purchaser Heterogeneity			Prices Adjusted for Purchaser and Store Heterogeneity		
	CEPI _c	VA _c	EPI _c	CEPI _c	VA _c	EPI _c	CEPI _c	VA _c	EPI _c
Ln(Population _c)	[1] 0.012*** [0.0035]	[2] -0.016*** [0.0028]	[3] -0.0025 [0.0050]	[4] 0.012*** [0.0034]	[5] -0.016*** [0.0028]	[6] -0.003 [0.0050]	[7] 0.0041** [0.0018]	[8] -0.015*** [0.0028]	[9] -0.011*** [0.0036]
Constant	0.82*** [0.051]	1.31*** [0.042]	1.11*** [0.074]	0.82*** [0.050]	1.31*** [0.042]	1.12*** [0.073]	0.94*** [0.027]	1.31*** [0.042]	1.24*** [0.053]
Observations	49	49	49	49	49	49	49	49	49
R-squared	0.206	0.394	0.005	0.201	0.394	0.007	0.096	0.388	0.167

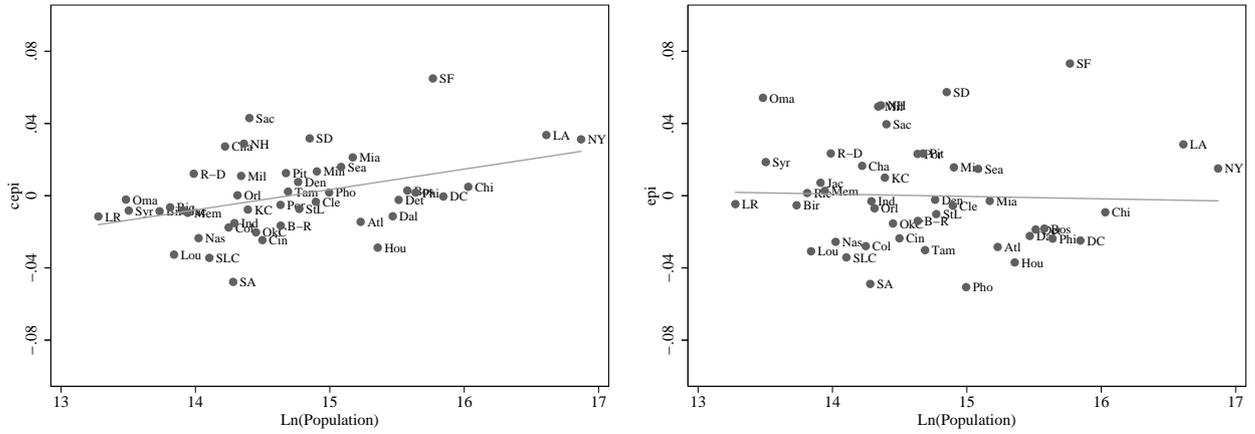
Standard errors in brackets; *** p<0.01, ** p<0.05, * p<0.1.

Notes:

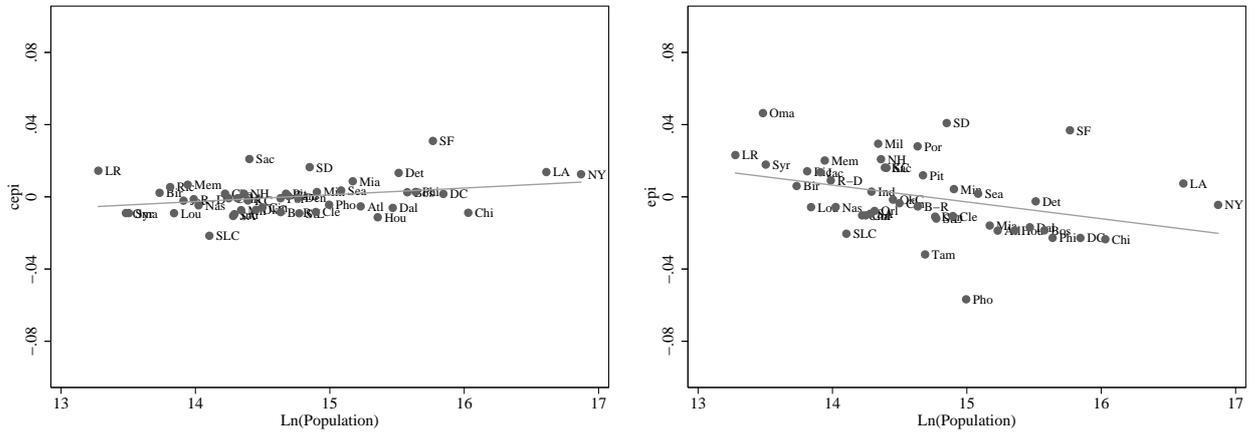
1. The dependent variables in the above regressions are indices. These indices are calculated using unadjusted prices or prices that have been adjusted as indicated above, using estimates from the regression in Column [4] of Panel A in Table 2.
2. $EPI_c = CEPI_c VA_c$ which implies that $\log(EPI_c) = \log(CEPI_c) + \log(VA_c)$. Note that the dependent variables in the above regressions are in levels, not logs, so the coefficients on log population in the CEPI_c and VA_c regressions do not add to the coefficient on log population in the EPI_c regression.

As we argued in Section 3.3, the unadjusted prices, however, are problematic because they do not correct for the store in which the goods are purchased or the type of household making the purchase. We therefore use adjusted prices in the subsequent columns. Our price adjustments are based on the estimates from Column 4 of Table 2. Adjusting for purchaser heterogeneity hardly had any impact in Table 2, and it barely affects our point estimates in columns 4-6. However, adjusting for retailer heterogeneity has a big impact. Columns 7-9 present results in which we also control for retailer heterogeneity. We still obtain a statistically significant positive association between city size and the common goods price index, but it is economically small in size. A person moving from Des Moines to New York City would experience the price

Figure 12: City Price Indexes vs. Log City Population
 Indexes Calculated using Unadjusted Prices



Indexes Calculated using Prices
 Adjusted for Purchaser and Store Heterogeneity



Notes:

1. Acronyms on plots reference the city represented, as listed in Table A.1.
2. City price indexes are normalized to be mean zero.

of commonly available groceries rise by 1.5 percent. However, this higher common goods price index would be more than offset by the fact that many goods are not available in Des Moines and are therefore effectively priced at their reservation level. The gains due to variety mean that the theoretically rigorous exact price index indicates that the price level in New York is actually 4.2 percent lower than Des Moines.

We can assess the importance of the heterogeneity and variety availability biases visually in Figure 12. The plots on the left shows how the common-goods exact price index varies across cities, and the plots on the right shows how the variety-adjusted exact price index varies. As we saw in Figure 2, there is virtually no upward relationship between the price index for commonly available UPCs that adjusts for heterogeneity biases and only a slight upward sloping

relationship when we do not. Similarly, the negative relationship between city population and the variety-adjusted exact price index is clearly not driven by any outliers.

Figure 12 also demonstrates how controlling for heterogeneity impacts the variance of the price level across cities. The common-goods exact price indexes plotted on the top left are calculated with unadjusted prices. The variance in these price indexes, that control for product heterogeneity but not purchaser or store effects, is already 87 percent lower than the variance in our replication of the ACCRA food price index. When we further control for purchaser and store heterogeneity in the indexes on plotted on the bottom left, the variance falls by an additional 10 percentage points to just 3 percent of the variance in the ACCRA-style index. In sum, the variance in the price level falls by 97 percent when we move from an *ad hoc* price index to a theoretically-justified one that corrects for both heterogeneity and substitution biases. This drop is almost identical to the 95 percent decline in variance that we observed comparing our replication of the ACCRA food price index with the city fixed effects estimated in Section 5.3. The methodological differences between the city fixed effects and the CES indexes, therefore, play only a small role in the drop in variance. We attribute most of the variance in measured PPP of commonly-available food products to heterogeneity, rather than substitution biases.³¹

It is interesting to compare the common-good exact price index, $CEPI_c$, in Table 6, with the coefficient on the city fixed effects, $\hat{\alpha}_c$, in Table 2. We can think of the city fixed effects as a price index in the sense that they summarize the relative price levels across locations, where relative prices are weighted by the transaction value weights used in the regressions. An important difference between the two is that the Sato-Vartia weights used to calculate the $CEPI_c$ are theoretically justified if one assumes a CES utility function, whereas the transaction weights used in the estimation of the $\hat{\alpha}_c$'s do not have a theoretical justification but rather serve an econometric purpose: to minimize the influence of errors associated with small transactions on the estimates. Nevertheless, it is reassuring that the associations between these variables and population are so close; the differences are typically around 0.002. This indicates that our results regarding the computation of the price level of available goods in a city are robust to a number of sensible aggregation methods.

Taken together with those in Section 3, these results suggest that conventional price indexes provide a very misleading picture of what the prices of tradables are in cities and also help to explain why. In Column 1 of Table 1, we saw that the elasticity of a conventional urban price index with respect to population was 4.2 percent. However, once we controlled for product heterogeneity, the fact that the standard urban price indexes don't compare the same goods, we see that the elasticity falls to 1.6 percent (Column 1 of Table 2): a loss of two-thirds of its magnitude. Controlling for purchaser and especially retailer heterogeneity more than halves the magnitude of this coefficient, taking it down to 0.6 percent. Further using the common exact

³¹These price indexes and their relative variances are listed in Appendix Table (A.2).

price index instead of an *ad hoc* one, which is the principle difference between $CEPI_c$ and $\hat{\alpha}_c$, reduces the coefficient to 0.4 percent (Column 7 of Table 6), which is less than one tenth of its initial magnitude. Finally, the exact price index, which differs from the $CEPI_c$ by correcting for the fact that some goods are dropped from $CEPI_c$ because they are not available (an extreme form of substitution bias), flips the sign, suggesting that the elasticity of urban price with respect to population is -1.1 percent (Column 9 of Table 6): a 5 percentage point differential. Thus, the standard result that the prices of tradables are higher in cities is due to a series of measurement issues. If one uses a theoretically-sound approach, the result reverses.

5.4 Robustness

One possible concern about these results is that people who live in large cities might not shop in all of the neighborhoods within the city's borders. This might bias our variety adjustment downwards because we may be counting varieties in, say, suburbs that are irrelevant for people who live downtown.

An easy solution to this problem is to make use of the fact that the Nielsen data records the household's county and the zip codes of the stores in which they purchase each UPC. We first assume that a household is located in the zip code in the county where its grocery expenditures are the highest. We then construct the population of the market where households in each zip code can choose to shop. For a set of households living in the same zip code, we sum the populations of the zip codes where we observe the set of households purchasing UPCs. This sum tells us the population of the market that is available to households located in that zip code. We average these market populations across the zip codes in each city to yield the average population of the neighborhoods where households in the city choose to shop. We refer to this number as the relevant population.

Table 7 presents results in which we replace the city's population with the relevant population in the neighborhoods in which households actually shop. The results are, if anything, stronger than the results using urban population as the measure of city size. The fact that the results are so similar indicates that understanding where households shop within cities is not important for understanding the relationship between city size and prices. The key finding is that the exact price index falls with city size, and this relationship is significant whenever we adjust prices to take into account store amenities and buyer characteristics.

6 Producer Prices in Cities

Thus, far we have been focused on consumer prices. However, given that we did not find much variation in retail markups (which is consistent with our CES structure), we push the

Table 7: Are price indexes higher in larger markets?

	Unadjusted Prices			Prices Adjusted for Purchaser Heterogeneity			Prices Adjusted for Purchaser and Store Heterogeneity		
	CEPI _c	VA _c	EPI _c	CEPI _c	VA _c	EPI _c	CEPI _c	VA _c	EPI _c
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]
Ln(Relevant Population _c)	0.0082*** [0.0030]	-0.016*** [0.0020]	-0.0066 [0.0041]	0.0079** [0.0030]	-0.016*** [0.0020]	-0.0069* [0.0040]	0.0027* [0.0016]	-0.015*** [0.0020]	-0.012*** [0.0027]
Constant	0.86*** [0.050]	1.34*** [0.034]	1.18*** [0.068]	0.86*** [0.049]	1.34*** [0.034]	1.19*** [0.067]	0.95*** [0.026]	1.34*** [0.034]	1.29*** [0.045]
Observations	49	49	49	49	49	49	49	49	49
R-squared	0.137	0.553	0.052	0.133	0.553	0.058	0.06	0.549	0.304

Standard errors in brackets; *** p<0.01, ** p<0.05, * p<0.1

Notes:

1. The dependent variables in the above regressions are indices. These indices are calculated using unadjusted prices or prices that have been adjusted as indicated above, using estimates from the regression in Column [4] of Panel A in Table 2.
2. $EPI_c = CEPI_c VA_c$ which implies that $\log(EPI_c) = \log(CEPI_c) + \log(VA_c)$. Note that the dependent variables in the above regressions are in levels, not logs, so the coefficients on log population in the CEPI_c and VA_c regressions do not add to the coefficient on log population in the EPI_c regression.

data and strip prices of retailing costs to estimate how producer prices vary across cities. This is an interesting exercise because many economic geography models (*e.g.*, Helpman (1998)) postulate that the price index for tradables should be lower in larger cities, and the price of land offsets this price. In these models, it might be the case that producer prices are lower in larger cities, even when consumer prices indexes, which include land costs, are not.³² Land costs may drive a similar wedge between retail and wholesale prices in our data. Our retail prices undoubtedly contain a non-traded land component reflecting the rent paid by the stores selling the goods. The fact that we observe PPP in these retail prices does not rule out the possibility that producer prices are lower in bigger cities. While our results provide the correct elasticity between consumer prices and the size of a city, to the extent that rents are higher in larger cities, they likely understate the relationship between the prices of traded goods and city size. Although it is not appropriate for welfare analysis, it is worth considering what happens to the traded component of goods prices as city size varies.

If we assume that all prices contain a non-traded rental component that rises in proportion to the rent, we can examine how the tradable component of the price index moves with population by controlling for a city's commercial real estate rental cost in the price regressions. This effect is easiest to see in comparing Panels A and B of Table 8. Panel A replicates the price-city size regressions across the 45 cities for which we have retail rent data. These results are almost identical to the estimates over the full sample: each of the coefficients and their associated standard errors are identical to their corresponding values in Panel B in Table 2 up to one significant figure. In Panel B, we include retail rents as an explanatory variable and estimate

³²In Helpman (1998), producer prices of identical tradable goods are lower, on average, in larger cities because more goods are produced in high population areas and these goods can be consumed locally without paying transport costs. There could, however, be other mechanisms driving our results here. We leave distinguishing between these explanations to future work.

$$\hat{\alpha}_c = \alpha + \gamma \ln(\text{Pop}_c) + \delta \ln(\text{Rent}_c) + \varepsilon_c, \quad (8)$$

where $\ln(\text{Rent}_c)$ is the log of commercial real estate rents in city c and greek letters are parameters to be estimated. The results indicate a very strong positive relationship between retail rents and urban price levels as is evidenced by the large positive and significant coefficient on log retail rents and the dramatic jump in R^2 . Once again, the coefficient on population now becomes negative (although not significantly so), indicating a weak negative association between city size and the prices of traded goods. Interestingly, the coefficient on retail rent is quite small in most specifications, which suggests that the elasticity of goods prices with respect to retail rents is less than 10 percent. This helps explain why the deviations from PPP are so small in figures 2 and 12.³³

We also can adjust our price index regressions in Table 6 for this force. In Panel B of Table 9, we regress each element of our exact price index on both log population and log rent. As one might expect, controlling for land rents has a big impact on the common-good exact price index but has almost no impact on the variety-adjustment, which has no clear theoretical tie to land prices. Overall, this suggests that the price index for the tradable component of goods prices falls about 50 percent faster than the overall index.

Although there is a weak positive relationship between city size and retail prices after controlling for product, purchaser, and retailer heterogeneity (but not product availability), these results suggest that this positive association is entirely due to the fact that retail prices incorporate non-traded goods prices, which vary across cities: if we focused on wholesale prices alone, we would expect to see, if anything, a negative association with city size.

7 Conclusion

In this paper we have shown how the lack of sophistication of standard spatial price indexes can lead to very misleading portraits of how prices vary across space. We find that most of the large dispersion in food prices across space can be attributed problems in the measurement of these prices. 97 percent of the variation in price levels across space in these indexes can be attributed to unobserved heterogeneity in the goods being compared. In particular, product and retailer heterogeneity play very large roles. We also show that standard indexes suffer severely from the spatial equivalent of the “new goods bias” that is well known in inflation measurement. Small cities offer consumers a much smaller array of available goods. Building a price index that

³³This result is robust across all of our specifications from Section 3. We have replicated the results in Tables 3 and 4 over the sample of cities with retail rents data and find almost no change in the coefficients and their statistical significance. In all cases, adding retail rents to these regressions flips the sign on the coefficient on population (although the point estimate is no-longer significant).

Table 8: Prices and City Size, Controlling for Rents

	City Fixed Effect Coefficient from Panel A of Table 2			
	[1]	[2]	[3]	[4]
UPC Fixed Effects	Yes	Yes	Yes	Yes
City Fixed Effects	Yes	Yes	Yes	Yes
Household Demographic Dummies	No	Yes	No	Yes
Store Dummies	No	No	Yes	Yes

Panel A: Cities with Non-Missing Retail Rent Data				
Ln(Population _c)	0.0133*** [0.00437]	0.0123*** [0.00433]	0.00649*** [0.00237]	0.00610** [0.00236]
Constant	-0.235*** [0.0644]	-0.218*** [0.0637]	-0.124*** [0.0350]	-0.117*** [0.0348]
Observations	45	45	45	45
R-Squared	0.177	0.159	0.148	0.134

Panel B: Controlling for Retail Rents				
Ln(Population _c)	-0.00648 [0.00421]	-0.00725* [0.00416]	-0.00236 [0.00266]	-0.00279 [0.00263]
Ln(Retail Rent _c)	0.101*** [0.0148]	0.0996*** [0.0146]	0.0450*** [0.00932]	0.0452*** [0.00923]
Constant	-0.216*** [0.0451]	-0.199*** [0.0445]	-0.115*** [0.0284]	-0.108*** [0.0282]
Observations ¹	45	45	45	45
R-Squared	0.608	0.601	0.452	0.449

Standard errors in brackets; *** p<0.01, ** p<0.05, * p<0.1

Notes:

1. Sample is limited to the 45 markets for which commercial retail rent data was available.

takes all of these biases into account, we reverse the conventional finding that tradable goods prices rise with city size. The elasticity of tradable goods prices with respect to city population is -0.11.

These results have important implications for a broad class of New Economic Geography models in which agglomeration is driven, at least in part, by pecuniary externalities. In Krugman (1991)'s model of monopolistic competition, these externalities arise when there are positive trade costs between locations: locally-produced goods are cheaper in the local market than they are elsewhere and larger cities produce more varieties than smaller ones, so consumers in these locations have access to more locally-produced goods and therefore face a lower price index.

More recent papers propose that other mechanisms are at play here. Behrens and Robert-Nicoud (2011) and Combes et al. (2012) demonstrate that, with quadratic linear preferences and fixed trade costs as in Melitz and Ottaviano (2008), three related phenomena yield lower prices in larger cities: (i) more efficient firms are more likely to locate in, or "select into,"

Table 9: Variety-Adjusted Price Indexes and City Size, Controlling for Rents

Panel A									
	Unadjusted Prices			Prices Adjusted for Purchaser Heterogeneity			Prices Adjusted for Purchaser and Store Heterogeneity		
	CEPI _c	VA _c	EPI _c	CEPI _c	VA _c	EPI _c	CEPI _c	VA _c	EPI _c
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]
Ln(Population _c)	0.011*** [0.0038]	-0.014*** [0.0030]	-0.0013 [0.0055]	0.011*** [0.0037]	-0.014*** [0.0030]	-0.0019 [0.0055]	0.0038** [0.0018]	-0.013*** [0.0030]	-0.0093** [0.0036]
Constant	0.83*** [0.056]	1.28*** [0.045]	1.09*** [0.082]	0.84*** [0.055]	1.28*** [0.045]	1.10*** [0.081]	0.94*** [0.026]	1.28*** [0.045]	1.22*** [0.054]
Observations	45	45	45	45	45	45	45	45	45
R-squared	0.172	0.317	0.001	0.160	0.316	0.003	0.094	0.31	0.132

Panel B									
	Unadjusted Prices			Prices Adjusted for Purchaser Heterogeneity			Prices Adjusted for Purchaser and Store Heterogeneity		
	CEPI _c	VA _c	EPI _c	CEPI _c	VA _c	EPI _c	CEPI _c	VA _c	EPI _c
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]
Ln(Population _c)	-0.0044 [0.0039]	-0.015*** [0.0042]	-0.020*** [0.0065]	-0.0046 [0.0038]	-0.015*** [0.0042]	-0.020*** [0.0064]	-0.0023 [0.0024]	-0.015*** [0.0042]	-0.017*** [0.0050]
Ln(Retail Rent _c)	0.078*** [0.013]	0.0072 [0.014]	0.091*** [0.022]	0.077*** [0.013]	0.0072 [0.014]	0.090*** [0.022]	0.033*** [0.0081]	0.0079 [0.014]	0.043** [0.017]
Constant	0.82*** [0.051]	0.85*** [0.042]	1.28*** [0.045]	1.12*** [0.070]	0.86*** [0.041]	1.28*** [0.045]	1.12*** [0.069]	0.94*** [0.026]	1.28*** [0.045]
Observations	45	45	45	45	45	45	45	45	45
R-squared	0.549	0.32	0.292	0.549	0.32	0.291	0.348	0.316	0.228

Standard errors in brackets; *** p<0.01, ** p<0.05, * p<0.1

Notes:

1. The dependent variables in the above regressions are indices. These indices are calculated using unadjusted prices or prices that have been adjusted as indicated above, using estimates from the regression in Column [4] of Panel A in Table 2.
2. $EPI_c = CEPI_c VA_c$ which implies that $\log(EPI_c) = \log(CEPI_c) + \log(VA_c)$. Since the dependent variables in the above regressions are in levels, not logs, the coefficients on log population in the CEPI_c and VA_c regressions do not add to the coefficient on log population in the EPI_c regression.

larger cities; (ii) firms in larger cities can take advantage of agglomeration economies in those locations; and finally, (iii) larger cities are more competitive so firms charge lower markups in these locations. Finally, Hummels et al. (2009) argue that scale economies in the shipping industry lower costs and increase competition on routes to large markets. Similar economies in domestic transportation could also lower prices in large cities.

While separating the contributions of each of these forces to understanding the spatial distribution of prices is beyond the scope of this paper, our results indicate that the New Economic Geography framework is correct to predict that the price level of tradable goods tend to be lower in larger cities.

References

- Albouy, David**, “The Unequal Geographic Burden of Federal Taxation,” *Journal of Political Economy*, 2009, 117 (41), 635–667.
- Anderson, Simon P., Andre de Palma, and Jacques-Francois Thisse**, “The CES is a Discrete Choice Model?,” *Economics Letters*, 1987, 24, 139–140.
- Aten, Bettina**, “Report on Interarea Price Levels,” *Bureau of Economic Analysis Working Papers*, November 2005.
- Atkin, David and Dave Donaldson**, “Who’s Getting Globalized? The Size and Nature of Intranational Trade Costs,” *Yale University and MIT, unpublished mimeo*, 2012.
- B., Figueroa-E. Aten and T. Martin**, “Notes on Estimating the Multi-Year Regional Price Parities by 16 Expenditure Categories: 2005-2009,” *Bureau of Economic Analysis Working Papers*, April 2012, (0071).
- Behrens, Kristian and Frederic Robert-Nicoud**, “Survival of the Fittest in Cities: Agglomeration, Selection, and Inequality,” *CEPR Discussion Papers*, February 2011, (7019).
- Bernard, Andrew B., J. Bradford Jensen, Stephen J. Redding, and Peter K. Schott**, “Firms in International Trade,” *The Journal of Economic Perspectives*, 2007, 21 (3), pp. 105–130.
- Broda, Christian and David E. Weinstein**, “Product Creation and Destruction: Evidence and Price Implications,” *American Economic Review*, 2010, 100, 691.
- , **Nuno Limão, and David E. Weinstein**, “Optimal Tariffs and Market Power: The Evidence,” *The American Economic Review*, 2008, 98 (5), 2032–2065.
- Burstein, Ariel and Gita Gopinath**, “International Prices and Exchange Rates,” *NBER Working Paper 18829*, 2013.
- **and Nir Jaimovich**, “Understanding Movements in Aggregate and Product-Level Real Exchange Rates,” *unpublished paper, UCLA and Stanford University*, 2009.
- Chevalier, Judith A.**, “Capital structure and product-market competition: Empirical evidence from the supermarket industry,” *The American Economic Review*, 1995, p. 415–435.
- Colwell, Robert K. and Jonathan A. Coddington**, “Estimating Terrestrial Biodiversity through Extrapolation,” *Philosophical Transactions: Biological Sciences*, July 1994, 345 (1311), 101–118.

- Combes, Pierre-Philippe, Gilles Duranton, Laurent Gobillon, Diego Puga, and Sébastien Roux**, “The productivity advantages of large cities: Distinguishing agglomeration from firm selection,” *Econometrica*, 2012, 80 (6), 2543–2594.
- Couture, Victor**, “Valuing the Consumption Benefits of Urban Density,” 2013.
- Crucini, Mario J. and Mototsugu Shintani**, “Persistence in law of one price deviations: Evidence from micro-data,” *Journal of Monetary Economics*, 2008, 55 (3), 629–644.
- DuMond, J. Michael, Barry T. Hirsch, and David A. Macpherson**, “Wage Differentials Across Labor Markets and Workers: Does Cost of Living Matter?,” *Economic Inquiry*, October 1999, 37 (4), 577–598.
- Eaton, Jonathan, Samuel Kortum, and Francis Kramarz**, “An anatomy of international trade: Evidence from French firms,” *Econometrica*, 2011, 79 (5), 1453–1498.
- Ellickson, Paul B. and Sanjog Misra**, “Supermarket pricing strategies,” *Marketing Science*, 2008, 27 (5), 811–828.
- Feenstra, Robert C.**, “New Product Varieties and the Measurement of International Prices,” *The American Economic Review*, 1994, 84 (1), 157–177.
- , *Product Variety and the Gains from International Trade*, MIT Press, 2010.
- **and David E. Weinstein**, “Globalization, Markups, and the U.S. Price Level,” *NBER Working Paper 15749*, February 2010, 35.
- , **Robert Inklaar, and Marcel Timmer**, “The Next Generation of the Penn World Table,” 2012.
- Fujita, Masahisa**, “A Monopolistic Competition Model of Spatial Agglomeration,” *Regional Science and Urban Economics*, 1988, 18, 87–124.
- Gopinath, Gita, Pierre-Olivier Gourinchas, Chang-Tai Hsieh, and Nicholas Li**, “International prices, costs, and markup differences,” *The American Economic Review*, 2011, 101 (6), 2450–2486.
- Handbury, Jessie**, “Are Poor Cities Cheap for Everyone? Non-Homotheticity and the Cost of Living Across U.S. Cities,” 2012.
- Helpman, Elhanan**, “The size of regions,” in Itzhak Zilcha David Pines, Efraim Sadka, ed., *Topics in Public Economics*, Cambridge University Press, 1998, pp. 33–54.

- Hummels, David and Peter J. Klenow**, “The Variety and Quality of a Nation’s Exports,” *The American Economic Review*, 2005, 95 (3), 704–723.
- , **Volodymyr Lugovskyy, and Alexandre Skiba**, “The trade reducing effects of market power in international shipping,” *Journal of Development Economics*, 2009, 89 (1), 84–97.
- Jimenez-Valverde, Alberto, Silvia Jimenez Mendoza, Jose Martin Cano, and Miguel L. Munguira**, “Comparing relative model fit of several species-accumulation functions to local Papilionoidea and Hesperioidea butterfly inventories of Mediterranean habitats,” *Biodiversity and Conservation*, 2006, 15, 163–176.
- Krugman, Paul**, “Increasing Returns and Economic Geography,” *Journal of Political Economy*, 1991, 99 (3), 483–499.
- Landry, Anthony**, “Borders and Big Macs,” *Economics Letters*, 2013.
- Mao, Chang Xuan, Robert K. Colwell, and Jing Chang**, “Interpolating, Extrapolating, and Comparing Incidence-Based Species Accumulation Curves,” *Ecology*, 2004, 85 (2717-2727), 433–441.
- , – , and – , “Estimating the Species Accumulation Curve Using Mixtures,” *Biometrics*, 2005, 61 (2), 433–441.
- Melitz, Marc J. and Gianmarco I. P. Ottaviano**, “Market Size, Trade, and Productivity,” *Review of Economic Studies*, January 2008, 75 (1), 295–316.
- Moretti, Enrico**, “Real Wage Inequality,” *American Economic Journal: Applied Economics*, 2013, 5 (1), 65–103.
- Ottaviano, Gianmarco I.P., Takatoshi Tabuchi, and Jacques-Francois Thisse**, “Agglomeration and Trade Revisited,” *International Economic Review*, May 2002, 43 (2), 409–435.
- Parsley, David C. and Shang-Jin Wei**, “Convergence to the Law of One Price Without Trade Barriers or Currency Fluctuations,” *The Quarterly Journal of Economics*, 1996, 111 (4), 1211–1236.
- Rivera-Batiz, Francisco L.**, “Increasing Returns, Monopolistic Competition, and Agglomeration Economies in Consumption and Production,” *Regional Science and Urban Economics*, 1988, 18, 125–153.
- Sato, Kazuo**, “The Ideal Log-Change Index Number,” *The Review of Economics and Statistics*, 1976, 58 (2), 223–228.

Schiff, Nathan, “Cities and product variety,” *University of British Columbia*, 2012.

Simonovska, Ina, “Income Differences and Prices of Tradables,” *UC Davis Working Paper*, 2010, *No. 10-15*.

Suedekum, Jens, “Agglomeration and Regional Costs of Living,” *Journal of Regional Science*, 2006, 46 (3), 529–543.

Tabuchi, Takatoshi, “On Interregional Price Differentials,” *The Japanese Economic Review*, March 2001, 52 (1), 104–115.

Vartia, Y., “Ideal Log-Change Index Numbers,” *Scandinavian Journal of Statistics*, 1976, 3 (3), 121–126.

Appendices

Table A.1: Sample Market Data

Market Name	Population ¹ (thousands)	Sample Household Count ²	Effective Retail Rent ³ (psf)	Normalized Herfindahl Index ⁴
Correlation with Population	0.20	0.54	0.00	0.18
Omaha (Oma)	717	116	11.70	0.116
Jacksonville (Jac)	1,100	130	12.95	0.157
Albany (Alb)	876	138	-	0.139
San Diego (SD)	2,814	139	23.52	0.116
Des Moines (DM)	456	145	-	0.176
Syracuse (Syr)	732	164	10.77	0.078
Milwaukee (Mil)	1,690	170	13.37	0.220
Hartford-New Haven (NH)	1,725	193	14.53	0.138
Richmond (Ric)	997	194	15.04	0.120
Grand Rapids (GR)	1,089	198	-	0.221
Kansas City (KC)	1,776	220	12.10	0.082
Raleigh-Durham (R-D)	1,188	239	15.25	0.113
Indianapolis (Ind)	1,607	245	12.82	0.109
Orlando (Orl)	1,645	247	15.74	0.203
Salt Lake City (SLC)	1,334	248	14.11	0.092
Portland, Or (Por)	2,265	263	16.76	0.110
Pittsburgh (Pit)	2,359	303	14.07	0.143
Cincinnati (Cin)	1,979	317	12.52	0.201
Memphis (Mem)	1,136	331	11.98	0.155
Little Rock (LR)	584	351	11.11	0.193
Cleveland (Cle)	2,946	362	13.88	0.095
Nashville (Nas)	1,231	403	12.99	0.169
New Orleans-Mobile (NO)	1,878	409	-	0.216
Louisville (Lou)	1,026	454	13.46	0.209
Oklahoma City-Tulsa (OkC)	1,887	528	9.83	0.147
Birmingham (Bir)	921	570	12.81	0.154
Dallas (Dal)	5,222	899	14.05	0.119
San Antonio (SA)	1,592	921	12.48	0.305
San Francisco (SF)	7,039	995	29.90	0.152
Detroit (Det)	5,456	1,001	15.66	0.135
Atlanta (Atl)	4,112	1,038	15.31	0.141
Sacramento (Sac)	1,797	1,052	19.84	0.064
Chicago (Chi)	9,158	1,064	16.92	0.127
Houston (Hou)	4,670	1,070	13.63	0.143
Miami (Mia)	3,876	1,086	20.08	0.286
Columbus (Col)	1,540	1,101	10.97	0.171
Charlotte (Cha)	1,499	1,115	14.89	0.132
Washington, DC-Baltimore (DC)	7,608	1,115	21.49	0.070
Denver (Den)	2,582	1,128	14.70	0.163
Buffalo-Rochester (B-R)	2,268	1,135	11.20	0.176
Seattle (Sea)	3,555	1,149	19.78	0.104
Los Angeles (LA)	16,400	1,151	24.47	0.100
Tampa (Tam)	2,396	1,173	13.11	0.198
St. Louis (StL)	2,604	1,183	12.84	0.107
Philadelphia (Phi)	6,188	1,183	17.02	0.066
Phoenix (Pho)	3,252	1,211	16.27	0.129
Minneapolis (Min)	2,969	1,236	14.99	0.165
Boston (Bos)	5,819	1,275	19.03	0.122
New York (NY)	21,200	1,477	20.13	0.088

Sources and notes:

1. 2000 U.S. Census.
2. 2005 Nielsen HomeScan sample.
3. 2005 Community/Neighborhood Shopping Center Effective Rents from reis.com.
4. Herfindahl indexes measure the concentration of retail chains in each market.

Table A.2: Price Indexes

Market Name	Population ¹ (thousands)	Price Indexes ²					
		Nielsen Food Index ³	Unadjusted City Fixed Effect ⁴	CEPI ⁶	Purchaser-Store Adjusted City Fixed Effect ⁴	CEPI ⁶	EPI ⁶
Correlation with Population	1.00	0.49	0.44	0.43	0.42	0.37	-0.19
Coefficient of Variation		6.8%	2.6%	2.2%	1.3%	1.1%	2.0%
Var(Index)/Var(Nielsen Food Index)			0.18	0.13	0.05	0.03	0.12
Des Moines (DM)	717	0.802	0.959	0.968	0.963	0.971	1.039
Little Rock (LR)	1,100	0.872	0.953	0.961	0.969	0.976	1.007
Omaha (Oma)	876	0.924	0.967	0.970	0.973	0.975	1.013
Syracuse (Syr)	2,814	0.973	1.004	1.001	0.996	0.995	1.033
Albany (Alb)	456	0.844	0.927	0.940	0.964	0.974	1.039
Birmingham (Bir)	732	0.847	0.953	0.961	0.967	0.972	1.014
Richmond (Ric)	1,690	0.871	0.969	0.980	0.955	0.965	1.015
Louisville (Lou)	1,725	0.976	0.998	0.998	0.987	0.985	1.019
Grand Rapids (GR)	997	0.849	0.954	0.963	0.976	0.983	1.007
Jacksonville (Jac)	1,089	0.833	0.930	0.938	0.978	0.981	1.016
Memphis (Mem)	1,776	0.840	0.953	0.962	0.975	0.980	1.014
Raleigh-Durham (R-D)	1,188	0.905	0.972	0.981	0.968	0.977	1.003
Nashville (Nas)	1,607	0.836	0.945	0.955	0.964	0.972	1.000
Salt Lake City (SLC)	1,645	0.907	0.961	0.970	0.971	0.975	0.985
Charlotte (Cha)	1,334	0.818	0.928	0.936	0.948	0.957	0.976
Columbus (Col)	2,265	0.834	0.963	0.964	0.978	0.979	1.022
San Antonio (SA)	2,359	0.827	0.973	0.982	0.977	0.982	1.008
Indianapolis (Ind)	1,979	0.844	0.934	0.946	0.969	0.977	0.996
Orlando (Orl)	1,136	0.813	0.946	0.960	0.976	0.987	1.016
Milwaukee (Mil)	584	0.802	0.946	0.958	0.985	0.995	1.019
Hartford-New Haven (NH)	2,946	0.859	0.958	0.966	0.968	0.973	0.988
Kansas City (KC)	1,231	0.806	0.934	0.947	0.965	0.976	0.992
Sacramento (Sac)	1,878	0.877	0.952	0.965	0.980	0.990	1.019
New Orleans-Mobile (NO)	1,026	0.827	0.927	0.938	0.964	0.973	0.993
Oklahoma City-Tulsa (OkC)	1,887	0.830	0.937	0.950	0.964	0.973	0.996
Cincinnati (Cin)	921	0.858	0.948	0.961	0.971	0.982	1.002
Portland, Or (Por)	5,222	0.861	0.945	0.958	0.963	0.974	0.981
Buffalo-Rochester (B-R)	1,592	0.784	0.909	0.923	0.953	0.964	0.982
Pittsburgh (Pit)	7,039	1.031	1.035	1.033	1.014	1.013	1.033
Tampa (Tam)	5,456	0.840	0.958	0.967	0.994	0.999	1.000
Denver (Den)	4,112	0.863	0.943	0.955	0.965	0.974	0.978
St. Louis (StL)	1,797	1.024	1.016	1.011	1.004	1.003	1.014
San Diego (SD)	9,158	0.902	0.968	0.974	0.961	0.966	0.970
Cleveland (Cle)	4,670	0.798	0.925	0.941	0.955	0.966	0.976
Minneapolis (Min)	3,876	0.932	0.981	0.990	0.980	0.986	0.979
Phoenix (Pho)	1,540	0.835	0.941	0.952	0.974	0.980	0.988
Seattle (Sea)	1,499	0.885	0.985	0.996	0.969	0.978	0.984
Miami (Mia)	7,608	0.911	0.965	0.969	0.975	0.978	0.972
Atlanta (Atl)	2,582	0.938	0.973	0.977	0.976	0.981	0.988
Houston (Hou)	2,268	0.859	0.941	0.953	0.965	0.973	0.993
Dallas (Dal)	3,555	0.905	0.987	0.985	0.982	0.983	0.997
Detroit (Det)	16,400	0.952	1.002	1.002	0.991	0.993	1.003
Boston (Bos)	2,396	0.905	0.963	0.972	0.972	0.976	0.963
Philadelphia (Phi)	2,604	0.832	0.952	0.962	0.961	0.969	0.983
San Francisco (SF)	6,188	0.938	0.968	0.971	0.984	0.986	0.979
Washington, DC-Baltimore (DC)	3,252	0.857	0.968	0.971	0.969	0.972	0.941
Chicago (Chi)	2,969	0.900	0.971	0.983	0.969	0.981	0.998
Los Angeles (LA)	5,819	0.937	0.965	0.972	0.977	0.983	0.980
New York (NY)	21,200	1.000	1.000	1.000	1.000	1.000	1.000

Sources and notes:

1. 2000 U.S. Census.
2. All price indexes are relative to the price index for New York.
3. Nielsen food index is replication of ACCRA food index, as described in Section 2.
4. Unadjusted city fixed effects are α_c estimates from equation (2) without purchaser demographic or storetype controls.
5. Purchaser-store adjusted city fixed effects are α_c estimates from equation (2) with purchaser demographic and storetype controls.
6. CEPI and EPI indexes are calculated according to methodology described in Section 5.

A Structural Variety Estimation

In Section 4, we used a non-structural “accumulation curve” methodology to estimate the number of UPCs available in each city. Here we demonstrate that a structural methodology yields very similar results for how the number of varieties varies across cities.

A.1 Methodology

We first consider a simple case in which all UPCs are purchased by households with the same probability in order to develop the intuition, and then move to a more realistic case in which each UPC has its own purchase probability.

In order to work out the math in the simple case assume that each household selects only one UPC out of the S UPCs available in the market. If we also assume that each UPC is purchased with a probability π that is identical for all UPCs in the market, then it follows that $\pi = 1/S$. Our task is to estimate S using the number of different UPCs purchased by a sample of H households. To do this, we make one additional assumption: stores have sufficient inventories of goods so that the purchase of a UPC by one household does not reduce the probability of another household buying the same UPC. If household purchases are independent in the cross-section, then the probability that we observe one of the H households in our sample selecting a particular UPC is equal to one minus the probability that none of the H households selects the UPC, or $1 - (1 - 1/S)^H$. The number of different UPCs that we expect to observe in the purchase records of the H households is simply the sum of these probabilities across all of the S available UPCs,

$$S(H) = S[1 - (1 - 1/S)^H] \quad (\text{A.1})$$

It is immediately apparent that equation (A.1) is an accumulation curve that has been derived from a particular set of assumptions. It would be straightforward to obtain an estimate for the number of varieties in the market in this simple case. By equating $S(H)$ to the sample UPC count, $\tilde{S}(H)$, we can derive an estimate for the number of available UPCs that satisfies equation (A.1). Note that the accumulation curve, $S(H)$ should follow the negative exponential function (which a special case of one of the functional forms we considered in Section 4:

$$S(H) = S \left(1 - e^{-\ln((1-1/S)H)} \right) \quad (\text{A.2})$$

This simple approach cannot be applied to the data for two reasons. First, households purchase more than one UPC in the course of a year. And second, some UPCs, like milk, are likely to be purchased at higher frequencies than other UPCs, like salt. Hence the probability that we observe the purchase of a UPC will vary across UPCs. We can deal with the first problem by

allowing the purchase probability, π , to differ from $1/S$. The second problem, however, is more complicated because solving it requires us to know the purchase frequencies not only of every observed UPC but also of the UPCs that we do not observe in our sample. In order to solve this problem, we follow Mao et al. (2005).

We allow for different products to have different purchase probabilities, but we restrict these probabilities to be identical for groups of UPCs that we will refer to as “incidence groups.” This is not that restrictive an assumption as we do not put a limit on the number of incidence groups but only assume that some UPCs from each incidence group are purchased in our data. Grouping UPCs into this way enables us to estimate the purchase frequencies for all of the UPCs in an incidence group even though we may only observe a fraction of the UPCs in the group being purchased. To put this concretely, suppose that each UPC u has a probability of π_{cu} of being selected by a household in city c , and that there are K different incidence groups, or values that π_{cu} can take in each city c , such that $\pi_{cu} = \pi_{ck}$ for each UPC in incidence group k . We define α_{ck} as the proportion of UPCs in incidence group k in city c , *i.e.*, the proportion of UPCs that are selected with probability π_{ck} in city c . For example, when residents of a particular city choose to purchase a good, the probability that they will purchase any given UPC can take one of K values. The incidence groups do not map directly into the product groups or product modules. The estimated purchase frequency associated with each UPC within a product group or brand-module will vary with the popularity of the UPC, which may be correlated with its brand, container, size, and other characteristics. The UPCs in the high frequency incidence group tend to be the most popular varieties of products that are frequently purchased, *e.g.* 12-packs of some soda varieties, which are purchased by almost a third of the households in our sample. Less popular varieties of soda tend to fall into the lower purchase frequency incidence groups, along with the UPCs in less frequently purchased product categories, *e.g.*, cake yeast.

Suppose that there are S_c UPCs available in city c . If we again assume that household purchases are independent in the cross-section, we can express the probability that any of the UPCs in incidence group k available in city c is purchased by one of H_c households as one minus the probability that none of the H_c households selects the UPC, or $1 - (1 - \pi_{ck})^{H_c}$. Taking the weighted average of these probabilities across incidence groups (using incidence group shares as weights), we can now express the unconditional probability that any one UPC available in city c is observed in the purchases of a sample of H_c households *without knowing its incidence group* as $\sum_{k=1}^K \alpha_{ck}(1 - (1 - \pi_{ck})^{H_c})$. The number of different UPCs that we expect to observe in the purchase records of H_c households in city c is the sum of these probabilities across all of the S_c available UPCs, or

$$S_c(H_c) = S_c \sum_{k=1}^K \alpha_{ck}(1 - (1 - \pi_{ck})^{H_c}). \quad (\text{A.3})$$

Mao et al. (2004) note that equation (A.3) can be re-written as:

$$S_c(H_c) = S_c \sum_{k=1}^K \alpha_{ck} (1 - \exp(-C_{ck} H_c)) \text{ where } C_{ck} = -\ln(1 - \pi_{ck})$$

Equation (A.3) is therefore referred to as the “generalized negative exponential” (GNE) model. It is straightforward to see that equation (A.2) is a specific case of equation (A.3) in which the selection probability is identical for all UPCs, *i.e.*, $K = 1$ and $\pi = 1/S$. Moreover, if we knew the share of all goods in each incidence group, α_{ck} , and the probability that each UPC was purchased, π_{ck} , (A.3) would just give us the equation of a particular accumulation curve for a city.

Fortunately, we can use maximum likelihood estimation to estimate S_c as well as the α_{ck} and π_{ck} for any choice of K . The number of households purchasing product u in city c , h_{cu} , follows a binomial distribution where the probability of observing h_{cu} purchases is given by:

$$P(h_{cu}) = \varphi(h_{cu}; \pi_{cu}) = \binom{H_c}{h_{cu}} (\pi_{cu})^{h_{cu}} (1 - \pi_{cu})^{(H_c - h_{cu})} \quad (\text{A.4})$$

where H_c is the total number of households in the sample for city c ,

$$\binom{H_c}{h_{cu}} \equiv \frac{H_c!}{h_{cu}!(H_c - h_{cu})!}$$

and once again $\pi_{cu} = \pi_{ci}$ for each UPC, u , in category k . Let $\{h_{cu}\}_{u \in U_c}$ be the observed counts of each UPC purchased by our sample of households in a city c , where U_c is the set of UPCs observed in the city c sample. Now, we can define the binomial mixture distribution as follows:

$$\Phi(h_{cu}) = \sum_{k=1}^K \alpha_{ck} \varphi(h_{cu}; \pi_{ck})$$

This distribution tells us the unconditional probability of observing h_{cu} purchases of *any* UPC u in our data, regardless of its incidence group, given the size and purchase probabilities of each of the incidence groups.

Mao et al. (2005) derive a maximum likelihood methodology for estimating the α 's and π 's for a given K using data on the number of samples (in our case, households) in which each variety is observed. The variable n_{cj} is defined to be the number of products that are purchased by j households in the dataset for city c , *i.e.*, for which h_{cu} equals j . In other words, if 100 UPCs are purchased by no households, 50 UPCs are purchased by 1 household, and 25 UPCs are purchased by 2 households, then we would have $n_{c0} = 100$; $n_{c1} = 50$; and $n_{c2} = 25$. The

joint likelihood of the total number of products available in the city, S_c , and the parameters of the mixture distribution is

$$L\left(S_c, \{\alpha_{ck}, \pi_{ck}\}_{k=1}^K\right) = \frac{S_c!}{\prod_{j=0}^{H_c} n_{cj}!} \prod_{j=0}^{H_c} \Phi(j)^{n_{cj}}$$

Note that from equation (A.3), we know that the number of available products, S_c , is a function of the number of observed products, $S_c(H_c) = \sum_{j=1}^{H_c} n_{cj}$, and the parameters of the mixture distribution, $\{\alpha_{ck}, \pi_{ck}\}_{k=1}^K$. Therefore, we only need to estimate the parameters of the mixture distribution to derive an estimate for the number of available products. To do so, we will maximize a conditional likelihood function. Let $\tilde{\varphi}(h_{cu}; \pi_{cu})$ be a zero-truncated binomial density, *i.e.*, the probability that a product is purchased by j households conditional on it being purchased by more than one household, and $\tilde{\Phi}(j) = \sum_{k=1}^K \alpha_{ck} \tilde{\varphi}(j; \pi_{ck})$ be the mixture distribution over these densities for K incidence groups. If we denote the total number of UPCs that are purchased by at least one household in the sample n_{c+} , the conditional likelihood function is

$$L\left(S_c, \{\alpha_{ck}, \pi_{ck}\}_{k=1}^K\right) = \frac{n_{c+}!}{\prod_{j=1}^{H_c} n_{cj}!} \prod_{j=1}^{H_c} \tilde{\Phi}(j)^{n_{cj}} \quad (\text{A.5})$$

Equation (A.5) gives us the likelihood function as a function of the number of incidence groups, K . In order to identify the correct number of these groups, we estimate the parameters of mixture distributions for a range of values for K using the conditional likelihood function. Each distribution implies a different estimate for the total number of UPCs available in a city. We then compute the Akaike Information Criterion (AIC) for each value of K in each city. We choose between the distributions by selecting the number of incidence groups for all cities equal to the K that maximizes the sum of the AICs across all cities. The intuition for this procedure is that each additional incidence group improves the fit but adds parameters, and the AIC provides a way of determining how many groups are likely to exist in the data.³⁴

Once we have our estimates for the α 's and π 's, we can use equation (A.3) to obtain an

³⁴We also assume that the sampling is sufficient so that we observe some UPCs purchased in each incidence group. We need this assumption because we cannot say anything about the number of goods in an incidence group that are purchased with such low a probability that no one in the sample ever buys them. In other words, we can only discuss the number of available varieties for observable classes of goods.

estimate of the total number of varieties as follows:

$$\hat{S}_c = \tilde{S}_c(H_c) \left[\sum_{k=1}^K \hat{\alpha}_{ck} (1 - (1 - \hat{\pi}_{ck})^{H_c}) \right]^{-1} \quad (\text{A.6})$$

where variables with circumflexes represent parameter estimates and $\tilde{S}_c(H_c)$ is equal to the sample count of distinct varieties in the city. It is useful to note that when the number of sampled households in the city, H_c approaches infinity, the fact that the α 's sum to one implies that our count of the number of distinct products purchased by these households becomes our estimate of the number of varieties.

A.2 Results

With 49 cities, the GNE approach involves the estimation of several hundred parameters, so we do not report all the values here. The AIC indicates that UPCs tend to fall within 10 incidence groups in terms of their purchase frequency. Table A.3 summarizes these estimates across our sample of 49 cities. We see that in all cities there are few UPCs that are purchased with very high frequency – on average, one in ten thousand UPCs are purchased with a frequency of 0.5 by a household. This would correspond to about 8 UPCs in the typical city having a purchase probability of 0.5 over the course of a year by a typical household. However, we also see that the vast majority of UPCs have extremely low purchase probability. 49 percent of UPCs have a purchase probability of approximately 1 in a thousand. Thus, the product space is characterized by a few UPCs with high purchase probabilities and a vast number of UPCs that are rarely purchased.

Table A.3: Summary Statistics for GNE Parameter Estimates

Incidence Group (k)	Probability of Purchase (π_{ck})		Share of UPCs (α_{ck})	
	Mean	Standard Deviation	Mean	Standard Deviation
1	0.496	0.096	0.00009	0.00006
2	0.332	0.103	0.00041	0.00029
3	0.226	0.085	0.00116	0.00071
4	0.152	0.066	0.003	0.00181
5	0.102	0.052	0.00757	0.0047
6	0.065	0.035	0.01953	0.01045
7	0.038	0.022	0.05083	0.01821
8	0.019	0.011	0.12115	0.01879
9	0.008	0.004	0.30803	0.03277
10	0.001	0.001	0.48823	0.03465

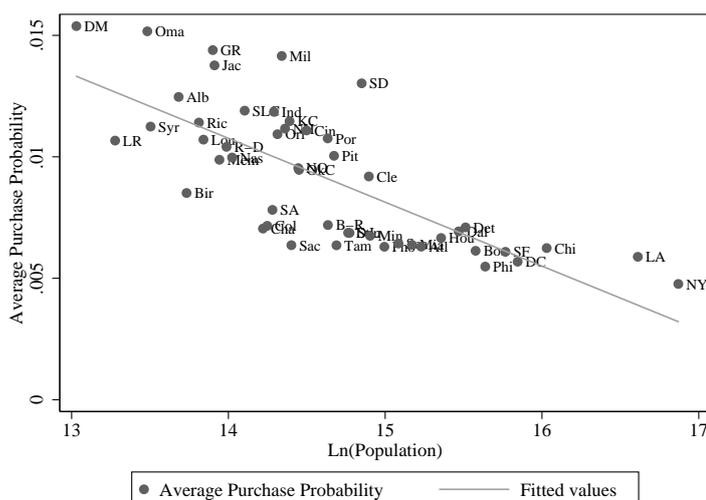
Another way of summarizing the estimates is to examine how the probability that a UPC

is purchased varies with city size. We would expect that the probability that any consumer purchases any one UPC would go down as the range of available UPCs in the city increases. Fortunately, this is easy to examine given that our GNE structure enables us to estimate the average probability that a household in city c purchases a UPC by simply calculating

$$\Pi_c = \sum_{k=1}^K \hat{\alpha}_{c,k} \hat{\pi}_{c,k}.$$

In Figure A.1, we see that the estimated average probability of purchase (which uses no population data in its estimation) decreases sharply with city size. A UPC sold in our smallest city, Des Moines, has three times the probability of being purchased by any individual household as a UPC sold in New York. The fact that households in larger cities are much less likely to buy any individual UPC is strongly suggestive of the fact that the parametric estimate of the UPCs available in a city is increasing with city size. We test this directly by using the GNE parameter

Figure A.1: Average Purchase Probability vs. Log City Population



Notes:

1. Average purchase probability is the average of the estimated probability of a household purchasing a UPC in each incidence group weighted by the estimated share of UPCs in the same incidence group, *i.e.*, Average Purchase Probability in City $c = \Pi_c = \sum_{k=1}^{10} \hat{\alpha}_{c,k} \hat{\pi}_{c,k}$.
2. Acronyms on plots reference the city represented, as listed in Table A.1.

estimates to calculate an estimate for total number of varieties in each city and considering how this varies with city size. Figure A.2a plots how the log estimated number of varieties in each city varies with city size. Visual inspection reveals that the relative relationship between the estimated number of varieties and city size using the Weibull and plotted in Figure A.2b is almost identical to that obtained with the GNE. We can see this even more clearly in Figure A.2c where we plot the Weibull asymptotes against the GNE asymptotes. The correlation between

the two is 0.99 and the slope is 1.07 indicating that both methodologies yield essentially the same relationship between city size and the number of available varieties.

The results in Table A.4 confirm this. The first three columns repeat the regression results reported in columns 4 through 6 of Table 5. The last three columns repeat these regressions with the log GNE asymptote instead of the log Weibull asymptote as the dependent variable. The coefficients on log population are almost identical in all three specifications.

Table A.4: Are prices higher in larger cities?

	Ln(Weibull Asymptote)			Ln(GNE Asymptote)		
	[1]	[2]	[3]	[4]	[5]	[6]
Ln(Population _c)	0.278*** [0.0364]	0.300*** [0.0572]	0.261*** [0.0821]	0.281*** [0.0340]	0.303*** [0.0534]	0.275*** [0.0769]
Ln(Per Capita Income _c)	-	-0.137 [0.288]	-0.060 [0.312]	-	-0.129 [0.269]	-0.074 [0.292]
Income Herfindahl Index	-	-0.630 [2.641]	-0.178 [2.745]	-	-0.321 [2.466]	0.000 [2.568]
Race Herfindahl Index	-	0.074 [0.347]	0.109 [0.353]	-	0.111 [0.324]	0.135 [0.330]
Birthplace Herfindahl Index	-	-0.012 [0.238]	0.004 [0.241]	-	-0.028 [0.222]	-0.017 [0.225]
Ln(Land Area _c)	-	-	0.060 [0.0898]	-	-	0.042 [0.0840]
Constant	6.912*** [0.533]	8.057*** [2.860]	7.240** [3.132]	6.937*** [0.498]	7.962*** [2.670]	7.381** [2.930]
Observations	49	49	49	49	49	49
R-squared	0.55	0.56	0.56	0.59	0.60	0.60

Standard errors in brackets; *** p<0.01, ** p<0.05, * p<0.1.

B Formulas for Sato-Vartia Weights

$$w_{uc} = \frac{\frac{s_{uc}-s_u}{\ln s_{uc}-\ln s_u}}{\sum_{u' \in U_b} \left(\frac{s_{u'c}-s_{u'}}{\ln s_{u'c}-\ln s_{u'}} \right)}, w_{bc} = \frac{\frac{s_{bc}-s_b}{\ln s_{bc}-\ln s_b}}{\sum_{b' \in B_g} \left(\frac{s_{b'c}-s_{b'}}{\ln s_{b'c}-\ln s_{b'}} \right)} \text{ and } w_{gc} = \frac{\frac{s_{gc}-s_g}{\ln s_{gc}-\ln s_g}}{\sum_{g' \in G} \left(\frac{s_{g'c}-s_{g'}}{\ln s_{g'c}-\ln s_{g'}} \right)}$$

s_{uc} , s_{bc} , and s_{gc} are city-specific expenditure shares defined as follows:

$$s_{uc} = \frac{v_{uc}}{\sum_{u' \in U_b} v_{u'c}}, s_{bc} = \frac{\sum_{u \in U_b} v_{uc}}{\sum_{b' \in B_g} \sum_{u' \in U_{b'}} v_{u'c}} \text{ and } s_{gc} = \frac{\sum_{b \in B_g} \sum_{u \in U_b} v_{uc}}{\sum_{g' \in G} \sum_{b' \in B_{g'}} \sum_{u' \in U_{b'}} v_{u'c}}$$

s_u , s_b , and s_g are national expenditure shares defined as follows:

$$s_u = \frac{v_u}{\sum_{u' \in U_b} v_{u'}}, s_b = \frac{\sum_{u \in U_b} v_u}{\sum_{b' \in B_g} \sum_{u' \in U_{b'}} v_{u'}} \text{ and } s_g = \frac{\sum_{b \in B_g} \sum_{u \in U_b} v_u}{\sum_{g' \in G} \sum_{b' \in B_{g'}} \sum_{u' \in U_{b'}} v_{u'}}$$

for $v_u = \sum_c v_{uc}$. G denotes set of product groups, and G_c denotes the set of product groups with products sold in city c . Mathematically, $G_c = \{g \in G | B_g \neq \emptyset\}$. All product groups are sold in every city, so $G_c = G$.

C How much do Index Formulas Matter?

One concern about comparing the ACCRA fixed-weight methodology with that of the CES exact price index is that there are a host of differences that are difficult to formalize. Two of these differences seem to be potentially important. First, the ACCRA index uses purposive sampling (i.e. choosing the most commonly purchased variety) while CEPI is constructed using the full sample of goods. Second the ACCRA index is a fixed-weight index, but the CEPI adjusts for substitution bias by using Sato-Vartia weights.

In Table A.5, we show that neither of these differences are quantitatively important for understanding our results. The first column of Table A.5 replicates the ACCRA methodology using the Nielsen data, as we did in the first column of Table 1, but now using the full set of 49 cities instead of the 47 cities for which the ACCRA data was available. Not surprisingly, the coefficient in the first column of Table A.5 almost exactly matches that in the first column of the earlier table. The second column of Table A.5 maintains the same sample of products used in Column 1 but now recomputes each city's price index weighting relative category unit values with Sato-Vartia weights, rather than ACCRA's fixed weights; essentially using the CEPI formula to compare category prices, rather than identical UPC prices. The fact that the elasticity of prices with respect to population hardly changes means that simply moving to the correct weighting is not an important factor in understanding why the ACCRA price index overstates the relationship between prices and city size.

The importance of product heterogeneity can be easily seen by comparing Columns 2 and 3. In Column 3, we continue to use the Sato-Vartia weights for each good category but instead of comparing the unit value paid for all UPCs in a category (*e.g.*, half gallons of whole milk) we compare the prices of a single UPC within that category across cities. This procedure eliminates the influence played by product heterogeneity in computing the price index, but enables us to continue to see the role played by the ACCRA sampling method. The large drop in the elasticity as we move from Column 2 to Column 3 mirrors the large drop that we saw as we moved from the first column of Table 1 to the first columns of Table 2 and Table 6, which proves that it is product heterogeneity that accounts for the big drop in the association between price and city size.

Columns 4-6 are drawn from Columns 1, 7, and 9 of Table 6. We include them again for heuristic reasons. In particular, we can see that moving whether we use the ACCRA purposive sample of UPCs (which yields the results in Column 3) or the full sample of UPCs to Column 4 hardly affects the results. However, as we saw earlier, adjusting for the other forms of heterogeneity bias and the variety bias ultimately flips the sign of the coefficient.

Table A.5: Transition from the Replication of the ACCRA Index to the CEPI

Index Methodology:	ACCRA Replication			Unadjusted CEPI			Store-Purchaser Adjusted CEPI			Store-Purchaser Adjusted EPI		
	Fixed ACCRA Within Category	Sato-Vartia Within Category	Sato-Vartia Within UPC	Sato-Vartia Within UPC	Sato-Vartia Within UPC	Sato-Vartia Within UPC	Sato-Vartia Within UPC	Sato-Vartia Within UPC	Sato-Vartia Within UPC	Sato-Vartia Within UPC	Sato-Vartia Within UPC	Sato-Vartia Within UPC
- Weights												
- Unit Value Comparisons												
- Sample												
Price Adjustments												
Variety Adjustment												
Ln(Population _c)	[1] 0.0413*** [0.0102]	[2] 0.0467*** [0.0147]	[3] 0.0131*** [0.00368]	[4] 0.012*** [0.0035]	[5] 0.0041** [0.0018]	[6] -0.011*** [0.0036]						
Constant	4.000*** [0.149]	-0.689*** [0.215]	-0.192*** [0.0539]	0.82*** [0.051]	0.94*** [0.027]	1.24*** [0.053]						
Observations	49	49	49	49	49	49						49
R-Squared	0.26	0.18	0.21	0.21	0.10	0.17						0.17

Standard errors in brackets; *** p<0.01, ** p<0.05, * p<0.1