The Sources of Firm Success*

Preliminary and Incomplete
Please do not cite without permission

Colin Hottman
Columbia University†

Stephen J. Redding
Princeton University and NBER‡

David E. Weinstein
Columbia University and NBER§

May 28, 2014

Abstract

This paper estimates a structural model of heterogeneous multiproduct firms and uses the estimated model to decompose the firm size distribution into the contributions of costs, qualities, product scope, and markups. We use detailed Nielsen barcode data to document that virtually all output in industries that use barcodes is produced by multiproduct firms, and these firms exhibit extreme heterogeneity in terms of sales. Taking our model to the data, we find that variation in firm quality and product scope explains at least three quarters of the variation in firm sales depending on the specification with the remainder mostly due to marginal cost variation. Most firms are well approximated by the monopolistic competition benchmark of constant markups, but the largest firms that account for most of aggregate sales departure substantially from this benchmark. Although the output of multiproduct firms is differentiated, cannibalization is quantitatively important for the largest firms. This imperfect substitutability of products within firms, and the fact that larger firms supply more products than smaller firms, implies that standard productivity measures are not independent of demand system assumptions and probably dramatically understate the relative productivity of the largest firms.

JEL CLASSIFICATION: L11, L21, L25, L60

KEYWORDS: firm heterogeneity, multiproduct firms, cannibalization effects

*We are grateful to Andrew Bernard, Swati Dhingra, Cecilia Fieler, Oleg Itskhoki, Bernard Salanié and seminar participants at Dartmouth and Princeton Universities for helpful comments. The usual disclaimer applies. All results are calculated based on data from The Nielsen Company (US), LLC and provided by the Marketing Data Center at The University of Chicago Booth School of Business.

†420 W. 118th Street, MC 3308, New York, NY 10027. Email: cjh2164@columbia.edu.

‡Fisher Hall, Princeton, NJ 08544. Email: reddings@princeton.edu.

§420 W. 118th Street, MC 3308, New York, NY 10027. Email: dew35@columbia.edu.

1 Introduction

Why are some firms larger than others? Some companies, such as the Coca Cola Corporation, generate billions of dollars of sales and dominate the markets in which they operate. Other companies account for only a small fraction of the sales of their larger competitors. What explains these vast differences in firm performance? Answering this question is important for companies seeking to maximize shareholder value, for discriminating between alternative economic models in international trade and macroeconomics, and for understanding the evolution of economic aggregates. Given these extreme differences in firm performance, the largest firms are unlikely to be measure zero relative to their markets, which implies that both firm pricing and product introduction decisions vary systematically with firm size.

Economists tend to point to three reasons for firm heterogeneity: differences in costs, differences in quality, and differences in product scope (*i.e.,* the number of products produced by firms). However, we know little about the relative importance of these three forces. This paper fills this gap by structurally estimating a model of heterogeneous multiproduct firms for the approximately 40,000 firms that produce goods with barcodes in the Nielsen HomeScan Database, and using the estimated model to evaluate the importance of different sources of firm heterogeneity. In standard models of monopolistic competition and constant elasticity of substitution (CES) preferences, quality and marginal cost enter firm revenue isomorphically. We are able to separate quality and marginal cost because we observe price and quantity data for every product supplied by these 40,000 firms. To do so, we use the exclusion restriction that marginal cost only affects firm sales through price. In contrast, quality is a demand shifter that shifts sales conditional on price.[1]

Our results point to quality differences, which we define as average consumer utility per physical unit of output, as being the principal reason why some firms are successful in the marketplace and others are not. Depending on the specification considered, we find that 50-80 percent of the variance in firm size can be attributed to quality differences, about 20 percent to differences in product scope, and only 3-26 percent to cost differences. We estimate that firms in the largest decile in terms of sales produce much higher quality products than firms in the smallest decile, but do not enjoy much of a cost difference. When we turn to examine time-series evidence, the results become even more stark. Virtually all firm growth can be attributed quality-upgrading. Changes in costs and product scope are unimportant sources of variation. These results suggest that most of what economists call differences in revenue productivity reflects differences in quality rather than cost.

Our main contribution is to develop and estimate a structural model of heterogeneous firms that allows firms to be large relative to the market, supply multiple products, and provide an exact decomposition of firm sales. We consider a nested constant elasticity of substitution (CES) utility system that allows the elasticity of substitution between varieties within a firm to differ from the elasticity

[1]Our definition of quality is standard in the literature. For example Sutton (1986) defines "'vertical' product differentiation...has the defining property that if two distinct products are offered at the same price, then all consumers prefer the same one (the higher-quality product)"

of substitution between varieties supplied by different firms. This setup naturally results in variable markups that vary with firm shares, but nests the standard CES-monopolistic competition framework as a special case. Thus, we can compare our results with those that would obtain in a standard trade or economic geography model. Our framework requires only price and expenditure data, and hence is widely applicable. We use these price and expenditure data to estimate the key parameters of the model, namely the elasticities of substitution within and between firms, and the elasticities of marginal cost with respect to output. Given these estimated parameters and the observed data, the model yields an exact decomposition of firm sales into the contributions of overall firm quality, the relative quality of firm products, marginal costs, the number of products a firm produces and the firm's markup. We also use the estimated model to quantify the variation in markups across firms, the strength of the cannibalization of sales of existing products by new products, and the magnitude of fixed costs for products and firms.

Our model also makes clear a conceptual problem in the estimation of firm productivity that is likely to bias existing estimates. Most productivity estimates rely on the concept of real output, which is calculated by dividing nominal output by a price index. However, the formula for any economically motivated price index, which is the same as a unit expenditure function, is dependent on implicit assumptions about how the output of firms enters utility. Thus, one cannot move from nominal output to real output without imposing assumptions about the structure of the demand system.[2] Our results show theoretically and empirically that the CES measure of a multiproduct firm's price is highly sensitive to how differentiated its output is and how many products it supplies. The sensitivity of the CES price index to demand parameters like the elasticity of substitution and whether multiproduct firms exist means that estimates of real output are equally sensitive to implied demand parameters. We show that if demand has a nested CES structure, conventional measures of real output will have a downward bias that rises with firm size. In other words, real output variation is significantly greater than nominal output variation. This bias also implies that true productivity differences are much larger than conventionally measured productivity differences.

The bias is driven by two features of reality that are typically ignored in most analyses. First, most analyses treat producers as single-product firms so they can avoid complications arising from the complexities of measuring the real output of multiproduct firms. However, our results indicate that multiproduct firms are the norm. For example, we document that 67 percent of firms produce more than one barcode and these firms account for more than 99 percent of output in their sectors. By contrast, U.S. Census of Manufactures data indicates that producers of multiple five-digit Standard Industrial Classification (SIC) products account for 37 percent of firms and 87 percent of shipments (Bernard, Redding and Schott 2010). Therefore truly single product firms account for a negligible share of sales in our data. Second, we show that if the output of multiproduct firms is differentiated, the common assumption that total firm output is simply the sum of the output of each variety of

[2] This point was stressed by Aristotle in his *Nicomachean Ethics* (Book V, Section 5): "Demand holds things together as a single unit.... In truth it is impossible that things differing so much should become commensurate, but with reference to demand they may become so." (see http://classics.mit.edu/Aristotle/nicomachaen.5.v.html)

output understates real output for multiproduct firms, and the degree of this downward bias rises in the number of products produced.

Our framework provides a new metric for quantifying the extent to which a firm's products are differentiated from those of its rivals. If a firm's products are perfectly substitutable with each other but not with that of other firms, then 100 percent of a new product's sales will come from the firm's existing sales. However, if a firm's market share is negligible (as it is for most firms) and its products are as differentiated from each other as they are with products produced by other firms, none of a new a new product's sales will come at the expense of its other products, so its cannibalization rate will be zero. We estimate that the cannibalization rate for the typical firm is about 0.45, indicating that although products produced by the same firm are more substitutable with each other than with those of other firms, it is not correct to think of them as perfect substitutes.

We find that the monopolistic competition of atomistic firms with constant markups provides a good approximation for the vast majority of firms. The reason is simple. Most firms have trivial market shares and hence are unable to exploit their market power. However, there is substantial variation in markups for the very largest firms that account for disproportionate shares of aggregate sales. In most sectors, the largest firm has a market share above 20 percent, which enables it to charge a markup that is a third higher than the median firm in the sector. We use the model to undertake counterfactuals, in which we show that these variable markups for the largest firms have quantitatively relevant effects on aggregate welfare and the firm sales distribution.

The remainder of the paper is structured as follows. Section 2 reviews the related literature. Section 3 discusses the data. Section 4 introduces the model. Section 5 uses the structure of the model to derive moment conditions to estimate elasticities of substitution and undertake our decomposition of firm sales. Section 6 presents our estimation results. Section 8 concludes.

2 Related Literature

Over the last decade, the field of international trade has undergone a transformation as the dissemination of micro datasets and the development of new theories has led to a shift in attention towards firm heterogeneity. Existing research has suggested a number of candidate explanations for differences in firm performance, including differences in production efficiency (e.g. Melitz 2003), product quality (e.g. Khandelwal 2011, Johnson 2011, Schott 2004), markups (e.g. De Loecker and Warzynski 2012, De Loecker, Goldberg, Khandelwal and Pavcnik 2013), fixed costs (e.g. Das, Roberts and Tybout 2009), and the ability to supply multiple products (e.g. Arkolakis and Muendler 2010, Bernard, Redding and Schott 2010, 2011, Eckel and Neary 2010, Mayer, Melitz and Ottaviano 2011). While existing research typically focuses on one or more of these candidate explanations, they are all likely to operate to some degree in the data. We develop a general theoretical model that incorporates each of these candidate explanations and estimate the model structurally using disaggregated data on prices and sales by firm and product. We use the estimated model to provide evidence on the

quantitative importance of each source of firm heterogeneity and the ways in which they interact with one another.

In much of the literature on firm heterogeneity and international trade (e.g. Melitz 2003) productivity and product quality are isomorphic. Under the assumption of CES preferences and monopolistic competition, productivity and product quality enter equilibrium firm revenue in exactly the same way. However, these different sources of firm heterogeneity have different implications for firm revenue conditional on prices (e.g. Berry, Levinsohn and Pakes 1994). While productivity only affects firm revenue through prices, product quality is a demand-shifter that shifts firm revenue conditional on prices. An advantage of our approach is that we observe prices and sales in our data, we are able to separate these two sources of dispersion of firm sales.

Much of the existing literature on firm heterogeneity in international trade has assumed that firms are atomistic and compete under conditions of monopolistic competition (e.g. Melitz 2003, Bernard, Redding and Schott 2011). In contrast, a small number of papers have allowed firms to be large relative to the markets in which they operate (e.g. Atkeson and Burstein 2008, Eckel and Neary 2010, Dhingra 2013, and Edmond, Midrigan and Xu 2012). When firm internalize the effects of their decisions on market aggregates, they behave systematically differently from atomistic firms. Even under CES demand, firms charge variable mark-ups, because each firm internalizes the effects of its pricing decisions on market price indices and these effects are greater for larger firms.[3] To the extent that large firms supply multiple products, they take into account that their pricing and product introduction decisions for one variety affect the demand for their other varieties. These cannibalization effects imply that product prices and the range of available varieties differ depending on whether products are supplied together within multiproduct firms or by separate single-product firms. We use our theoretical framework and data to provide quantitative evidence on how important it is to introduce such cannibalization effects into theoretical models of firm behavior.

Most research on firm heterogeneity and international trade has measured products using production classification codes (e.g. around 1,500 five-digit Standard Industrial Classification (SIC) categories in Bernard, Redding and Schott 2010) or trade classification code (e.g. around 10,000 Harmonized System codes in Bernard, Jensen and Schott 2009). In contrast, we measure products at a much finer level of resolution using Universal Product Codes (UPCs) ("barcodes") in scanner data. This measure corresponds closely to the level at which product choice decisions are made by firms, because it is rare for an observable change in product attributes to occur without the introduction of a new UPC.

Our paper is related to the literature on firm heterogeneity following Melitz (2003), as surveyed in Melitz and Redding (2013). Although we consider a nested Constant Elasticity of Substitution (CES) demand system, we allow firms to be of positive measure relative to the market and hence

[3]Our model generates variable mark-ups without a "choke price" above which demand is zero. Therefore this model lies outside the classes considered by Arkolakis, Costinot and Rodriguez-Clare (2012) and Arkolakis, Costinot, Donaldson and Rodriguez-Clare (2012), in which aggregate statistics such as the trade share and trade elasticity are sufficient statistics for the welfare gains from trade.

internalize the effect of their price choices on market price indices, which results in variable markups as in Amiti, Itskhoki and Konings (2012), Atkeson and Burstein (2008), Edmond, Midrigan and Xu (2012). Since firms are of positive measure, idiosyncratic shocks to these "granular" firms can affect aggregate outcomes, as in Gabaix (2011) and di Giovanni and Levchenko (2012). In contrast to these studies, our theoretical framework captures the multiproduct nature of the firms in our data, and we use our structural estimates to quantify the cannibalization effect and isolate different sources of firm heterogeneity.

Our analysis also relates to a recent theoretical research in international trade and macroeconomics on multiproduct firms. Most of this work assumes monopolistic competition and atomistic firms, as in Agur (2010), Allanson and Montagna (2005), Arkolakis and Muendler (2012), Bernard, Redding and Schott (2010, 2011), Mayer, Melitz and Ottaviano (2012) and Nocke and Yeaple (2006). In contrast, Eckel and Neary (2008), Feenstra and Ma (2008) and Dhingra (2013) develop models of multiproduct firms that feature granular firms and cannibalization effects, but these studies do not undertake a structural estimation of a model of multiproduct firms or present evidence on the quantitative importance of cannibalization effects.

Our econometric approach builds on the literature estimating elasticities of substitution and quantifying the contribution of new varieties to welfare following Broda and Weinstein (2004), Feenstra (2004), Sato (1976) and Vartia (1976). We extend this estimation approach to allow for multiproduct firms that are of positive measure relative to the market and we show how this extended approach can be used to recover demand heterogeneity, marginal cost heterogeneity, variable markups, and cannibalization effects. While a number of other studies have used scanner data, including Broda and Weinstein (2010) and Chevalier et al. (2003), so far these data have not been used to estimate a structural model of heterogeneous multiproduct firms.

Our work also relates to the industrial organization literature on discrete choice models of demand with multiproduct firms, including Berry, Levinsohn and Pakes (1995), Goldberg (1995), Nevo (2001), Dube (2004) and the references therein. This industrial organization approach requires detailed information on product characteristics to estimate discrete choice demand models. As a result, this literature is typically required to focus on a single industry and to abstract from general equilibrium interactions between industries through the assumption of an outside alternative. In contrast, our focus is on the sources of firm heterogeneity for economic aggregates across a broad range of industries. Our approach is tailored to this focus and has the advantage of requiring only price and quantity data, which are available for the broad range of industries considered in this study.[4]

[4]Our CES formulation of market demand can be derived from a discrete choice model of the demands of individual consumers, as shown in Anderson, Palma and Thisse (1992). Although demand is CES, firms charge variable mark-ups because they internalize the effect of their price choices on market price indices.

3 Data

Our data source is the Nielsen Homescan database which enables us to observe price and sales information for millions of products with a barcode[5]. Barcode data have a number of advantages for the purpose of our analysis. First, since barcodes are inexpensive but provide sellers access to stores with scanners as well as internet sales, producers have a strong incentive to purchase barcodes for all products that have more than a trivial amount of sales.[6] This feature of the data means that it is likely we observe all products produced by firms. Second, since assigning more than one product to a single barcode can interfere with a store's inventory system and pricing policy, firms have a strong incentive not to reuse barcodes. This second feature of the data ensures that in general identical goods do not have different barcodes. Thus, a barcode is the closest thing we have empirically to the theoretical concept of a good. Finally, since the cutoff size for a firm is to make a sale rather than an arbitrary number of workers, we actually observe something close to the full distribution of firms.

Nielsen collects its barcode data by providing handheld scanners to on average 55,000 households per year to scan each good purchased that has a barcode.[7] Prices are either downloaded from the store in which the good was purchased, or they are hand entered, and the household records any deals used that may affect the price. These households represent a demographically balanced sample of households in 42 cities in the United States. Overall, the database covers around 30 percent of all expenditure on goods in the CPI.[8] We collapse the household dimension in the data and collapse the weekly purchase frequency in order to construct a national quarterly database by UPC on the total value sold, total quantity sold, and average unit value.

Defining products as goods with barcodes has a number of advantages over defining goods by industry classifications. While industry classifications aggregate products produced by a firm within an industry, our data reveals that large firms typically sell hundreds of different products within even a narrowly defined sector. In other words, while prior work equates multiproduct firms with multi-industry firms, our data hews extremely closely to what an economist would call a multiproduct firm. In principle, it could be appropriate to aggregate all output of a firm within an industry into a single good (Is whole milk the same as skim? Is a six-pack of soda the same as a two-liter bottle?) However, as we will show theoretically, the validity of this procedure depends on the elasticity of substitution between the products produced by a firm, which is an empirical question.

Instead of relying on product data for a single industry, we observe virtually the entire universe of goods purchased by households in the sectors that we examine. Our database covers approximately

[5]Our results are calculated based on data from The Nielsen Company (US), LLC and provided by the Marketing Data Center at The University of Chicago Booth School of Business. Information on availability and access to the data is available at http://research.chicagobooth.edu/nielsen

[6]GS1 provides a company with up to 10 barcodes for a $250 initial membership fee and a $50 annual fee. There are deep discounts in the per barcode cost for firms purchasing larger numbers of them (see http://www.gs1us.org/get-started/im-new-to-gs1-us)

[7]The data for 2004 through 2006 come from a sample of 40,000 households, while the data for 2007 through 2011 come from a sample of 60,000 households.

[8]For further discussion of the Nielsen data, see Broda and Weinstein (2010).

Table 1: Sample Statistics

	Avg	Median	Std. Dev.	10th Percentile	90th Percentile	Largest Firm	Largest Barcode
# Firms in Product Group	560.2	455.0	332.0	218.0	1061.0	na	na
Firm sales (thousands)	3860.9	2927.2	24744.4	4.4	4238.6	401509.3	na
Log Firm Sales	11.63	11.58	2.6	8.2	15.0	19.4	na
# UPCs per Firm	12.6	11.7	31.2	1.0	30.7	407.1	na
UPC sales (thousands)	306.9	266.6	1207.9	2.2	623.4	na	36727.5

Note: Calculated based on data from The Nielsen Company (US), LLC and provided by the Marketing Data Center at The University of Chicago Booth School of Business.

1.4 million goods purchased at some point by households in our sample. The data were weighted by Nielsen to correct for sampling error. For example, if the response rate for a particular demographic category is low relative to the census, Nielsen re-weights the averages so that the price paid and the quantity purchased is representative of the United States as a whole.

Nielsen organizes the UPCs into product groups according to where they would likely be stocked in a store. The five largest of our 100 product groups are carbonated beverages, pet food, paper products, bread and baked goods, and tobacco. Output units are common within a product group: typically volume, weight, area, length, or counts. Importantly, we deflate by the number of units in the barcode, so prices are expressed in price per unit (*e.g.,* price per ounce). When the units are in counts, we also deflate by the number of goods in a multipack, so, for instance, we would measure price per battery for batteries sold in multipacks. While about two thirds of these barcoded items correspond to food items, the data also contains significant amounts on information about non-food items like medications, housewares, detergents and electronics. The first six digits of the UPC identifies the manufacturer.

Table 1 presents descriptive statistics for our sample of firms and barcodes in the 100 product groups. We weight the data by the sales of the product group in each year and average across years because sectors like carbonated beverages are much larger economically than sectors like "feminine hygiene." There are on average 560 firms in each product group with 90 percent of the product groups having more than 218 firms. We see enormous range in firm product scope. The median number of products produced by a firm is 12 and the average is 13. On average, 706,000 different UPCs were sold each quarter.

One of the most striking facts displayed in this table is the degree of firm heterogeneity. This is manifest in the skewness of the size and barcode distributions. The largest firm in an industry typically sells almost 137 times more than the median firm. We see similar patterns in terms of product scope and sales per product. The firm with the most products typically has 35 times more products than the firm with the median number of barcodes, and the barcode with the most sales on average generates almost 138 times more revenue than the revenue of the median barcode.

We see in Table 2 that almost 90 percent of sales in a product group was produced by firms with sales in the top decile of sales. Table 3 provides a more detailed description of this firm heterogeneity

Table 2: Size Distribution by Decile

Decile (1 is largest)	Number Firms	Decile's share value (%)	Firm's Avg [Median] # UPCs	Avg Firm Market Share (%)	Avg Log Firm Sales	Avg Firm's Std. Dev. of Log UPC Sales
10	56.5	0.01	1.1 [1.0]	0.0003	7.2	0.54
9	56.0	0.03	1.5 [1.1]	0.001	8.8	0.88
8	56.1	0.08	2.0 [1.4]	0.003	9.8	1.11
7	56.0	0.16	2.7 [2.1]	0.005	10.6	1.23
6	56.0	0.32	3.7 [2.9]	0.01	11.3	1.31
5	56.2	0.61	5.4 [4.2]	0.02	11.9	1.38
4	56.1	1.21	8.3 [6.6]	0.04	12.6	1.46
3	56.0	2.59	13.5 [10.9]	0.1	13.4	1.55
2	56.1	6.83	24.2 [19.5]	0.3	14.4	1.61
1	55.6	88.16	64.8 [42.5]	2.5	16.3	1.83

Note: Calculated based on data from The Nielsen Company (US), LLC and provided by the Marketing Data Center at The University of Chicago Booth School of Business.

Table 3: Size Distribution by Firm Rank

Firm Rank (1 is largest)	Firm's share of value (%)	Log Firm Sales	# UPCs per Firm
10	1.9	17.0	74.4
9	2.1	17.1	86.6
8	2.4	17.2	97.2
7	2.8	17.4	110.5
6	3.3	17.6	110.8
5	4.1	17.8	111.8
4	5.4	18.1	118.8
3	7.8	18.4	155.3
2	12.4	18.8	195.1
1	22.3	19.4	298.5

Note: Calculated based on data from The Nielsen Company (US), LLC and provided by the Marketing Data Center at The University of Chicago Booth School of Business.

by focusing on the ten largest firms in each product group (where we weight the averages by the sales of the product group). Table 3 reveals an almost fractal nature of firm sales. Almost three-fourths of all of the sales of firms in the top decile is produced by the ten largest firms (which on average only account for 18 percent of firms in this decile). While on average half of all output in a product group is produced by just five firms, 98 percent of firms have market shares of less than 2 percent. Thus, the typical sector is characterized by a few large firms and a vast competitive fringe comprised of firms with trivial market shares. A second striking feature of the data is that even the largest firms are not close to being monopolists. The largest firm in a product group on average only has a market share of 22 percent. Finally, the data reveal that firms in the top decile of sales are all multiproduct firms, firms on average producing 65 different goods, and the largest firms producing hundreds of goods.

The extent of multiproduct firms can be seen more clearly in Table 4, which shows the results of splitting the data by the number of UPCs produced by a firm. While single product firms constitute about one third of all firms on average, these firms account for less than one percent of all output.

Table 4: Size Distribution by Number UPCs

Number UPCs	Number Firms	Bin's share value (%)	Avg Sales per UPC (thousands)	Median UPC sales (thousands)	Avg Firm's Std. Dev. of Log UPC Sales
1	184.4	0.9	70.3	15.6	na
2-5	185.5	3.7	315.9	22.9	1.24
6-10	66.6	4.9	1181.7	30.8	1.45
11-20	49.8	8.9	2891.3	41.4	1.50
21-50	45.4	19.2	7818.5	57.7	1.61
51-100	18.4	18.6	23032.2	72.7	1.74
>100	10.1	43.6	122915.6	137.8	1.89

Note: Calculated based on data from The Nielsen Company (US), LLC and provided by the Marketing Data Center at The University of Chicago Booth School of Business.

In other words virtually all output is produced by multiproduct firms. Moreover, the fact that over ninety percent of output is sold by firms selling 11 or more varieties and two-thirds of all output is produced by firms selling more than 50 varieties, suggests that single product firms are more the exception than the rule.

Large firms not only sell more products but they also sell a lot more of each product. The last columns of Table 4 document that while the typical barcode sold by a single product firm only only brings in $15,600 in revenue, the typical barcode sold by a firm selling over one hundred barcodes brings in nine times more. In other words, large firms not only produce more products but they sell more of each product. If firms differed only in the fixed cost of adding new varieties, one would not expect to see large firms sell more of each variety. The fact that they do strongly suggests that large firms must also differ in the marginal cost or quality of the output they produce.

In sum, our overview of the data reveals some key features that we model in our empirical exercise. First, the vast majority of firms have trivial market shares, which means that if we believe that all firms may have some market power, we need to work with demand systems that do not imply that firms with trivial market shares have trivial markups. Second, the fact that most economic output is produced by mulitproduct firms impels us to build this feature directly into the estimation system and allow for both differences in the fixed cost of developing new products as well as cost and quality differences among products. Finally, the fact that there are firms with non-trivial market shares implies that at least some firms are unlikely to behave as if they are measure zero relative to the market as a whole. To the extent that these large firms internalize the effects of their price choices on market aggregates, this concentration of market shares will induce departures from the monopolistic competition benchmark.

4 Theoretical Framework

Our choice of functional forms is motivated not only by the data issues we identified above but also by some theoretical concerns. Since much of the theoretical literature in international trade and economic geography has worked with CES models, we want our results to nest this case (at least

within product groups), so that our results easily can be compared with existing work in trade and regional economics. However, we also need a framework that allows for the elasticity of substitution of products produced by the same firm to be different than that for products produced by different firms, thereby to allow for the possibility of cannibalization effects. Finally, we need a setup that can be applied to firm-level data without imposing implausible assumptions or empirical results. This last requirement rules out two common demand systems: linear demand and the symmetric translog. A linear demand system would be problematic in our setting because one needs to impose the assumption that income elasticities are equal to zero and the estimation of marginal costs derived from a linear demand system can often result in negative values. While the symmetric translog demand system improves on the linear demand system in this regard, it is a difficult system to implement at the firm level because it has the undesirable result that firms with negligible market shares have negligible markups.

Our estimation strategy therefore is based on an upper level Cobb-Douglas demand system across product groups with CES nests below it. The upper-level Cobb-Douglas assumption dramatically simplifies the estimation process because it allows us to assume that firms producing different classes of products, e.g. yogurt and insecticides, do not interact strategically. However, the nested CES structure within broadly defined "product groups" allows for strategic interactions among firms producing similar products. Within this structure the real consumption of each product group is composed of the real consumption of each firm's output, which itself is made up of the consumption of each of the varieties produced by the firm.

4.1 Demand

In order to implement this approach, we assume that utility, U_t, at time t is a Cobb-Douglas aggregate of real consumption of each product group, C_{gt}:

$$U_t = \prod_{g \in G} C_{gt}^{\varphi_{gt}}, \qquad \sum_{g \in G} \varphi_{gt} = 1,$$

where g denotes each product group, φ_{gt} is the share of expenditure on product group g at time t.[9] The consumption indices for product groups and firms take the nested CES form and can be written as follows for tier of utility $j \in \{g, f\}$:

$$C_{jt} = \left[\sum_{k \in K_t} \left(\varphi_{kt} C_{kt} \right)^{\frac{\sigma_K - 1}{\sigma_K}} \right]^{\frac{\sigma_K}{\sigma_K - 1}}, \qquad \sigma_k > 1,\, \varphi_{kt} > 0, \tag{1}$$

where k is the tier of utility below j (i.e.,if j corresponds to firm f, k corresponds to UPCs u, and if j corresponds to a product group g, k corresponds to firms f); K_t is the set of varieties k; C_{kt} denotes consumption of variety k; φ_{kt} is the perceived quality of variety k; σ_K is the constant elasticity of

[9] While we allow the Cobb-Douglas parameters φ_{gt} to change over time, we find that product-group expenditure shares are relatively constant over time, which suggests that a Cobb-Douglas functional form with time-invariant parameters would provide a reasonable approximation to the data.

substitution across varieties k. In other words, the real consumption in any product group, g, is a function of the consumption of each firm's output, C_{ft}, weighted by the quality consumers assign to that firm's physical output at time t, φ_{ft}, and adjusted for the substitutability of the output of each firm, σ_F.[10] Similarly, we can write the sub-utility derived from the consumption of a firm's output, C_{ft}, as a function of the consumption of each UPC (i.e. barcode) produced by that firm, C_{ut}, multiplied by the quality assigned to the physical output of that barcode, φ_{ut}, and adjusted by the substitutability between the various varieties produced by the firm, σ_U.[11]

There are a few features of this specification that are worth noting. First, if the elasticity of substitution across varieties produced by a firm, σ_U, is finite, then the real output of a multiproduct firm is not equal to the sum of the outputs of each product. For example, if the only reason firms differ in size is that larger firms produce more varieties than smaller firms, then assuming firm real output is the sum of the output of each variety will tend to understate the relative size of larger firms. This size bias is a topic that we will explore in much more detail later.

Second, in order to keep things tractable, we will assume that all firms that sell within the same product group, g, have a common elasticity of substitution, σ_F, but that this elasticity can vary across product groups. Similarly, we will assume that the elasticity of substitution among UPCs produced by the same firm, σ_U, does not vary across firms within a product group but does vary across product groups. Third, we would expect that the elasticity of substitution across varieties is larger within firms than across firms, *i.e.*, $\sigma_U \geq \sigma_F$. When the two elasticities are equal, our system will collapse to a standard CES at the product-group level, and if the inequality is strict, we will show that our setup features cannibalization effects.

Third, we allow firms to be large relative to product groups (and hence internalize their effects on the consumption and price index for the product group). But we assume that the number of product groups is sufficiently large that each firm remains small relative to the economy as a whole (and hence takes aggregate expenditure E_t as given). Finally, since the utility function is homogeneous of degree one in quality it is impossible to have a firm-quality, φ_{ft}, that is independent of the quality of the varieties produced by that firm, φ_{ut}. We therefore need to choose a normalization. It will prove convenient to normalize the geometric means of the φ_{ut} for each firm and the φ_{ft} for each product group to equal one:

[10] A small number of firms operate across multiple product groups. We assume that consumers view each of these firm-product-groups as a separate firm and that pricing and UPC introduction decisions are made for each firm-product-group separately. Hence, from now onwards, we refer to firm-product-groups as firms. These assumptions are consistent with product groups being quite distinct from one another (e.g. Carbonated Beverages versus Office Supplies) and the vast majority of firms being active in only one product group.

[11] Our definition of quality is the utility per common physical unit (e.g. utility per ounce of a firm's output). However, variation in quality could either be interpreted as a difference in utility per physical unit or as variation in the number of identical-quality unobservable sub-units within a physical unit. For example, it is isomorphic to say that Firm A produces products with twice the utility per ounce as Firm B and to say that 1/2 an ounce of Firm A's product generates as much utility as an ounce of Firm B's product. In the latter case, an ounce of Firm A's product would contain two "1/2 ounce sub-units" each of which has identical quality to Firm B's product. We think our interpretation of utility per physical unit is the most natural for our data.

$$\left(\prod_{u\in U_{ft}} \varphi_{ut}\right)^{\frac{1}{N_{ft}}} = \left(\prod_{f\in F_{gt}} \varphi_{ft}\right)^{\frac{1}{N_{gt}}} = 1, \tag{2}$$

where U_{ft} is the set of varieties produced by firm f at time t, and N_{ft} is the number of elements of this set. Thus, firm quality (φ_{ft}) corresponds to a demand-shifter that affects all products produced by the firm *proportionately*, while product quality (φ_{ut}) is a demand-shifter that determines the *relative* sales of individual products within the firm.

We can gain some intuition for this framework by considering a simple example. Aggregate utility depends on the share (given by φ_{gt}) and amount of consumption of goods in the "carbonated beverages" product group, (given by C_{gt}). The utility derived from the consumption of carbonated beverages depends on the quality of Coke versus Perrier (φ_{ft}), the amounts of each firm's real output consumed (C_{ft}), and the degree of substitutability between Coke and Perrier (σ_F). Finally, the real amount of Coke or Perrier consumed (C_{ft}) depends on the number of different types of soda produced by each company ($|K|$), the quality of each of these types of soda (φ_{ut}), the consumption of each variety of soda (C_{ut}) and how similar varieties of Coke (or Perrier) products are with other varieties offered by the same company (σ_U).

It turns out that it will also be useful to also work with the exact price index for consumption:

$$P_{jt} = \left[\sum_{k\in K} \left(\frac{P_{kt}}{\varphi_{kt}}\right)^{1-\sigma_k}\right]^{\frac{1}{1-\sigma_k}}. \tag{3}$$

Without loss of generality, we index varieties k in a given year t from the largest in terms of sales to the smallest, and we denote the variety with the largest sales in a given year by $\underline{k}$.

Using the properties of CES demand, the expenditure share of variety k within tier j (S_{kt}) is equal to the elasticity of the price index for tier j with respect to the price of variety k and is given by the following expression:

$$S_{kt} = \frac{\left(P_{kt}/\varphi_{kt}\right)^{1-\sigma_K}}{\sum_{k\in K_t}\left(P_{kt}/\varphi_{kt}\right)^{1-\sigma_K}} = \frac{dP_{jt}}{dP_{kt}}\frac{P_{kt}}{P_{jt}}. \tag{4}$$

Equation (4) makes clear exactly how we conceive of quality in this setup. Holding fixed prices, a good with higher quality will have higher market shares. Similarly, holding fixed prices, firms that produce higher quality goods will have greater market shares.

The role of both firm and product quality can also be seen by writing down the demand for the output of each variety:

$$C_{ut} = \varphi_{ft}^{\sigma_F-1}\varphi_{ut}^{\sigma_U-1}E_{gt}P_{gt}^{\sigma_F-1}P_{ft}^{\sigma_U-\sigma_F}P_{ut}^{-\sigma_U}, \tag{5}$$

where E_{gt} denotes total expenditure on product group g at time t. Equation (5) is critical in determining how we can use this framework for understanding the different roles played by costs and quality for understanding the sales of a firm. While quality has a direct effect on consumer demand independent of price, cost only affects consumer demand through price. Thus, the specification of how cost and quality affect firm pricing decisions is crucial for our identification strategy.

4.2 Technology

We allow the variable costs of production to vary across UPCs, firm-product-groups and firms, which encompasses both heterogeneity in productivity across firms (as in Melitz 2003) and heterogeneity in productivity within firms (as in Bernard, Redding and Schott 2011). All costs are incurred in terms of a composite factor input that is chosen as our numeraire. We assume that the variable cost function is separable across UPCs and that supplying Y_{ut} units of output of UPC u incurs a total variable cost of $A_{ut}\left(Y_{ut}\right) = a_{ut} Y_{ut}^{1+\delta}$, where a_{ut} is a cost shifter and $\delta > 0$ parameterizes the convexity of marginal costs with respect to output. In addition, each firm faces a fixed market entry cost of $H_{ft} > 0$ (e.g. the fixed costs of headquarters operations) and a fixed market entry cost for each UPC supplied of $h_{ft} > 0$ (e.g. the fixed costs of product development and distribution).

4.3 Profit Maximization

In our baseline specification, we assume that firms choose prices under Bertrand competition, though we also report results in which firms instead choose quantities under Cournot competition. Under our assumption of CES preferences, the decisions of any one firm only affect the decisions of other firms through the product group price indices (P_{gt}). Each firm chooses the set of UPCs $u \in \left\{\underline{u}_{ft}, \ldots, \overline{u}_{ft}\right\}$ to produce and their prices $\left\{P_u\right\}$ to maximize its profits:

$$\max_{N_{ft},\left\{P_{ut}\right\}} \Pi_{ft} = \sum_{u=\underline{u}_{ft}}^{\overline{u}_{ft}} \left[P_{ut}Y_{ut} - A_u\left(Y_{ut}\right)\right] - N_{ft}h_{ft} - H_{ft}, \tag{6}$$

where we index the UPCs supplied by the firm from the largest to the smallest in sales, and the total number of goods produced by the firm is denoted by N_{ft}, where $\overline{u}_{ft} = \underline{u}_{ft} + N_{ft}$.

multiproduct firms that are large relative to the market internalize the effect of their decisions for any one variety on the sales of their other varieties. From the first-order conditions for profit maximization, we can derive the firm markup for each UPC, as shown in Appendix A.1:

$$\mu_{ft} = \frac{\varepsilon_{ft}}{\varepsilon_{ft} - 1}, \tag{7}$$

where we define the firm's perceived elasticity of demand as

$$\varepsilon_{ft} = \sigma_F - \left(\sigma_F - 1\right) S_{ft} = \sigma_F\left(1 - S_{ft}\right) + S_{ft}, \tag{8}$$

and the firm's pricing rule as

$$P_{ut} = \mu_{ft}\gamma_{ut}, \qquad \gamma_{ut} = \left(1+\delta\right) a_{ut} Y_{ut}^{\delta}, \tag{9}$$

where γ_{ut} denotes marginal cost.

One of the surprising features of this setup is that *markups only vary at the firm level*.[12] The intuition is that the firm internalizes that it is the monopoly supplier of its real output, which in our model equals real consumption of the firm's bundle of goods, C_{ft}. Hence its profit maximization

[12] Markups would continue to only vary at the firm level if we substituted nested logit demand for nested CES.

problem can be thought of in two stages. First, the firm chooses the price index (P_{ft}) to maximize the profits from supplying real consumption (C_{ft}), which implies a markup at the firm level over the cost of supplying real output. Second, the firm chooses the price of each UPC to minimize the cost of supplying real output (C_{ft}), which requires setting the relative prices of UPCs equal to their relative marginal costs. Together these two results ensure the same markup across all UPCs supplied by the firm.

The firm's perceived elasticity of demand (8) is less than the consumer's elasticity of substitution between firms, σ_F. The reason is that each firm is large relative to the market and hence internalizes the effect of its pricing choices on market price indices. When the firm raises the price of a UPC (P_{ut}) and hence the firm price index (P_{ft}), it reduces demand for that UPC and raises demand for the firm's other UPCs, as captured by the first and second terms on the right-hand side of (5). However, the rise in the firm price index also increases the product-group price index (P_{gt}), which raises demand for each UPC and for the firm as a whole, as captured in the third term on the right-hand side of (5). Finally, our assumption of a Cobb-Douglas functional form for the upper tier of utility together with the assumption that each firm is small relative to the aggregate economy ensures that total expenditure on each product group (E_{gt}) in demand (5) is constant and unaffected by firm decisions.

Although consumers have constant elasticity of substitution preferences (σ_F), each firm perceives a variable elasticity of demand (ε_{ft}) that is decreasing in its expenditure shares (S_{ft}), as in Atkeson and Burstein (2008) and Edmond, Midrigan and Xu (2012). As a result, the firm's equilibrium pricing rule (9) involves a variable markup (μ_{ft}) that is increasing in its expenditure shares (S_{ft}). For a positive equilibrium price (9), we require that the perceived elasticity of demand (ε_{ft}) is greater than one (firms produce substitutes), which requires that the elasticity of substitution between firms (σ_F) is sufficiently large. As a firm's sales becomes small relative to the product group ($S_{ft} \to 0$), the mark-up (7) collapses to the standard constant mark-up of price over marginal cost under monopolistic competition with atomistic multiproduct firms and nested CES preferences (see for example Allanson and Montagna 2005 and Arkolakis and Muendler 2012).

Using the above equilibrium pricing rule, overall UPC profits (Π_{ut}) are equal to UPC variable profits ($\pi_{ut}\left(N_{ft}\right)$) minus fixed costs. UPC variable profits in turn can be written in terms of UPC revenues ($P_{ut}Y_{ut}$), the markup (μ_{ft}), and the elasticity of marginal costs with respect to output (ζ_u):

$$\Pi_{ut} = \pi_{ut}\left(N_{ft}\right) - h_{ft}, \tag{10}$$

$$\pi_{ut}\left(N_{ft}\right) = P_{ut}Y_{ut} - A_u\left(Y_{ut}\right) = \left(\frac{\zeta_u\mu_{ft}-1}{\zeta_u\mu_{ft}}\right)P_{ut}Y_{ut}, \text{ where } \zeta_u = \frac{dA_u\left(Y_{ut}\right)}{dY_{ut}}\frac{Y_{ut}}{A_u\left(Y_{ut}\right)} = 1+\delta. \tag{11}$$

In other words, $\pi_{ut}\left(N_{ft}\right)$ denotes the variable profits from UPC u when the firm supplies N_{ft} UPCs.

4.4 Cannibalization Effects

The number of UPCs produced by each firm, N_{ft}, is determined by the requirement that the increase in profits from introducing an additional UPC, $\overline{u}_{ft}+1$, minus the reduction in profits from reduced

sales of existing UPCs $u \in \left\{\underline{u}_{ft}, \ldots, \overline{u}_{ft}\right\}$ is less than the fixed cost of introducing the new UPC, h_{ft}. If a firm makes N_{ft} products in equilibrium, then it must be the case that if it were to introduce a new good, its profits would fall, *i.e.*,

$$\begin{aligned} \sum_{u=\underline{u}_{ft}}^{\overline{u}_{ft}+1} \pi_{ut}\left(N_{ft}+1\right) - \left(N_{ft}+1\right) h_{ft} \quad &< \quad \sum_{u=\underline{u}_{ft}}^{\overline{u}_{ft}} \pi_{ut}\left(N_{ft}\right) - N_{ft} h_{ft} \\ \Leftrightarrow \pi_{\overline{u}_{ft}+1,t}\left(N_{ft}+1\right) - \sum_{u=\underline{u}_{ft}}^{\overline{u}_{ft}} \left\{\pi_{ut}\left(N_{ft}\right) - \pi_{ut}\left(N_{ft}+1\right)\right\} \quad &< \quad h_{ft}. \end{aligned} \tag{12}$$

In the case where the number of UPCs is large and can be approximated by a continuous variable, we obtain after some manipulation (see Appendix A.2) an expression for the "cannibalization rate":

$$-\frac{\partial Y_{ut}}{\partial N_{ft}} \frac{N_{ft}}{Y_{ut}} \quad = \quad \left[\left(\frac{\sigma_U - \sigma_F}{\sigma_U - 1}\right) + \left(\frac{\sigma_F - 1}{\sigma_U - 1}\right) S_{ft}\right] S_{N_{ft}} N_{ft} > 0, \; for\, \sigma_u \geq \sigma_f > 1. \tag{13}$$

where $S_{N_{ft}}$ is the share of firm revenues from the new product when firm produces N_{ft} products, and S_{ft} is the share of consumer expenditure on firm f. This cannibalization rate is defined as the partial elasticity of the sales of existing products with respect to the number of products. This partial elasticity captures the direct effect of the introduction of a new product on the sales of existing products through firm and product group price indices, holding constant the prices and marginal costs of these existing products.

The first term on the right hand-side captures the cannibalization rate within firms: The introduction of a UPC reduces the firm price index (P_{ft}), which reduces the revenue of existing UPCs if varieties are more substitutable within firms than across firms ($\sigma_U > \sigma_F$). The second term captures the cannibalization rate across firms: The introduction of the new UPC reduces the product-group price index (P_{gt}), which reduces the revenue of existing UPCs from other firms if varieties are more substitutable within product-groups than across product-groups ($\sigma_F > 1$).

There are two useful benchmarks for understanding the magnitude of the cannibalization rate in equation (13). In both cases, it useful to think of the introduction of a "standardized" product that has a market share equal to the average market share of the firms other goods ($S_{N_{ft}} N_{ft} = 1$). If firms are monopolistic competitors, $S_{ft} \approx 0$. Moreover, if we also assume that all products produced by a firm are as differentiated among themselves as they are with the output of other firms, *i.e.*, $\sigma_U = \sigma_F$, then the cannibalization rate will be zero because all sales revenue arising from introducing a new product will come from the sales of goods produced by other firms. Thus, a world with monopolistic competition and equal product differentiation is a world with no cannibalization. Clearly, the cannibalization rate will rise if firms cease being small $S_{ft} > 0$ or if goods produced by the same firm are more substitutable with each other than goods produced by different firms $\sigma_U > \sigma_F$. At the other extreme, we can assume that goods are "perfect substitutes" within firms, *i.e.*, $\sigma_U = \infty > \sigma_f$, so that varieties are differentiated across firms but there is no difference between varieties produced by the same firm. In this case the cannibalization rate will be 1 because any sales of a new product will

be exactly offset by a reduction in the sales of existing products. Thus the cannibalization rate provides a measure of where in the spectrum ranging from perfect substitutes to equal differentiation the output of a firm lies.

4.5 The Sources of Firm Success

In this section, we return to the question posed in the title of the paper and use the model to quantify the contribution of the different sources of firm heterogeneity to the dispersion in sales across firms. Nominal firm sales, E_{ft}, is the sum of sales across UPCs supplied by the firm:

$$E_{ft} \equiv \sum_{u \epsilon U_{ft}} P_{ut} C_{ut}.$$

Using CES demand (5), firm sales can be re-written as:

$$E_{ft} = \varphi_{ft}^{\sigma_F - 1} E_{gt} P_{gt}^{\sigma_F - 1} P_{ft}^{\sigma_U - \sigma_F} \sum_{u \epsilon U_{ft}} \left(\frac{P_{ut}}{\varphi_{ut}}\right)^{1-\sigma_U}. \tag{14}$$

Using the firm price index (3) to substitute for $P_f^{\sigma_U - \sigma_F}$ and taking logarithms, we obtain:

$$\ln E_{ft} = (\sigma_F - 1) \ln \varphi_{ft} + \ln E_{gt} + (\sigma_F - 1) \ln P_{gt} + \left(\frac{1-\sigma_F}{1-\sigma_U}\right) \ln\left(\sum_{u \epsilon U_{ft}} \left(\frac{P_{ut}}{\varphi_{ut}}\right)^{1-\sigma_U}\right),$$

where the final term summarizes the impact of productivity, markups, UPC quality, and number of products on firm sales. This final term can be further decomposed into the contributions of the number of UPCs produced by a firm (N_{ft}), the firm markup (μ_{ft}) and average quality-adjusted marginal costs:

$$\begin{aligned} \ln E_{ft} = &\left\{\ln E_{gt} + (\sigma_F - 1) \ln P_{gt}\right\} \\ &+ \left\{(\sigma_F - 1) \ln \varphi_{ft} + \left(\frac{1-\sigma_F}{1-\sigma_U}\right) \ln N_{ft}\right\} \\ &+ \left\{\left[-(\sigma_F - 1)\, \tilde{\gamma}_{ft}\right] + \left(\frac{1-\sigma_F}{1-\sigma_U}\right) \ln\left(\frac{1}{N_{ft}} \sum_{u \epsilon U_{ft}} \left(\frac{\gamma_{ut}/\tilde{\gamma}_{ft}}{\varphi_{ut}}\right)^{1-\sigma_U}\right)\right\} + (1-\sigma_F) \ln \mu_{ft}, \end{aligned} \tag{15}$$

where

$$\tilde{\gamma}_{ft} = \left(\prod_{u \in U_{ft}} \gamma_{ut}\right)^{\frac{1}{N_{ft}}}$$

and $\gamma_{ut} = (1+\delta)\, a_{ut} Y_{ut}^{\delta}$ denotes marginal cost.

Equation (15) decomposes firm sales into seven terms that capture the various margins through which firms can differ in sales. Clearly these margins are related to one another, since both the number of products and the markup are endogenous to firm quality, product quality and marginal cost. Nonetheless the decomposition isolates the direct effects of firm quality, product quality and marginal cost from their indirect effects through the number of products and the markup.

Although our decomposition is exact for any co-movement in variables, it is easiest to obtain intuition for this equation if we consider one-at-a-time, small movements in each variable so that we can safely ignore interactions between different variables. We therefore will explain the intuition for this equation in terms of small movements in each variable, and note that in general and in our empirical implementation we will allow all variables to move simultaneously.

The seven terms in equation (15) can be grouped into four main elements. The first element, contained in the first set of braces, captures market size and relative pricing. Our demand system is homogeneous of degree one in product group expenditures, so firm sales rise one to one with aggregate expenditures. The second term captures the impact of the general price index in the sector. Holding fixed firm characteristics, an increase in the prices of all other firms will cause a firm's sales to rise by $(\sigma_F - 1)$ percent. Here, the elasticity of substitution between firm's output and the output of other firms, σ_F, plays the crucial role of explaining how much a relative price movement affects firm sales.

Firm quality is captured by the second two terms in braces, which capture the role played by the average quality of output of a firm and the second term captures the role played by product scope (holding fixed average quality). Consider two firms that produce the same number of products, but one firm produces higher quality varieties (measured in utility per unit), meaning that $\varphi_{ft} > \varphi_{f't}$ and $N_{ft} = N_{f't}$. Firm f will then have a higher market share; how much depends on the elasticity of substitution between firm output, σ_F. For a larger value of this elasticity, a given difference in firm quality will translate into a larger difference in firm market share.

Now consider two firms that produce products of identical quality but one firm produces more UPCs than another ($N_{ft} > N_{f't}$). Here, it is easiest to think about this term in a symmetric world in which all goods and firms have identical quality ($\varphi_{ut} = \varphi_{ft} = 1$) and identical marginal cost ($\gamma_{ut} = \gamma_t$), so we can just focus on the role played by product scope. Although firms have identical qualities , they do not have identical market shares because they differ in the number of products they offer, *i.e.*, $\ln N_{ft} > \ln N_{f't}$. For example, if consumers treated all UPCs identically regardless of which firm produced them, *i.e.*, $\sigma_U = \sigma_F$, firm f would sell $\ln\left(N_{ft}/N_{f't}\right)$ percent more output than firm f'. More generally, if the products produced by a firm are more substitutable with each other than with those of other firms, $\sigma_U > \sigma_F$, the percentage gain in sales accruing to a firm that adds a product will be less than one reflecting the fact that a new product will cannibalize the sales of its existing products. Indeed the degree of cannibalization will depend on the magnitude of σ_U; as this elasticity approaches infinity, the cannibalization rate will approach one, and all sales of new products will come from the sales of the firm's existing products. Hence, adding product scope will have no impact on sales.

The terms in third set of braces captures the role played average marginal costs ($\tilde{\gamma}_{ft}$) and quality-adjusted marginal costs. Average marginal costs ($\tilde{\gamma}_{ft}$) capture the standard effect of cost on sales—high cost producers have lower sales in equilibrium. The second term captures the fact that a firm sells more in equilibrium as the *dispersion* in the cost-to-quality ratio across its products. This feature

of the model arises naturally from the concavity of the firm's cost function in the costs of each of each variety it produces. The intuition is straightforward. If $\sigma_U > 1$, *i.e.*, expenditure shares on goods depend on prices, the cost of producing a given level of output (C_{ft}) will fall faster than the average drop in marginal costs because the firm can produce C_{ft} using more of the cheaper input and less of the more expensive one.

Finally, the last term captures the role played by firm markups (μ_{ft}), which are themselves a function of the elasticity of substitution between firms (σ_F) and firm market share (S_{ft}).

4.6 Firm Sales Decompositions

We now apply two different methods to the decomposition (15) to understand cross-sectional and time series variation in firm size. The first approach is based on Eaton, Kortum, and Kramarz's (2004) (EKK) variance decomposition in the international trade literature. We regress each of the components of log firm sales on log firm sales as follows:

$$
\begin{aligned}
(\sigma_F - 1)(\ln \varphi_{ft} - \ln \varphi_{f't}) &= \alpha_\varphi(\ln \Psi_{ft} - \ln \Psi_{f't}) + \varepsilon_\varphi, \quad (16)\\
-(\sigma_F - 1)\left(\ln \tilde{\gamma}_{ft} - \ln \tilde{\gamma}_{f't}\right) &= \alpha_\gamma(\ln \Psi_{ft} - \ln \Psi_{f't}) + \varepsilon_\gamma, \quad (17)\\
\left(\frac{1-\sigma_F}{1-\sigma_U}\right)\left(\ln\left[\frac{1}{N_{ft}}\sum_{u\epsilon U_{ft}}\left(\frac{\gamma_{ut}/\tilde{\gamma}_{ft}}{\varphi_{ut}}\right)^{1-\sigma_U}\right] - \ln\left[\frac{1}{N_{f't}}\sum_{u\epsilon U_{f't}}\left(\frac{\gamma_{ut}/\tilde{\gamma}_{f't}}{\varphi_{ut}}\right)^{1-\sigma_U}\right]\right) &= \alpha_a(\ln \Psi_{ft} - \ln \Psi_{f't}) + \varepsilon_a, \quad (18)\\
\left(\frac{1-\sigma_F}{1-\sigma_U}\right)(\ln N_{ft} - \ln N_{f't}) &= \alpha_N(\ln \Psi_{ft} - \ln \Psi_{f't}) + \varepsilon_N, \quad (19)\\
(1-\sigma_F)(\ln \mu_{ft} - \ln \mu_{f't}) &= \alpha_\mu(\ln \Psi_{ft} - \ln \Psi_{f't}) + \varepsilon_\mu, \quad (20)
\end{aligned}
$$

where each firm's log sales is differenced relative to the largest firm in its product group. By the properties of OLS, $\alpha_\varphi + \alpha_\gamma + \alpha_a + \alpha_N + \alpha_\mu = 1$. The values for each of the α's provide us with a measure of how much of the variance in the distribution of firm sales can be attributed to each factor.

The second approach is based on the Blinder-Oaxaca (BO) quantile decompositions from the labor economics literature. Suppose that there are M_{it} firms in quantile i of the firm sales distribution in time t and denote the mean for that quantile by $\bar{X}_{it} = \frac{1}{M_{it}}\sum_{f\in i} X_{ft}$. Using the linearity of our decomposition (15), we can write average firm sales for that quantile as:

$$
\begin{aligned}
\overline{\ln E}_{it} = {} & (\sigma_F - 1)\overline{\ln \varphi}_{it} - (\sigma_F - 1)\overline{\ln \tilde{\gamma}}_{it} + \ln E_{gt} + (\sigma_F - 1)\ln P_{gt}\\
& + \left(\frac{1-\sigma_F}{1-\sigma_U}\right)\left[\overline{\ln N}_{it} + (1-\sigma_U)\,\overline{\ln \mu}_{it} + \overline{\ln\left(\frac{1}{N_{uft}}\sum_{u\epsilon U_{ft}}\left(\frac{\gamma_{ut}/\tilde{\gamma}_{ft}}{\varphi_{ut}}\right)^{1-\sigma_U}\right)}_{it}\right].
\end{aligned}
$$

We can then decompose the difference in average sales between firms in any two quantiles as:

$$\overline{\ln E_{it}} - \overline{\ln E_{i't}} = (\sigma_F - 1)\left\{\overline{\ln \varphi_{it}} - \overline{\ln \varphi_{i't}}\right\} + \left(\frac{1-\sigma_F}{1-\sigma_U}\right)\left[\overline{\ln N_{it}} - \overline{\ln N_{i't}}\right] \\ -(\sigma_F - 1)\left\{\overline{\ln \tilde{\gamma}_{it}} - \overline{\ln \tilde{\gamma}_{i't}}\right\} + (1-\sigma_F)\left(\overline{\ln \mu_{it}} - \overline{\ln \mu_{i't}}\right) \\ + \left(\frac{1-\sigma_F}{1-\sigma_U}\right)\left[\overline{\ln\left(\frac{1}{N_{uft}}\sum_{u\epsilon U_{ft}}\left(\frac{\gamma_{ut}/\tilde{\gamma}_{ft}}{\varphi_{ut}}\right)^{1-\sigma_U}\right)_{it}} - \overline{\ln\left(\frac{1}{N_{uft}}\sum_{u\epsilon U_{ft}}\left(\frac{\gamma_{ut}/\tilde{\gamma}_{ft}}{\varphi_{ut}}\right)^{1-\sigma_U}\right)_{i't}}\right]. \tag{21}$$

The five terms in equation (21) tell us the importance of firm quality, firm scope, markups, and adjusted marginal costs in understanding differences in sales between quantiles of the firm sales distribution. We can also undertake a similar decomposition of differences in sales between a pair of firms in a given time period and between a pair of time periods for a given firm by using equation (15) to obtain an expression the difference in log sales between two firms f and f' or for the change in log sales of a given firm between two time periods t and t'.

4.7 Decomposing Changes in Firm Quality

Having shown the role of firm quality in accounting for differences in firm size, we now further decompose differences in firm quality. We focus our exposition on decomposing changes in firm quality within firms over time. To understand the sources of variation in firm quality, note that our normalization for UPC quality implies that the geometric mean of the φ_{ut} equals one for each firm in each period, *i.e.*,

$$\varphi_{ft} = \varphi_{ft}\left(\prod_{u\in U_{ft}} \varphi_{ut}\right)^{\frac{1}{N_{ft}}} \tag{22}$$

We can then can take log differences of this equation over time to yield:

$$\Delta \ln \varphi_{ft} \equiv \ln \varphi_{ft} - \ln \varphi_{f,t-1} = \Delta \ln \varphi_{ft} + \frac{1}{N_{ft}}\sum_{u\in U_{ft}} \ln \varphi_{ut} - \frac{1}{N_{f,t-1}}\sum_{u\in U_{f,t-1}} \ln \varphi_{u,t-1}, \tag{23}$$

where our normalization (2) implies that the final two terms are both equal to zero.

To isolate the sources of changes in firm quality, we distinguish between UPCs that are supplied in both periods versus those that are supplied in only one of the two periods. Let $I_{ft} = U_{ft} \cap U_{f,t-1}$ denote the set of UPC's that are produced by firm f in both periods t and $t-1$. Similarly, define U_{ft}^{+} to be the set of newly introduced UPCs, *i.e.*, the set of UPCs in U_{ft} but not in $U_{f,t-1}$, and U_{ft}^{-} to be the set of disappearing UPCs, *i.e.*, the set of UPCs in $U_{f,t-1}$ but not in U_{ft}.

Noting that $\left|U_{ft}\right| = \left|I_{ft}\right| + \left|U_{ft}^{+}\right|$ and $\left|U_{ft-1}\right| = \left|I_{ft}\right| + \left|U_{ft}^{-}\right|$, we can decompose the change in overall firm quality into quality changes for a constant set of products and quality changes from the adding and dropping of UPCs:

$$\Delta \ln \varphi_{ft} = \left\{ \frac{|I_{ft}|}{|U_{ft}|} \left[\frac{1}{|I_{ft}|} \sum_{u \in I_{ft}} \ln \varphi_{ut} \right] - \frac{|I_{ft}|}{|U_{f,t-1}|} \left[\frac{1}{|I_{ft}|} \sum_{u \in I_{ft}} \ln \varphi_{u,t-1} \right] + \Delta \ln \varphi_{ft} \right\} \tag{24}$$
$$+ \left\{ \frac{\left|U_{ft}^{+}\right|}{|U_{ft}|} \left[\frac{1}{\left|U_{ft}^{+}\right|} \sum_{u \in U_{ft}^{+}} \ln \varphi_{ut} \right] - \frac{\left|U_{ft}^{-}\right|}{|U_{f,t-1}|} \left[\frac{1}{\left|U_{ft}^{-}\right|} \sum_{u \in U_{ft}^{-}} \ln \varphi_{u,t-1} \right] \right\}.$$

The first term in braces captures quality changes for a constant set of products, and includes the change in firm quality ($\Delta \ln \varphi_{ft}$), as well as the difference between average product quality for the common set of products (in square brackets) weighted by their relative importance in the overall set of products in each year. This difference adjusts changes in our measure of firm quality for changes in our normalization due to the entry and exit of new goods. We refer to the first term in braces in equation (24) as the "marketing effect". It represents the change in consumers perceived quality of the UPCs present in both periods. If the firm does not add or eliminate any products, the summation terms will be zero since our normalization (2) means that average log quality must be zero in all periods.

However, if the set of products changes, the summation terms will not necessarily be zero. Product turnover can now influence sales either through product upgrading, which is measured in the second term in braces in equation (24), as well as through changes in the number of products, which enters separately into our decomposition of log firms sales (15) above. The second term in braces captures the "product upgrading effect," *i.e.,* quality changes arising from the adding and dropping of products. Product upgrading depends on average product quality for the entering and exiting products (in square brackets) weighted by their relative importance in the overall set of products in each year. For example, suppose a firm introduces a new high-quality product and retires a low-quality product leaving the number of products unchanged. In this case, the second term in braces will be positive reflecting the fact that the firm upgraded the average quality of its product mix.

5 Structural Estimation

Our structural estimation of the model has two components. First, given data on expenditure shares and prices $\{S_{ut}, S_{ft}, P_{ut}\}$ and known values of the elasticities of substitution $\{\sigma_U, \sigma_F\}$, we show how the model can be used to determine unique values of firm quality (φ_{ft}), product quality (φ_{ut}), and marginal cost shocks (a_{ut}) up to our normalization of quality. These correspond to structural residuals of the model that are functions of the observed data and parameters and ensure that the model exactly replicates the observed data. We use these structural residuals to implement our decomposition of firm sales from Section 4.5 with our observed data on expenditure shares and prices. Second, we estimate the elasticities of substitution$\{\sigma_U, \sigma_F\}$ using a generalization of Feenstra (1994) and Broda and Weinstein (2006) to allow allow firms to be large relative to the markets in which they operate (which introduces variable markups) and to incorporate multiproduct firms (so that firm

pricing decisions are made jointly for all varieties). This estimation uses moment conditions in the double-differenced values of the structural residuals $\{\varphi_{ft}, \varphi_{ut}, a_{ut}\}$ and also has a recursive structure. In a first step, we first estimate the elasticity of substitution across UPCs within firms for each product group $\{\sigma_U\}$. In a second step, we use these estimates for UPCs to estimate the elasticity of substitution across firms for each product group $\{\sigma_F\}$.

5.1 Structural Residuals

We begin by showing that there is a one-to-one mapping from the observed data on expenditure shares and prices $\{S_{ut}, S_{ft}, P_{ut}\}$ and the model's parameters $\{\sigma_U, \sigma_F, \delta\}$ to the unobserved structural residuals $\{\varphi_{ut}, \varphi_{ft}, a_{ut}\}$.

Given known values for the model's parameters $\{\sigma_U, \sigma_F, \delta\}$ and the observed UPC expenditure shares and prices, we can use the expression for the expenditure share given in equation (4) to determine UPC qualities $\{\varphi_{ut}\}$ up to our normalization that the geometric mean of UPC qualities is equal to one. These solutions for UPC qualities and observed UPC prices can be substituted into the CES price index (3) to compute firm price indices $\{P_{ft}\}$. These solutions for firm price indices and observed firm expenditure shares can be combined with the CES expenditure share (4) to determine firm qualities $\{\varphi_{ft}\}$ up to our normalization that the geometric mean of firm qualities is equal to one. Furthermore observed firm expenditure shares and the CES markup (7) are sufficient to recover firm markups $\{\mu_{ft}\}$. These solutions for markups and observed UPC prices and expenditures can be substituted into the CES pricing rule (9) to determine the marginal cost shock (a_{ut}).

Finally, our solutions for markups and observed UPC expenditures can be combined with CES variable profits (11) to obtain upper upper bounds to the fixed costs of supplying UPCs (h_{ft}) and the fixed costs of operating a firm (H_{ft}). The upper bound for UPC fixed costs for each product group (h_{ut}) is defined by the requirement that variable profits for the least profitable UPC within a product group must be greater than this fixed cost. Similarly, the upper bound for firm fixed costs for each product group (h_{ft}) is defined by the requirement that variable profits for the least profitable firm must be greater than this fixed cost.

This mapping from the observed data on expenditure shares and prices $\{S_{ut}, S_{ft}, P_{ut}\}$ and the model's parameters $\{\sigma_U, \sigma_F, \delta\}$ to the unobserved structural residuals $\{\varphi_{ut}, \varphi_{ft}, a_{ut}\}$ does not impose assumptions about the functional forms of the distributions for the structural residuals or about their correlation with one another. When we estimate the model's parameters $\{\sigma_U, \sigma_F, \delta\}$ below, we impose some identifying assumptions on the double-differenced values of these structural results but not upon their levels. Therefore, having recovered these structural residuals, we can examine the functional form of their distributions and their correlation with one another.

5.2 UPC Moment Conditions

We now discuss our methodology for estimating the elasticities of substitution $\{\sigma_U, \sigma_F\}$, which uses moment conditions in double-differenced values of the structural residuals $\{\varphi_{ft}, \varphi_{ut}, a_{ut}\}$. This esti-

mation again has a recursive structure. In a first step, we first estimate the elasticity of substitution across UPCs within firms for each product group $\{\sigma_U\}$. In a second step, we use these estimates for UPCs to estimate the elasticity of substitution across firms for each product group $\{\sigma_F\}$.

In the first step, we double-difference log UPC expenditure shares (4) over time and relative to the largest UPC within each firm to obtain the following equation for relative UPC demand:

$$\triangle^{k,t} \ln S_{ut} = (1 - \sigma_U)\triangle^{k,t} \ln p_{ut} + \omega_{ut}, \tag{25}$$

where$\triangle^{k,t}$ is the double-difference operator such that $\triangle^{k,t} \ln S_{ut} = \triangle \ln S_{ut} - \triangle \ln S_{kt}$; u is a UPC produced by the firm; k corresponds to to the largest UPC produced by the same firm (as measured by the sum of expenditure across the two years); S_{ut} and p_{ut} are directly observed in our data; $\omega_{ut} = (1 - \sigma_U) [\triangle \ln \varphi_{kt} - \triangle \ln \varphi_{ut}]$ is a stochastic error. Since we double difference the market shares of two UPCs produced by the same firm, we eliminate all demand shocks that are common across a firm's UPCs, which leaves only demand shocks that affect the sales of one a firm's UPCs relative to another. For example, random demand shocks and the timing of holidays, weekends, and meteorological events might affect the success of certain products in a firm's lineup relative to others.

Double-differencing the UPC pricing rule (9) enables us to obtain an equation for relative UPC supply. Using the cost function ($A_u (Y_u) = a_{ut} Y_{ut}^{1+\delta}$), and the relationship between output and revenue ($Y_{ut} = S_{ut}/P_{ut}$), the UPC pricing rule given in equation (9) can be re-written as:

$$P_{ut} = \mu_{ft}\gamma_{ut} = \mu_{ft}^{\frac{1}{1+\delta}} (1 + \delta)^{\frac{1}{1+\delta}} a_{ut}^{\frac{1}{1+\delta}} S_{ut}^{\frac{\delta}{1+\delta}}.$$

Taking double-differences, we obtain the following equation for relative UPC supply:

$$\triangle^{k,t} \ln p_{ut} = \frac{\delta}{1 + \delta} \triangle^{k,t} \ln S_{ut} + \kappa_{ut}, \tag{26}$$

where the markup μ_f has differenced out because it is the same across UPCs within the firm; p_{ut} is again directly observed in our data; $\kappa_{ut} = \frac{1}{1+\delta} [\triangle \ln a_{ut} - \triangle \ln a_{\underline{u}t}]$ is a stochastic error. Once again the double differencing within a firm eliminates shocks that are common across a firm's products, and the fact that a barcode uniquely identifies a product means that changes in product quality will be manifest in a change of barcode but will not be manifest in a change in a_{ut}. Therefore the supply-side shocks κ_{ut} correspond to factors that affect the plants or assembly lines that affect one variety produced by firm but not another. These might be problems in individual plants, exchange rate movements in the countries supplying those products, etc. We assume that these idiosyncratic shocks to supply, κ_{ut}, are orthogonal to to the idiosyncratic demand shocks, ω_{ut}.

Following Broda and Weinstein (2006), the orthogonality of idiosyncratic demand and supply shocks defines a set of moment conditions (one for each UPC):

$$G(\beta_g) = \mathbb{E}_{\mathbb{T}} \left[v_{ut}(\beta_g) \right] = 0, \tag{27}$$

where $\beta_g = \begin{pmatrix} \sigma_U \\ \delta \end{pmatrix}$ and $v_{ut} = \omega_{ut}\kappa_{ut}$. For each product group, we stack all the moment conditions to form the GMM objective function and obtain:

$$\hat{\beta}_g = \arg\min_{\beta_g} \left\{ G^*(\beta_g)' W G^*(\beta_g) \right\} \qquad \forall g, \tag{28}$$

where $G^*(\beta_g)$ is the sample analog of $G(\beta_g)$ stacked over all UPCs in a product group and W is a positive definite weighting matrix. As in Broda and Weinstein (2006), we weight the data for each UPC by the number of raw buyers for that UPC to ensure that our objective function is more sensitive to UPCs purchased by larger numbers of consumers.

The moment condition (27) for each UPC involves the expectation of the product of the double-differenced demand and supply shocks: $v = \omega_{ut}\kappa_{ut}$. From relative demand (25) and relative supply (26), this expectation depends on the variance of prices, the variance of expenditure shares, the covariance of prices and expenditure shares, and parameters. Our identifying assumption that this expectation is equal to zero defines a rectangular hyperpola in (σ_U, δ) space for each UPC, along which a higher value of σ_U has to be offset by a lower value of δ in order for the expectation to be equal to zero (Leontief 1929). Therefore, this rectangular hyperbola places bounds on the demand and supply elasticities for each UPC, even in the absence of instruments for demand and supply. Furthermore, if the variances for the double-differenced demand and supply shocks are heteroscedastic across UPCs, the rectangular hyperbolas are different for each pair of UPCs, and their intersection can be used to separately identify the demand and supply elasticities (Feenstra 1994). Consistent with these identifying assumptions in Broda and Weinstein (2006, 2010) and Feenstra (1994), we find evidence that the double-differenced demand and supply shocks are indeed heteroscedastic.[13]

5.3 Firm Moment Conditions

We use our estimates of the UPC elasticities of substitution $\{\sigma_U\}$ from the first step to solve for product quality $\{\varphi_{ut}\}$ and compute the firm price indices $\{P_{ft}\}$ using equation (4). In the second step, we double difference log firm expenditure shares over time and relative to the largest firm within each product-group, $\underline{f}$, to obtain the following equation for relative firm market share:

$$\Delta^{\underline{f},t} \ln S_{ft} = (1 - \sigma_F)\, \Delta^{\underline{f},t} \ln P_{ft} + \omega_{ft}, \tag{29}$$

where the stochastic error can be written as $\omega_{ft} \equiv -(\sigma_F - 1)\, \Delta^{\underline{f},t} \ln \varphi_{ft}$.

Estimating equation (29) using ordinary least squares would be problematic, because changes in firm price indices could be correlated with changes in firm quality: $\mathrm{Cov}\left(\Delta^{\underline{f},t} \ln \widetilde{P}_{ft}, \Delta^{\underline{f},t} \ln \varphi_{ft}\right) \neq 0$,. To find a suitable instrument for changes in firm price indices, we use the structure of the model to write changes in firm price indices in terms of the underlying UPC characteristics of the firm. Using the CES expenditure shares (4), we can write relative UPC expenditures in terms of relative UPC prices and relative UPC qualities:

$$\frac{S_{ut}}{\tilde{S}_{ft}} = \left(\frac{P_{ut}/\varphi_{ut}}{\tilde{P}_{ft}/\tilde{\varphi}_{ft}} \right)^{1-\sigma_U}, \qquad u \in U_f, \tag{30}$$

[13] In a White test for heteroscedasticity, we are able to reject the null hypothesis of homoscedasticity at conventional significance levels for 92 percent of product groups.

where here we compare each UPC to the geometric mean of UPCs within the firm, which is denoted by a tilde so that $\tilde{S}_{ft} = \exp\left\{\frac{1}{N_{uft}}\sum_{u\in U_{ft}}\ln S_{ut}\right\}$. Using this expression for relative expenditure shares to substitute for product quality (φ_{ut}) in the CES price index (3), we can write the firm price index solely in terms of observed relative expenditures and the geometric mean of UPC prices:

$$\ln P_{ft} = \ln \tilde{P}_{ft} + \frac{1}{1-\sigma_U}\ln\left[\sum_{u\in U_{ft}}\frac{S_{ut}}{\tilde{S}_{ft}}\right], \tag{31}$$

where we have used our normalization that $\tilde{\varphi}_{ft} = 1$.

Equation (32) decomposes the price level into two terms. The first term is entirely conventional: the geometric mean of the prices of all goods produced by the firm. When researchers approximate this using firm-level unit values or the prices of representative goods produced by firms they are essentially capturing this component of the firm price index. The second term is novel and arises because the conventional price indexes are not appropriate for the measurement of multiproduct firms. The log component of the second term is a variant of the Theil index of dispersion.[14] If the shares of all products are equal, the Theil index will equal $\ln N_{ft}$, which is increasing the number of products produced by the firm, N_{ft}. Since the Theil index is multiplied by $(1-\sigma_U)^{-1} \leq 0$, the firm's price index falls as the number of goods produced by the firm rises. This Theil index also increases as the dispersion of the market shares of each good it produces rises.

One can obtain some intuition for this formula by comparing it to the conventional price indexes commonly used in economics. Firm price indexes are typically constructed by making at least one of two critical assumptions—firms only produce one product ($N_{ft} = 1$), or the goods produced by firms are perfect substitutes (*i.e.*, $\sigma_U = \infty$). Either of these assumptions is sufficient to guarantee that the firm's price level equals its average price level. However, both of these assumptions are likely violated in reality. The average price of a firm's output prices overstates the price level for a multiproduct firms because consumers derive more utility per dollar spent on a firm's production if that production bundle contains more products. *Thus, although using (geometric) average firm prices is a theoretically rigorous way to measure the price level of single product firms, it overstates the prices of multiproduct firms, and this bias will tend to rise as the number of products produced by the firm increases.*

The structure of the model implies that the dispersion of the shares of UPCs in firm expenditure (S_{ut}) only affects the shares of firms in product group expenditure (S_{ft}) through the firm price indices (P_{ft}). Double-differencing equation (31) for the log firm price index over time and relative to the largest firm within each product-group, we obtain:

$$\Delta^{\underline{f},t}\ln P_{ft} = \Delta^{\underline{f},t}\ln \widetilde{P}_{ft} + \frac{1}{1-\sigma_U}\Delta^{\underline{f},t}\ln\left[\sum_{u\in U_{ft}}\frac{S_{ut}}{\widetilde{S}_{ft}}\right], \tag{32}$$

where the model implies that the second term on the right-hand side in the shares of UPCs in firm expenditure is a valid instrument for the double-differenced firm price index in (29). We find that

[14] The standard Theil index uses shares relative to simple average shares, while ours expresses shares relative to the geometric mean.

Table 5: Distribution of 100 GMM Estimates

Percentile	σ_u	σ_f	$\sigma_u - \sigma_f$	δ
1	4.41	2.25	0.18	0.02
5	4.74	2.51	1.27	0.04
10	5.03	2.83	1.53	0.06
25	5.4	3.34	1.87	0.09
50	6.93	4.27	2.4	0.15
75	8.5	5.41	3.57	0.20
90	13.73	7.55	6.09	0.31
95	17.64	8.98	9.32	0.43
99	33.79	14.66	28.21	0.62

Note: Calculated based on data from The Nielsen Company (US), LLC and provided by the Marketing Data Center at The University of Chicago Booth School of Business.

this instrument is powerful in the first-stage regression (32), with a first-stage F-statistic for the statistical significance of the excluded exogenous variable that is substantially above the recommended threshold of 10 from Stock, Wright and Yogo (2002).

6 Estimation Results

We present our results in several stages. First, we present our elasticity and quality estimates and show that they are reasonable. Second we use these estimates to examine cannibalization, and finally we present our results on the sources of firm heterogeneity.

6.1 Estimated Elasticities of Substitution

Because we estimate 100 σ_U's and σ_F's, it would needlessly clutter the paper to present all of them individually. Table 5 shows goods produced by the same firm are imperfect substitutes. For UPCs, the estimated elasticity of substitution ranges from 4.7 at the 5th percentile to 17.6 at the 95th percentile with a median elasticity of 6.93. These numbers are large compared with trade elasticities reflecting the fact that products produced by the same firm are closer substitutes than products produced by different firms. The median elasticity implies that a one percent price cut causes the market share of that UPC to rise by 5.9 percent.

The fact that we estimate different products produced by a given firm to be imperfect substitutes for each other has profound implications for how we should understand firm pricing and productivity. All productivity estimates are based on a concept of real output, which, in modern index number theory, equals nominal output divided by the minimum expenditure necessary to generate a unit of utility. According to equation (1) the quality-adjusted flow of consumption from a firm's output is $\varphi_{ft}C_{ft}$, so the corresponding expenditure function for a unit ofquality-adjusted consumption is P_{ft}/φ_{ft}. However, for multiproduct firms, the formula for P_{ft} depends on the demand system, and therefore the formula for a firm's real output also depends on the demand system. *In other words,*

for multiproduct firms, the concept of real output is not independent of the demand system, so all attempts to measure productivity based on a real output concept contain an implicit assumption about the structure of the demand system.

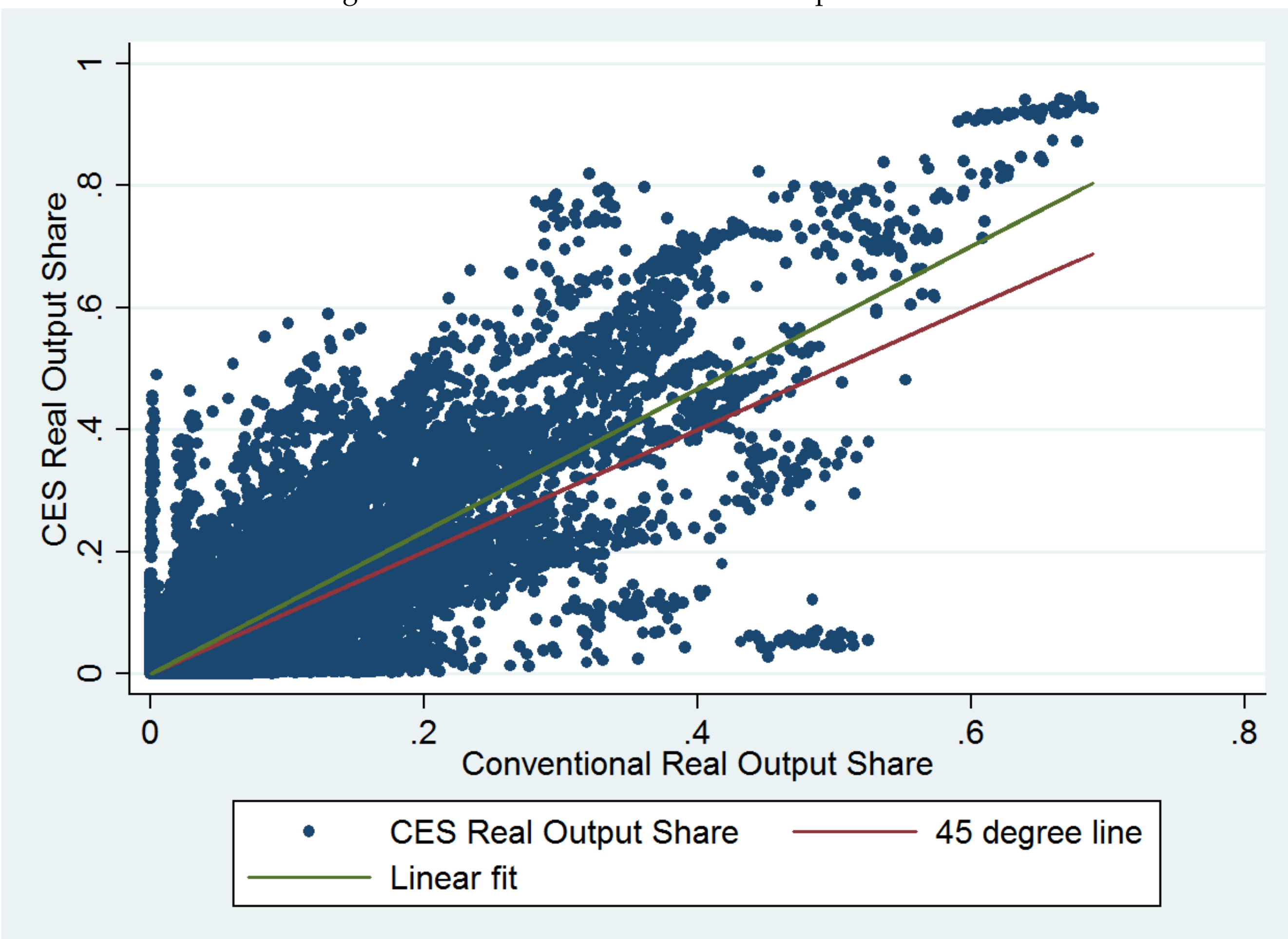
This problem is more pernicious than simply saying firm prices are measured with error; the errors are likely to be systematic. Equation (31) indicates that in the CES system the use of average goods prices to measure firm-level prices will overstate the price level (and understate real output) more for large multiproduct firms relative to small single-product ones. Therefore, if the true model is CES but a researcher uses a quantity-weighted average of the prices of the goods produced by the firm to measure firm prices—what we term a "conventional" price index or P_{ft}^{Conv}—the results are likely to underestimate economies of scale.

But how big is this problem in practice? If we denote the consumer's expenditure on a firm's output by E_{ft}, we can write the conventional measure of real output, Q_{ft}^{Conv}, as E_{ft}/P_{ft}^{Conv}, and the CES measure of real output, Q_{ft}^{CES}, as $E_{ft}/\left(P_{ft}/\varphi_{ft}\right)$. These measures will vary across sectors due to the units, so in order to compare size variation across sectors we will work with unitless shares of each firm in total output (*i.e.*, $Q_{ft}^{CES}/\sum_{f\in g} Q_{ft}^{CES}$ and $Q_{ft}^{Conv}/\sum_{f\in g} Q_{ft}^{Conv}$). Figure 1 plots these to measures of real output. As one can see in Figure 1, the choice of price index matters enormously for the computation of real output. Moreover, our estimate of σ_U suggests that there are large downward biases in the measurement of real output of large firms. While, on average, the conventional measure of real output is quite close the CES measure for small firms it systematically understates the value of real output for firms with large market shares. We can econometrically measure the magnitude of this bias by regressing $\ln\left(P_{ft}^{Conv}/\left(P_{ft}/\varphi_{ft}\right)\right)$ on $\ln E_{ft}$ (with product group fixed effects) and examining the coefficient on log sales. If there were no bias, we would expect a coefficient of zero, but we actually obtain a coefficient of 0.3050 (s.e. 0.0001, $R^2 = 0.3$) indicating that failing to take into account the multiproduct nature of firms causes us to significantly underestimate the size of large firms relative to a CES world. This estimate is economically significant as well. It implies that holding fixed a firm's conventional price index, every one percent increase in firm sales is associated with a 1.3 percent increase in its real output in a CES model. The difference between the two estimates reflects the importance of what demand assumption one makes when constructing a firm price index.

A second striking feature of our estimated results is that the elasticity of substitution among varieties produced by a firm is always larger than that that between firms (*i.e.*, $\sigma_U > \sigma_F$). It is important to remember that this is not a result that we imposed on the data. Moreover, most of the elasticities are precisely estimated—in 75 percent of the cases, we can statistically reject the hypothesis that $\sigma_U = \sigma_F$ at the 5% level. The higher elasticities of substitution across UPCs than across firms imply that that varieties are more substitutable within firms than across firms, which implies cannibalization effects from the introduction of new varieties by firms. Moreover, the fact that the estimated elasticities of substitution are always greater than one implies that firms' varieties are substitutes, which is required for positive markups of price over marginal costs (*c.f.*, equation (9)).

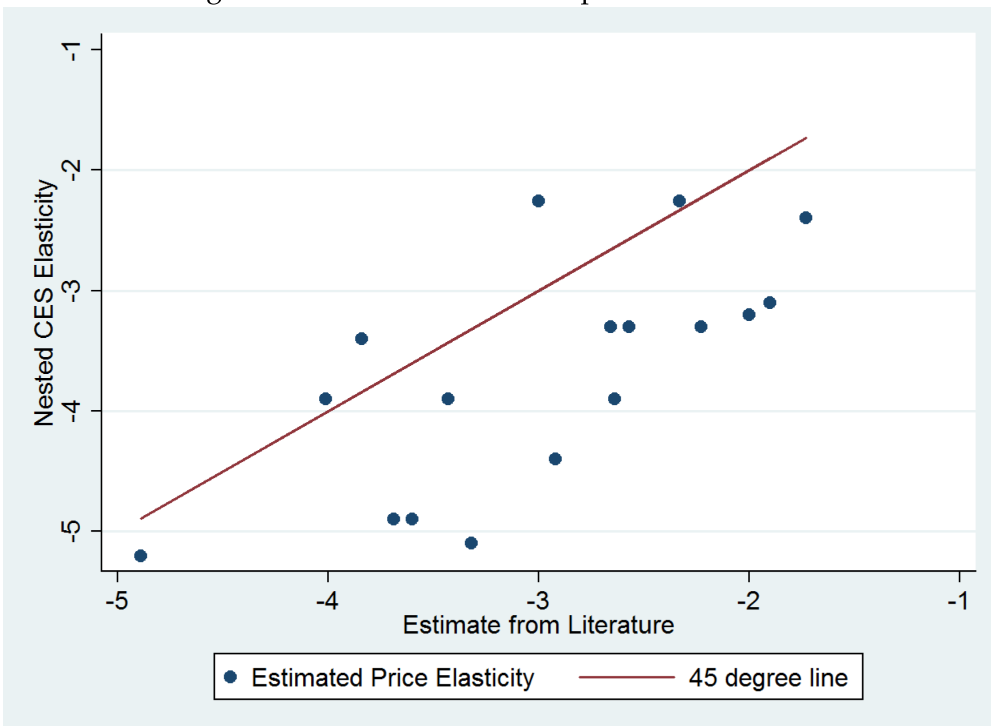
In order to assess whether our elasticities are plausible, it is useful to compare our estimates with

Figure 1: Bias in Conventional Real Output Share



Note: Calculated based on data from The Nielsen Company (US), LLC and provided by the Marketing Data Center at The University of Chicago Booth School of Business.

Figure 2: Price Elasticities Compared to Literature



Note: Calculated based on data from The Nielsen Company (US), LLC and provided by the Marketing Data Center at The University of Chicago Booth School of Business.

those of other papers. In order to do this, we restricted ourselves to comparing our results with studies that used US scanner data and estimated elasticities for the same product groups as ours. We do this because, as Broda and Weinstein (2006) show, elasticity estimates for aggregate data can look quite different than those for disaggregate data. Unfortunately, we did not find studies estimating the elasticity of substitution within firms, but we did find several studies that examined elasticities akin to our cross-firm elasticity, σ_F. Figure 2 plots our estimates of σ_F against other estimates from three studies: Chevalier et al. 2003, Pancras et al (2013), and Gordon et al. (2013). The results of this comparison indicate our estimates are very similar with a statistically significant correlation of 0.70 and the points arrayed fairly close to the 45-degree line. Therefore, while our empirical approach has a number of novel features in modeling multiproduct firms that are of positive measure relative to the markets in which they operate, the empirical estimates generated by our procedure are reasonable compared to the benchmark of findings from other empirical studies.

6.2 Quality

The quality parameters are a second key parameter that we will use for our decomposition. Our estimation procedure allows us to estimate a different φ_{ut} for every quarter in our dataset. We have strong priors that the quality of each UPC should be fairly stable across time. One way to gauge this stability is to regress these quality parameters on UPC fixed effects in order to determine how much of a UPC's quality is common across time periods. When we do this and include time fixed effects

to control for inflation and other common demand shocks, we find that the R^2 is 0.84, which implies that very little of the variation in our quality measures arise from changes in UPC quality. We should also expect firm quality, φ_{ft}, to exhibit a strong firm component but to be less stable over time. This is exactly what the data reveals. Running the same regression for firm quality, we find that 82 percent of the variance can be explained by firm fixed effects. Consistent with these results, we also estimate that firm quality exhibits more variation overall (not just across time) than UPC quality or marginal cost. This can be seen in Figure 3, which plots kernel density estimates of UPC quality, firm quality, and marginal cost.

6.3 Markups

Using our estimated elasticities of substitution, we can compute implied firm markups. As discussed above, multiproduct firms should internalize the fact that they supply a firm consumption index and hence choose a price for each UPC that takes into account interdependencies in pricing decisions across their products and depends on the perceived elasticity of demand for the firm consumption index as a whole (ε_{ft}). Although demand exhibits a constant elasticity of substitution, this perceived elasticity of demand differs across firms because they internalize the effect of their pricing decisions on market price indices. For any firm with a strictly positive expenditure share, the perceived elasticity of demand (ε_{ft}) is strictly less than consumers' elasticity of substitution across varieties ($\varepsilon_{ft} < \sigma_F$). Larger firms that account for greater shares of expenditure have lower perceived elasticities of demand and hence charge higher markups of price over marginal costs (μ_{ft}), as summarized in equations (7) and (8) that are reproduced below:

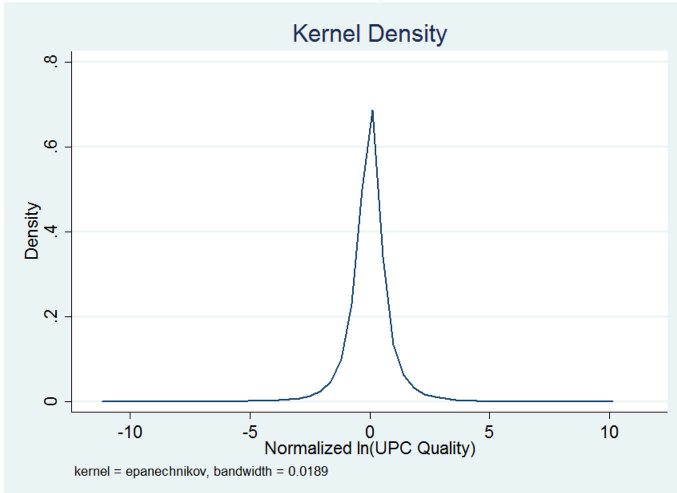
$$\mu_{ft} = \frac{\varepsilon_{ft}}{\varepsilon_{ft} - 1},$$

$$\varepsilon_{ft} = \sigma_F - (\sigma_F - 1)\, S_{ft}.$$

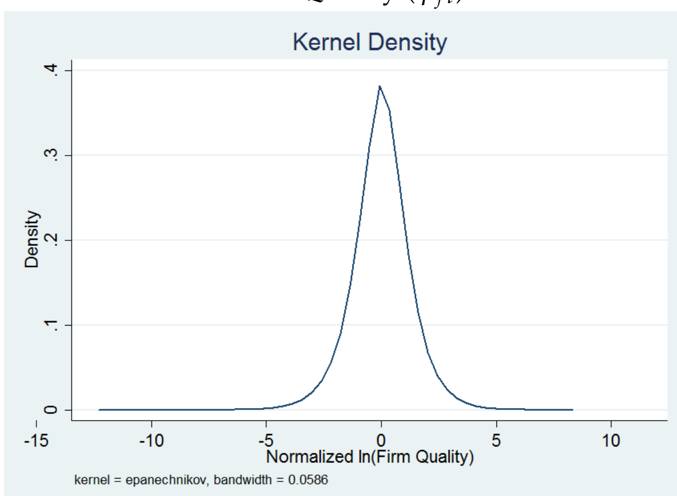
There are two important dimensions across which markups vary. The first is cross-sector variation with captures the fact that σ_F varies systematically across product groups, and the second is within-sector variation which captures larger firms will have higher markups within a sector than small firms. Our baseline results assume that firms are Bertrand competitors. This does not affect any of the estimation, but it does affect the markup formula. The derivation of the Cournot markup is presented in Appendix A.3, but the key point for our purposes is that Cournot competition tends to produce substantially higher markups for larger firms than Bertrand competition if goods are substitutes. Table 6 presents data on the distribution of markups across all firms in our 100 product groups. The median markup is 29 percent, which is slightly lower than Domowitz et al. (1988) estimate of 36 percent for U.S. consumer goods, and slightly above the median markups estimated by De Loecker and Warzinski (2012) for Slovenian data which range from 17 to 28 percent. This suggests that our markup estimates are reasonable in the sense that they do not differ greatly from those found in prior work.

Figure 3: Kernel Densities of Quality and Cost Estimates

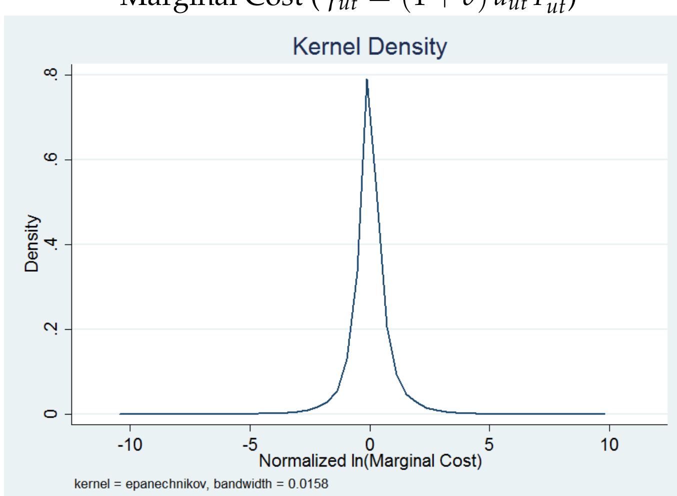
UPC Quality (φ_{ut})



Firm Quality (φ_{ft})



Marginal Cost ($\gamma_{ut} = (1 + \delta)\, a_{ut} Y_{ut}^{\delta}$)



Note: The UPC quality (φ_{ut}) is normalized to geometric mean of UPC qualities within the firm and taken in logarithm; the firm quality (φ_{ft}) is normalized to the geometric mean of firm qualities within the product group and taken in logarithm; The marginal cost is computed as $\gamma_{ut} = (1 + \delta)\, a_{ut} Y_{ut}^{\delta}$ normalized by its geometric mean within the firm and taken in logarithm. Calculated based on data from The Nielsen Company (US), LLC and provided by the Marketing Data Center at The University of Chicago Booth School of Business.

Table 6: Distribution of Markups

Percentile	Cournot Markup using ε_f	Bertrand Markup using ε_f	Markup using σ_F	Markup using σ_U
10	0.14	0.14	0.14	0.09
25	0.22	0.22	0.22	0.14
50	0.29	0.29	0.29	0.16
75	0.39	0.39	0.39	0.23
90	0.51	0.49	0.49	0.24

Note: Markup = (Price-Marginal Cost)/Marginal Cost. Calculated based on data from The Nielsen Company (US), LLC and provided by the Marketing Data Center at The University of Chicago Booth School of Business.

Table 7: Distribution of Markups Relative to Product Group Avg Markup

Percentile	Largest firm - Cournot	2nd Largest firm - Cournot	Largest firm - Bertrand	2nd Largest firm - Bertrand
10	1.40	1.25	1.10	1.07
25	1.60	1.34	1.16	1.09
50	1.98	1.53	1.24	1.13
75	3.12	1.86	1.41	1.18
90	4.89	2.38	1.68	1.28

Note: Results in the table are reported as (Firm Markup)/(Avg Markup in Product Group). Calculated based on data from The Nielsen Company (US), LLC and provided by the Marketing Data Center at The University of Chicago Booth School of Business.

While the markup of most firms is essentially the same regardless of whether we assume firms compete through prices or quantities. The markups of the largest firms is substantially different. Table 7 shows the distribution of the markups for the largest and second largest firms by product group. In each case, we report the markup relative to the average markup in the product group, so that one can see how different the markup of the largest firm is. The largest firm in a product group typically has a markup that is 24 percent larger than average if we assume competition is Bertrand and almost double the average if we assume competition is Cournot. However, as one can also see from this table, these markups drop off quite rapidly for the second largest firm. These results suggest that both models of competition suggest that very few firms can exploit their market power.

6.4 Cannibalization Rate

Our estimated elasticities of substitution also determine the magnitude of cannibalization effects in the model. As discussed above, when a multiproduct firm chooses whether to introduce a new variety, it takes into account the impact of this production introduction on the sales of its existing varieties. The elasticity of the revenue of existing UPCs with respect to the introduction of a UPC is given by (13), which is reproduced below:

$$\text{Cannibalization rate} \equiv \left[\left(\frac{\sigma_U - \sigma_F}{\sigma_U - 1} \right) + \left(\frac{\sigma_F - 1}{\sigma_U - 1} \right) S_{ft} \right] S_{N_{ft}} N_{ft}$$

As we mentioned earlier, a cannibalization level of 0 corresponds to equal differentiation of their

Table 8: Markups and Cannibalization by Decile

Decile (1 is largest)	Avg Cournot Markup	Avg Bertrand Markup	Avg Cannibalization
10	0.335	0.335	0.453
9	0.335	0.335	0.453
8	0.335	0.335	0.453
7	0.335	0.335	0.453
6	0.335	0.335	0.453
5	0.335	0.335	0.453
4	0.335	0.335	0.453
3	0.335	0.335	0.453
2	0.336	0.335	0.454
1	0.356	0.340	0.460

Note: Results are reported for the weighted average across product groups. Calculated based on data from The Nielsen Company (US), LLC and provided by the Marketing Data Center at The University of Chicago Booth School of Business.

products from those of their competitors and a level of 1 corresponds to a world in which the products of firms are perfect substitutes. In Table 8, we present the average estimated markups and cannibalization levels for firms by decile of size. As one can see from the table, these are remarkably stable, which reflects the fact that even in the upper decile of firms most firms have trivial market shares ($S_f \approx 0$). Interestingly, we see that cannibalization levels lie almost exactly between the benchmarks of perfect substitutes and equal differentiation. This cannibalization level implies that about half of the sales of a new product introduced by a firm comes from the sales of existing products and half from the sales of other firms. The fact that these cannibalization rates are much less than one, however, underscores the fact that it is not appropriate to treat the goods produced by a multiproduct firm as perfect substitutes nor is it appropriate to treat multiproduct firms as if the introduction of a new good had no impact on the sales of existing goods.

Moreover, there is a slight tendency for the markups and cannibalization levels to rise with firm size reflecting the fact that large firms have higher markups and take into account the fact that more of their sales of new products comes at the expense of their existing products. We can see this more clearly in Table 9, where we focus on the ten largest firms in the sector. We estimate that markups of the two largest firms are on average 14 and 31 percent larger than markups of firms in the first nine deciles. Similarly, cannibalization levels tend to rise for the the very largest firms. We estimate that when the largest firm in a sector, which has an average market share of 22 percent, introduces a new product, 58 percent of the sales of that product comes from from the sales of its existing products. Since we saw from Table 3 that the largest firms typically have the most products per firm, this implies that cannibalization is likely to be a first order issue for them.

6.5 Decomposing Firm Sales

We can estimate the EKK variance decompositions described in equations (16) to (20) by product group, which results in 100 estimates of each α. Table 10 presents the results from this exercise. On

Table 9: Markups and Cannibalization by Firm Rank

Firm Rank (1 is largest)	Firm's Cournot Markup	Firm's Bertrand Markup	Firm's Cannibalization
10	0.364	0.342	0.464
9	0.370	0.343	0.465
8	0.373	0.344	0.467
7	0.375	0.344	0.468
6	0.383	0.347	0.471
5	0.396	0.351	0.476
4	0.413	0.355	0.482
3	0.455	0.364	0.495
2	0.528	0.382	0.519
1	0.761	0.439	0.577

Note: Results are reported for the weighted average across product groups. Calculated based on data from The Nielsen Company (US), LLC and provided by the Marketing Data Center at The University of Chicago Booth School of Business.

Table 10: Variance Decomposition

Decomposition		Quality	Average MC	Cost Dispersion	Scope	Markup	Normalization Changes	Upgrading
EKK: Cross-Sectional	Mean	0.77	-0.06	0.07	0.21	-0.004	na	na
	Std. Dev.	0.34	0.33	0.03	0.07	0.005	na	na
BO: 50 Largest Firms	Mean	0.51	0.18	0.08	0.26	-0.024	na	na
	Standard Error	0.13	0.06	0.03	0.01	0.003	na	na
EKK: Firm Growth	Mean	0.88	-0.11	0.07	0.16	-0.001	0.001	-0.001
	Std. Dev.	0.19	0.20	0.03	0.07	0.002	0.02	0.02
BO: Firm Growth: 50 Largest Firms	Mean	0.59	0.10	0.07	0.24	-0.001	0.02	-0.02
	Standard Error	0.21	0.20	0.03	0.03	0.002	0.05	0.05

Note: Calculated based on data from The Nielsen Company (US), LLC and provided by the Marketing Data Center at The University of Chicago Booth School of Business.

average 77 percent of the variation in firm sales is driven by firm quality differences. Interestingly, marginal cost differences only explain almost none (1 percent) of the variation in firm sales with 21 percent explained by differences in product scope. Markup variation, which is a mixture of interaction effects, is entirely unimportant, explaining virtually none of the firm-size dispersion in all sectors.

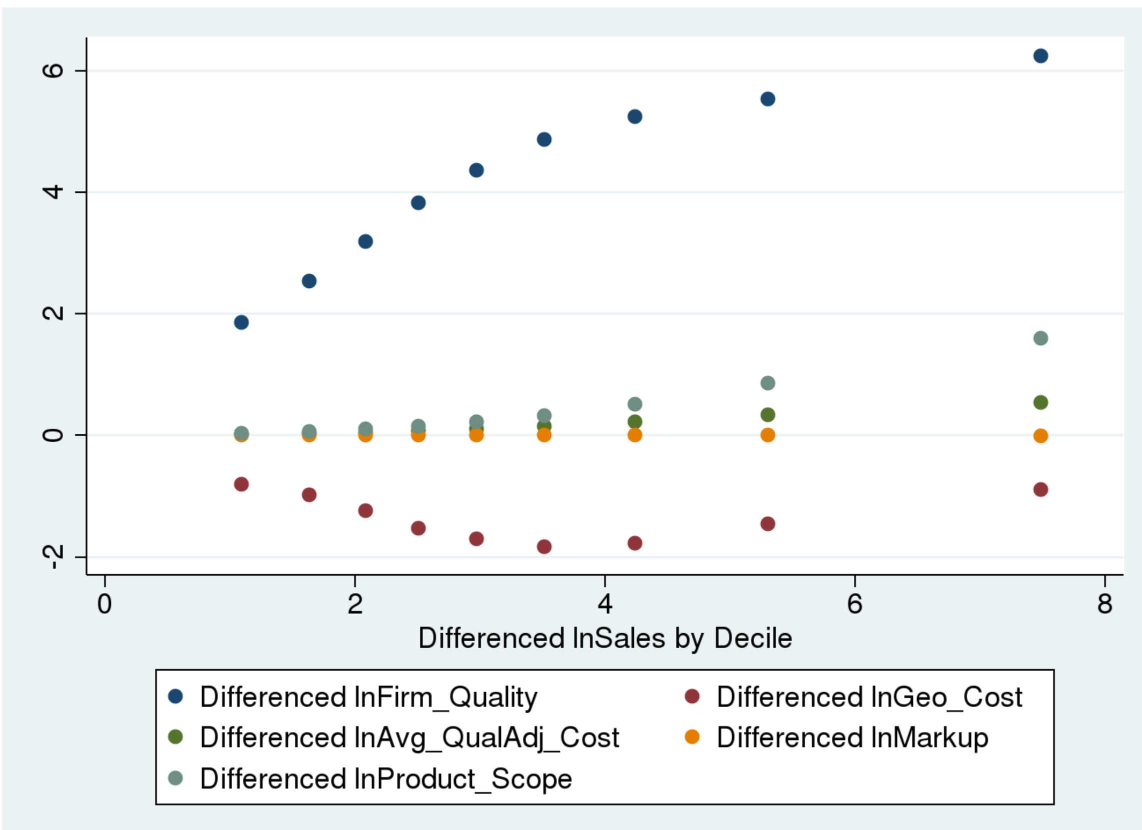
As one can see from Table 10, the importance of costs in determining firm size is rather distributed around zero with very few sectors having costs account for more than a quarter of the variance in firm sales. Similarly, quality accounts for well over half of the variance of firm sales in the vast majority of sectors with product scope accounting for about a quarter of the variance in the remaining sectors. Since product scope is essentially another form of firm quality if consumers value variety, it is clear that quality (broadly defined) is the fundamental determinate of firm size.

Figure 4 presents the results of the exact decomposition discussed in Section 4.5. For each product group, we performed the exact decomposition by decile and then used the sales weights of each product group to average the contributions. Each point represents the component of sales for firms

in a given decile relative to firms in the lowest decile. For example, firms in the first decile are on average about 7.5 log units larger than firms in the tenth decile. Of this 7.5 log unit difference in size, about 6 log units can be attributed to quality differences and most of the remainder is attributable to product scope.

An interesting feature of this decomposition is that the smallest firms appear to have slightly higher average marginal costs than larger firms and smaller dispersion in marginal costs per unit quality. However, we also can see from the figure that the relationship between firm size and marginal cost is not upward sloping in general. Larger firms have higher costs on average than the smallest firms but lower costs on average than mid-sized firms. Thus, unlike the clear upward sloping relationship between firm size and quality and product scope, the relationship with marginal cost is much less clear.

Figure 4: Sales Decomposition by Decile

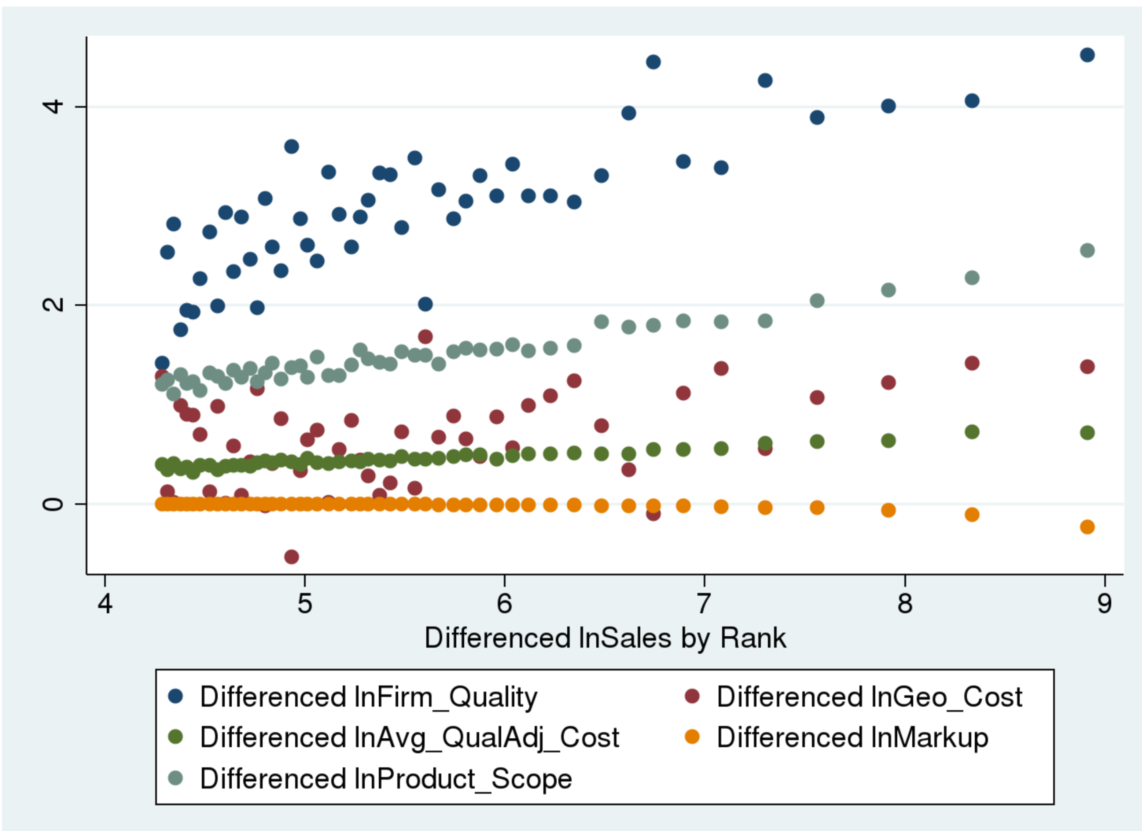


Note: The graph contains the weighted average across product groups of the Blinder-Oaxaca decile decomposition of firm sales into estimated firm quality, marginal cost, markup and product scope. All the variables are taken in logarithm and are differenced relative to the lowest decile. Calculated based on data from The Nielsen Company (US), LLC and provided by the Marketing Data Center at The University of Chicago Booth School of Business.

One of the issues with this decomposition is that it does not tell us much about the very largest firms, who produce the vast majority of sales in a sector. Figure 5 repeats this analysis for the largest 50 firms in each product group relative to the average firm in the average firm. These results are

similar to the decile-by-decile results for the contributions of quality and product scope. As we saw before, both firm quality and scope rise consistently with firm size, and quality consistently is the most important explanator, followed by product scope. However, what is most interesting in this decomposition is that for the seven largest firms at least, there appears to be clear positive association between lower costs (productivity) and firm size. In other words, the most successful firms tend to have lower marginal costs and higher quality. This result is strongly suggestive of managerial models in with good management can not only produce relatively quality goods but also can produce these products at low cost.

Figure 5: Sales Decomposition by Firm Rank



Note: The graph contains the weighted average across product groups of the Blinder-Oaxaca decomposition of firm sales into estimated firm quality, marginal cost, markup and product scope for the top 50 firms by size. All the variables are taken in logarithm and are differenced relative to the average firm. Calculated based on data from The Nielsen Company (US), LLC and provided by the Marketing Data Center at The University of Chicago Booth School of Business.

Table 10 also presents the Blinder-Oaxaca variance decomposition for the 50 largest firms. The table shows that on average, 51 percent of a firm's size is associated with quality differences, with the remainder split between costs and product scope. In contrast to what we saw in the cross-section EKK, the 50 firm EKK variance decomposition attributes a somewhat larger role for costs—larger firms have slightly lower costs using this methodology. These results are broadly consistent with the earlier comparison we did across deciles. In all specifications we see robust, important roles played

by product quality and scope; large firms differ from small ones in that they produce a diversified set of very high-quality products. By contrast, the impact of marginal cost variations is less than 24 percent with the sign and magnitude changing depending on the sample of firms used and the decomposition methodology employed.

Figure 6: Growth Decomposition by Firm Rank

Note: The graph contains the weighted average across product groups of the Blinder-Oaxaca decomposition of the change in firm sales into estimated firm quality, marginal cost, markup and product scope for the top 50 firms by 2011 size. All the variables are taken in logarithm, are normalized relative to the average firm and differenced across the period 2004 quarter 4 - 2011 quarter 4. Calculated based on data from The Nielsen Company (US), LLC and provided by the Marketing Data Center at The University of Chicago Booth School of Business.

As we showed in Section 4.6, we can also examine the time series determinants of firm growth. Since these decompositions are all within firm we do not need to worry about the possibility that differences in units across firms might be contaminating our results. These results, therefore, tell us how important each element of our decomposition is for understanding firm growth. In these specifications, we difference each firm's sales from the fourth quarter of 2011 with the fourth quarter of 2004 and with that of the average firm in the product group. We explore these decompositions in the lower two rows of Table 10. Here, the "normalization changes" refer to the first two terms in the first row of equation (24) and "upgrading" tells us the importance of product turnover for firm growth. Interestingly, these two factors play very small roles in understanding firm growth. Most

firm growth arises from the "marketing effect" or shifts in demand for the existing set of products produced by a firm.

Figure 6 presents the results from this decomposition for the fifty firms whose sales in 2011 were the highest. Consistent with our earlier results on the cross-section, firm quality growth appears to be the most important determinant of firm sales growth followed by the creation of new products, and finally by cost reductions. In the EKK decomposition, over 80 percent of firm growth can be explained by quality growth while the Blinder-Oaxaca decomposition indicates that 59 percent of firm growth is attributable to improvements in quality with the remainder about evenly split between scope and quality. Since scope in our setup is another form of quality, we can say that quality, broadly defined, accounts for a minimum of 77 percent of the growth and size dispersion of firms.

7 Counterfactuals

We now use the structure of the model to undertake counterfactuals to assess the quantitative relevance of our estimated departures from monopolistic competition for aggregate welfare. We consider two types of counterfactuals. First, we consider the implications for consumers from a regulator preventing large firms from exercising their market power thereby reducing each firm to monopolistically competitive competitor. Second, we consider the implications of allowing firms to produce a number of differentiated products. Thus, the second type of counterfactuals examines the gains to consumers from the different varieties produced by each firm.

In each of these counterfactual exercises, we examine the impact on consumers under both Bertrand and Cournot competition. In all cases, we hold constant aggregate expenditure (E_t), the sets of active UPCs and firms $\{U_{ft}, F_{gt}\}$, and the set of firm qualities, UPC qualities and cost shifters $\{\varphi_{ft}, \varphi_{ut}, a_{ut}\}$. Therefore we abstract from general equilibrium effects on aggregate expenditure or on factor prices that enter the cost shifters.

In the first set of counterfactuals (Counterfactuals I-B and I-C), we compare the actual equilibrium in which firms exploit their market power with a counterfactual equilibrium in which a price regulator requires all firms to charge the same constant markup over marginal cost equal to the monopolistically competitive firm markup:

$$\mu_{ft} = \frac{\sigma_F}{\sigma_F - 1}. \tag{33}$$

Moving from the actual to the counterfactual markup reduces the prices of successful firms with high market shares and hence redistributes market shares towards these firms. Therefore, we expect these counterfactuals to increase welfare and the dispersion of firm sales, but the magnitude of these effects depends on the estimated parameters $\{\sigma_U, \sigma_F, \delta\}$ and the sets of firm qualities, product qualities and cost shocks $\{\varphi_{ft}, \varphi_{ut}, a_{ut}\}$.

Given the assumed constant markup (33), we solve for the counterfactual equilibrium using an iterative procedure to solve for a fixed point in a system of five equations for expenditure shares and price indices $\{S_{ut}, P_{ut}, S_{ft}, P_{ft}, P_{gt}\}$. These five equations are the UPC expenditure share (given by

Table 11: Counterfactual Relative to Actual Values

Counterfactural	Coefficient Variation Firm Sales	Aggregate Welfare
I-B Bertrand to Monopolistic Competition	1.130	1.037
I-C Cournot to Monopolistic Competition	1.450	1.156
II-B multiproduct to Single Product Bertrand	0.803	0.679
II-C multiproduct to Single Product Cournot	0.847	0.685

Note: Each statistic is expressed as the value in the counterfactual relative to the value in the observed data. Therefore relative coefficient variation sales is the coefficient of variation of firm sales in the counterfactual divided by the coefficient of variation of firm sales in the observed data. Calculated based on data from The Nielsen Company (US), LLC and provided by the Marketing Data Center at The University of Chicago Booth School of Business.

(4) for $k = u$), the UPC pricing rule (9), the firm expenditure share (given by (4) for $k = f$), the firm price index (given by (3) for $k = f$) and the product group price index (given by (3) for $k = g$). We use the resulting solutions for product group price indices together with aggregate expenditure to compute counterfactual aggregate welfare:

$$W_t = \frac{E_t}{\prod_{g=1}^{G} P_{gt}^{\varphi_{gt}}}. \tag{34}$$

In Counterfactuals II-B and II-C, we examine the quantitative relevance of multiproduct firms. We compare the actual equilibrium in which firms supply multiple products to a counterfactual equilibrium in which firms are restricted to supply a single UPC (their largest). Counterfactual II-B undertakes this comparison for the actual equilibrium under Bertrand competition, while Counterfactual II-C undertakes the same comparison for the actual equilibrium under Cournot competition. We again solve for the counterfactual equilibrium by solving for a fixed point in the system of five equations for expenditure shares and price indices discussed above.

Table 11 reports the results. Each row of the table corresponds to a different counterfactual. The second column reports the coefficient of variation of firm sales in the counterfactual relative to that in the actual data. The third column reports aggregate welfare in the counterfactual relative to that in the actual data.

In Counterfactual I-B, we find that moving from Bertrand to monopolistic competition increases aggregate welfare by around 4 percent, which is comparable to standard estimates of the welfare gains from trade for an economy such as the United States. In Counterfactual I-C, we find much larger welfare effects of moving from Cournot to monopolistic competition, which increases aggregate welfare by around 16 percent. These much larger welfare effects reflect the fact that large firms charge much higher markups under Cournot than under Bertrand in Table 7, which implies larger price reductions from moving to monopolistically competitive markups, and hence larger increases in welfare. In both Counterfactuals I-B and I-C, the relative prices of the largest firms fall and demand is elastic. Therefore, the coefficient of variation of firms sales increases, by around 13 percent

for Bertrand competition and around 45 percent for Cournot competition. Again the larger effects for Cournot reflect the greater variation in markups under this mode of competition.

In both Counterfactuals II-B and II-C, we find quantitatively large effects of multiproduct firms. Restricting firms to supply a single product under Bertrand competition reduces aggregate welfare by around 32.1 percent and reduces the coefficient of variation of firm sales by around 20 percent. The corrresponding reductions in aggregate welfare and the coefficient of variation of firm sales under Cournot competition are 31.5 percent and 15 percent respectively. These results suggest that consumers benefit enormously from the range of products produced within a single firm.

All of these counterfactuals are subject to a number of caveats. In particular, they do not allow for firm entry or exit, because they hold the set of firms constant, and they abstract from general equilibrium effects by holding aggregate expenditure and factor prices constant. Nonetheless, these counterfactuals highlight that our departures from the case of atomistic monopolistic competition with single product firms are quantitatively important.

8 Conclusions

We develop and estimate a structural model of heterogeneous firms that allows firms to supply multiple products and be large relative to markets. Our framework requires only price and expenditure data and hence is widely applicable. We use these price and expenditure data to estimate the key parameters of the model, namely the elasticities of substitution between and within firms, and the elasticities of marginal cost with respect to output. The estimated model permits an exact decomposition of differences in firm sales into the contributions of overall firm quality, the relative quality of firm products, marginal costs, the number of products a firm produces and the firm's markup. We use the estimated model to undertake counterfactuals to quantify the effect of firms being large relative to their markets on aggregate welfare.

Our results point to quality differences as being the principal reason why some firms are successful in the marketplace and others are not. Depending on the specification considered, we find that 50-80 percent of the variance in firm size can be attributed to quality differences, about 20 percent to differences in product scope, and only 3-26 percent to cost differences. If we use a broad measure of firm quality, which encompasses both quality and scope, we find that quality accounts for at least three quarters of firm size differences. We estimate substantially higher elasticities of substitution between varieties within firms than between firms (median elasticities of 6.9 and 4.3 respectively), implying that a firm's introduction of new product varieties cannibalizes the sales of existing varieties. We estimate that the cannibalization rate for the typical firm is 0.45, roughly half way in between the extreme of no cannibalization (equal elasticities of substitution within and between firms and atomistic producers), and the extreme of complete cannibalization (varieties perfectly substitutable within firms).

Our findings that firms supply multiple imperfectly substitutable varieties have important im-

plications for the measurement of firm productivity, because the true ideal price index for the firm depends on the number of its products, and larger firms systematically supply more products than smaller firms. We find that most firms charge markups close to the monopolistic competition benchmark of constant markups, because most firms have trivial market shares, and hence are unable to exploit their market power. However, the largest firms that account for up to around 20 percent of sales within sectors charge markups up to one third higher than the median firm. We find that these departures from the monopolistically competitive benchmark have quantitatively substantial effects on aggregate welfare.

References

[1] Agur, Itai (2010) "Trade Liberalization, Firm Selection and Variety Growth," *Review of International Economics*, 18(3), 582-594.

[2] Allanson, Paul and Catia Montagna (2005) "Multiproduct Firms and Market Structure: An Explorative Application to the Product Life Cycle," *International Journal of Industrial Organization*, 23, 587-597.

[3] Amiti, Mary, Oleg Itskhoki and Jozef Konings (2012) "Importers, Exporters and Exchange Rate Disconnect," Princeton University, mimeograph.

[4] Anderson, Simon P., André de Palma and Jacques-François Thisse (1992) *Discrete Choice Theory of Product Differentiation*, Cambridge MA: MIT Press.

[5] Arkolakis, Costas and Marc-Andreas Muendler (2010) "The Extensive Margin of Exporting Products: A Firm-Level Analysis," *NBER Working Paper*, 16641.

[6] Atkeson, Andrew and Ariel Burstein (2008) "Pricing to Market, Trade Costs and International Relative Prices," *American Economic Review*, 98(5), 1998-2031.

[7] Baldwin Richard E. and James R. Harrigan (2011) "Zeros, Quality, and Space: Trade Theory and Trade Evidence," *American Economic Journal: Microeconomics*, 3(2), 60-88.

[8] Bernard, Andrew B., J. Bradford Jensen and Peter K. Schott (2009) "Importers, Exporters, and Multinationals: A Portrait of Firms in the U.S. that Trade Goods," in *Producer dynamics: New Evidence from Micro Data*, (Eds) T Dunne, JB Jensen and MJ Roberts, Chicago: University of Chicago Press.

[9] Bernard, Andrew B., Stephen J. Redding and Peter K. Schott (2010) "multiproduct Firms and Product Switching," *American Economic Review*, 100(1), 70-97.

[10] Bernard, Andrew B., Stephen J. Redding and Peter K. Schott (2011) "multiproduct Firms and Trade Liberalization,"*Quarterly Journal of Economics*, 126(3), 1271-1318.

[11] Berry, Steven, James Levinsohn and Ariel Pakes (1995) "Automobile Prices in Market Equilibrium," *Econometrica*, 63(4), 841-890.

[12] Broda, Christian and David E. Weinstein (2006) "Globalization and the Gains from Variety," *Quarterly Journal of Economics*, 121(2), 541-86.

[13] Broda, Christian and David E. Weinstein (2010) "Product Creation and Destruction: Evidence and Price Implications," *American Economic Review*, 100, 691–723.

[14] Chevalier, Judith A., Anil K. Kashyap, and Peter E. Rossi (2003) "Why Don't Prices Rise during Periods of Peak Demand? Evidence from Scanner Data," *American Economic Review*, 93(1), 15-37.

[15] Das M, Mark J. Roberts, James R. Tybout (2007) "Market Entry Costs, Producer Heterogeneity and Export Dynamics," *Econometrica*, 75(3), 837-73.

[16] De Loecker, Jan and Frederic Warzynski (2012) "Markups and Firm-Level Export Status," *American Economic Review* 2012, 102(6): 2437–2471.

[17] De Loecker, Jan, Penny Goldberg, Amit Khandelwal and Nina Pavcnik (2012) "Prices, Markups and Trade Reform," Princeton University, mimeo.

[18] Dhingra, Swati (2013) "Trading Away Wide Brands for Cheap Brands," *American Economic Review*, 103(6), 2554-84.

[19] Di Giovanni, Julian and Andrei Levchenko (2012) "Country Size, International Trade, and Aggregate Fluctuations in Granular Economies," *Journal of Political Economy*, forthcoming.

[20] Domowitz, Ian, R. Glenn Hubbard, and Bruce C. Petersen (1988), "Business Cycles and the Relationship between Concentration and Price-Cost Margins," *The RAND Journal of Economics*, 17(1), 1-17.

[21] Dube, Jean-Pierre (2004) "Multiple Discreteness and Product Differentiation: Demand for Carbonated Soft Drinks," *Marketing Science*, 23(1), 66-81.

[22] Eckel, Carsten and J. Peter Neary (2010) "multiproduct Firms and Flexible Manufacturing in the Global Economy," *Review of Economic Studies*, 77(1), 188-217.

[23] Edmond, Chris, Virgiliu Midrigan and Daniel Yi Xu (2012) "Competition, Markups, and the Gains from International Trade," *NBER Working Paper*, 18041.

[24] Foster, Lucia, John Haltiwanger and Chad Syverson (2008) "Reallocation, Firm Turnover, and Efficiency: Selection on Productivity or Profitability?" *American Economic Review*, 98(1), 394-425.

[25] Feenstra, Robert C. (1994) "New Product Varieties and the Measurement of International Prices," *American Economic Review*, 84(1), 157-177.

[26] Feenstra, Rob and Hong Ma (2008) "Optimal Choice of Product Scope for Multiproduct Firms," in (eds) Elhanan Helpman, Dalia Marin and Thierry Verdier, *The Organization of Firms in a Global Economy*, Cambridge MA: Harvard University Press.

[27] Gabaix, Xavier (2011) "The Granular Origins of Aggregate Fluctuations," *Econometrica*, 79, 733-772.

[28] Goldberg, Penny (1995) "Product Differentiation and Oligopoly in International Markets: The Case of the U.S. Automobile Industry," *Econometrica*, 891-951.

[29] Gordon, Brett R., Avi Goldfarb, and Yang Li (2013), "Does Price Elasticity Vary with Economic Growth? A Cross-Category Analysis," *Journal of Marketing Research*, Vol. L (Feb.), 4-23.

[30] Johnson, Robert C. (2012) "Trade and Prices with Heterogeneous Firms," *Journal of International Economics*, 86 (1), 43-56.

[31] Khandelwal, Amit K. (2010) "The Long and Short (of) Quality Ladders," *Review of Economic Studies*, 77(4), 1450-1476.

[32] Leontief, Wassily (1929) "Ein Versuch Zur Statistichen Analyse von Angebot und Nachfrage," *Weltwirtschaftliches Archiv*, 30(1), 1-53.

[33] Mayer, Thierry, Marc Melitz and Gianmarco I. P. Ottaviano (2012) "Market Size, Competition, and the Product Mix of Exporters," *American Economic Review*, forthcoming.

[34] Melitz, Marc J. (2003) "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity," *Econometrica*, 71, 1695-725.

[35] Melitz, Marc J. and Gianmarco I. P. Ottaviano (2008) "Market Size, Trade, and Productivity," *Review of Economic Studies*, 75(1), 295-316.

[36] Melitz, Marc J. and Stephen J. Redding (2013) "Heterogeneous Firms and Trade," *Handbook of International Economics*, Volume 4, forthcoming.

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

[38] Pancras, Joseph, Dinesh K. Gauri, and Debabrata Talukdar (2013), "Loss leaders and cross-category retailer pass-through: A Bayesian multilevel analysis,", *Journal of Retailing*, 89(2), 140-157.

[39] Sato, Kazuo (1976) "The Ideal Log-Change Index Number," *Review of Economics and Statistics*, 58(2), 223-228.

[40] Schott, Peter K. (2004) "Across-product Versus Within-product Specialization in International Trade," *Quarterly Journal of Economics*, 119, 647-678.

[41] Stock, James H., Jonathan H. Wright and Motohiro Yogo (2002) "A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments," *Journal of Business and Economic Statistics*, 20(4), 518-529.

[42] Sutton, John (1986) "Vertical Product Differentiation: Some Basic Themes," *The American Economic Review Papers and Proceedings*, 76(2), 393-398.

[43] Vartia, Yrjo (1976) "Ideal Log-Change Index Numbers," *Scandinavian Journal of Statistics*, 3(3), 121-126.

[44] Vives, Xavier (2001) *Oligopoly Pricing: Old Ideas and New Tools*, MIT Press.

A Appendix A: Derivations

A.1 Derivation of Equations(7)-(9)

The first-order condition with respect to the price of an individual UPC implies:

$$Y_{ut}+\sum_{k=\underline{u}_{ft}}^{\underline{u}_{ft}+N_{ft}}\left(P_{kt}\frac{dY_{kt}}{dP_{ut}}-\frac{dA_k\left(Y_{kt}\right)}{dY_{kt}}\frac{dY_{kt}}{dP_{ut}}\right)=0. \tag{35}$$

Using equation 5 and setting UPC supply equal to demand we have

$$\frac{\partial Y_{kt}}{\partial P_{ut}}=\left(\sigma_f-1\right)\frac{Y_{kt}}{P_{gt}}\frac{\partial P_{gt}}{\partial P_{ut}}+\left(\sigma_u-\sigma_f\right)\frac{Y_{kt}}{P_{ft}}\frac{\partial P_{ft}}{\partial P_{ut}}-\sigma_u\frac{Y_{kt}}{P_{ut}}\frac{\partial P_{kt}}{\partial P_{ut}}.$$

We now can use equation 4 to solve for the elasticities and rewrite $\frac{\partial Y_k}{\partial P_u}$ as

$$\begin{aligned}\frac{\partial Y_{kt}}{\partial P_{ut}} &= \left(\sigma_f-1\right)\left(\frac{\partial P_{gt}}{\partial P_{ft}}\frac{P_{ft}}{P_{gt}}\right)\left(\frac{\partial P_{ft}}{\partial P_{ut}}\frac{P_{ut}}{P_{ft}}\right)\frac{Y_{kt}}{P_{ut}}+\left(\sigma_u-\sigma_f\right)\left(\frac{\partial P_{ft}}{\partial P_{ut}}\frac{P_{ut}}{P_{ft}}\right)\frac{Y_{kt}}{P_{ut}}-\sigma_u\frac{Y_{kt}}{P_{ut}}1_{\{u=k\}}\\ &= \left(\sigma_f-1\right)S_{ft}S_{ut}\frac{Y_{kt}}{P_{ut}}+\left(\sigma_u-\sigma_f\right)S_{ut}\frac{Y_{kt}}{P_{ut}}-\sigma_u\frac{Y_{kt}}{P_{ut}}1_{\{u=k\}}.\end{aligned} \tag{36}$$

If we now substitute equation 36 into equation 35 and dividing both sides by Y_u, we get

$$\begin{aligned}1+\sum_k\left(\sigma_f-1\right)S_{ft}S_{ut}\frac{P_{kt}Y_{kt}}{P_{ut}Y_{ut}}+\sum_k\left(\sigma_u-\sigma_f\right)S_{ut}\frac{P_{kt}Y_{kt}}{P_{ut}Y_{ut}}-\sigma_u\\ -\sum_k\left(\sigma_f-1\right)S_{ft}S_{ut}\frac{\frac{\partial A_k(Y_{kt})}{\partial Y_{kt}}Y_{kt}}{P_{ut}Y_{ut}}-\sum_k\left(\sigma_u-\sigma_f\right)S_{ut}\frac{\frac{\partial A_k(Y_{kt})}{\partial Y_{kt}}Y_{kt}}{P_{ut}Y_{ut}}+\sigma_u\frac{\frac{\partial A_k(Y_{kt})}{\partial Y_{kt}}}{P_{ut}} &= 0.\end{aligned} \tag{37}$$

We define the markup at the firm or UPC level as $\mu_k \equiv P_k \big/ \frac{\partial A_k(Y_k)}{\partial Y_k}$. Since $S_u\frac{1}{P_uY_u}=\frac{1}{\sum_k P_kY_k}$ and therefore $\sum_k S_u\frac{P_kY_k}{P_uY_u}=1$, we can rewrite equation 37 as

$$1+\left(\sigma_f-1\right)S_{ft}+\left(\sigma_u-\sigma_f\right)-\sigma_u-\left(\sigma_f-1\right)S_{ft}\frac{\sum_k\frac{\partial A_k(Y_{kt})}{\partial Y_{kt}}Y_{kt}}{\sum_k P_{kt}Y_{kt}}-\left(\sigma_u-\sigma_f\right)\frac{\sum_k\frac{\partial A_k(Y_{kt})}{\partial Y_{kt}}Y_{kt}}{\sum_k P_{kt}Y_{kt}}+\sigma_u\frac{1}{\mu_{ut}}=0.$$

Because we assume that σ_u is the same for all u produced by a firm, μ_{ut} is the only u-specific term in this expression. Hence, μ_{ut} must be constant for all u produced by firm f in time t; in other words, *markups only vary at the firm level*. Together these two results ensure the same markup across all UPCs supplied by the firm.

We can now solve for μ_{ft} by

$$\begin{aligned}1+\left(\sigma_f-1\right)S_{ft}+\left(\sigma_u-\sigma_f\right)-\sigma_u-\left(\sigma_f-1\right)S_{ft}\frac{1}{\mu_{ft}}-\left(\sigma_u-\sigma_f\right)\frac{1}{\mu_{ft}}+\sigma_u\frac{1}{\mu_{ft}} &= 0\\ \Rightarrow \mu_f=\frac{\sigma_f-\left(\sigma_f-1\right)S_{ft}}{\sigma_f-\left(\sigma_f-1\right)S_{ft}-1}.\end{aligned}$$

A.2 Derivation of Equation (13)

Equilibrium prices P_{gt} and P_{ft}, but not P_{ut}, are functions of this N_{ft}. Recall $Y_{ut} = \varphi_{ft}^{\sigma_f - 1} \varphi_{ut}^{\sigma_u - 1} E_{gt} P_{gt}^{\sigma_f - 1} P_{ft}^{\sigma_u - \sigma_f} P_{ut}^{-\sigma_u}$.

$$\begin{aligned}
\frac{\partial Y_{ut}}{\partial N_{ft}} &= (\sigma_f - 1)\, \frac{Y_{ut}}{P_{gt}} \frac{\partial P_{gt}}{\partial N_{ft}} + (\sigma_u - \sigma_f)\, \frac{Y_{kt}}{P_{ft}} \frac{\partial P_{ft}}{\partial N_{ft}}. \\
&= (\sigma_f - 1)\, \frac{Y_{ut}}{P_{gt}} \frac{N_{ft}}{P_{ft}} \frac{P_{ft}}{N_{ft}} \frac{\partial P_{gt}}{\partial P_{ft}} \frac{\partial P_{ft}}{\partial N_{ft}} + (\sigma_u - \sigma_f)\, \frac{Y_{kt}}{P_{ft}} \frac{N_{ft}}{N_{ft}} \frac{\partial P_{ft}}{\partial N_{ft}} \\
&= (\sigma_f - 1) \left(\frac{\partial P_{gt}}{\partial P_{ft}} \frac{P_{ft}}{P_{gt}} \right) \left(\frac{\partial P_{ft}}{\partial N_{ft}} \frac{N_{ft}}{P_{ft}} \right) \frac{Y_{ut}}{N_{ft}} + (\sigma_u - \sigma_f) \left(\frac{\partial P_{ft}}{\partial N_{ft}} \frac{N_{ft}}{P_{ft}} \right) \frac{Y_{kt}}{N_f t}
\end{aligned}$$

If we treat the number of goods sold by the firm as continuous and assume that δ is sufficiently close to zero that we can ignore the impact of new product on the marginal cost of other products, we can use equation (3) defined for the firm price index and the expression for a UPC's share as a function of prices and qualities described in equation (4) to write,

$$\frac{\partial P_{ft}}{\partial N_{ft}} \frac{N_{ft}}{P_{ft}} = \frac{\left(\frac{P_{N_{ft}}}{\varphi_{N_{ft}}} \right)^{1-\sigma_u}}{1 - \sigma_u} P_{ft}^{\sigma_u} \frac{N_{ft}}{P_{ft}} = \frac{N_{ft}}{1 - \sigma_u} \frac{\left(\frac{P_{N_{ft}}}{\varphi_{N_{ft}}} \right)^{1-\sigma_u}}{\sum_{n=1}^{N_{ft}} \left(\frac{P_{ut}}{\varphi_{ut}} \right)^{1-\sigma_u}} = \frac{N_{ft}}{1 - \sigma_u} S_{N_{ft}}.$$

Remembering that shares are also price elasticities we obtain,

$$\begin{aligned}
\text{Cannibalization} \equiv - \frac{\partial Y_{ut}}{\partial N_{ft}} \frac{N_{ft}}{Y_{ut}} &= - \quad (\sigma_f - 1)\, S_{ft} \frac{N_{ft}}{1 - \sigma_u} S_{N_{ft}} - (\sigma_u - \sigma_f)\, \frac{N_{ft}}{1 - \sigma_u} S_{N_{ft}} \\
&= \left[\left(\frac{\sigma_u - \sigma_f}{\sigma_u - 1} \right) + \left(\frac{\sigma_f - 1}{\sigma_u - 1} \right) S_{ft} \right] S_{N_{ft}} N_{ft} > 0, \qquad \text{for } \sigma_u \geq \sigma_f > 1.
\end{aligned}$$

A.3 Cournot Quantity Competition

In our baseline specification in the main text above, we assume that firms choose prices under Bertrand competition. In this appendix, we discuss a robustness test in which we assume instead that firms choose quantities under Cournot competition. Each firm chooses the number of UPCs (N_{ft}) and their quantities (Y_{ut}) to maximize its profits::

$$\max_{N_{ft}, \{Y_{ut}\}} \Pi_{ft} = \sum_{u = \underline{u}_f}^{\underline{u}_f + N_{ft}} P_{ut} Y_{ut} - A_u (Y_{ut}) - N_{ft} h_{ut} - h_{ft}, \tag{38}$$

where we again index the UPCs supplied by the firm from the largest to the smallest in sales. From the first-order conditions for profit maximization, we obtain the equilibrium markup:

$$\mu_{ft} = \frac{\varepsilon_{ft}}{\varepsilon_{ft} - 1}, \tag{39}$$

where the firm's perceived elasticity of demand is now:

$$\varepsilon_{ft} = \frac{1}{\frac{1}{\sigma_f} - \left(\frac{1}{\sigma_f} - 1 \right) S_{ft}}, \tag{40}$$

and the firm's pricing rule is:

$$P_{ut} = \mu_{ft} \frac{dA_u\left(Y_{ut}\right)}{dY_{ut}} = \mu_{ft}\left[\left(1+\delta\right)a_{ut}Y_{ut}^{\delta}\right]. \tag{41}$$

Therefore the analysis with Cournot competition in quantities is similar to that with Bertrand competition in prices, except that firms' perceived elasticities of demand (ε_{ft}) and hence their markups differ between these two cases. Since firms internalize their effects on product group aggregates, markups are again variable. Furthermore, markups only vary at the firm level for the reasons discussed in the main text above.

Our estimation procedure for $\{\sigma_U, \sigma_F, \delta\}$ remains entirely unchanged because the firm markup differences out when we take double differences between a pair of UPCs within the firm over time. Therefore our estimates of the parameters $\{\sigma_U, \sigma_F, \delta\}$ are robust to the assumption of either Bertrand or Cournot competition. Our solutions for firm and product quality $\{\varphi_U, \varphi_F, \delta\}$ are also completely unchanged, because they depend solely on observed expenditure shares and prices $\left\{S_{ut}, S_{ft}, P_{ut}\right\}$ and our estimates of the parameters $\{\varphi_U, \varphi_F, \delta\}$. Therefore whether we assume Cournot or Bertrand competition is only consequential for the decomposition of observed UPC prices (P_{ut}) into markups (μ_{ut}) and marginal costs (a_{ut}). Markups are somewhat more variable in the case of Cournot competition, which strengthens our finding of substantial departures from the atomistic monopolistically competitive markup for the largest firms within each product group.