論　文

The Contribution of Quality and Product Variety to Retail Growth in Japan[†]

Masahiro SATO[a,*], Taisuke KAMEDA[b], Shigeru SUGIHARA[b], Colin HOTTMAN[c]

Abstract

This paper examines the contribution of service quality to service sector output in Japan taking retail services as an example. Specifically, it examines how much of the variation in retail sales across retail firms is due to differences in the quality of services, including the product variety offered, and how the variation in real output of the retail sector would change if it were deflated by price indices incorporating differences in quality instead of conventional price indices. We address the methodological difficulties in quantitatively evaluating service quality by using a massive dataset of barcode-level purchase records providing detailed information about purchases at individual retail firms, constructing a structural model of consumer demand for the real output of each retail firm, and introducing a benchmark product to normalize the quality parameters. This approach enables us to assess the contribution of service quality independently from product quality. Our results show that 57% of the variation in retail firms' sales is attributable to differences in firm-level service quality and 26% to differences in product group variety. Estimates based on the conventional price index understate the real output of large firms by a quarter relative to small firms.

JEL Classification Code: L11, L13, L81

Keywords: Service Sector, Productivity, Product Variety, Structural Estimation

[†] We thank to Kyoji Fukao, Kohei Kawaguchi and other seminar participants at ESRI for their insightful comments and suggestions.

[a] Cabinet Office, Government of Japan (Present Affiliation: Tohoku University)

[b] Cabinet Office, Government of Japan

[c] Federal Reserve Board of Governors, United States of America

[*] Corresponding author: Tohoku University, Graduate School of International Cultural Studies, 41 Kawauchi, Aoba-ku, Sendai, Miyagi 980-8576, Japan

日本の小売業の成長におけるサービスの質と製品多様性の貢献

佐藤正弘・亀田泰佑・杉原　茂・Colin Hottman

〈要 旨〉

本論文では、小売業を例に、サービスの質が日本のサービス産業の成長にどれだけ貢献しているのかを検証する。具体的には、小売企業間の売上規模の違いのうち、製品多様性を含むサービスの質の違いに起因する割合はどの程度か、質を考慮した価格指数を用いると小売企業間の実質産出額の違いはどう変化するかを検討する。分析手法としては、小売企業毎の商品バーコード・レベルの購買履歴データを用いて、質のパラメーターを含む消費者需要の構造モデルを推計する。その際、ベンチマーク財を設定して質のパラメーターを標準化することで、サービスの質の貢献を製品自体の質から分離して評価する。検証の結果、小売企業間の売上規模の違いの 57%は企業レベルのサービスの質、26%は製品多様性によるものであること、また、質を考慮しない従来型の価格指数では、小売企業間の実質産出額の違いは 4 分の 1 程度過少評価されることが明らかになった。

JEL 分類コード: L11, L13, L81
キーワード：サービスセクター、生産性、製品多様性、構造推定

1. Introduction

Growth of the service sector, which accounts for about 70% of Japan's gross domestic product and employment, is essential for the sustainable growth of its economy. In particular, productivity increases through improvements in service quality will need to play a stronger role than ever, since the decline in labor input due to the persistent population decline is expected to be a persistent drag on the economy.

Against this background, the aim of this paper, taking retail services as an example, is to examine the contribution of service quality to service sector output in Japan in order to derive lessons regarding the service sector's role as a driver of macro-economic growth. Specifically, we examine the following two questions: (1) how much of the variation in retail sales across retail firms is due to differences in the quality of services, including the product variety offered; and (2) how much the variation in real output of the retail sector would change if it were deflated by price indices incorporating differences in quality instead of conventional price indices.

However, due to the economic characteristics of services and the lack of appropriate data, there are methodological difficulties in quantitatively assessing service quality and its contribution to output. First, certain qualitative aspects of retail services cannot be measured using objective indicators, since they are essentially subjective in nature.[1] Examples are the friendliness of sales staff and the sophistication of shop displays. Second, the intangibility and simultaneity of production and consumption of retail services make it difficult for outsiders to grasp the quality of services provided. For instance, precise explanations of a product by sales staff cannot be measured in the same way as the user interface of electronic devices is tested in a laboratory. Moreover, it is difficult to grasp the quality of a service independent from that of the product itself, since consumers' purchase decisions may be affected by both factors. Third, partly due to the above characteristics, detailed information on consumers' purchase behavior at each retail store that would be needed to obtain consumers' assessment of service quality is usually unavailable. Even relatively basic information

[1] Of course, some quality components can be measured using objective indicators, and there are several studies employing such indicators to examine service quality. Matsuura and Sunada (2009), for instance, quantify the effects of service quality improvements in the Japanese food retail industry on consumer welfare using store-level data such as business hours, the area of floor space per employee, the inventory-to-sales ratio, and the cash settlement ratio.

that links the products a retailer supplies with the prices and quantities at which they are sold is unavailable.[2]

This study attempts to overcome these limitations by relying on three closely related methodological devices. First, we use a massive dataset of barcode-level purchase records at individual retail firms. Specifically, we extract combinations of price and quantity information for a quarterly average of 157,147 products purchased at 1,517 retail firms from a large-scale dataset collected through barcode scans by about 50,000 consumers.

Second, we construct a structural model of consumer demand for the real output of each retail firm to estimate structural parameters capturing consumers' implicit valuation of service quality employing the massive dataset of barcode-level transactions. The theoretical framework in this respect is largely based on Hottman *et al.* (2016) and Hottman (2016). Hottman *et al.* (2016) develop a three-level nested constant elasticity of substitution (CES) utility system and decompose the firm-size distribution of heterogeneous multiproduct manufacturers into the contribution of cost, appeal, markup, and product scope. Hottman (2016) uses a similar framework to examine the mechanisms through which the retail sector in the United States affects the attractiveness of cities and consumer welfare. We extend their three-level nested CES utility system to investigate the variation in sales and real output of Japanese retail firms.

However, as will be explained later, if we were to directly employ the approach developed by Hottman *et al.* (2016) and Hottman (2016) we would not be able to assess the contribution of service quality independently from product quality. Therefore, as the third element of our approach, we introduce a benchmark product and a benchmark product group to normalize the quality parameters of each product and product group across retail firms, which makes a direct comparison of service quality of the same product sold at different retail firms possible and enables us to extract firm-level service quality, which homogeneously affects the sales of all products at a specific retail firm.

---

[2] For instance, the *Census of Commerce* (Ministry of Economy, Trade and Industry), whose store-level data are often used in quality analyses of the Japanese retail sector, does not contain comprehensive price information. Matsuura and Sunada (2009) overcome this limitation by assigning average product-level price data for each retail category taken from the National Survey of Prices (Ministry of Internal Affairs and Communications) as a proxy of price data for each retail store in the *Census of Commerce*.

The remainder of the study is organized as follows. Section 2 explains the data we use and provides some descriptive statistics. Section 3 presents our structural model of the consumer problem and the retail firm problem. Section 4 then explains the estimation methods. Through structural estimations, we obtain unobserved service quality, unobserved markups, marginal costs, and the elasticity of substitution. Next, Section 5 presents the estimation results and examines the contribution of service quality to variations in sales and real output. Finally, Section 6 concludes.

2. Data

Our data source is INTAGE SCI, which is a massive dataset of purchase records of about 50,000 respondents.[3] Respondents scan the barcode (Japan Article Number, JAN[4]) of products using a portable barcode scanner or smartphone at the time they purchase the item and enter the purchase details (where and when they bought the item, how much it cost, how much they spent shopping in total) by the end of the day. There are more than 10 million purchase records per quarter.[5] The products fall into 14 broad categories and 302 product groups based on the classification of the Japan Item Code File Service (JICFS).[6] The major categories covered are staple food, processed food, beverages, household goods, cosmetics, and drugs. Each purchase record contains information on the retail store belonging to one of 1,517 retail firms at which the product was sold.[7] We classify retail firms into seven retail categories: supermarkets, convenience stores, 99 or 100 Yen stores, hardware/discount stores, pharmacies/drugstores, liquor stores, and department stores.

---

[3] The respondents are recruited via quota sampling by area (10 areas), gender, marital status, and age bracket, and the data is weighted to reflect the ratios in the general population, which ensures that data are nationally representative and it is possible to investigate even items which are purchased infrequently. Our dataset includes the data from about 20,000 respondents in April 2010-March 2011, about 27,000 respondents in April 2011-December 2011, and about 50,000 respondents in each of the years from January 2012 to December 2014.

[4] The JAN code is the product code employed in Japan and is used for barcode representation in point of sale (POS) systems, ordering systems, and inventory control systems. It forms part of the GTIN (Global Trade Item Number) system, the globally standardized product identification system for trade in retail and other supply chains, which includes the European Article Number (EAN) used in Europe and the Universal Product Code (UPC) employed in the United States and Canada.

[5] For instance, the number of records for October-December 2014 is 11,557,641.

[6] The JICFS is a service run by GS1 Japan, which is responsible for assigning JAN codes and collecting and maintaining product data.

[7] Our dataset does not contain records collected by respondents in Okinawa Prefecture.

Table 1：Summary Statistics

	Mean	Median	Std. dev.	10th percentile	90th percentile	Max. value
Firm sales (1,000 Yen)	10,089,385	5,203,766	16,789,878	966,293	23,821,732	162,056,106
Product-group sales per firm (1,000 Yen)	45,211	27,974	61,573	7,459	99,469	5,262,860
Product sales per retail firm (1,000 Yen)	1,941	1,666	1,184	850	3,419	6,778
No. of product groups per firm	174	175	54	103	243	284
No. of products per firm	3,037	2,037	54	618	6,365	26,519

Note: Weighted by the sales of each retail category in each quarter.

Based on the original dataset, we construct a nationwide quarterly database from the second quarter of 2010 to the fourth quarter of 2014 containing the quarterly average prices and total quantities of each barcode-level item sold at each retail firm. Note that, as explained in the following section, we screen the original dataset and limit it to firms that sold the benchmark product of the benchmark product group in each quarter, which we use to compare the quality parameters of each product and product-group across retail firms. As a result, the database on average covers 157,147 different products sold per quarter.[8] Table 1 provides summary statistics of our database. We weight the data by the sales of each of the seven retail categories mentioned above in each quarter and average across quarters. As one can see in Table 1, there are large variations in retail firms' sales. The top firm sells over thirty times as much as the median firm and nearly 170 times as much as the firm at the 10th percentile.

Next, Table 2 shows the distribution of retail firms' sales by quintile. The table indicates that retail firms in the top quintile of sales account for about 70% of total sales. On average, the firms in the top quintile have about 19 times more products than firms in the bottom quintile, indicating that product variety may affect sales variation across retail firms to a certain extent. Table 3 provides a more detailed description of this firm heterogeneity by focusing on the ten largest firms in each of the seven retail categories (where we weight the averages by the sales of the category). The table shows that even the largest firms are not close to being monopolists. On average, the largest firms have a market share of 5.4%, and around one-fifth of all of the sales of firms in the top quintile accrue to the ten largest firms.

[8] This is the quarterly average during April 2010-December 2014.

Table 2：Size Distributions by Quintile

Ranked quintile	Quintile market share	Average firm market share	Average firm sales (1,000 yen)	Average no. of product groups per firm	Mean no. of products per firm
5	70.79%	6.21%	29,995,824	201	7,623
4	16.20%	3.34%	12,129,724	176	2,998
3	7.97%	1.92%	6,803,100	189	1,727
2	3.60%	0.70%	2,607,222	193	1,043
1	1.44%	0.34%	1,208,336	116	402

Note: Weighted by the sales of each retail category in each quarter.

Table 3：Size Distribution by Firm Rank

Firm rank	Firm market share	Firm sales (1,000 yen)	No. of product groups per firm	No. of products per firm
1	5.4%	162,056,106	283	26,336
2	2.7%	82,164,779	273	19,157
3	2.2%	64,941,582	274	13,677
4	1.8%	53,393,314	270	13,465
5	1.4%	41,820,093	271	14,752
6	1.3%	37,985,449	260	10,063
7	1.2%	35,452,648	267	12,076
8	1.1%	32,492,155	256	11,722
9	1.0%	28,987,622	239	9,715
10	0.9%	27,022,369	261	10,711

Note: Weighted by the sales of each retail category in each quarter.

3. Model

3.1 Preferences

3.1.1 Utility function

We consider a nested constant elasticity of substitution (CES) utility system that allows the elasticity of substitutions and quality parameters of the three levels, i.e., firms, product groups, and products, to differ from each other. Utility U_t at time t is given by the following CES function:

$$U_t=\left[\sum_{r\in R_t}(\varphi_{rt}C_{rt})^{\frac{\sigma_R-1}{\sigma_R}}\right]^{\frac{\sigma_R}{\sigma_R-1}}, \tag{1}$$

where C_{rt} denotes the subutility derived from the real consumption of the product groups sold by retail firm r; R_t is the set of retail firms at time t; φ_{rt} is the perceived overall service quality of retail firm r; and σ_R is the elasticity of substitution between retail firms.

Below the utility described above, we assume two CES nests for the product-group level and the barcode level. For the first nest, the subutility from the real consumption of the product groups supplied by retail firm r, C_{rt}, is given as a CES consumption index of product groups:

$$C_{rt}=\left[\sum_{g\in G_{rt}}(\varphi_{grt}C_{grt})^{\frac{\sigma_G-1}{\sigma_G}}\right]^{\frac{\sigma_G}{\sigma_G-1}},$$

where C_{grt} denotes the subutility derived from the real consumption of product group g; G_{rt} is the set of product groups sold at retail firm r; φ_{grt} is the perceived quality of, or the consumer's taste for, product group g at retail firm r; and σ_G is the elasticity of substitution between product groups. In other words, C_{rt} is a composite good composed of the various product groups supplied by retail firm r, adjusted by the substitutability between product groups, where each product group is weighted by its perceived quality.

For the third nest, the subutility from the real consumption of product group g supplied by retail firm r is given as a CES consumption index of products:

$$C_{grt}=\left[\sum_{u\in U_{grt}}(\varphi_{urt}C_{urt})^{\frac{\sigma_{Ug}-1}{\sigma_{Ug}}}\right]^{\frac{\sigma_{Ug}}{\sigma_{Ug}-1}},$$

where C_{urt} denotes the real consumption of product u; U_{grt} is the set of products within product group g supplied by r; φ_{urt} is the perceived quality of, or the consumer's taste for, product u at retail firm r; and σ_{Ug} is the elasticity of substitution between products within product group g. Thus, C_{grt} is a composite good of the various products within product group g supplied by retail firm r.

We assume that the elasticity of substitution at each level, σ_R, σ_G, and σ_{Ug}, is constant over time, and σ_G, and σ_{Ug} are constant across retail firms.

3.1.2 Consumer problem

We assume a representative consumer and solve for his budgeting decisions via backward induction. In the lowest tier of demand, the representative consumer allocates expenditure across products in a given product group at a particular retail firm. The share of consumer spending on products with barcode u in product group g at retail firm r is given by

$$S_{urt}=\frac{\left(\frac{P_{urt}}{\varphi_{urt}}\right)^{1-\sigma_{Ug}}}{\sum_{v\in U_{grt}}\left(\frac{P_{vrt}}{\varphi_{vrt}}\right)^{1-\sigma_{Ug}}},\qquad \sigma_{Ug}>0,\quad \varphi_{vrt}>0, \tag{2}$$

where P_{urt} is the retail price of product u at retail firm r. The corresponding price index for product group g, P_{grt}, is then given by

$$P_{grt}=\left[\sum_{u\in U_{grt}}\left(\frac{P_{urt}}{\varphi_{urt}}\right)^{1-\sigma_{Ug}}\right]^{\frac{1}{1-\sigma_{Ug}}}. \tag{3}$$

With the price indices for each product group known, we solve for the allocation of expenditure across product groups at a given retail firm. The share of consumer spending on product group g at retail firm r is given by

$$S_{grt}=\frac{\left(\frac{P_{grt}}{\varphi_{grt}}\right)^{1-\sigma_G}}{\sum_{k\in G_{rt}}\left(\frac{P_{krt}}{\varphi_{krt}}\right)^{1-\sigma_G}}, \qquad \sigma_G>0, \qquad \varphi_{krt}>0. \tag{4}$$

The corresponding price index for retail firm r is then given by

$$P_{rt}=\left[\sum_{g\in G_{rt}}\left(\frac{P_{grt}}{\varphi_{grt}}\right)^{1-\sigma_G}\right]^{\frac{1}{1-\sigma_G}}. \tag{5}$$

Similarly, the share of consumer spending at retail firm r is given by

$$S_{rt}=\frac{\left(\frac{P_{rt}}{\varphi_{rt}}\right)^{1-\sigma_R}}{\sum_{s\in R_t}\left(\frac{P_{st}}{\varphi_{st}}\right)^{1-\sigma_R}}, \qquad \sigma_R>0, \qquad \varphi_{st}>0. \tag{6}$$

The corresponding price index for the total consumption of the representative consumer at time t is then given by

$$P_t=\left[\sum_{r\in R_t}\left(\frac{P_{rt}}{\varphi_{rt}}\right)^{1-\sigma_R}\right]^{\frac{1}{1-\sigma_R}}.$$

We can now solve for the quantity of each product demanded at each retail firm. The sales of product u at retail firm r are given by

$$E_{urt}=S_{urt}S_{grt}S_{rt}E_t, \tag{7}$$

where E_t is the total expenditure of the representative consumer at time t. The quantity demanded of product u at retail firm r, Q_{urt}, is, by substituting equation (7) into $Q_{urt}=E_{urt}/P_{urt}$ and rewriting, given by

$$Q_{urt}=\varphi_{rt}^{\sigma_R-1}\varphi_{grt}^{\sigma_G-1}\varphi_{urt}^{\sigma_{Ug}-1}E_tP_t^{\sigma_R-1}P_{rt}^{\sigma_G-\sigma_R}P_{grt}^{\sigma_{Ug}-\sigma_G}P_{urt}^{-\sigma_{Ug}}. \tag{8}$$

3.1.3 Normalization of quality parameters

Service quality perceived by consumers can be divided into the perceived service quality relating to each tier of demand. Product-level service quality is the quality of services related to supplying a specific product. This kind of service quality matters especially in the case of food and drink. Even if the quality of the product itself is the same across different retail firms, their efforts to sell items with a longer sell-by date may affect their sales. Next, group-level service quality refers to the quality of services that are common to all products within the same product group. Maintaining the appropriate temperature in the cold drinks shelf is an example of such efforts. Finally, firm-level service quality refers to the quality of services that are common to all products and product groups at a retail firm. Accessibility of a firm's stores from consumers' home or from a station, wide parking spaces, hospitality of the sales staff, and sophisticated shop displays are major examples.

In our framework, we expect these different types of service quality to be reflected in the quality parameters for each tier of demand. However, the values of the parameters cannot necessarily be attributed to service quality alone: the product-level parameter φ_{urt} is determined not only by the quality of the service but also by the quality of the product. Similarly, the group-level parameter φ_{grt} is also affected by consumers' relative tastes with regard to different product groups, which are mostly unrelated to retail firms' efforts.

In order to focus on service quality, we assume that the quality of a product itself is the same across retail firms. Clearly, in practice this is not true for some types of goods. The quality of vegetables, fruit, meat, and fish, for example, differs significantly across stores and sometimes even over time in the same store. However, such goods are the exception in our dataset, since it contains only products with a JAN code, the quality of which is highly standardized in most cases. Similarly, we assume that consumers' tastes with respect to the same product or the same product group are the same across retail firms irrespective of retail firms' efforts. Given these assumptions, we expect that differences across retail firms in the quality parameters for the same product or product group are entirely attributable to differences in product-level or group-level service quality.

However, a simple application of the approach employed by Hottman *et al.* (2016) and Hottman (2016) does not allow us to directly measure these differences. Hottman *et al.* (2016) and Hottman (2016) normalize the quality parameters by setting their geometric means to one. In the context of our paper, such normalization would be expressed as follows:

$$\left(\prod_{g\in G_{rt}} \varphi_{grt}\right)^{\frac{1}{N_{G_{rt}}}}=\left(\prod_{u\in U_{grt}} \varphi_{urt}\right)^{\frac{1}{N_{U_{grt}}}}=1,$$

where N_{Grt} is the number of product groups supplied by retail firm r and $N_{U_{grt}}$ is the number of products supplied by retail firm r that belong to product group g. Under this normalization, the quality ratio of product u sold at retail firm r and r' is given by

$$\varphi_{urt}/\varphi_{ur't}=\frac{P_{urt}/\tilde{P}_{U_{grt}}}{P_{ur't}/\tilde{P}_{U_{gr't}}}\cdot\left(\frac{S_{urt}/\tilde{S}_{U_{grt}}}{S_{ur't}/\tilde{S}_{U_{gr't}}}\right)^{\frac{1}{\sigma_{Ug}-1}},$$

where $\tilde{P}_{U_{grt}}=\left(\prod_{v\in U_{grt}}P_{vrt}\right)^{\frac{1}{N_{U_{grt}}}}$ and $\tilde{S}_{U_{grt}}=\left(\prod_{v\in U_{grt}}S_{vrt}\right)^{\frac{1}{N_{U_{grt}}}}$. That is, the quality ratio is determined by the ratio of relative prices measured by the average price of the products in group g at each firm (the ratio of $P_{urt}/\tilde{P}_{U_{grt}}$ and $P_{ur't}/\tilde{P}_{U_{gr't}}$) and the ratio of relative sales measured by the average sales of the products in group g at each firm (the ratio of $S_{urt}/\tilde{S}_{U_{grt}}$ and $S_{ur't}/\tilde{S}_{U_{gr't}}$). Since the relative prices and relative sales are evaluated in relation to the average prices and sales of all the products in group g sold at each firm, it is not possible to isolate the quality ratio of product u from the prices and sales–and hence indirectly from the product and service quality–of other products.

Therefore, we instead use a benchmark approach in the normalization. We choose a product group, say, group $\bar{g}$, that is sold in the broadest range of retail firms as the benchmark group, and a product in each product group, say, product $\bar{u}$, that is sold in the broadest range of retail firms as the benchmark product of that group. We assume that the service quality related to product $\bar{u}$ and group $\bar{g}$ is the same across the retail firms, namely $\varphi_{\bar{u}rt}=\varphi_{\bar{u}r't}=1$ and $\varphi_{\bar{g}rt}=\varphi_{\bar{g}r't}=1$ for all r and r'.[9] The quality parameters of the other groups and products are determined in relation to the quality of the benchmark group and product. Thus, the quality ratio is given by

$$\varphi_{urt}/\varphi_{ur't}=\frac{P_{urt}/P_{\bar{u}rt}}{P_{ur't}/P_{\bar{u}r't}}\cdot\left(\frac{S_{urt}/S_{\bar{u}rt}}{S_{ur't}/S_{\bar{u}r't}}\right)^{\frac{1}{\sigma_{Ug}-1}},$$

which is determined by the ratio of relative prices measured by the price of the benchmark product $\bar{u}$ at each firm (the ratio of $P_{urt}/P_{\bar{u}rt}$ and $P_{ur't}/P_{\bar{u}r't}$) and the ratio of relative sales measured by the sales of $\bar{u}$ at each firm (the ratio of $S_{urt}/S_{\bar{u}rt}$ and $S_{ur't}/S_{\bar{u}r't}$). Unlike in the approach by Hottman *et al.* (2016) and Hottman (2016), the

[9] These may be strong assumptions for some types of groups and products. However, the benchmark group in our dataset is snack food, and the benchmark product of that group is Calbee's Potato Chips, Lightly Salted, 60g. Thus, it is relatively unlikely that retail firms add substantial value to the quality of this product (group). A shortcoming of our approach is that we cannot evaluate the service quality of retail firms that did not sell the benchmark product in a particular quarter. The total number of such firms is about 28% of all retail firms. The quarterly purchase data of the products that were sold at these firms and are therefore excluded from the following analysis makes up about 15% (about 5 million) of all observations (about 33 million).

product and service quality of other products does not affect the quality ratio of product u. The quality ratio is purely determined by the difference in service quality related to product u.

Similarly, the quality ratio of product group g at retail firm r and r' is given by

$$\frac{\varphi_{grt}}{\varphi_{gr't}}=\frac{P_{\bar{u}grt}/P_{\bar{u}\bar{g}rt}}{P_{\bar{u}gr't}/P_{\bar{u}\bar{g}r't}}\cdot\frac{\left(S_{\bar{u}grt}{}^{\frac{1}{\sigma_{Ug}-1}}S_{grt}{}^{\frac{1}{\sigma_G-1}}\right)/\left(S_{\bar{u}\bar{g}rt}{}^{\frac{1}{\sigma_{U\bar{g}}-1}}S_{\bar{g}rt}{}^{\frac{1}{\sigma_G-1}}\right)}{\left(S_{\bar{u}gr't}{}^{\frac{1}{\sigma_{Ug}-1}}S_{gr't}{}^{\frac{1}{\sigma_G-1}}\right)/\left(S_{\bar{u}\bar{g}r't}{}^{\frac{1}{\sigma_{U\bar{g}}-1}}S_{\bar{g}r't}{}^{\frac{1}{\sigma_G-1}}\right)}.$$

If, for simplicity, we assume for the moment that $\sigma_G=\sigma_{Ug}=\sigma_{U\bar{g}}=2$, we have

$$\frac{\varphi_{grt}}{\varphi_{gr't}}=\frac{P_{\bar{u}grt}/P_{\bar{u}\bar{g}rt}}{P_{\bar{u}gr't}/P_{\bar{u}\bar{g}r't}}\cdot\frac{(S_{\bar{u}grt}S_{grt})/(S_{\bar{u}\bar{g}rt}S_{\bar{g}rt})}{(S_{\bar{u}gr't}S_{gr't})/(S_{\bar{u}\bar{g}r't}S_{\bar{g}r't})},$$

which is determined by the ratio of the relative prices of product $\bar{u}$ measured by the price of the benchmark product $\bar{u}$ of the benchmark group $\bar{g}$ at each firm (the ratio of $P_{\bar{u}grt}/P_{\bar{u}\bar{g}rt}$ and $P_{\bar{u}gr't}/P_{\bar{u}\bar{g}r't}$) and the ratio of relative sales measured by the sales of the benchmark product $\bar{u}$ of the benchmark group $\bar{g}$ at each firm (the ratio of $(S_{\bar{u}grt}S_{grt})/(S_{\bar{u}\bar{g}rt}S_{\bar{g}rt})$ and $(S_{\bar{u}gr't}S_{gr't})/(S_{\bar{u}\bar{g}r't}S_{\bar{g}r't})$).

Finally, the quality ratio of retail firm r and r' is given by

$$\frac{\varphi_{rt}}{\varphi_{r't}}=\frac{P_{\bar{u}rt}(S_{\bar{u}rt})^{\frac{1}{\sigma_{Ug}-1}}(S_{\bar{g}rt})^{\frac{1}{\sigma_G-1}}(S_{rt})^{\frac{1}{\sigma_R-1}}}{P_{\bar{u}r't}(S_{\bar{u}r't})^{\frac{1}{\sigma_{Ug}-1}}(S_{\bar{g}r't})^{\frac{1}{\sigma_G-1}}(S_{r't})^{\frac{1}{\sigma_R-1}}}.$$

If we again assume $\sigma_G=\sigma_{Ug}=\sigma_{U\bar{g}}=2$ for simplicity, we have

$$\frac{\varphi_{rt}}{\varphi_{r't}}=\frac{P_{\bar{u}rt}S_{\bar{u}rt}S_{\bar{g}rt}S_{rt}}{P_{\bar{u}r't}S_{\bar{u}r't}S_{\bar{g}r't}S_{r't}}=\frac{P_{\bar{u}rt}E_{\bar{u}rt}}{P_{\bar{u}r't}E_{\bar{u}r't}}.$$

Since the barcode-level product and service quality of product $\bar{u}$ are the same across retail firms, this expression captures the difference in firm-level quality that homogeneously affects all products within the firm. For example, if retail firm r sells more of product $\bar{u}$ than retail firm r' ($E_{\bar{u}rt}>E_{\bar{u}r't}$) even though the prices are the same or even higher ($P_{\bar{u}rt}\geq P_{\bar{u}r't}$), we assume that firm r's firm-level quality is greater than that of firm r'.

3.2 The retail firm's problem

3.2.1 Technology

We allow the costs of supplying products to the market to vary across barcodes and retail firms. Retail firm r at time t has a variable cost for supplying product u in product group g of

$$V_{urt}(Q_{urt})=z_{urt}Q_{urt}^{1+\delta_g},\quad z_{urt}>0,\quad \delta_g>0,$$

where Q_{urt} is the total quantity supplied of product u by retail firm r; δ_g determines the convexity of marginal costs with respect to output for products in product group g; and z_{urt} is a firm-product-specific shifter of the cost function. Costs are incurred in terms of a composite input that is chosen as the numéraire. One reason why we assume that $\delta_g>0$ is the presence of fixed factors in the retail production function. This type of convex cost function is also generated by inventory-capacity problems (Gallego *et al.*, 2006). The same kind of cost function at the product level is used by Burstein and Hellwig (2007) and Broda and Weinstein (2010). The marginal cost of supplying product u is given by

$$m_{urt}=(1+\delta_g)z_{urt}Q_{urt}^{\delta_g}.$$

In addition, each retail firm must also pay a fixed market access cost of $H_t>0$.

3.2.2 The retail firm's problem

The total profit of retail firm r at time t is thus given by

$$\pi_{rt}=\sum_{g\in G_{rt}}\sum_{u\in U_{grt}}[P_{urt}Q_{urt}-V_{urt}(Q_{urt})]-H_t.$$

In the case of Bertrand competition, each retail firm chooses its prices $\{P_{urt}\}$ to maximize profits. The first order conditions take the following form:

$$Q_{urt}+\sum_{g\in G_{rt}}\sum_{v\in U_{grt}}\left[P_{vrt}\frac{\partial Q_{vrt}}{\partial P_{urt}}-\frac{\partial V_{vrt}(Q_{vrt})}{\partial Q_{vrt}}\frac{\partial Q_{vrt}}{\partial P_{vrt}}\right]=0$$

Solving the first order conditions allowing retail firms to internalize their impact on the price index, the optimal price is given by

$$P_{urt}=\mu_{rt}m_{urt}. \tag{9}$$

μ_{rt} is the markup over marginal cost and is given by

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

where ε_{rt} is retail firm r's perceived elasticity of demand,

$$\varepsilon_{rt}=\sigma_R-(\sigma_R-1)S_{rt}. \tag{10}$$

In the case of Cournot competition, the perceived elasticity becomes

$$\varepsilon_{rt}=\frac{1}{1/\sigma_R-(1/\sigma_R-1)S_{rt}}. \tag{11}$$

One of the surprising features of this setup is that markups prove to be the same across all products sold by retail firm r. This is a generic property of nested demand systems. The intuition is that the retail firm internalizes that it is the monopoly supplier of its real output, which in our model equals real consumption of the retail firm's bundle of product groups, C_{rt}, and products, C_{grt}. Hence, the retail firm's profit maximization problem can be thought to consist of two stages. First, the retail firm chooses the price index (P_{rt} and P_{grt}) to maximize its profits from supplying C_{rt} and C_{grt}, which implies a markup at the firm level over the cost of supplying real output. Second, the retail firm chooses the price of each product to minimize the cost of supplying real output within the product group (C_{grt}), which requires setting the relative prices of products equal to their relative marginal costs. Together, these two results ensure the same markup across all product groups and products sold by the retail firm. Nonetheless, markups vary across retail firms.

In the Bertrand case, retail firms with a higher market share (S_{rt}) face a lower perceived elasticity of demand and thus set a higher markup, as in previous studies (Atkeson and Burstein, 2008; Edmond *et al.,* 2012; Hottman, 2016). Although consumers have constant elasticity of substitution preferences (σ_R), each firm perceives a variable elasticity of demand (ε_{rt}), which is decreasing in its market share (S_{rt}). As a result, the retail firm's equilibrium pricing rule (9) involves a variable markup that is increasing in its market share. For a positive equilibrium price (9), the perceived elasticity of demand (ε_{rt}) needs to be greater than one (retail firm supply substitutes), which requires the elasticity of substitution between retail firms (σ_R) to be sufficiently large. As a firm's market share approach zero, the markup approaches the standard CES markup of $\mu_{rt}=\frac{\sigma_R}{\sigma_R-1}$.

3. 3 Decomposition of retail firms' sales

In this section, we use the model to quantify the contribution of the different sources of retail firm heterogeneity to the dispersion in sales across retail firms. The nominal sales of retail firm r, E_{rt}, are given by

$$E_{rt}=\sum_{g\in G_{rt}}\sum_{u\in U_{grt}}P_{urt}Q_{urt}.$$

Using the quantity demanded of each product at each retail firm (8), the price indexes (3) and (5), and the equilibrium pricing rule (9), and taking the logarithm, we obtain

$$\ln E_{rt} = \ln E_t + (\sigma_R - 1)\ln P_t + (\sigma_R - 1)\ln \varphi_{rt} + \left(\frac{\sigma_R - 1}{\sigma_G - 1}\right)\ln N_{Grt} - (\sigma_R - 1)\ln m_{rt} - (\sigma_R - 1)\ln \mu_{rt}, \qquad (12)$$

where m_{rt} is a marginal cost index of retail firm r, which is defined as

$$m_{rt} = \left[\frac{1}{N_{Grt}} \sum_{g \in G_{rt}} \left(\frac{m_{grt}}{\varphi_{grt}}\right)^{1-\sigma_G}\right]^{\frac{1}{1-\sigma_G}},$$

where $m_{grt} = \left[\sum_{u \in U_{grt}} \left(\frac{m_{urt}}{\varphi_{urt}}\right)^{1-\sigma_{Ug}}\right]^{\frac{1}{1-\sigma_{Ug}}}$. m_{rt} is a composite index of marginal costs weighted by both the quality of each product (φ_{urt}) and the quality of each product group (φ_{grt}).

Equation (12) decomposes the retail firm's sales into seven terms that capture various factors through which firms can differ in sales. The first two terms capture macro- or semi-macroeconomic factors, namely, the size of the market (E_t) for and price level (P_t) of the commodities that all retail firms supply. Our demand system is homogeneous of degree one in consumer spending at each retail firm, so firm' sales rise one to one with an increase in aggregate expenditures. The second term captures the impact of the price index for total consumption, which summarizes the prices of the output of competing retail firms. An increase in the price index of one percent will cause the firm's sales to rise by $(\sigma_R - 1)$ percent. Here, the elasticity of substitution between retail firms plays a crucial role in determining how much a price movement affects a retail firm's sales.

The next two terms capture the impact of the quality aspects of each retail firm. The third term captures the impact of firm-level service quality (φ_{rt}). Again, the elasticity of substitution between retail firms determines how much the movement of overall quality affects a retail firm's sales.

The fourth term captures the contribution of product scope at the product-group level. Let us first consider two retail firms that provide the same level of overall service quality ($\varphi_{rt} = \varphi_{r't}$), but one retail firm supplies more product groups than the other ($N_{Grt} > N_{Gr't}$). For simplicity, assume all product groups and all products are of identical quality ($\varphi_{gst} = \varphi_{ust} = 1$ for all $g \in G_{st}$ and $u \in U_{gst}$) and have identical marginal costs ($z_{urt} = z$ for all $g \in G_{st}$ and $u \in U_{gst}$) for both retail firms ($s = r, r'$). For example, if consumers treated all product groups identically regardless of which retail firm supplies them, i.e., $\sigma_G = \sigma_R$, retail firm r would sell $\ln(N_{Grt}/N_{Gr't})$ percent more than retail firm r'. More generally, if the product groups supplied by a retail firm are more substitutable for one another than for those of other retail firms, i.e., $\sigma_G > \sigma_R$, the percentage gain in sales accruing to a retail firm that adds a product group will be less than one, reflecting the fact that the new product group will cannibalize the sales of its existing product groups.

The sixth and seventh terms capture the contribution of the firm-level relative price, which is decomposed into the marginal cost (m_{rt}) and the markup (μ_{rt}). Again, the elasticity of substitution between retail firms determines how much the movement of marginal costs and markups affects firm sales.

In order to remove the effect of aggregate expenditure and price movements, and thereby focus on the cross-sectional determinants of a retail firm's sales and sales growth, we first decompose a retail firm's sales relative to the average sales of the retail industry:

$$\Delta^R \ln E_{rt} = (\sigma_R - 1)\Delta^R \ln \varphi_{rt} + \left(\frac{\sigma_R - 1}{\sigma_G - 1}\right)\Delta^R \ln N_{G_{rt}} - (\sigma_R - 1)\Delta^R \ln m_{rt} - (\sigma_R - 1)\Delta^R \ln \mu_{rt} \tag{13}$$

where Δ^R is the difference operator relative to the geometric mean for all retail firms of all retail categories, such that $\Delta^R \ln E_{rt} = \ln E_{rt} - (1/N_{Rt}) \sum_{k \in R_t} \ln E_{kt}$, where N_{Rt} is the number of retail firms at time t.

We can now decompose the cross-sectional variation in retail firms' sales using a procedure analogous to the variance decomposition developed by Eaton et al. (2004) and widely used in the international trade literature. Specifically, we regress each of the components of the log of a retail firm's sales in the decomposition (13) on the log of the retail firm's sales as follows:

$$(\sigma_R - 1)\Delta^R \ln \varphi_{rt} = \alpha^{\varphi} \Delta^R \ln E_{rt} + \varepsilon^{\varphi}_{rt} \tag{14a}$$

$$\left(\frac{\sigma_R - 1}{\sigma_G - 1}\right)\Delta^R \ln N_{G_{rt}} = \alpha^{N_G} \Delta^R \ln E_{rt} + \varepsilon^{N_G}_{rt} \tag{14b}$$

$$-(\sigma_R - 1)\Delta^R \ln m_{rt} = \alpha^{m} \Delta^R \ln E_{rt} + \varepsilon^{m}_{rt} \tag{14c}$$

$$-(\sigma_R - 1)\Delta^R \ln \mu_{rt} = \alpha^{\mu} \Delta^R \ln E_{rt} + \varepsilon^{\mu}_{rt} \tag{14d}$$

We allow the coefficients (α^{φ}, α^{N_G}, α^{m}, α^{μ}) to differ across retail categories. Due to the properties of ordinary least squares (OLS) estimation, this decomposition allocates the covariance terms between the components of retail firms' sales equally across those components, and implies $\alpha^{\varphi} + \alpha^{N_G} + \alpha^{m} + \alpha^{\mu} = 1$. The values for each of the α's provide us with a measure of how much of the variation in retail firms' sales can be attributed to each component.

4. Structural Estimation

Our structural estimation has two components for each tier. First, given the elasticity of substitution at each tier (σ_{Ug}, σ_G, σ_R) and the data on expenditure shares and prices (S_{urt}, S_{grt}, S_{rt}, P_{urt}), we show how we can recover the unobserved service quality (φ_{urt}, φ_{grt}, φ_{rt}). Second, given the elasticity of substitution across retail firms, we show how we recover the unobserved markups and marginal costs. Finally, we explain the strategy for estimating the elasticity of substitution at each tier of demand.

4.1 Recovering Unobserved Service Quality, Retail Markups, and Marginal Costs

4.1.1 Quality

Consider the lowest-tier of the demand system. Given the elasticity of substitution between products, σ_{Ug}, and the observed data on product-level expenditure shares and prices (S_{urt}, P_{urt}), the CES expenditure share of each product (equation (2)), together with $\varphi_{\bar{u}rt}=1$, the unobserved quality of each product is given by

$$\varphi_{urt}=\frac{P_{urt}}{P_{\bar{u}rt}}\cdot\left(\frac{S_{urt}}{S_{\bar{u}rt}}\right)^{\frac{1}{\sigma_{Ug}-1}}.$$

Substituting the unobserved quality and observed barcode price of each product into equation (3) yields the price indices for each product group, P_{grt}. Similarly, given the elasticity of substitution between product groups, σ_G, the observed data on group expenditure shares, S_{grt}, and the estimated price indices for each product group, P_{grt}, the CES expenditure share of each product group (equation (4)) yields the unobserved quality of each product group, φ_{grt}. The price index for each retail firm, P_{rt}, and the unobserved retail-level service quality, φ_{rt}, can be calculated through similar steps.

4.1.2 Retail Markups and Marginal Costs

Given the elasticity of substitution between retail firms, σ_R, equation (10) or (11) yields the retail firm's perceived elasticity of demand, ε_{rt}, which can then be used to compute the retail firm's markup, μ_{rt}. Marginal costs, m_{urt}, can then be computed from the observed barcode price using equation (9).

4. 2 Estimating the elasticities of substitution

We now explain our methodology for estimating the elasticities of substitution at each tier ($\sigma_{Ug}, \sigma_G, \sigma_R$) and the elasticity of marginal costs (δ_g).

4.2.1 Lower tier of demand

Estimation of the elasticity of substitution in the lowest tier follows the approach employed by Broda and Weinstein (2006, 2010), which in turn is based on Feenstra (1994). The identification is as follows. In our setting, the slopes of the demand and supply curves for a given product group, σ_{Ug} and δ_g, are assumed to be constant across products and over time, but their intercepts are allowed to vary across products and time. As Leontief (1929) points out, if the supply and demand intercepts for a given product are orthogonal, there is a rectangular hyperbola in (σ_{Ug}, δ_g) space which best fits the observed price and share data of that product. The orthogonality assumption alone does not provide identification: a higher value of σ_{Ug} but lower value of δ_g, for example, will keep the expectation at zero. If the variances of the supply and demand intercepts are heteroskedastic across products in a product group, then the hyperbolas that fit the data are different for each product. Since the slopes of the demand and supply curves are the same, the intersection of the hyperbolas of the different products in the product group separately identifies the demand and supply elasticities (Feenstra 1994).

The rest of this subsection defines the orthogonality conditions for each product in terms of its double-differenced supply and demand intercepts and outlines the generalized method of moments (GMM) procedure for estimating the slopes of the demand and supply curves for each product group.

We take the double-difference of the log of the expenditure share of each product (equation (2)) over time and relative to the average share of all products in the same product group supplied by retail firm r:

$$\Delta^{U,t} \ln S_{urt} = -(\sigma_{Ug}-1)\,\Delta^{U,t} \ln P_{urt} + \omega_{urt},$$

where the unobserved error term is $\omega_{urt} = -(\sigma_{Ug}-1)(\Delta^t \overline{\ln \varphi_{urt}} - \Delta^t \ln \varphi_{urt})$. $\Delta^{U,t}$ is the double-difference operator across products and over time such that $\Delta^{U,t} \ln S_{urt} = \Delta^t \ln S_{urt} - \Delta^t \overline{\ln S_{urt}}$ and $\overline{\ln S_{urt}} = (1/N_{U_{grt}}) \sum_{v \in U_{grt}} \ln S_{vrt}$, while Δ^t is the first-difference operator over time such that $\Delta^t \ln S_{urt} = \ln S_{urt} - \ln S_{urt-1}$. Since we difference the product expenditure share relative to the average share, we eliminate all demand shocks that are common across products in the same product group supplied by retail firm r, which leaves only demand shocks that affect the sales of a certain product relative to other products.

Transforming the equilibrium pricing rule (9) using $Q_{urt}=(S_{urt}\sum_{v\in U_{grt}}E_{vrt})/P_{urt}$ and taking the double-difference of its log over time and relative to the average price of all products in the same product group supplied by retail firm r, we obtain

$$\Delta^{U,t}\ln P_{urt}=\frac{\delta_g}{1+\delta_g}\Delta^{U,t}\ln S_{urt}+\kappa_{urt},$$

where the unobserved error term is $\kappa_{urt}=\frac{1}{1+\delta_g}(\Delta^t\ln z_{urt}-\Delta^t\overline{\ln z_{urt}})$. Note that since the retail markup μ_{rt} has been differenced out through the double-difference operation, our estimation approach is the same under either price or quantity competition, and hence is robust across these different forms of competition.

Following Broda and Weinstein (2006), the orthogonality of the double-differenced demand and supply shocks defines the set of moment conditions for each product:

$$G(\beta_g)=\mathbb{E}_{\mathrm{T}}[x_{urt}(\beta_g)]=0,$$

where $\beta_g=\begin{pmatrix}\sigma_{Ug}\\ \delta_g\end{pmatrix}$, $x_{urt}=\omega_{ut}\kappa_{urt}$, and $\mathbb{E}_{\mathrm{T}}$ is the expectations operator over time.

This condition assumes orthogonality of idiosyncratic demand and supply shocks at the product level. This orthogonality is plausible for the following reasons. First, differencing across products within the same product group supplied by retail firm r eliminates common firm-level shocks (e.g., changes in management) and common group-level shocks (e.g., renewal of the wine cellar) that affect both costs and appeal (quality or taste) across all products within the firm or within the product group. Similarly, differencing over time eliminates time-invariant heterogeneity between products caused by different production technologies, which could affect both costs and appeal (quality or taste) in all time periods. Therefore our double differencing nets out both these types of shocks. In other words, our identification is based only on relative differences in the demand for and supply of individual products. Another main potential threat to identification is a change in observable product characteristics that affects both relative costs and relative appeal, but this endogeneity concern does not arise with barcode data. Fortunately, any substantive change in product characteristics is accompanied by the introduction of a new bar code. Therefore, by using the variation within bar codes over time, we hold constant observable product characteristics. As a result, it is much harder to think of reasons why double-differenced changes in costs that leave the observable characteristics of a product constant (e.g., changes in buying-in prices caused by a strike at a manufacturer or changes in raw material prices) should affect double-differenced consumer demand for that product conditional on price.

For each product group, we stack all the moment conditions to form the GMM object function and obtain

$$\hat{\beta}_g = \underset{\beta_g}{\operatorname{argmin}} \left\{ G^*(\beta_g)' \cdot W \cdot G^*(\beta_g) \right\} \quad \forall g,$$

where $G^*(\beta_g)$ is the sample analog of $G(\beta_g)$ and W is a positive definite weighting matrix. As in Broda and Weinstein (2006), we weight the data for each product by the number of consumers who bought the product in each quarter.[10]

4.2.2 Middle tier of demand

We take the double-difference of the log of the product group expenditure share (equation (4)) over time and relative to the average share of all product groups supplied by retail firm r:

$$\Delta^{G,t} \ln S_{grt} = -(\sigma_G - 1)\, \Delta^{G,t} \ln P_{grt} + \omega_{grt}, \qquad (15)$$

where the unobserved error term is $\omega_{grt} = -(\sigma_G - 1)\Delta^{G,t} \ln \varphi_{grt}$. $\Delta^{G,t}$ is the double-difference operator across product groups and over time such that $\Delta^{G,t} \ln S_{grt} = \Delta^t \ln S_{grt} - \Delta^t \overline{\ln S_{grt}}$ and $\overline{\ln S_{grt}} = (1/N_{Grt}) \sum_{h \in G_{rt}} \ln S_{hrt}$, while Δ^t is the first-difference operator over time such that $\Delta^t \ln S_{grt} = \ln S_{grt} - \ln S_{grt-1}$.

OLS estimation of equation (15) may be biased due to endogeneity, since the unobserved error term is likely to be correlated with the double-differenced product group price index. This correlation occurs because a relative increase in product group quality raises the quantity demanded of the products within the product group and thus raises the product group price index, since product supply curves are upward sloping.

We therefore estimate σ_G using an instrumental variable (IV) approach. The double-differenced product group price index can be written as:

[10] Since product-level data for individual firms is sometimes thin and intermittent, we use only the quarterly purchase data of products that were sold in at least 10 sequential quarters of the total 19 quarters for the GMM estimation. As a result, the quarterly purchase data of products that we use here make up about 15% (about 5 millions) of all observations (about 33 millions).

$$\Delta^{G,t}\ln P_{grt}=\Delta^{G,t}\ln\tilde{P}_{U_{grt}}-\Delta^{G,t}\left[\frac{1}{\sigma_{Ug}-1}\ln\left(\sum_{u\in U_{grt}}\frac{S_{urt}}{\tilde{S}_{U_{grt}}}\right)\right]-\Delta^{G,t}\ln\left(\prod_{u\in U_{grt}}\varphi_{urt}\right)^{\frac{1}{N_{U_{grt}}}}. \quad (16)$$

The first term on the right-hand side of equation (16) is the natural log of the geometric mean of product prices within the product group. This term is the reason why the product group price index is correlated with the error term in equation (15). That is, increases in product group prices from movements along upward sloping product supply curves due to increases in product group demand are fully captured in this term. The log component of the second term is a variant of the Theil index of dispersion,[11] which reflects the dispersion of the shares of individual products within the product group. If the shares of all products are equal, the term will equal ln $N_{U_{grt}}$, which is increasing in the number of products within product group g supplied by retail firm r. Since the term is multiplied by $-1/(\sigma_{Ug}-1)\leq 0$, the product group price index falls as the number of products supplied by the retail firm rises. The product group index also falls as the dispersion of market shares across products within the product group increases.

We estimate σ_G using the second term of equation (16) as an instrument for the product price index in equation (15). The moment condition for the instrumental variables is:

$$\mathbb{E}\left\{\omega_{grt}\cdot\Delta^{G,t}\left[\frac{1}{\sigma_{Ug}-1}\ln\left(\sum_{u\in U_{grt}}\frac{S_{urt}}{\tilde{S}_{U_{grt}}}\right)\right]\right\}=0.$$

4.2.3 Upper tier of demand

We take the double-difference of the log of the retail firm's expenditure share (equation (6)) over time and relative to the average share of all retail firms:

$$\Delta^{R,t}\ln S_{rt}=-(\sigma_R-1)\,\Delta^{R,t}\ln P_{rt}+\omega_{rt}, \quad (17)$$

where the unobserved error term is $\omega_{rt}=-(\sigma_R-1)\Delta^{R,t}\ln\varphi_{rt}$. $\Delta^{R,t}$ is the double-difference operator across retail firms and over time such that $\Delta^{R,t}\ln S_{rt}=\Delta^{t}\ln S_{rt}-\Delta^{t}\overline{\ln S_{rt}}$ and $\overline{\ln S_{rt}}=(1/N_{Rt})\sum_{s\in R_t}\ln S_{st}$, while Δ^{t} is the first-difference operator over time such that $\Delta^{t}\ln S_{rt}=\ln S_{rt}-\ln S_{rt-1}$.

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

As in the middle tier of demand, we use an IV approach for the estimation of σ_R. The retail firm's price index can be written as:

$$\begin{aligned}\Delta^{R,t}\ln P_{rt} = & -\frac{1}{\sigma_G - 1}\Delta^{R,t}\ln\left(\sum_{g\in G_{rt}}\frac{S_{grt}}{\tilde{S}_{Grt}}\right) \\ & -\Delta^{R,t}\left\{\frac{1}{N_{Grt}}\sum_{g\in G_{rt}}\left[\frac{1}{\sigma_{Ug}-1}\ln\left(\sum_{u\in U_{grt}}\frac{S_{urt}}{\tilde{S}_{Ugrt}}\right)\right]\right\} \\ & +\frac{1}{N_{Grt}}\Delta^{R,t}\ln\prod_{g\in G_{rt}}\left[\tilde{P}_{Urt}\cdot\left(\prod_{u\in U_{grt}}\varphi_{urt}\right)^{-\frac{1}{N_{Ugrt}}}\right] \\ & -\Delta^{R,t}\ln\left(\prod_{g\in G_{rt}}\varphi_{grt}\right)^{\frac{1}{N_{Grt}}}.\end{aligned} \tag{18}$$

We use the sum of the first two terms on the right-hand side of equation (18) as an instrument for the retail firm's price index in equation (17). The moment condition for the instrumental variables is

$$\mathbb{E}\left\langle\omega_{grt}\cdot\Delta^{R,t}\left\{\frac{1}{\sigma_G-1}\ln\left(\sum_{g\in G_{rt}}\frac{S_{grt}}{\tilde{S}_{Grt}}\right)+\frac{1}{N_{Grt}}\sum_{g\in G_{rt}}\left[\frac{1}{\sigma_{Ug}-1}\ln\left(\sum_{u\in U_{grt}}\frac{S_{urt}}{\tilde{S}_{U_{grt}}}\right)\right]\right\}\right\rangle=0.$$

5. Estimation Results

5.1 Estimated Elasticities of Substitution

Table 5 shows that products and product groups supplied by the same firm are imperfect substitutes. At the product level, the estimated elasticity of substitution ranges from 27.56 at the 5th percentile to 4.76 at the 95th percentile with a median elasticity of 10.57. The elasticities of products below the 10th percentile are smaller than the firm-level elasticity, reflecting the fact that products supplied by different retail firms are closer substitutes than products supplied by the same retail firm. The

Table 5：Distribution of estimated elasticities of substitution

Ranked percentile	σ_{Ug}	σ_G	σ_R
1%	47.79		
5%	27.56		
10%	20.17		
25%	15.51		
50%	10.57	18.95	20.78
75%	8.24		
90%	6.15		
95%	4.76		
99%	2.86		

median elasticity implies that a one percent price cut leads to a rise in the sales of that product of about 11%. Similarly, a one percent overall price cut in a product group (or at a retail firm) leads to a rise in the sales of that group (or retail firm) of about 19% (or 21%).

5.2 Group-and Product-level Quality

Table 6 (a) presents the correlations between the relative level of sales, prices, quality, and marginal costs across the products supplied by each firm, where Δ^U denotes the first difference relative to the geometric mean for the product group supplied by each firm. The table provides several important insights. First, there is a strong positive correlation between relative product-level quality ($\Delta^U \ln \varphi_{urt}$) and relative marginal costs ($\Delta^U \ln m_{urt}$). Specifically, the correlation is 0.978, meaning that on average it is costlier for retail firms to sell more appealing products. This high correlation is not that surprising, since much of the variation in appeal across a single retail firm's products probably reflects procurement differences including wholesale prices that are reflected in marginal costs.

Next, Table 6 (b) is useful for understanding the determinants of retail firms' sales. While relative firm-level quality ($\Delta^R \ln \varphi_{rt}$) is moderately correlated with relative firm sales ($\Delta^R \ln E_{rt}$), the correlation between relative marginal costs ($\Delta^R \ln m_{rt}$) and

Table 6 (a): Sales component correlations (Product level)

	$\Delta^U \ln E_{urt}$	$\Delta^U \ln P_{urt}$	$\Delta^U \ln \varphi_{urt}$	$\Delta^U \ln m_{urt}$
$\Delta^U \ln E_{urt}$	1			
$\Delta^U \ln P_{urt}$	0.3224	1		
$\Delta^U \ln \varphi_{urt}$	0.4908	0.978	1	
$\Delta^U \ln m_{urt}$	0.3224	1	0.978	1

Table 6 (b): Sales component correlations (Firm level)

	$\Delta^R \ln E_{rt}$	$\Delta^R \ln P_{rt}$	$\Delta^R \ln \varphi_{rt}$	$\Delta^R \ln N_{grt}$	$\Delta^R \ln \mu_{rt}$	$\Delta^R \ln m_{rt}$
$\Delta^R \ln E_{rt}$	1					
$\Delta^R \ln P_{rt}$	-0.228	1				
$\Delta^R \ln \varphi_{rt}$	0.2761	0.8729	1			
$\Delta^R \ln N_{grt}$	0.8765	-0.2394	0.2029	1		
$\Delta^R \ln \mu_{rt}$	0.5535	-0.068	0.2103	0.357	1	
$\Delta^R \ln m_{rt}$	0.1644	0.9188	0.9893	0.1309	0.1404	1

relative sales is smaller. How can it be that marginal costs are strongly associated with firm-level quality, which in turn is positively correlated with sales, but marginal costs are correlated with sales to a lesser extent? There are two forces at work in these correlations. First, there is the direct effect of marginal costs on sales through the price channel: higher marginal costs lead to higher prices and lower sales. However, we also saw that the correlation between quality and marginal costs is positive, which arises from the fact that higher marginal costs are associated with higher quality service, which tends to be associated with higher sales. The fact that these forces tend to cancel each other out explains why sales are correlated with quality on the one hand and with marginal costs to a lesser extent on the other.

5.3 Decomposing Firm Sales

Table 7 presents the cross-sectional decomposition described by equations (14a) to (14d). The table suggests the quality of a retail firm is an important determinant of its size. The variance decomposition indicates that 57% of the overall size distribution can be attributed to firm-level service quality. 26% of the distribution is attributable to the variety of product groups offered: larger firms differentiate themselves by offering a broader range of product groups they supply. Thus, total firm-level service quality including product variety accounts for 83% of the variation in firm sales. The markup plays almost no role in determining firm size, since markup rates do not vary very much across retail firms. The marginal cost index term accounts for 17% of the overall size distribution.[12] It is worth noting that costs play a smaller role than quality in determining firm size.

Table 7：Variance decomposition

	Quality	Scope	Markup	Average MC
All firms	0.5726	0.2573	-0.0018	0.1720
(Std. error)	(0.0191)	(0.0016)	(0.00003)	(0.0190)

5.4 Firm Output and Productivity

All productivity estimates are based on the concept of real output, which in theory equals nominal output divided by a price index. However, which price index should be

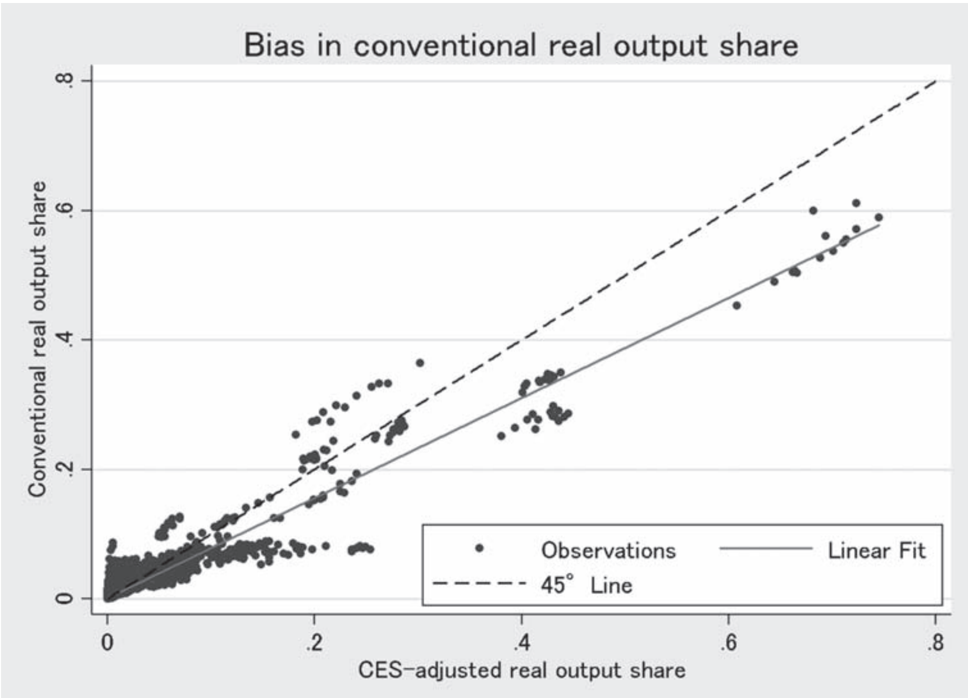
[12] A positive coefficient means that larger firms have lower costs.

chosen is not arbitrary: it is determined by the utility function. In other words, the concept of real output is not independent of the demand system, so that all attempts to measure productivity based on the real output concept contain an implicit assumption about the structure of the demand system.

Given these considerations, we compare retail firms' real output using two types of price index as deflators. The first is a conventional price index, $\widehat{P}_{rt}$, which is the quantity-weighted average price of a firm's output. The second is the CES price index, which 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 $\varphi_{rt} C_{rt}$, so the corresponding expenditure function for a unit of consumption is P_{rt}/φ_{rt}. We can write the conventional measure of real output, $\widehat{Q}_{rt}$, as $E_{rt}/\widehat{P}_{rt}$, and the CES measure of real output, Q_{rt}, as $E_{rt}/(P_{rt}/\varphi_{rt})$.

Figure 1 plots these two measures of real output. The dotted line is the 45° line. Note that we standardize the two measures by working with the unitless share of each firm (i.e., $Q_{rt}/\sum_{s\in R_t} Q_{st}$ and $\widehat{Q}_{rt}/\sum_{s\in R_t} \widehat{Q}_{st}$, where R_t is the set of retail firms). As one can see, the choice of price index matters enormously for the computation of real output. If we

Figure 1：Bias in conventional real output share



regress $\widehat{Q}_{rt}/\sum_{s\in R_f}\widehat{Q}_{st}$ on $Q_{rt}/\sum_{s\in R_f}Q_{st}$, we obtain an R^2 of 0.89 with a coefficient of 0.75 (s.e. 0.0062).

These results imply that more than 10% of the variation in firm-level real output obtained using the conventional price index is simply due to the fact that the conventional price index is not quality-adjusted and assumes that firms produce homogeneous output: an assumption that can be easily rejected. Moreover, using the conventional price index understates the real output of large firms by a quarter relative to small firms.

6. Conclusion

Taking retail services as an example, this study examined the contribution of service quality to the output of service companies in Japan. Specifically, it looked at how much of the variation in retail firms' sales is determined by quality factors including the variety of products offered, and how the variation in real output would change if output was deflated using price indices that incorporate quality factors instead of conventional price indices.

Due to the economic characteristics of services and the lack of appropriate data, quantitative analyses of service quality face methodological difficulties. We addressed these difficulties by (1) using a massive dataset of barcode-level purchase records, (2) developing a structural model of consumer demand for the real output of each retail firm to estimate structural parameters that gauge consumers' implicit valuation of service quality, and (3) introducing a benchmark approach in the normalization of the quality parameters to assess the contribution of service quality independent from product quality.

Our results show that 57% of the variation in retail firm sales is attributable to differences in firm-level service quality and 26% to differences in product group variety. A possible interpretation of the results is that most of the large contribution of firm-level service quality reflects differences in the convenience of access to the stores of retail chains across Japan. This may seem odd from the perspective of individual consumers, who typically choose retail stores close to where they live. There may be many cases where stores belonging to local retail firms are located in areas as convenient as those belonging to national retail chains. For example, assume there are three retail firms, A, B, and C that have stores in two cities, D and E. Firm A is a national retail chain that has a store in each of the two cities, while firms B and C are local retail firms and firm B has a store only in city D and firm C has a store only in city E. Assume further that the stores are identical in terms of service quality and the products they provide, and consumers in the two cities have identical preferences and budgets. Then, the market share of A would be twice as large as that of the other two firms, although the service quality of the three firms for local consumers is exactly the same.

However, mainly due to the shortage of purchase data at the store level, our model does not reflect differences across consumers in terms of where they make their purchases. Instead, it is based on a nationwide representative consumer, who chooses retail stores across the country. What matters for this representative consumer is nationwide store access. The fact that national retail chains have branches across the country, especially in areas of high population density, itself can, therefore, be regarded as an important component of firm-level service quality. We leave analyses reflecting heterogeneity of consumers, especially differences across consumers in terms of where they make their purchases, as a subject of future studies.

Finally, the comparison of retail firms' real output deflated by a conventional price index and a CES price index showed that more than 10% of the variation in firm-level real output obtained using the conventional price index is simply due to the fact the conventional price index is not quality-adjusted and assumes that firms produce homogeneous output. Moreover, estimates based on the conventional price index understate the real output of large firms by a quarter relative to small firms. In sum, our results suggest that the use of conventional price indexes overstates the price level (and understates real output) to a greater extent for large retail firms with higher quality and a broader product scope than for small firms with lower quality and a narrower product scope.

References

Atkeson, Andrew and Ariel Burstein (2008). Pricing to market, trade costs, and international relative prices, *American Economic Review,* 98 (5), pp. 1998-2031.

Berry, Steven T. (1994). Estimating discrete-choice models of product differentiation, *The RAND Journal of Economics*, 25 (2), pp. 242-262.

Broda, Christian and David E. Weinstein (2006). Globalization and the gains from variety, *The Quarterly Journal of Economics*, 121 (2), pp. 541-585.

Broda, Christian and David E. Weinstein (2010). Product creation and destruction: Evidence

and price Implications, *American Economic Review*, 100 (3), pp. 691–723.
Burstein, Ariel and Christian Hellwig (2007). Prices and market shares in a nenu cost model, NBER Working Paper, 13455.
Eaton, Jonathan, Samuel Kortum, and Francis Kramarz. (2004). Dissecting trade: Firms, Industries, and export destinations, *American Economic Review*, 94 (2), pp. 150–154.
Edmond, Chris, Virgiliu Midrigan, and Daniel Yi Xu. (2012). Competition, markups, and the gains from international trade, NBER Working Paper, 18041.
Feenstra, Robert C. (1994). New product varieties and the measurement of international prices, *American Economic Review*, 84 (1), pp. 157–177.
Feenstra, Robert C. (2014). Restoring the product variety and pro-competitive gains from trade with heterogeneous firms and bounded productivity, NBER Working Paper, 19833.
Gallego, Guillermo, Woonghee Tim Huh, Wanmo Kang, and Robert Phillips. (2006). Price competition with the attraction demand model: Existence of unique equilibrium and its stability, *Manufacturing and Service Operations Management*, 8 (4), pp. 359–375.
Hottman, Colin, Stephen J. Redding, and David E. Weinstein. (2016). Quantifying the sources of firm heterogeneity, *The Quarterly Journal of Economics*, 131 (3), pp. 1291–1364.
Hottman, Colin. (2016). Retail markups, misallocation, and store variety in the US, mimeo, available online: http: //www.columbia.edu/~cjh2164/HottmanJMP.pdf.
Leontief, Wassily. (1929). Ein Versuch zur Statistischen Analyse von Angebot und Nachfrage, *Weltwirtschaftliches Archiv*, 30 (1), pp. 1–53.
Manova, Kalina and Zhiwei Zhang. (2012). Export prices across firms and destinations, *The Quarterly Journal of Economics*, 127 (1), pp. 379–436.
Matsuura, Toshiyuki and Mitsuru Sunada. (2009). Kourigyo ni okeru Kyoso to Shohisha Kosei no Keisoku [Competition in retail industry and measurement of consumer welfare], CPRC Discussion Paper Series, 40-J, Competition Policy Research Center, Japan Fair Trade Commission.
Melitz, Marc J. (2003). The impact of trade on intra-industry reallocations and aggregate industry productivity, *Econometrica*, 71 (6), pp. 1695–1725.