Discussion Paper No. 2018-001

The Contribution of Unobserved Contemporaneous Preference Structures on Consumer Spending Using Micro-Panel Data

Market Quality Discussion Series

Project by
Complex Dynamic Analysis on Economic Crisis and Social Infrastructure
Grant-in-Aid for Specially Promoted Research (# 23000001)

market-quality.net
THE MARKET QUALITY ECONOMICS PORTAL

The Contribution of Unobserved Contemporaneous Preference Structures on Consumer Spending Using Micro-Panel Data

MQ Discussion Series 2018-001

(Updated, August 2022)

Masahiro Sato [a]*, Makoto Yano [b]

[a] Tohoku University, 41 Kawauchi, Aoba-ku, Sendai, Miyagi 9808576, Japan

[b] Research Institute of Economy, Trade and Industry (RIETI), 1-3-1, Kasumigaseki Chiyoda-ku, Tokyo, 100-8901 Japan

* Corresponding author, TEL: +81-90-9855-1674, E-mail: contact@intergreen.jp (M. Sato)

Abstract

This paper examines the contribution of unobserved contemporaneous preference structures over different products on consumer spending. Previous studies that dealt with idiosyncratic preference shocks only focus on time preference of individual consumers. But for variety-loving consumers, the differences in preferences structure over various products between consumers and its change over times might affect consumer spending through different income and price elasticity. We use micro panel data on barcode-level purchase records to recover unobserved preference parameters of consumers belonging to different clusters by estimating a structural model of consumer demand. Then a fixed effects model estimation is conducted on the micro panel data. The results show that using the price index adjusted by contemporaneous preference structure improve the explanatory power of the model significantly compared to other reference price indices such as Stone price index with true price, Stone price index with product-group CPI, and all item CPI.

JEL classification:

Keywords: micro-panel data, structural estimation, big data,

1. Introduction

In this paper, we examine the contribution of within-period preference structures over multiple products on inter-temporal consumption dynamics. Since a pioneering study of Hall & Mishkin (1982), a large body of empirical literature has examined the life-cycle model and permanent income hypothesis using micro data (e.g. Attanasio and Weber, 1995; Cochrane, 1991; Dynan, 2000; Attanasio and Borella, 2014). But none of these studies, with a very few exceptions (Blundell, Browning, and Meghir, 1994; Attanasio and Weber, 1995), have investigated the correspondence between within-period preference structures of households over multiple products and their inter-temporal consumption dynamics. Most of the studies assume that households allocate consumption of a single identical good over time. What only matters here is an intertemporal preference structure, which is captured by a time preference parameter.

However, as we will show bellow, the correspondences of a household's taste distribution over different products with their relative prices can differentiate the real values of each period's consumption over time. Moreover, within-period preference structures and their dynamics might greatly differ across households, which must also be a major source of heterogeneity in consumption dynamics across households. Ignoring these facts may make empirical explanation power extremely weak.

This paper use micro panel data on barcode-level purchase records to recover unobserved preference parameters of consumers belonging to different clusters by estimating a structural model of consumer demand. Then a fixed effects model estimation is conducted on the micro panel data to examine how within-period preference structures over multiple products on inter-temporal consumption dynamics.

2. Theoretical backgrounds

This section presents theoretical issues lying behind the empirical analyses in the following sections using the simplest settings.

A typical strategy to describe intertemporal consumption decisions with multiple commodities is two-stage budgeting: at the higher stage, expenditure is allocated to small groups of commodities, and at the lower stage, group expenditures are allocated to individual commodities within the group. In an intertemporal context, the former stage can be interpreted as an intertemporal allocation and the latter as within-period allocations. In the two-stage budgeting, information required for the choices of each stage is only that specific to the stage: at the higher stage, total expenditure and well-defined group prices, and at the lower stage, group expenditure and prices of commodities within that group.

But except for the case in which we can rely on an external condition where relative prices within a group are fixed, further restrictions on preferences are required for the two-stage budgeting to lead to the same result as the one when the allocation were made in a single step. Gorman (1959) has shown that the lower-stage consistency requires weak separability of preferences, which can be expressed as

$$U = u(c_1^1, c_2^1, \dots, c_l^1, c_{l+1}^2, \dots) = f[u^1(c_1^1, c_2^1, \dots c_l^1, \varphi^1), u^2(c_{l+1}^2, \dots, \varphi^2), \dots], \tag{1}$$

where c_i^t is consumption of good i at time t, both u and f are some increasing function, u^t is a subutility function for time t, and φ^t is a vector representing preference structure specific to time t[1]. Also, the upper stage requires stronger restrictions to describe the optimization problem with a single price index, instead of the knowledge of all of the individual prices. Gorman (1959) has shown preferences for commodities within a group must be homothetic, or the group indirect utility function must take the Gorman generalized polar form while preferences are strongly separable between groups[2]. In the former case, group indirect utility functions can be written as

$$u^t = v^t(p^t, x^t, \varphi^t) = x^t / P^t(p^t, \varphi^t), \tag{2}$$

where p^t is a vector of relative prices of commodities. Then the overall utility becomes

$$U = f\left(x^1/P^1(p^1, \varphi^1), x^2/P^2(p^2, \varphi^2), \dots\right) \tag{3}$$

Similarly, in the case of the Gorman generalized polar form:

$$u^t = v^t(p^t, x^t, \varphi^t) = \psi^t\left[\frac{x^t}{P^t(p^t, \varphi^t)}\right] + \chi^t(p^t, \varphi^t) \tag{4}$$

for some monotone increasing function ψ^t, the utility function becomes

$$U = u^1(c_1^1, c_2^1, \dots c_l^1, \varphi^1) + u^2(c_{l+1}^2, \dots \varphi^2) + \dots = \sum_t \psi^t\left(\frac{x^t}{P^t(p^t, \varphi^t)}\right) + \sum_t \chi^t(p^t, \varphi^t) \tag{5}$$

In both cases, the price index is a function of the period-specific preference structure φ^t.

Let us see how the period-specific preference structure can affect consumption dynamics using the simplest settings with a homothetic within-period preference and a CRRA intertemporal utility. A within-period utility of a household is given by

$$u_t(c_t, \varphi_t) = \left[\sum_{i \in I_t} (\varphi_{it} c_{it})^{\frac{\sigma-1}{\sigma}}\right]^{\frac{\sigma}{\sigma-1}}, \tag{6}$$

where I_t is a set of brands in the market at time t, $c_t = \{c_{it} | i \in I_t\}$ and $\varphi_t = \{\varphi_{it} | i \in I_t\}$ are a consumption vector and a tastes vector over the brands, respectively, and σ is the elasticity of substitution. Then, the within-period indirect utility function is given by

$$\begin{gathered} v_t(x_t, P_t^{\Phi}) = x_t / P_t^{\Phi}, \\ P_t^{\Phi} = \left[\sum_{i \in I_t} \left(\frac{p_{it}}{\varphi_{it}}\right)^{1-\sigma}\right]^{\frac{1}{1-\sigma}}, \end{gathered} \tag{7}$$

where x_t is the total expenditure at time t and P_t^{Φ} is the preference-adjusted price index determined by the relative prices p_{it} $(i \in I_t)$ and taste structure φ_{it} $(i \in I_t)$. For intertemporal allocation, let us use a period-specific CRRA utility:

$$U_t = U(C_t) \equiv \frac{{C_t}^{1-\theta} - 1}{1-\theta}, \tag{8}$$

where θ represents the curvature of instantaneous utility and C_t is the instantaneous utility obtained by consumption at time t, which can be replaced by $C_t \equiv v_t = x_t / P_t^{\Phi}$. Let describe the intertemporal optimization problem by:

$$\begin{gathered} \max E[U_t] = E_0\left[\sum_{t=1}^{T} \beta^t U(C_t)\right], \\ \text{s.t. } A_{t+1} \leq (1 + r_t) A_t + Y_t - C_t, \;\; \mathrm{A_T} \geq 0. \end{gathered} \tag{9}$$

Then, the Euler equation is given by

$$E_t[\ln C_{t+1} - \ln C_t] = \frac{1}{\theta}(r_t - \rho), \qquad \rho \equiv \frac{1-\beta}{\beta}. \tag{10}$$

Ignoring uncertainty for simplicity, we get

$$\Delta \ln x_{t+1} = \Delta \ln P_{t+1}^{\Phi} + \frac{1}{\theta}(r_t - \rho). \tag{11}$$

Therefore, the within-period preference structure can affect consumption dynamics through a preference-adjusted price index.

3. Empirical Analysis

4. Data and methodology

3.1 Panel data

The monthly panel data on consumers' purchase profile from April 2010 to May 2015 (62 months) is constructed from INTAGE SCI, which is a scanner panel data of barcode-level daily purchase records of about 50,000 respondents. The major product categories covered are staple food, processed food, beverages, household goods, cosmetics, and drugs. We categorize the products into 128 different product groups mostly in accordance with the expenditure classification of the Family Income and Expenditure Survey, instead of using the INTAGE's original product-group category.

3.2 Consumer clustering

To recover unobserved preference structure of different types of consumers, we group the consumers in the panel data into several clusters according to their purchase patterns. The clustering is conducted in the following steps.

First, we reduce the dimensions of the feature quantities we use for clustering by converting the purchase profile into a set of mutually linearly-uncorrelated values using principal component analysis. As the original data before conversion, we use the share of individual consumer's spending on each of 128 product groups. The shares are calculated for every month, each of which uses the observations of 3 months including proceeding and following month. From the results of the principal components analysis, we choose the first 29 principal components as the feature quantities for clustering, each of which has a contribution more than 1 % and whose cumulated contribution is 25%.

Then a hierarchical clustering analysis is conducted over a set of 29-dimmentional vectors of each consumer's main component scores for every month. As distance metrics and linkage criteria, we use Euclidean distance and Ward' method. Taking the number of observations that we can use for the structural estimation in the next step into consideration, we cut the tree by seven clusters. Table 1 shows an example of the cross section of the seven clusters with regards to some characteristics of consumers.

Table 2

Cluster	Num of consumers	Age (%)						Sex (%)		Marital status (%)	
		-19	20-29	30-39	40-49	50-59	60-	Female	Male	Married	Unmarried
1	5,133	0.1	3.6	13.4	21.4	28.3	33.2	75	25	82	18
2	18,962	0.2	8.0	28.1	32.2	20.6	11.0	77	23	82	18
3	4,494	1.5	19.7	30.2	24.0	16.7	7.8	69	31	50	50
4	2,359	0.1	4.2	16.5	28.2	30.8	20.2	72	28	71	29
5	8,162	3.7	15.7	25.2	29.8	18.3	7.3	23	77	53	47
6	5,903	0.0	3.2	16.3	31.1	30.6	18.8	23	77	70	30
7	1,054	4.5	38.3	32.6	14.0	7.5	3.0	99	1	22	78

Cluster	Employment status (%)							
	Part time	Student	Self-employed	Regular employee/ Public service	Housewife/ Househusband	Dispatched employee/ Contract worker	Other job	Unemployed
1	24.8	0.6	6.9	18.5	29	5.3	3.6	11.3
2	26.7	1.4	4.8	26.1	26.7	6.7	2.3	5.4
3	16.5	5.3	4.6	44.2	10.6	11.6	2.2	5.1
4	23.3	0.6	8.8	26.7	20	9	2.8	8.8
5	10.8	8.2	5.5	56.9	4.1	7.9	1.4	5
6	9.5	0.3	11.5	53.5	6.2	8.5	2.1	8.4
7	13.9	10.6	1.8	48.8	5.4	14.4	1.6	3.4

3.2 Structural estimation of unobserved preference

We estimate unobserved preference parameters of each cluster by constructing a structural model of consumer demand, which is based on Hottman et al. (2016), Hottman (2016), and Sato et al. (2017). The following estimation process is conducted for the representative consumer of each of the seven consumer clusters, although we do not attach any subscript or superscript to show the clusters.

3.2.1. Model

The preference structure of the consumer is given by a nested constant elasticity of substitution (CES) utility system of the three levels, i.e., product groups, firms, and products. The total utility U_t at time t is given by the CES function:

$$U_t = \left[\sum_{g \in G_t} \left(\varphi_{gt} C_{gt} \right)^{\frac{\sigma_G - 1}{\sigma_G}} \right]^{\frac{\sigma_G}{\sigma_G - 1}}, \qquad \sigma_G > 0, \qquad \varphi_{gt} > 0, \tag{12}$$

where C_{gt} denotes the subutility derived from the consumption of product group g. G_t is the set of product groups consumed at time t. φ_{gt} is the consumer's taste for product group g and σ_G is the elasticity of substitution between product groups. The subutility from the consumption of product group g at time t is given as a CES consumption index of firms:

$$C_{gt} = \left[\sum_{f \in F_{gt}} (\varphi_{fgt} C_{fgt})^{\frac{\sigma_{Fg}-1}{\sigma_{Fg}}} \right]^{\frac{\sigma_{Fg}}{\sigma_{Fg}-1}}, \qquad \sigma_{Fg} > 0, \quad \varphi_{fgt} > 0, \tag{13}$$

where C_{fgt} is the subutility derived from the consumption of products within product group g supplied by firm f, F_{gt} is the set of firms supplying products within product group g at time t, φ_{fgt} is the consumer's taste for firm f, and σ_{Fg} is the elasticity of substitution between firms supplying products within product group g. The subutility from the consumption of products within product group g supplied by firm f at time t is given by:

$$C_{fgt} = \left[\sum_{u \in U_{fgt}} (\varphi_{ut} C_{ut})^{\frac{\sigma_{Ug}-1}{\sigma_{Ug}}} \right]^{\frac{\sigma_{Ug}}{\sigma_{Ug}-1}}, \qquad \sigma_{Ug} > 0, \qquad \varphi_{ut} > 0, \tag{14}$$

where C_{ut} is the amount of consumption of product u, U_{fgt} is the set of products within product group g supplied by firm f, φ_{ut} is the consumer's taste for product u, and σ_{Ug} is the elasticity of substitution between products within product group g.

In sum, we express the preference structure of the consumer in each cluster by two types of the structural parameters at each level, the consumer's tastes, φ_{gt}, φ_{fgt}, and φ_{ut}, and the elasticities of substitution, σ_G, σ_{Fg}, and σ_{Ug}. We allow the former ones to vary over time but assume the latter ones are constant over time.

By solving the utility-maximization problem of the product level in equation (14), we obtain the relationship between the product prices and the share of consumer spending on each product. Namely, the share of consumer spending on product u within product group g supplied by firm f at time t can be expressed as

$$S_{ut} = \frac{(P_{ut}/\varphi_{ut})^{1-\sigma_{Ug}}}{\sum_{v \in U_{fgt}} (P_{vt}/\varphi_{vt})^{1-\sigma_{Ug}}}, \tag{15}$$

where P_{ut} is the observed price of product u. We define the price index for a composite good of the various products within product group g supplied by firm f at time t as $P_{fgt} \equiv E_{fgt}/C_{fgt}$, where E_{fgt} is the total expenditure on those products. The solution of the utility-maximization problem enables us to rewrite P_{fgt} as

$$P_{fgt} = \left[\sum_{u \in U_{fgt}} \left(\frac{P_{ut}}{\varphi_{ut}} \right)^{1-\sigma_{Ug}} \right]^{\frac{1}{1-\sigma_{Ug}}}, \tag{16}$$

Similarly, solving the utility-maximization problem of the firm level expressed in equation (13) gives the share of consumer spending on products within product group g supplied by firm f at time t

$$S_{fgt} = \frac{\left(P_{fgt}/\varphi_{fgt}\right)^{1-\sigma_{Fg}}}{\sum_{k \in F_{gt}} \left(P_{kgt}/\varphi_{kgt}\right)^{1-\sigma_{Fg}}} \tag{14}$$

If we define the price index for product group g at time t as $P_{gt} \equiv E_{gt}/C_{gt}$, where E_{gt} is the total expenditure on the product group, it can be rewritten as

$$P_{gt} = \left[\sum_{f \in F_{gt}} \left(\frac{P_{fgt}}{\varphi_{fgt}} \right)^{1-\sigma_{Fg}} \right]^{\frac{1}{1-\sigma_{Fg}}}. \tag{15}$$

Similarly, we obtain the share of consumer spending on product group g at time t

$$S_{gt} = \frac{\left(P_{gt}/\varphi_{gt}\right)^{1-\sigma_G}}{\sum_{h \in G_t} (P_{ht}/\varphi_{ht})^{1-\sigma_G}} \tag{16}$$

The price index for the total consumption at time t, $P_t \equiv E_t/U_t$, can be written as

$$P_t = \left[\sum_{g \in G_t} \varphi_{gt}^{\sigma_G - 1} P_{gt}^{1-\sigma_G} \right]^{\frac{1}{1-\sigma_G}}, \tag{17}$$

where E_t is the total expenditure on all products at time t.

The taste parameters φ_{ut} expresses the relative preferences of that particular product with respect to other products supplied by the same firm. Therefore, one cannot directly compare the level of parameter with those of products supplied by different firms. Similarly, one cannot directly compare the level of φ_{fgt} with those of the firms belonging to different product groups. But, since the utility functions at each level display homogeneous of degree one in those parameters, we need to normalize them to avoid possible distortions that might occur when summing up the consumption of each level into a composite good of the upper level. Following Hottman et al. (2016) and Hottman (2016), we adopt a normalization strategy:

$$\left(\prod_{u\in U_{fgt}} \varphi_{ut}\right)^{\frac{1}{N_{Ufgt}}} = \left(\prod_{f\in F_{gt}} \varphi_{fgt}\right)^{\frac{1}{N_{Fgt}}} = \left(\prod_{g\in G_t} \varphi_{gt}\right)^{\frac{1}{N_{Gt}}} = 1, \tag{18}$$

where N_{Ufgt} is the number of products within product group g supplied by firm f, N_{Fgt} is the number of firms supplying products within product group g, and N_{Gt} is the number of product groups. We use the same normalization approach for the product-group level simply for a computational convenience, but it does not matter for our main results.

On the production side, we assume a firm supplying product u incur a variable cost V_{ut} at time t

$$V_{ut}(Q_{ut}) = z_{ut}Q_{ut}^{1+\delta_g}, \qquad z_{ut} > 0, \tag{19}$$

where Q_{ut} is the total quantity supplied of product u, δ_g is the time-invariant convexity parameter of marginal costs for product group g, and z_{ut} is the firm-product-specific shifter of the cost function. The total profit of firm f for supplying all products within product group g at time t is given as

$$\pi_{fgt} = \sum_{u\in U_{fgt}} [P_{ut}Q_{ut} - V_{ut}(Q_{ut})] - H_{fgt}, \tag{20}$$

where H_{fgt} is the fixed market access cost at time t.

3.2.2. Estimating the elasticities of substitution

First, we explain our methodology for estimating the elasticities of substitution at each level (σ_G, σ_{Fg}, and σ_{Ug}) and the elasticity of marginal costs (δ_g).

1) Product level

On the demand side, we take the double-difference of the log of the share of consumer spending on each product over time and relative to the average share of all products in the same product group supplied by firm f

$$\Delta^{U,t} \ln S_{ut} = -(\sigma_{Ug} - 1)\, \Delta^{U,t} \ln P_{ut} + \omega_{ut}, \tag{21}$$

where $\omega_{ut} = -(\sigma_{Ug} - 1)(\Delta^t \overline{\ln \varphi_{ut}} - \Delta^t \ln \varphi_{ut})$ is the unobserved error term. $\Delta^{U,t}$ is the double-difference operator across products and over time such that $\Delta^{U,t} \ln S_{ut} = \Delta^t \ln S_{ut} - \Delta^t \overline{\ln S_{ut}}$ and $\overline{\ln S_{ut}} = (1/N_{U_{fgt}}) \sum_{v \in U_{fgt}} \ln S_{vt}$, where Δ^t is the first-difference operator over time such that $\Delta^t \ln S_{ut} = \ln S_{ut} - \ln S_{ut-1}$. 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 firm f, which leaves only idiosyncratic demand shocks that affect the sales of a certain product relative to other products.

On the supply side, transforming the equilibrium pricing rule (See Appendix 1) 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 firm f, we obtain

$$\Delta^{U,t} \ln P_{ut} = \frac{\delta_g}{1 + \delta_g}\, \Delta^{U,t} \ln S_{ut} + \kappa_{ut}, \tag{22}$$

where $\kappa_{ut} = \frac{1}{1+\delta_g}(\Delta^t \ln z_{ut} - \Delta^t \overline{\ln z_{ut}})$ is the unobserved error term.

In the following, we assume the double-differenced idiosyncratic demand and supply shocks are orthogonal at the product level. This orthogonality is plausible for the following reasons. First, differencing across products within the same product group supplied by firm f eliminates common firm-level shocks (e.g., changes in management) and common group-level shocks (e.g., renewal of the package that is common to all products) that affect both production costs and product appeal to consumers (product quality or consumers' 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 in all time periods. Our identification is, therefore, 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 is very unlikely to arise with barcode data. Any substantive change in product characteristics is accompanied by the introduction of a new barcode in most cases. In other words, even if there are double-differenced changes in costs, it should affect double-differenced consumer demand for that product conditional on price, since it leaves the observable characteristics of a product constant.

Equation (21) and (22) gives

$$Y_{ut} = \theta_{g1} X_{1ut} + \theta_{g2} X_{2ut} + u_{ut}, \tag{23}$$

where $Y_{ut} = (\Delta^{U,t} \ln P_{ut})^2$, $X_{1ut} = (\Delta^{U,t} \ln S_{ut})^2$, $X_{2ut} = \Delta^{U,t} \ln P_{ut} \cdot \Delta^{U,t} \ln S_{ut}$, $u_{ut} = \omega_{ut}\kappa_{ut}$, $\theta_{g1} = \frac{\delta_g}{(1+\delta_g)(\sigma_{Ug}-1)}$, and $\theta_{g1} = \frac{\delta_g(\sigma_{Ug}-2)-1}{(1+\delta_g)(\sigma_{Ug}-1)}$. Let N_u denote the number of observations of product u that also have an observation at the previous period, and $N \equiv \sum_{u \in U_g} N_u$, where U_g is the set of products within product group g observed over at least two successive periods. Let Y denote the $N \times 1$ vector with components Y_{ut}, X denote the $N \times 2$ vector with rows (X_{1ut}, X_{2ut}), u the $N \times 1$ vector with components u_{ut}, and $\theta_g = (\theta_{g1}, \theta_{g2})'$. The column components of Y, X, and u are ordered in a sequence of product and time such as $(Y_{11}, Y_{12}, Y_{13}, \ldots, Y_{21}, Y_{22}, \ldots)'$. Then we get, for each product group,

$$Y = X\theta_g + u \tag{24}$$

Since the expenditure shares and prices are correlated with the errors ω_{ut} and κ_{ut} respectively, X_{1ut} and X_{2ut} are correlated with u_{ut}. Let Z_u denote a $N_u \times 1$ vector of 1's, and define Z as the $N \times N_{Ug}$ matrix of dummy variables for each product,

$$Z = \begin{bmatrix} Z_1 & 0 & \ldots & 0 \\ 0 & Z_2 & 0 & 0 \\ \vdots & 0 & \ddots & \vdots \\ 0 & \ldots & \ldots & Z_{N_{Ug}} \end{bmatrix},$$

where N_{Ug} is the number of products within product group g observed over at least two successive periods. Since $\operatorname{plim} \frac{1}{N} Z'u$ is the $N_{Ug} \times 1$ vector with components $\operatorname{plim} \frac{1}{N} \sum_t \omega_{ut} u_{ut}$, the orthogonality assumption of idiosyncratic demand and supply shocks gives $\operatorname*{plim}_{N\to\infty} \frac{1}{N} Z'u = 0$, that is, Z is asymptotically uncorrelated with u.

We assume there are some differences in the relative variances of the demand and supply shock across products, that is, there exist u and $u' \in U_g$ such that:

$$\frac{\eta_{\omega u}^2}{\eta_{\omega u'}^2} \neq \frac{\eta_{\kappa u}^2}{\eta_{\kappa u'}^2}, \tag{25}$$

where $\eta_{\omega u}^2$ ans $\eta_{\kappa u}^2$ are the variances of ω_{ut} and κ_{ut} respectively. This condition ensures that the two columns of the matrix $\operatorname{plim} \frac{1}{T} Z'X$ are linearly independent, so that it has full column rank.
Using Z as an instrument, we obtain a GMM estimator

$$\hat{\theta}_g = (X'ZWZ'X)^{-1}X'ZWZ'Y, \tag{26}$$

where W is a positive definite weighting matrix. Under the assumption of (18), $\text{plim}[(X'Z/N)(W/N)(Z'X/N)]$ has full rank of 2 and is invertible. Then the orthogonality assumption of idiosyncratic demand and supply shocks gives $\text{plim}\,\hat{\theta}_g = \theta_g$.

2) Firm level

We take the double-difference of the log of the share of consumer spending on products within product group g supplied by firm f over time and relative to the average share of all firms in the same product group

$$\Delta^{F,t} \ln S_{fgt} = -(\sigma_{Fg} - 1)\, \Delta^{F,t} \ln P_{fgt} + \omega_{fgt}, \tag{27}$$

where $\omega_{fgt} = -(\sigma_{Fg} - 1)\Delta^{F,t} \ln \varphi_{fgt}$ is the unobserved error term. As before, $\Delta^{F,t}$ is the double-difference operator across firms in the same product group and over time such that $\Delta^{F,t} \ln S_{fgt} = \Delta^{t} \ln S_{fgt} - \Delta^{t}\, \overline{\ln S_{fgt}}$ and $\overline{\ln S_{fgt}} = (1/N_{Fgt}) \sum_{k \in F_{gt}} \ln S_{kgt}$.

The unobserved error term is likely to be correlated with the double-differenced firm-level price index, because a relative increase in firm-level quality raises the quantity demanded of the products supplied by the firm and thus raises the firm-level price index. We therefore estimate σ_{Fg} using an instrumental variable approach. The relative product expenditures in terms of relative product prices and relative product demand shifters is

$$\frac{S_{ut}}{\tilde{S}_{Ufgt}} = \left(\frac{P_{ut}/\varphi_{ut}}{\tilde{P}_{Ufgt}/\tilde{\varphi}_{Ufgt}}\right)^{1-\sigma_{Ug}}, \tag{28}$$

where $\tilde{S}_{Ufgt} \equiv \left(\prod_{v \in U_{fgt}} S_{vt}\right)^{\frac{1}{N_{Ufgt}}}$, $\tilde{P}_{Ufgt} \equiv \left(\prod_{v \in U_{fgt}} P_{vt}\right)^{\frac{1}{N_{Ufgt}}}$, and $\tilde{\varphi}_{Ufgt} \equiv \left(\prod_{v \in U_{fgt}} \varphi_{vt}\right)^{\frac{1}{N_{Ufgt}}}$. We can rewrite the double-differenced firm-level price index as

$$\Delta^{F,t} \ln P_{fgt} = \Delta^{F,t} \ln \tilde{P}_{Ufgt} - \Delta^{F,t} \left[\frac{1}{\sigma_{Ug} - 1} \ln\left(\sum_{u \in U_{fgt}} \frac{S_{ut}}{\tilde{S}_{Ufgt}}\right)\right]. \tag{29}$$

The moment condition for the instrumental variables is

$$\mathbb{E}\left\{\omega_{fgt}\cdot\Delta^{F,t}\left[\frac{1}{\sigma_{Ug}-1}\ln\left(\sum_{u\in U_{fgt}}\frac{S_{ut}}{\tilde{S}_{Ufgt}}\right)\right]\right\}=0 \tag{30}$$

3) Product-group level

We take the double-difference of the log of the expenditure share of each product group in the total expenditure over time and relative to the average share of all product groups

$$\Delta^{G,t}\ln S_{gt}=-(\sigma_G-1)\,\Delta^{G,t}\ln P_{gt}+\omega_{gt}, \tag{31}$$

where $\omega_{gt}=-(\sigma_G-1)\Delta^{G,t}\ln\varphi_{gt}$ is the unobserved error term, $\Delta^{G,t}$ is the double-difference operator across product groups and over time such that $\Delta^{G,t}\ln S_{gt}=\Delta^{t}\ln S_{gt}-\Delta^{t}\,\overline{\ln S_{gt}}$ and $\overline{\ln S_{gt}}=(1/N_{Gt})\sum_{h\in G_t}\ln S_{gt}$.

As before, the relative firm expenditures in terms of relative firm prices and relative firm demand shifters is

$$\frac{S_{fgt}}{\tilde{S}_{Fgt}}=\left(\frac{P_{fgt}/\varphi_{fgt}}{\tilde{P}_{Fgt}/\tilde{\varphi}_{Fgt}}\right)^{1-\sigma_{Fg}}, \tag{32}$$

where $\tilde{S}_{Fgt}\equiv\left(\prod_{k\in F_{gt}}S_{kgt}\right)^{\frac{1}{N_{Fgt}}}$, $\tilde{P}_{Fgt}\equiv\left(\prod_{k\in F_{gt}}P_{kt}\right)^{\frac{1}{N_{Fgt}}}$, and $\tilde{\varphi}_{Fgt}\equiv\left(\prod_{k\in F_{gt}}\varphi_{kt}\right)^{\frac{1}{N_{Fgt}}}$. The double-differenced product-group-level price index can be expressed as

$$\Delta^{G,t}\ln P_{gt}=\Delta^{G,t}\ln\tilde{P}_{Fgt}-\Delta^{G,t}\left[\frac{1}{\sigma_{Fg}-1}\ln\left(\sum_{f\in F_{gt}}\frac{S_{fgt}}{\tilde{S}_{Fgt}}\right)\right]. \tag{33}$$

The moment condition for the instrumental variables is

$$\mathbb{E}\left\{\omega_{gt}\cdot\Delta^{G,t}\left[\frac{1}{\sigma_{Fg}-1}\ln\left(\sum_{f\in F_{gt}}\frac{S_{fgt}}{\tilde{S}_{Fgt}}\right)\right]\right\}=0 \tag{34}$$

3.2.3. Consumer's taste

After estimating the elasticities of substitution at each level (σ_G, σ_{Fg}, and σ_{Ug}), we can easily recover φ_{ut}, the unobserved consumer's taste for product u, since, using equation (15) and condition (18), it can be rewritten as

$$\varphi_{ut} = P_{ut}(S_{ut})^{\frac{1}{\sigma_{Ug}-1}} \left[\prod_{v \in U_{fgt}} \frac{1}{P_{vt}(S_{vt})^{\frac{1}{\sigma_{Ug}-1}}} \right]^{\frac{1}{N_{Ufgt}}} \tag{35}$$

We can recover φ_{fgt} and φ_{gt} in a similar manner.

$$\varphi_{fgt} = P_{fgt}\left(S_{fgt}\right)^{\frac{1}{\sigma_{Fg}-1}} \left[\prod_{k \in F_{gt}} \frac{1}{P_{kgt}\left(S_{kgt}\right)^{\frac{1}{\sigma_{Fg}-1}}} \right]^{\frac{1}{N_{Fgt}}} \tag{36}$$

$$\varphi_{gt} = P_{gt}\left(S_{gt}\right)^{\frac{1}{\sigma_{G}-1}} \left[\prod_{h \in G_t} \frac{1}{P_{ht}(S_{ht})^{\frac{1}{\sigma_{G}-1}}} \right]^{\frac{1}{N_{Gt}}} \tag{37}$$

3.3 Contribution of preference structure

We examine the contribution of preference structure on consumer spending. In our previous setting, the contemporaneous utility is given by

$$U_t = \left\{ \sum_{g \in G_t} \varphi_{gt}^{\sigma_G - 1} \left[\sum_{f \in F_{gt}} \varphi_{fgt}^{\sigma_{Fg}-1} \left(\sum_{u \in U_{fgt}} \varphi_{ut}^{\sigma_{Ug}-1} P_{ut}^{1-\sigma_{Ug}} \right)^{\frac{1-\sigma_{Fg}}{1-\sigma_{Ug}}} \right]^{\frac{1-\sigma_G}{1-\sigma_{Fg}}} \right\}^{\frac{1}{\sigma_G - 1}} E_t. \tag{38}$$

Thus the contemporaneous utility can be expressed as a linear function of the total expenditure with the slope of some preference structure. The preference-adjusted price index of the contemporaneous utility is therefore given by

$$P_t^{\Phi} = \left[\sum_{g \in G_t} \varphi_{gt}^{\sigma_G - 1} P_{gt}^{1-\sigma_G}\right]^{\frac{1}{1-\sigma_G}} = \left\{\sum_{g \in G_t} \varphi_{gt}^{\sigma_G - 1}\left[\sum_{f \in F_{gt}} \varphi_{fgt}^{\sigma_{Fg}-1}\left(\sum_{u \in U_{fgt}} \varphi_{ut}^{\sigma_{Ug}-1} P_{ut}^{1-\sigma_{Ug}}\right)^{\frac{1-\sigma_{Fg}}{1-\sigma_{Ug}}}\right]^{\frac{1-\sigma_G}{1-\sigma_{Fg}}}\right\}^{\frac{1}{1-\sigma_G}}. \tag{39}$$

We computed (39) for each of the seven clusters:

$$P_t^{\Phi,c} = \left\{\sum_{g \in G_t^c} {\varphi_{gt}^c}^{\sigma_G - 1}\left[\sum_{f \in F_{gt}^c} {\varphi_{fgt}^c}^{\sigma_{Fg}-1}\left(\sum_{u \in U_{fgt}^c} \left(\frac{P_{ut}^c}{\varphi_{ut}^c}\right)^{1-\sigma_{Ug}^c}\right)^{\frac{1-\sigma_{Fg}^c}{1-\sigma_{Ug}^c}}\right]^{\frac{1-\sigma_G^c}{1-\sigma_{Fg}^c}}\right\}^{\frac{1}{1-\sigma_G^c}}. \tag{40}$$

Using (40), we estimate the panel data by a fixed-effect model with instrument variables:

$$\begin{aligned} \ln E_{it} - \ln P_t = {} & \beta_1 kakekomi_t + \beta_2 hando_t + \beta_{3\ldots13} month_{dummies_t} + \beta_{14} real_{interest_{rate_t}} \\ & + \beta_{15\ldots18} income_{class_{dummies_t}} + \beta_{19} head's_{job_{status_t}} + \beta_{20} spuse's_{job_{status_t}} \\ & + \beta_{21} \Delta children_t + \beta_{22} baby_t + \varepsilon_{it} \end{aligned} \tag{41}$$

For the price index P_t, we use, as well as a) the preference-adjusted price index $P_t^{\Phi,c}$, the following indices for comparison:

b) Stone price index with true price

$$P_t^{Stone1,c} = \prod_{u \in U_{fgt}^c, F_{gt}^c, G_t^c} {P_{ut}^c}^{S_{ut}^c S_{fgt}^c S_{gt}^c}$$

c) Stone price index with product-group CPI

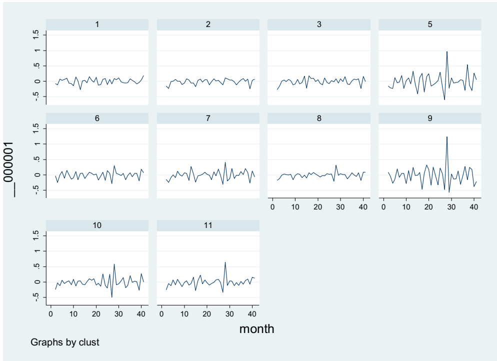
$$P_t^{Stone2,c} = \prod_{g \in G_t^c} \tilde{P}_{gt}^{S_{gt}^c}$$

d) All item CPI

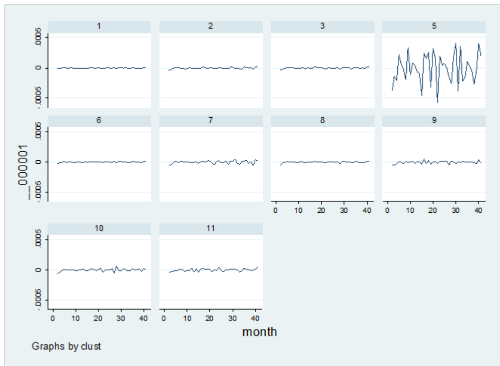
$$\tilde{P}_t.$$

4. Results

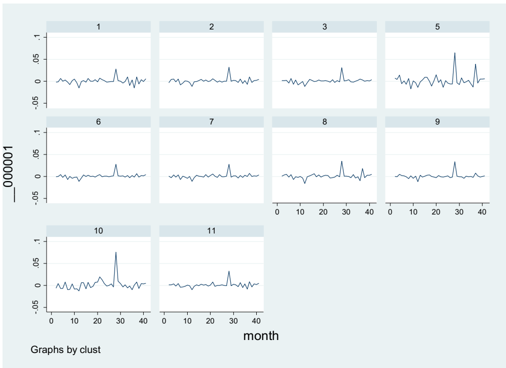
Figure 1 shows the transitions of the preference-adjusted price index and the other three price indices.



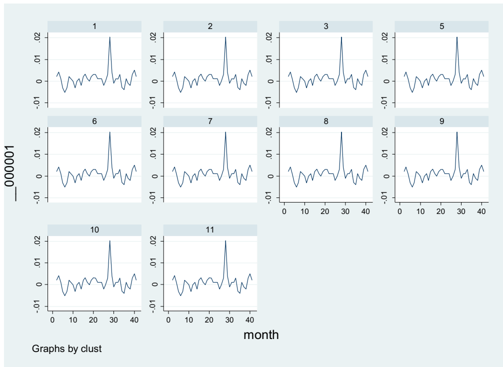
1) Preference-adjusted price index



2) Stone price index with true price



3) Stone price index with product-group CPI



4) All item CPI

Figure 1 Transition of price indices

Table 2 shows the results of the panel data estimation. We can see the model explanation power is the strongest for the result of preference-based price index, whereas the other three reference indices show similar result.

Table 2 Panel analysis (fixed-effect model with instrument variables)

	(1) Preference	(2) Stone1	(3)Stone2	(4)CPI
Real interest rate	.0385* (.0044)	-.0153* (.0043)	-.0114* (.0044)	-.0142* (.0044)
Income class 400-549	-.0032 (.0047)	-.0035 (.0046)	-.0035 (.0046)	-.0035 (.0046)
550-699	-.0009 (.0057)	.0003 (.0056)	.0002 (.0056)	.0001 (.0056)
700-899	-.0023 (.0067)	-.0011 (.0065)	-.0011 (.0065)	-.0012 (.0065)
900-	.0018	.0029	.0030	.0028

	(.0084)	(.0082)	(.0082)	(.0082)
Head's job status	-.0008 (.0057)	-.0003 (.0055)	-.0006 (.0055)	-.0004 (.0054)
Spouse's job status	.0032 (.0070)	.0032 (.0069)	.0032 (.0069)	.0032 (.0069)
Δ Children	-.0377* (.0100)	-.0153 (.0091)	-.0153 (.0092)	-.0154 (.0092)
Baby	-.0077 (.0049)	-.0074 (.0048)	-.0075 (.0048)	-.0073 (.0048)
R^2	.1170	.0431	.0470	.0446
Prob > F	.0000	.0000	.0000	.0000
Underidentification test (p-value)	5.0e+04 (.0000)	5.0e+04 (.0000)	5.0e+04 (.0000)	5.0e+04 (.0000)
Overidentification test (p-value)	.054 (.9732)	.064 (.9684)	0.065 (.9681)	.063 .9688

Other explanatory variables: *Kakekomi, Hando*, and *month dummies*
Instruments: The 1st and 2nd lag of *family income class, head's job status*, and *spouse's job status*.
Heteroscedasticity-robust standard errors are in parentheses
Underidentification test is by Kleibergen-Paap rk LM statistic (H_0: the equation is underidentified)
Overidentification test is by Hansen J statistic (H_0: the instruments are valid instruments, i.e., uncorrelated with the error term, and that the excluded instruments are correctly excluded from the estimated equation.)

Appendix 1

The marginal cost of supplying product u at time t is given by

$$m_{ut} = \left(1 + \delta_{fg}\right) z_{ut} Q_{ut}^{\delta_g}. \tag{A1}$$

The total profit of firm f for product group g at time t is given by

$$\pi_{fgt} = \sum_{u \in U_{fgt}} \left[P_{ut} Q_{ut} - V_{ut}(Q_{ut})\right] - H_{fgt}. \tag{A2}$$

Solving the profit maximization problem, we get the first-order condition

$$Q_{ut} + \sum_{v \in U_{fgt}} \left[P_{vt} \cdot \frac{\partial Q_{vt}}{\partial P_{ut}} - \frac{\partial V_{vt}(Q_{vt})}{\partial Q_{vt}} \cdot \frac{\partial Q_{vt}}{\partial P_{ut}}\right] = 0. \tag{A3}$$

The optimal pricing rule is given by

$$P_{ut} = \mu_{fgt} m_{ut}. \tag{A4}$$

where the markup over marginal cost is

$$\mu_{fgt} = \frac{\varepsilon_{fgt}}{\varepsilon_{fgt} - 1}, \qquad \varepsilon_{fgt} = \sigma_{Fg} - \left(\sigma_{Fg} - 1\right) S_{fgt}, \tag{A4}$$

where ε_{fgt} is firm f's perceived elasticity of demand.

The quantity demanded of product u at time t is given by

$$C_{ut} = \varphi_{fgt}^{\sigma_{Fg}-1} \varphi_{ut}^{\sigma_{Ug}-1} E_{gt} P_{gt}^{\sigma_{Fg}-1} P_{fgt}^{\sigma_{Ug}-\sigma_{Fg}} P_{ut}^{-\sigma_{Ug}}. \tag{A5}$$

Using (A1), (A4), and (A5), we obtain (22).

	(1) Preference	(2) Stone1	(3)Stone2	(4)CPI
Real interest rate	.5441* (.0135)	-.0662* (.0128)	-.1007* (.0128)	-.0858* (.0128)
Income class 400-549	-.0103 (.0072)	-.0129 (.0068)	-.0130 (.0068)	-.0130 (.0068)
550-699	.0028 (.0089)	-.0083 (.0084)	-.0082 (.0084)	-.0085 (.0084)
700-899	.0126 (.0104)	-.0036 (.0099)	-.0033 (.0099)	-.0038 (.0099)
900-	.0172 (.0131)	-.0034 (.0124)	-.0030 (.0124)	-.0037 (.0124)
Head's job status	.0031 (.0087)	.0038 (.0082)	.0039 (.0082)	.0036 (.0082)
Spouse's job status	.0212* (.0076)	.0084 (.0070)	.0088 (.0070)	.0081 (.0070)
Δ Children	.0076 (.0087)	.0033 (.0084)	.0034 (.0084)	.0032 (.0084)
Baby	-.0330* (.0065)	-.0201* (.0063)	-.0204* (.0063)	-.0198* (.0063)
R^2	.0381	.0240	.0246	.0245
Prob > F	.0000	.0000	.0000	.0000
Underidentification test (p-value)	9678.771 (.0000)	9678.771 (.0000)	9678.771 (.0000)	9678.771 (.0000)
Overidentification test (p-value)	.022 (.9892)	.247 (.8840)	.248 (.8832)	.234 .8894

Other explanatory variables: *Kakekomi, Hando*, and *season dummies*

Instruments: The 1st and 2nd lag of *family income class, head's job status*, and *spouse's job status*.

Heteroscedasticity-robust standard errors are in parentheses

Underidentification test is by Kleibergen-Paap rk LM statistic (H_0: the equation is underidentified)

Overidentification test is by Hansen J statistic (H_0: the instruments are valid instruments, i.e., uncorrelated with the error term, and that the excluded instruments are correctly excluded from the estimated equation.)

References

Attanasio, O., and Weber, G., Is Consumption Growth Consistent with Intertemporal Optimization ? Evidence from the Consumer Expenditure Survey. *J. Polit. Econ.* 103, 1121–1157 (1995).

Attanasio, O., and Borella, M., Stochastic components of individual consumption: a time series analysis of grouped data, *NBER Working Paper* (12456, 2006), , doi:10.3386/w12456.

Blundell, R., Browning, M., and Meghir, C., Consumer Demand and the Life-Cycle Allocation of Household Expenditures. *Rev. Econ. Stud.* 61, 57–80 (1994).

Cochrane, J. H., A Simple Test of Consumption Insurance. *J. Polit. Econ.* 99, 957–976 (1991).

Dynan, B. K. E., Habit Formation in Consumer Preferences : Evidence from Panel Data. *Am. Econ. Rev.* 90, 391–406 (2000).

Hall, R. E., and Mishkin, F. S., The Sensitivity of Consumption to Transitory Income: Estimates from Panel Data on Households. *Econometrica*. 50, 461–481 (1982).

Hottman, C., Stephen J. R., and Weinstein, D. E., Quantifying the Sources of Firm Heterogeneity, *The Quarterly Journal of Economics*, 131(3), pp.1291–1364 (2016).

Hottman, C., Retail Markups, Misallocation, and Store Variety in the US, mimeo, (2016), available online: http://www.columbia.edu/~cjh2164/HottmanJMP.pdf.

Sato, Masahiro, Taisuke Kameda, Shigeru Sugihara, and Colin Hottman, 2017, The contribution of quality and product variety to retail growth in Japan, *経済分析*, 194号, 65-92 (2017).

[1] The plausibility of this assumption in an intertemporal context may depend on the length of time adopted (Deaton, 1980) and characteristics of goods such as extent of durability. Habit formation is another possible factor that might break the separability assumption.

[2] Even if none of these upper-stage conditions is satisfied, one can rely on some approximate solutions (Deaton, 1980; Edgerton, 1997).