PLOS ONE

RESEARCH ARTICLE

Untangling the complexity of market competition in consumer goods—A complex Hilbert PCA analysis

Makoto Mizuno[1☯]*, **Hideaki Aoyama**[2,3☯], **Yoshi Fujiwara**[4☯]

1 School of Commerce, Meiji University, Tokyo, Japan, **2** Research Institute of Economy, Trade and Industry, Tokyo, Japan, **3** RIKEN iTHEMS, Wako, Saitama, Japan, **4** Graduate School of Simulation Studies, University of Hyogo, Kobe, Japan

☯ These authors contributed equally to this work.
* makmizuno@gmail.com

Citation: Mizuno M, Aoyama H, Fujiwara Y (2021) Untangling the complexity of market competition in consumer goods—A complex Hilbert PCA analysis. PLoS ONE 16(2): e0245531. https://doi.org/10.1371/journal.pone.0245531

Editor: Thilo Gross, University Of Bristol, UNITED KINGDOM

Received: September 5, 2020

Accepted: January 4, 2021

Published: February 3, 2021

Data Availability Statement: The data underlying the results presented in the study are available form INTAGE Inc. (https://www.intage.co.jp/english/).

Funding: YF was awarded by Grant-in-Aid for Scientific Research (KAKENHI) of JSPS Grant Numbers 17H02041, 20H02391. YF and HA were awarded by MEXT as Exploratory Challenges on Post-K Computer (Studies of Multilevel Spatiotemporal Simulation of Socioeconomic Phenomena: Macroeconomic Simulations), https://www.jsps.go.jp/english/e-grants/index.html.

Abstract

Today's consumer goods markets are rapidly evolving with significant growth in the number of information media as well as the number of competitive products. In this environment, obtaining a quantitative grasp of heterogeneous interactions of firms and customers, which have attracted interest of management scientists and economists, requires the analysis of extremely high-dimensional data. Existing approaches in quantitative research could not handle such data without any reliable prior knowledge nor strong assumptions. Alternatively, we propose a novel method called complex Hilbert principal component analysis (CHPCA) and construct a synchronization network using Hodge decomposition. CHPCA enables us to extract significant comovements with a time lead/delay in the data, and Hodge decomposition is useful for identifying the time-structure of correlations. We apply this method to the Japanese beer market data and reveal comovement of variables related to the consumer choice process across multiple products. Furthermore, we find remarkable customer heterogeneity by calculating the coordinates of each customer in the space derived from the results of CHPCA. Lastly, we discuss the policy and managerial implications, limitations, and further development of the proposed method.

Introduction

In rapidly evolving, highly competitive consumer goods markets, firms are faced with a number of factors affecting business. These factors often veer away from conventional rules of thumb or theory and can be mutually interacting, providing the challenge of detecting the meaningful relationships between such factors without any strong assumptions. This challenge represents the complexity of consumer choice process caused by the proliferation of viable marketing instruments and competing products. Traditionally, consumer purchase could have been a function of price, radio or TV advertising. Recently, with the penetration of the Internet and mobile devices the possible set of marketing instruments have been expanding. Some consumers might search for information using mobile phones while others might gain detailed

Competing interests: The authors have declared that no competing interests exist.

information by visiting related websites after watching TV advertisements. If the strategy for mixing these instruments is heterogeneous across competitive firms or products, the increased number of competitors would intensify the resulting complexity further.

To grasp the time sequence of exposures to multiple marketing instruments, the concept of 'customer journey map' is gaining popularity with practitioners and researchers [1, 2], for which some quantitative models have been proposed [3, 4]. This approach focuses on the consumer decision process for a single product/firm only, ignoring competition among multiple products. Hence, it fails to reveal real consumer behaviors for the market with multiple competitive products. In order to deal with dynamic market competition across products, multivariate time-series analyses are highly popular; and have been developed in econometrics and intensively applied to assess the long-term effectiveness of marketing instruments such as pricing or advertising. This method includes a variety of models such as the vector auto-regressive model [5], dynamic linear model [6], varying parameter model [7], and the Kalman filter [8]. These models are evaluated in terms of whether they satisfy the following conditions [6]:

- Endogeneity: The recent quantitative marketing studies inspired by economics has increasingly treated variables of marketing instruments as endogenous variables, not as exogenous variables freely determined by firms. Suppose customers who are more likely to purchase a focal product are targeted in an advertising campaign for the product. In this case, observed advertising data would reflect the potential purchases of the product. With no consideration for this endogeneity, the effects of advertising on purchases would be overestimated.
- Performance feedback: Even if firms do not foresee future performance, their marketing actions would be constrained by past performance due to performance-based budget allocation. Hence the lagged effect of the performance to the activity level of marketing instruments could exist.
- Competitive reactions: Firms that compete in the same market, are sensitive to competitors' actions, often causing retaliation. This has been studied as a hot issue in marketing science, using a variety of data and analytical methods [9].

Even multivariate time-series analyses often face serious difficulties in handling the complexity of current consumer markets, because of the limited number of tractable parameters. Although most consumer goods markets are composed of many more firms/products and marketing instruments, the above extant models have only been applied to cases with only three to five competitors and a few marketing instruments. The competition analyzed in this study consists of 18 products and 4 marketing instruments (price, TV advertising, search via mobile devices or PCs, and web visits via mobile devices or PCs). Applying the conventional time-series analyses for such phenomena would cause an explosive increase in possible parameters to be estimated [10]. To avoid this, researchers have relied on their experience or research tradition and have been forced to impose strong restrictions at the risk of losing crucial information on the complexity of the focal phenomenon.

Alternatively, we propose a novel method called complex Hilbert principal component analysis (CHPCA) [11, 12] to unravel the complexity of the consumer choice process across multiple competitive products. CHPCA was developed originally in econophysics as an extension of Principal Component analysis (PCA) to uncover temporal comovements among variables observed in the macro economy [13] or foreign exchange markets [14]. CHPCA can handle massive high-dimensional time-series data without any strong assumptions about the phenomenon of interest. In addition, this method satisfies the above-mentioned three conditions for evaluating marketing models with multiple competitive products [6]: CHPCA satisfies the first condition (endogeneity) since it treats all variables (marketing actions by

instrument and performance indexes) as comoving variables. Additionally, this method satisfies the second condition (feedback from past outcomes) and the third condition (competitive reactions) since these are incorporated as part of possible comovements.

Another advantage of CHPCA is its practicality. Practitioners can solve real problems under time and effort constraints. First, as with ordinal PCA, users of CHPCA may not need any prior knowledge or assumptions regarding the phenomenon. By using random rotation simulation (RRS), significant eigenmodes (principal components in PCA) can be selected automatically in a theoretically justifiable manner (see [12, Chapter 5] and references therein). Second, the results of CHPCA are easily interpreted on the complex plane corresponding to each eigenmode following stylized procedures. Conveniently, the information obtained is integrated and visualized by a synchronized network with Hodge decomposition [15]. Finally, this method provides a skeleton of consumer choice process across competitive products using aggregate marketing data, which is widely available for the packaged consumer goods markets. The model also provides information on heterogeneity in customer profile.

We recommend CHPCA for conducting exploratory analyses. Similar to the division of roles between exploratory and confirmatory factor analyses, CHPCA can be complementary to traditional multivariate time-series analyses. CHPCA is used to generate hypotheses on possible causalities among a huge number of variables while multivariate time-series analyses are used to rigorously test hypotheses. If a few critical relationships are detected as a result of the analysis using CHPCA, we can apply quasi-experimental methods such as a regression discontinuity design [16] or a propensity score method [17], which have been used to prove causality in observational data.

The remainder of this paper is organized as follows: In section 2, we describe the data for the beer market in Japan. We propose the application of CHPCA to understand the consumer choice process across multiple products in section 3 and the procedure for depicting the skeleton of the consumer choice process via a synchronization network in section 4. The procedure to detect customer heterogeneity and the results are reported in section 5. In section 6, we conclude our paper with a discussion on the policy and managerial implications, the remaining problems, and avenues for further research.

Data

Data collection

In this study, we analyze the comovement of consumer purchases (quantities and prices paid) of beer and related marketing communication activities. The reason for our focus on this market is that it is a typical monopolistic-competitive market, where a few firms compete with differentiated products using a full range of marketing instruments such as TV advertising, web/mobile marketing, price promotion, etc., attracting attentions of economic policymakers and marketers. Hence, we use INTAGE Single-source Panel (i-SSP) data, which is the most comprehensive consumer database commercially operated in Japan measuring daily purchases of a wide variety of consumer package goods and consumer marketing communication activities (exposure to TV ads, visits to web sites via mobile device/PC, and search activities via mobile device/PC. For the current analysis, we use these data for 365 days from April 1, 2013 to March 31, 2014 (inclusive). The abbreviation and description of each time series is shown in Table 1.

These data initially capture individual-level behaviors of the panel. For this analysis, however, we aggregated the data over all customers due to the limited size of the data. The potential heterogeneity among customers is represented as the location of a few-dimensional space (see the Customer Profile Section. The purchase data are documented at the store-keeping unit

Table 1. Abbreviation and the description of the five kinds of time-series.

Abbreviation	Description
P	Price per unit quantity (yen/mℓ)
Q	Quantity purchased (mℓ)
Visit	Visit to related web sites via mobile device or PC (seconds)
TVAd	Exposure to TV advertising (seconds)
Search	Search frequency via mobile device or PC (times)

https://doi.org/10.1371/journal.pone.0245531.t001

(SKU) level while the advertising and other communication variables are documented at the product level. In general, one product is composed of multiple SKUs. When merging the two types of data, the purchase data are summed across SKUs by their corresponding product.

There are 163 beer products from 14 firms in the original data. We selected the top 18 products according to the total quantity in the data during the stated period: Fig 1 is the rank-size plot of some of the top products. The top 18 products selected for the current analysis are shown with filled dots. As is apparent in this plot, these products form a distinctive top group with a large gap between this group and the followers (open dots). Furthermore, this top group roughly obeys the power-law indicated by the thin dashed line, with [Rank] $\propto$ [Total Quantity]$^{-1.109}$.

The data for the top 18 products cover 64.9% of all the sales. Fig 2 shows the daily total quantities for both all products (in light gray) and the selected 18 products (in dark gray).

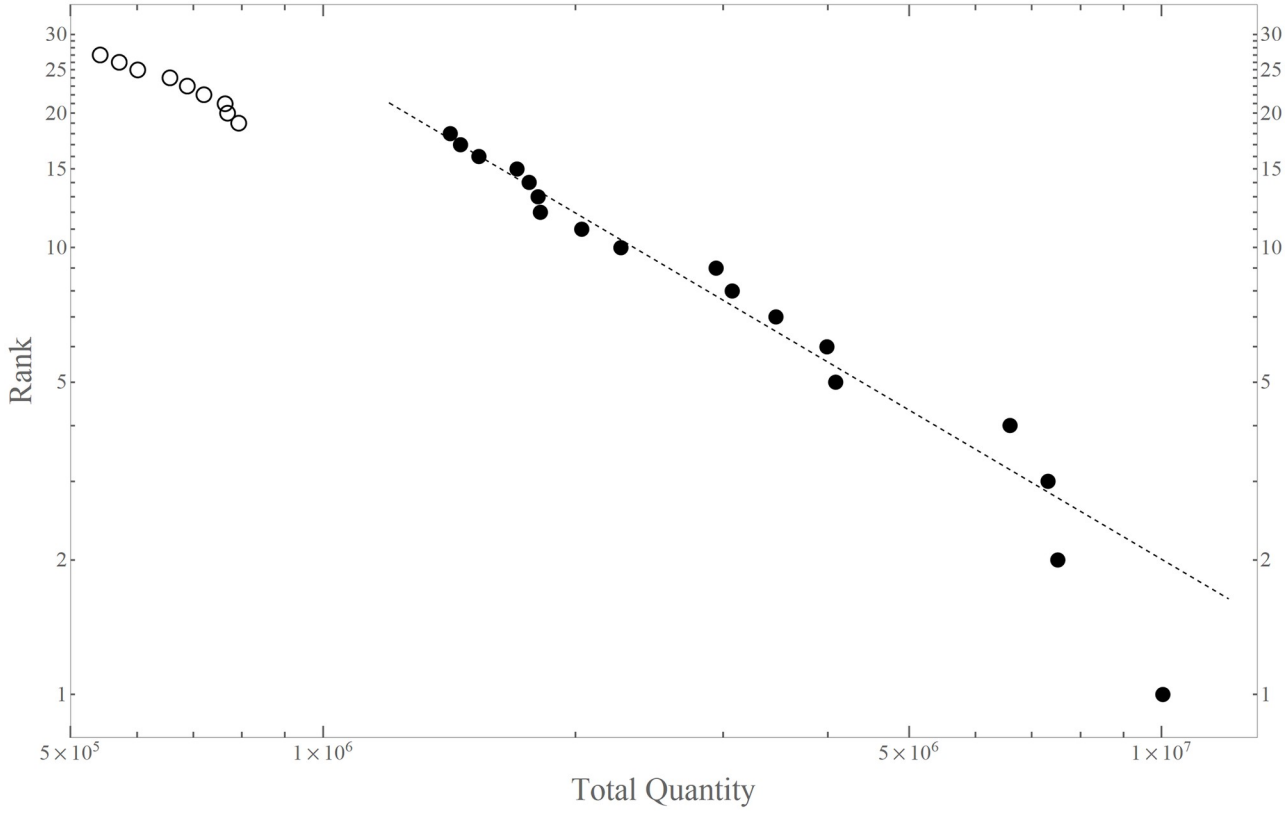

Fig 1. Rank-size plot of the top selling products. Each dot represents each products with the quantity Q of sales (horizontal) and the rank in descending order of Q (vertical). Filled dots denote the selected 18 brands. Note that both axes are in logarithmic scales.

https://doi.org/10.1371/journal.pone.0245531.g001

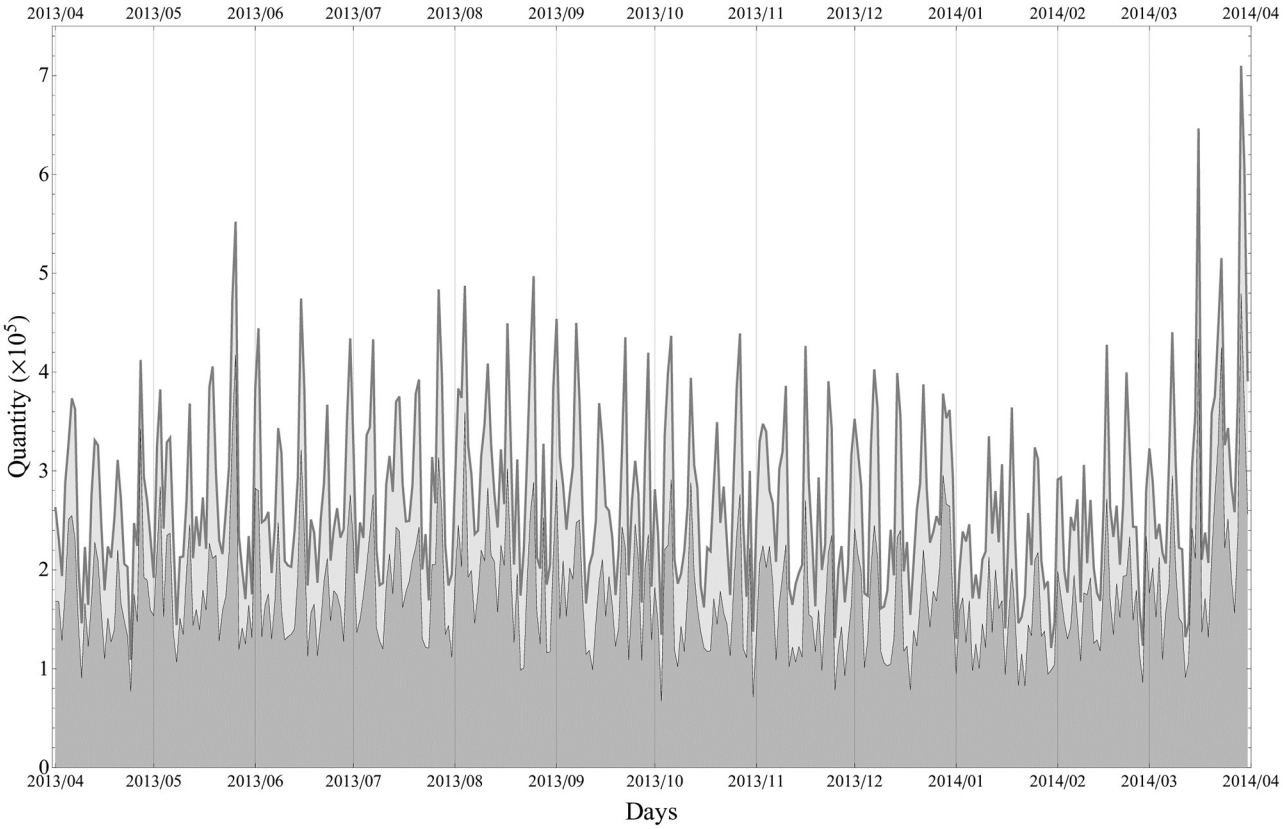

Fig 2. Daily total quantities. Light gray: all products, Dark gray: Selected 18 products. Apparent periodic peaks correspond to Sundays.

https://doi.org/10.1371/journal.pone.0245531.g002

Periodic peaks apparent in this plot occur at weekends when quantity rises on Saturdays and peaks on Sundays. The high peak structure at the end of the period; *that is*, at the end of March 2014, is explained by the VAT hike from 5% to 8% on April 1, 2014. It is remarked that we have data beyond this day for another several months, but we decided to take this one-year period to avoid bias from seasonal dependence and the strong influence from the increase in quantity (and the downfall on and after April 1, 2014).

The products are shown with codes, the first letter of which corresponds to the firm, and the second letter (digit) corresponds to the product: For example, the code "A1" means that the product is from firm 'A' and the product '1'. With five types of data for each product as listed in Table 1, we have $18 \times 5 = 90$ time series altogether. However as no communication activity was observed for some products, we set the threshold to 51 days: we used only the time-series data with 51 or more days of entry (more than or equal to once a week). With this threshold, we have 65 time series for purchases and related communication activities, which are listed in Table 2.

Descriptive statistics

Table 3 summarizes the means and standard deviations for the time series. Symbol "—" implies too sparse data due to no communication activity. We observe that these time series are highly volatile in temporal change. Let us denote the time series by $\tilde{x}_\alpha(t_i)$ where $\alpha = 1, \cdots, N(= 65)$ is the label for the time series, and $t_i = 1, \cdots, 365$ denotes the number of days. We

Table 2. Top-selling 18 products and availability of data in days.

Rank	Code	Total Sales	P	Q	Visit	TVAd	Search
1	D2	1.00×10^7	365	365	158	328	209
2	A4	7.52×10^6	365	365	25	287	163
3	B1	7.32×10^6	365	365	164	306	153
4	C3	6.59×10^6	365	365	29	173	76
5	B3	4.09×10^6	354	354	363	354	73
6	A3	3.99×10^6	357	357	29	249	34
7	A1	3.47×10^6	360	360	63	185	117
8	B5	3.08×10^6	365	365	62	274	19
9	D3	2.94×10^6	339	339	3	0	0
10	D1	2.27×10^6	355	355	284	334	270
11	B4	2.03×10^6	258	258	0	0	0
12	A6	1.82×10^6	330	330	0	0	0
13	B2	1.80×10^6	298	298	172	228	41
14	C1	1.76×10^6	341	341	82	219	168
15	A2	1.70×10^6	304	304	5	111	5
16	C2	1.53×10^6	342	342	39	216	25
17	A5	1.46×10^6	276	276	0	0	0
18	C4	1.42×10^6	198	198	0	0	0

https://doi.org/10.1371/journal.pone.0245531.t002

depict, as a sample of $\tilde{x}_\alpha(t_i)$, the five types of time series for the top product "D2" ($\alpha = 1, \cdots, 5$) in Fig 3. Price per unit quantity (P) is mostly stable but has spikes of increasing or decreasing price change. Quantity (Q) has volatility due to growing and sluggish sales. The frequency of site visits via PC or mobile device (Visit) have tranquility with sudden and short activities. The

Table 3. Descriptive statistics for the top-selling 18 products.

Rank	Code	P	Q	Visit	TVAd	Search
1	D2	0.287(0.011)	27483.0(14758.6)	18.9(66.6)	4770.1(6042.2)	1.92(3.59)
2	A4	0.294(0.014)	20596.9(14640.2)	—	4164.6(6126.4)	1.30(3.07)
3	B1	0.494(0.026)	20052.2(13850.3)	75.1(291.5)	4004.4(6704.8)	2.15(5.15)
4	C3	0.287(0.015)	18064.9(13970.4)	—	1271.7(3159.3)	0.54(1.76)
5	B3	0.295(0.023)	11192.3(10534.8)	175.6(365.3)	4792.1(5725.9)	0.50(1.92)
6	A3	0.350(0.025)	10926.4(11359.7)	—	2964.1(4648.4)	—
7	A1	0.506(0.035)	9499.7(8786.6)	9.5(38.5)	4232.1(7763.3)	1.10(2.56)
8	B5	0.301(0.023)	8424.9(8117.4)	19.3(118.3)	2861.0(3955.7)	—
9	D3	0.285(0.020)	8060.4(8510.2)	—	—	—
10	D1	0.567(0.039)	6205.9(5934.8)	73.7(184.8)	10051.1(10428.0)	3.81(6.15)
11	B4	0.296(0.018)	5573.2(8665.4)	—	—	—
12	A6	0.304(0.024)	4975.0(6037.5)	—	—	—
13	B2	0.362(0.028)	4943.8(6515.0)	46.2(115.2)	1769.2(3712.4)	—
14	C1	0.555(0.046)	4823.9(5755.2)	10.6(42.9)	2486.8(4732.9)	1.46(3.41)
15	A2	0.362(0.039)	4665.9(6790.5)	—	1359.0(4166.9)	—
16	C2	0.508(0.041)	4202.3(5325.6)	—	2019.6(5922.4)	—
17	A5	0.298(0.021)	3995.4(6606.3)	—	—	—
18	C4	0.296(0.028)	3884.6(6819.0)	—	—	—

Means and standard deviations (in parentheses). The symbol "—" corresponds to discarded data because of sparsity (see text).

https://doi.org/10.1371/journal.pone.0245531.t003

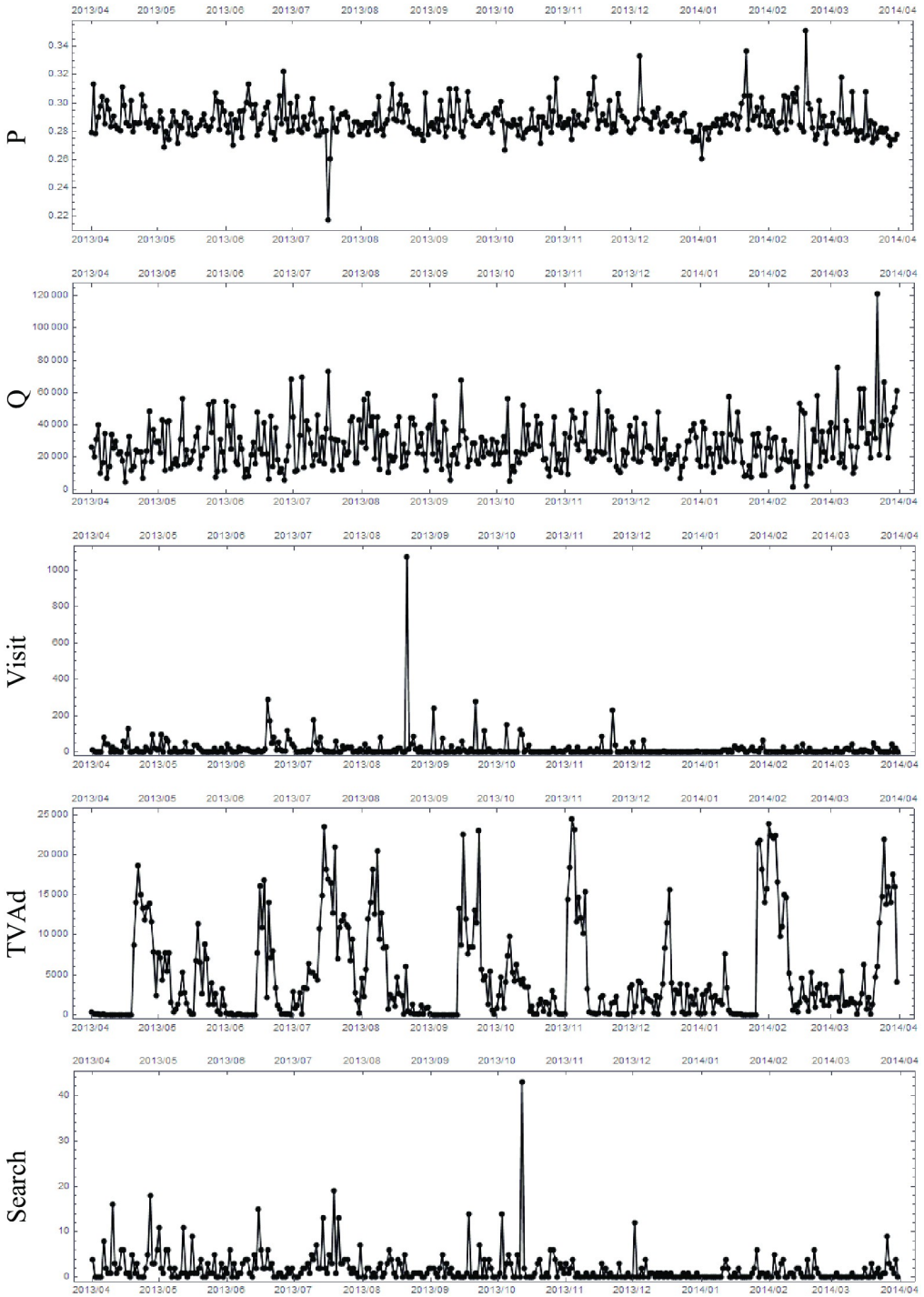

Fig 3. Sample of time-series. Time-series $\tilde{x}_{\alpha}(t_i)$ for the top product "D2" and for the five kinds of time-series, namely, P, Q, Visit, TVAd, and search are shown from top to bottom.

https://doi.org/10.1371/journal.pone.0245531.g003

frequency of TV advertising exposure (TVAd) has weak periodic behavior presumably due to the TV advertising activities of the product's firm and corresponding exposure to customers. The frequency of searches via PC or mobile device (Search) have continuous activities with bursts.

Then we use the standard method of subtracting the mean and dividing each time-series by the standard deviation. We denote the resulting time series by $x_\alpha(t_i)$. So the mean and standard deviation of $x_\alpha(t_i)$ are 0 and 1 respectively, for which we apply our methodology explained in the next section.

Complex Hilbert Principal Component Analysis (CHPCA)

Method

Any set of real world time-series data contains information on the behavior of individual time series and the inter correlations in the time series. In this study, we are interested in inter correlations in the time series. To identify the structure and dynamics of the customer choice process, we extract information on the inter correlation between price, sales, and other media approaches. Such comovement in the data set involves time lead and delay. Some time series follow other time series because of direct and indirect causal relationships. Here, our aim is to set up a methodology suitable for detecting inter relationships with time delay.

Principal component analysis (PCA) fulfills our goal partially. For this method, we calculate correlations between time series and identify the eigenmodes of the correlation matrix, which are independent comovements in the system. The larger the eigenvalue, the more significant the presence of the eigenmode. Some of the eigenmodes, however, are simply the result of random movements in the system. To identify which modes are significant real comovements, people often apply random matrix theory, which predicts the eigenvalues from the random time series. This method has several shortcomings.

i. When seeking comovements with time lead/delay, the time series is shifted relative to other time series to maximize the absolute value of the correlation coefficient. This is feasible for two time series but not so for a large number of time series. With 100 time series, for example, pair wise calculation is required for nearly 5,000 pairs. Then, there is the problem of combining them to obtain system-wide comovements.

ii. Random matrix theory (RMT) is practical on that the length of the time series (T) and the number of the time series (N) are both infinite with their ratio (T/N) kept finite, and all the time series has trivial auto correlation, none of which may be satisfied by the real data.

To overcome these difficulties, we use CHPCA and rotational random simulation RRS.

The former was originally introduced in [18–22] using the Hilbert transformation developed in [23–27] among others. The approach has been successfully applied in several areas of natural science and economics [13, 14, 28, 29]. We further introduced improvements on CHPCA by [12].

In CHPCA, we complexify each of the time series' Hilbert transformation as an imaginary part and then calculate the complex correlation matrix. We provide a pedagogical explanation of the merits of this method. The Hilbert transformation, simply put, transforms each of the Fourier components in the manner $\cos \omega t \rightarrow -\sin \omega t$ and $\sin \omega t \rightarrow \cos \omega t$. Therefore, the complexification converts $\cos \omega t$ to $e^{-i\omega t}$ and $\sin \omega t$ to $ie^{-i\omega t}$; clockwise rotation on its complex plane. Furthermore, the Hilbert transformation converts

$$\cos \omega(t + t_0) = \cos \omega t \cos \omega t_0 - \sin \omega t \sin t_0 \rightarrow e^{-i\omega(t+t_0)}. \quad (1)$$

We denote the complex time series obtained from $x_\alpha(t)$ and standarized (so that its means is equal to zero and its standard deviation is equal to one) as $z_\alpha(t)$. Complex correlation coefficients (CCC) are defined as inner products of one (complex and normalized) time series

$(z_\alpha(t))$ with another:

$$C_{\alpha\beta} := \sum_t z_\alpha(t) z_\beta^*(t). \tag{2}$$

Here and hereafter, $\cdot^*$ denotes the complex conjugate of $\cdot$. If the time series α and β are made of Fourier components of the same ω but with time constants t_α and t_β, the CCC has a phase factor proportional to $t_\alpha - t_\beta$, the time-difference between the two time series. If the time series contain multiple Fourier components, the phase of the CCC gives a nonlinear mean of the time differences of each combination of the Fourier components. Thus, analysis of the resulting complex correlation matrix enables us to obtain a view of comovements with system time-lag. By definition, this is one calculation that avoids any pairwise optimization analysis required by PCA. The eigenmode $\mathbf{e}_n$ of the complex correlation matrix $\mathbf{C} = \{C_{\alpha\beta}\}$ is defined by the following:

$$\mathbf{C}\mathbf{e}_n = \lambda_n \mathbf{e}_n, \tag{3}$$

where the subscript n is defined as the eigenvalues λ_n in descending order, $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N$. The eigenvalues satisfy an identity

$$\sum_{n=1}^{N} \lambda_n = N. \tag{4}$$

The time series are expanded in terms of the eigenmodes:

$$\mathbf{x}(t) = \sum_{n=1}^{N} s_n(t) \mathbf{e}_n, \tag{5}$$

where the coefficient $s_n(t)$ is called the *mode signal*, satisfying

$$\lambda_n = \sum_{t=1}^{T} |s_n(t)|^2. \tag{6}$$

In this sense, the eigenvalue λ_n is the strength of the presence of the corresponding eigenmode $\mathbf{e}_n$.

To avoid using the RMT, we employ RRS, introduced by [30]. This is done by (1) "rotating" each time series in time-direction (by attaching its end to the beginning) randomly, thus destroying the inter correlation between the time series while preserving the autocorrelation; (2) calculating the CCC and its eigenvalues several times ($10^4 \sim 10^5$ times typically); (3) grouping the eigenvalues by rank in descending order and obtaining the range (parametrized by the mean and the standard deviation) of each distribution; (4) comparing the true eigenvalue (in descending order) with the distribution of the same rank obtained by the previous step. If the true eigenvalue is larger than the specified range of the distribution, the corresponding eigenmode is regarded as significant. This method overcomes the shortcoming of the RMT by allowing us to deal with data with nontrivial auto correlation and T and N not so large.

In this sense, the methodology of CHPCA with RRS is ideal for our purpose, which is to identify the customer choice process in our data.

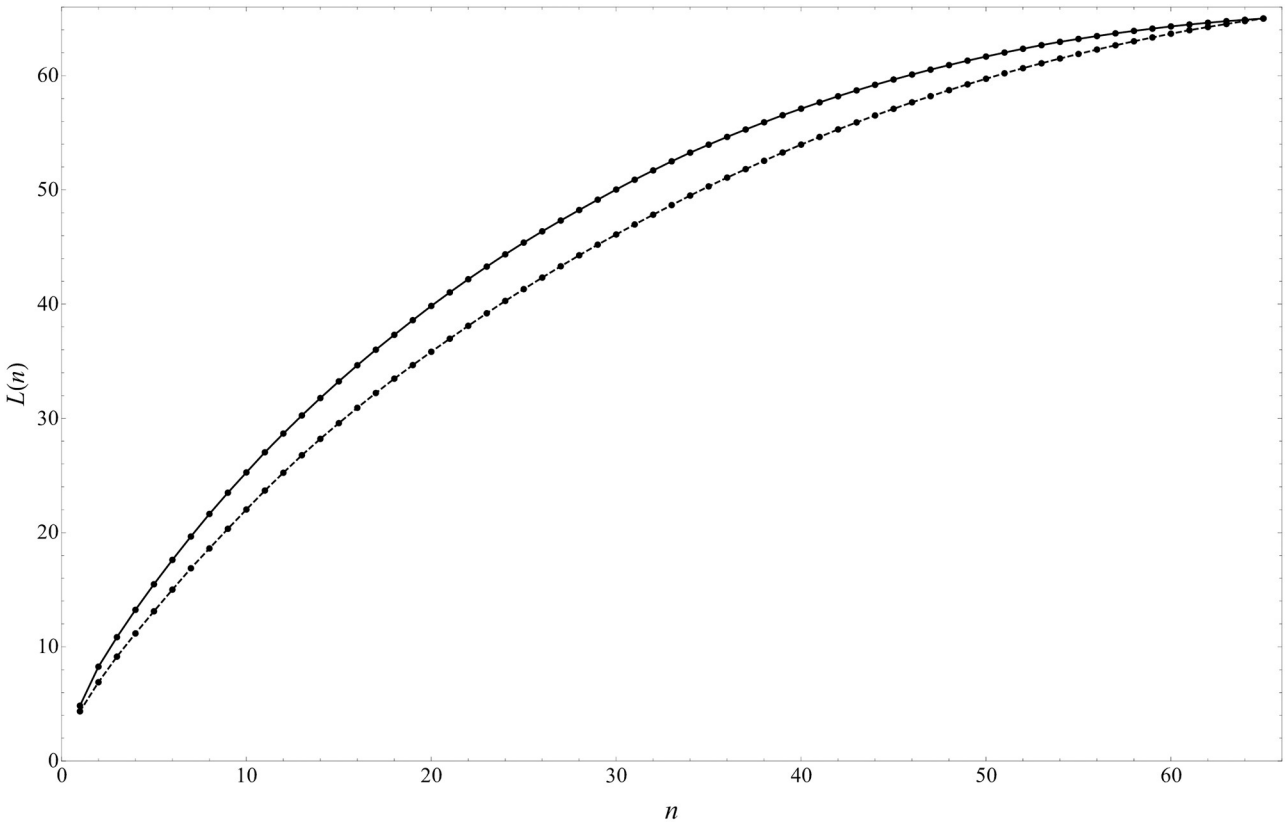

Fig 4. Cumulative eigenvalue $L(n)$ defined in Eq (7). This shows how the first n eigenmodes can explain the time-series with respect to the cumulative eigenvalue $L(n)$, in other words, the strength of eigenmode (see also Eq (6)).

https://doi.org/10.1371/journal.pone.0245531.g004

Results

The eigenvalue distribution is shown in Fig 4, where the ordinate is the cumulative eigenvalue

$$L(n) \coloneqq \sum_{k=1}^{n} \lambda_n. \tag{7}$$

The green dots are for CHPCA and blue for PCA. As explained, the eigenvalue shows the rate of the presence of the corresponding eigenmode in the data. Therefore, this plot shows that eigenmodes of CHPCA are more significant in explaining the variation in data than the corresponding eigenmodes in PCA. This is natural since PCA misses comovements with lead/lag.

The result of the RRS analysis of 10^4 times of the simulation is summarized in Fig 5 for the eigenvalues $n = 1, 2, 3$ from top to bottom. In each plot, the actual eigenvalue is shown by the thick vertical ticks. The distribution shown in gray is the distribution of the corresponding RRS eigenvalues, whose mean is shown by the short vertical line, and the 2σ range is shown by the horizontal error bars. Since the eigenvalues #1 and 2 are well above the 2σ range and #3 is not, we find that the top two eigenmodes are significant, inter-correlating comovements.

Since the largest two eigenvalues are

$$\lambda_1 = 4.845, \qquad \lambda_2 = 3.416, \tag{8}$$

respectively, these top eigenmodes take the share of $\sqrt{(\lambda_1 + \lambda_2)/65} \simeq 0.3565$; *that is*,, 35.7% of the data are due to comovements.

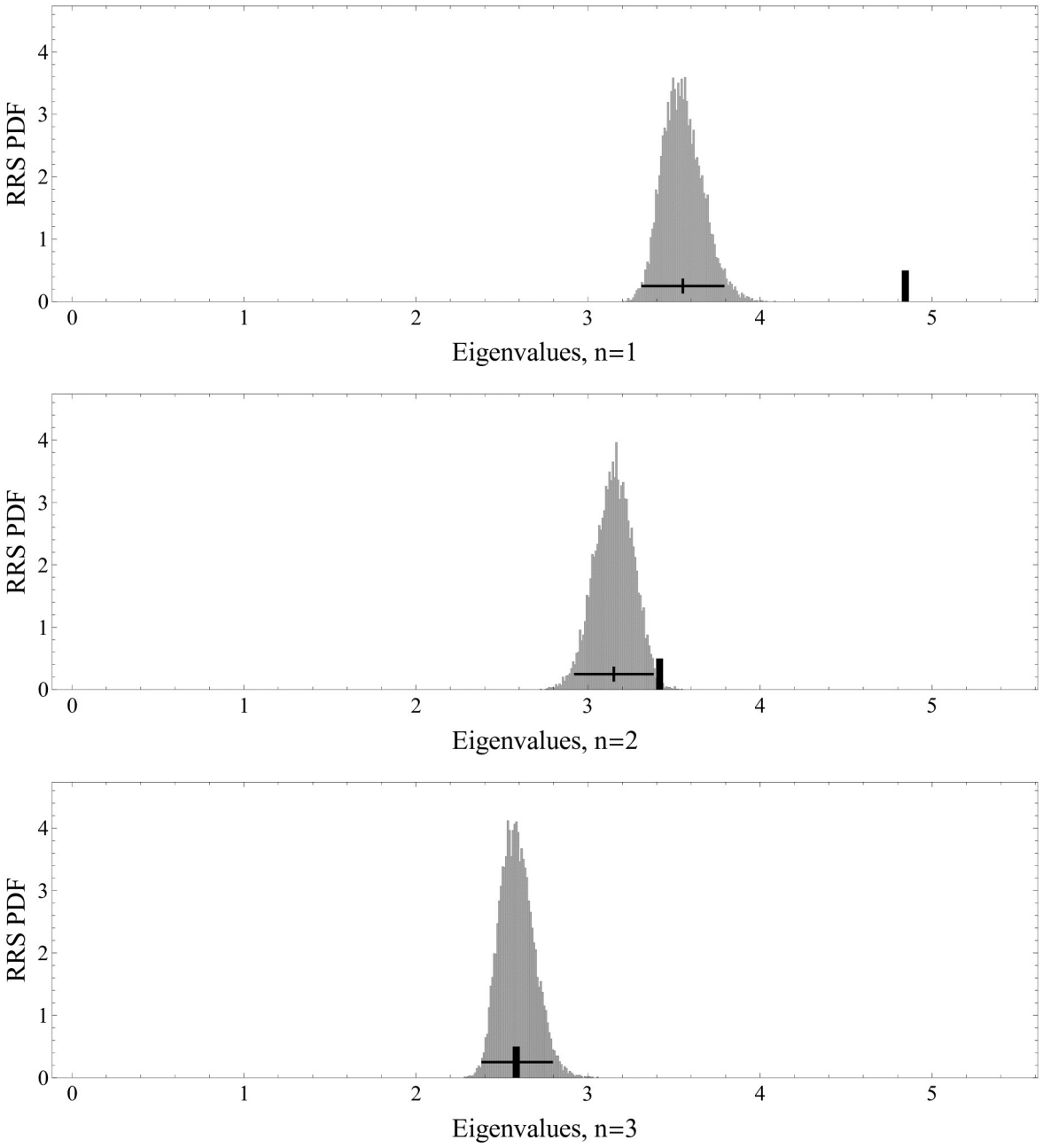

Fig 5. The top three eigenvalues and the corresponding RRS eigenvalue distributions. The eigenvalues are shown with thick ticks and the RRS eigenvalue distributions are shown with shaded bell-shaped curves. Horizontal bars show the 2σ ranges. The eigenvalues #1 and #2 are above the RRS 2σ range and, therefore, are significant. The eigenvalues #3 and below are not.

https://doi.org/10.1371/journal.pone.0245531.g005

The top and the second eigenvector components are shown in Fig 6, where each component is shown by a marker specified by the product code at its top and the style shown in the legend. (The details of these eigenmodes are given in the S1 File) The horizontal axis is its phase, and the vertical axis its absolute value. The arbitrary overall phase in the eigenvector $\mathbf{e}_n$ is chosen so that the components representing purchase quantities are toward the right-hand side of the plots. By the definition of the complexification, the phase corresponds to the time-variation; the components on the left move first, and the components to the right follow. We

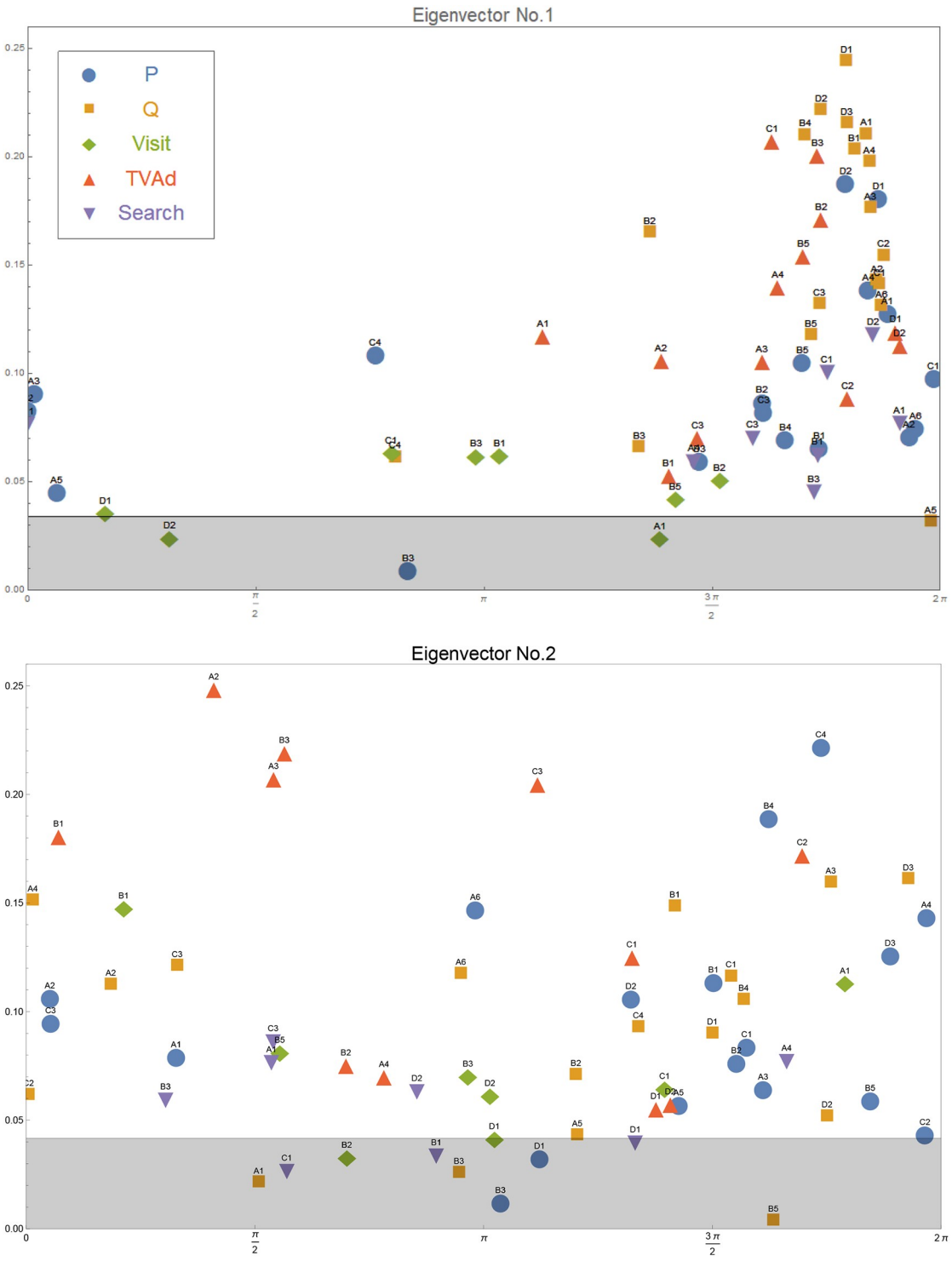

Fig 6. The components of the first (upper) and the second (lower) eigenvectors. Each component is indicated by a marker with the label of product code and style. Horizontal position is each component's phase, and vertical position is its absolute magnitude. See more in the main text.

https://doi.org/10.1371/journal.pone.0245531.g006

have changed the phase of prices by π to be consistent with the common knowledge that when the price goes down, quantity goes up. In these plots, we also show the significance level of the absolute values of the components by the gray bands. Components with less absolute values have less significance in the respective eigenmode. To clarify this significance level, we add a random time series to the original data set and measure its absolute value in the first and second eigenvectors. Repeating this simulation 100 times, we identify the distribution of the absolute value. The shaded gray area bounded by a solid horizontal line is 2σ range. Therefore, the components above the gray zones are the components with significant presence in the respective comovements.

The comovement of marketing instruments and purchase quantities across products represented in Fig 6 may still be too complicated for marketers to interpret. Thus, we offer an additional method to reduce information obtained from CHPCA focusing on synchronization of multiple time series.

Hodge decomposition and synchronization network

Method

The complex correlation coefficient, $C_{\alpha\beta}$, represents how strongly a pair of α and β are correlated possibly with lead and lag. However, as we have seen above, only two eigenmodes are significant and the rest are random noises. Therefore, we need to pick up the part of $C_{\alpha\beta}$ that are significant. This is done as follows: First we note that the complex correlation matrix $\mathbf{C}$ is written as

$$\mathbf{C} = \sum_{n=1}^{N} \lambda_n \mathbf{e}_n \mathbf{e}_n^{\dagger} \tag{9}$$

(see Eq (3)). We hereby define the significant part of the complex correlation matrix by limiting the sum to the significant modes:

$$\mathbf{C}^{(\mathrm{sig})} = \sum_{n=1}^{2} \lambda_n \mathbf{e}_n \mathbf{e}_n^{\dagger} \tag{10}$$

The strength of the correlation is given by the magnitude of the matrix elements of this matrix

$$\rho_{\alpha\beta} \coloneqq |C_{\alpha\beta}^{(\mathrm{sig})}| \ , \tag{11}$$

and the lead and lag can be measured by the phase

$$\theta_{\alpha\beta} \coloneqq \arg C_{\alpha\beta}^{(\mathrm{sig})} \ . \tag{12}$$

Note that α leads β if $\theta_{\alpha\beta} < 0$, and α lags β if $\theta_{\alpha\beta} > 0$ because we defined the direction of time by $e^{-i\omega t}$ (see Eq (1)).

If we consider all the pairs in the complex correlation $C_{\alpha\beta}^{(\mathrm{sig})}$, we have a complete graph in which every node α is connected to all the other nodes. It is difficult to understand how individual α leads or lags others in a more systematic way. To overcome this difficulty, we select *comoving* pairs with strong correlation in the following way, and then use the so-called *Hodge decomposition* of a flow on a directed and weighted network, which we call *synchronization network*.

First, we select pairs of α and β with

- comovement: $0 < \theta_{\alpha\beta} < \pi/2$,

- significant correlation: $\rho_{\alpha\beta} > \rho_*$ where ρ_* is a threshold given below.

In the first condition, we consider only the region $0 < \theta_{\alpha\beta} < \pi/2$, because the correlation matrix satisfies the Hermite conjugate relation; that is, $C^{(\mathrm{sig})}_{\beta\alpha} = C^{(\mathrm{sig})^*}_{\alpha\beta}$, so that the pairs in the region $-\pi/2 < \theta_{\alpha\beta} < 0$ are always in the region $0 < \theta_{\alpha\beta} < \pi/2$. In the second condition, we determine the threshold ρ_* as follows. If ρ_* is too large, the number of pairs satisfying the condition is too small and, eventually, the graph becomes disconnected; if ρ_* is too small, the graph is almost fully connected. In both cases, it would be difficult to understand the lead/lag relation. Therefore, we select ρ_* that connects the graph at its largest value. The resulting graph includes 65 nodes and 1,391 edges.

Second, we use a mathematical method of ranking nodes according to their location in terms of upstream and downstream flow in a directed network to identify which nodes are leading and lagging in the entire relation. In our case, a *flow* is said to be present from α to β if $0 < \theta_{\beta\alpha} < \pi/2$ and $\rho_{\beta\alpha} = \rho_{\alpha\beta} > \rho_*$ with the amount of flow or weight, $\rho_{\alpha\beta}$.

We briefly recapitulate the method (see [15] for example), which is called *Hodge decomposition*. Denote the adjacency matrix of the binary and weighted network by

$$A_{\alpha\beta} = \begin{cases} 1 & \text{if there is a directed edge from } \alpha \text{ to } \beta, \\ 0 & \text{otherwise,} \end{cases} \tag{13}$$

and

$$B_{\alpha\beta} = \begin{cases} f_{\alpha\beta} & \text{if there is a directed edge with a flow,} \\ 0 & \text{otherwise,} \end{cases} \tag{14}$$

where $f_{\alpha\beta}$ is a flow from α to β, and it is assumed that $f_{\alpha\beta} > 0$. Note that there can be such a pair of nodes that has both $A_{\alpha\beta} = 1$ and $A_{\beta\alpha} = 1$ and also that has both $f_{\alpha\beta} > 0$ and $f_{\beta\alpha} > 0$.

Then, the net *flow* from α to β is defined by

$$F_{\alpha\beta} = B_{\alpha\beta} - B_{\beta\alpha} \ . \tag{15}$$

Let us also define the net *weight* between α and β by

$$W_{\alpha\beta} = A_{\alpha\beta} + A_{\beta\alpha} \ . \tag{16}$$

Note that $F_{\alpha\beta}$ is anti-symmetric while $W_{\alpha\beta}$ is symmetric.

Hodge decomposition is given by

$$F_{\alpha\beta} = W_{\alpha\beta}\,(\phi_\alpha - \phi_\beta) + F^{(\mathrm{loop})}_{\alpha\beta} \ , \tag{17}$$

where $F^{(\mathrm{loop})}_{\alpha\beta}$ is a loop flow; that is, divergence-free:

$$\sum_\beta F^{(\mathrm{loop})}_{\alpha\beta} = 0 \tag{18}$$

by definition. ϕ_α is called *Hodge potential* of node α.

Rewriting Eq (18), we have for each $\alpha = 1, \cdots, N$,

$$\sum_\beta L_{\alpha\beta}\,\phi_\beta = \sum_\beta F_{\alpha\beta} \ , \tag{19}$$

Here, $L_{\alpha\beta}$ is the so-called graph Laplacian defined by

$$L_{\alpha\beta} = \delta_{\alpha\beta}\sum_{\gamma} W_{\alpha\gamma} - W_{\alpha\beta} \ , \tag{20}$$

where $\delta_{\alpha\beta} = 1$ if $\alpha = \beta$ and $\delta_{\alpha\beta} = 0$ otherwise. Given $F_{\alpha\beta}$ and $W_{\alpha\beta}$, Eq (19) are simultaneous linear equations to determine the Hodge potential ϕ_α of all the nodes α.

Note that simultaneous linear Eq (19) are not independent of each other. In fact, the summation over α of (19) is zero, as is easily shown, corresponding to the fact that there is a freedom to fix the origin of potential. It is not difficult to prove that if the network is weakly connected; that is, connected when considered an undirected graph, the potential can be determined uniquely up to the choice of the origin [31]. In the following, we use the convention that the mean is zero:

$$\sum_{\alpha}\phi_\alpha = 0 \ . \tag{21}$$

Thus, if we delete the loop flow, the remaining flow can be represented by a flow caused by the difference in potential between a pair of nodes. The Hodge potentials, therefore, can reveal which nodes are located in upstream or downstream sides in the relative relationship of the directed network. We emphasize that such information cannot be obtained simply by looking at the pairwise correlation among nodes because the entire connectivity of all the links is required to discard the loop flow and to determine the potentials.

Results and interpretation

Fig 7 shows a layout of the synchronization network. The vertical position of each node corresponds to its Hodge potential as a constraint in the force-directed algorithm of the graph layout. Upstream (leading) nodes are located toward the top while downstream (lagging) nodes are toward the bottom.

To depict a customer choice process covering all competitive products, Hodge potentials are used to constitute the fundamental time sequence of the exposures to marketing instruments (TV Ad, web/mobile site visit, search and price) and the purchase quantity for each product. In Fig 8, the time is passing from top to bottom along the vertical axis. The distance of this axis can be expressed in a time scale such that the distance of one corresponds to 1.55 days. The horizontal axis is nominal, where 18 products are arrayed in an arbitrary order. For instance, for Product 1, just after its price decreases and the search behavior increases, both of which occur almost simultaneously, the purchase quantity (sales) increases followed by an increase in exposure to web/mobile sites and TV advertising.

First, we trace the time sequence of variables within each product. For more than half of the products (11 out of 18), a decrease (increase) in price leads to an increase (decrease) in quantity to a certain extent. For some products, a change in price occurs almost simultaneously with a change in quantity. For these two groups, the behaviors pf prices and quantities are consistent with standard economic theory. On the other hand, for Products 15 and 18, an increase (decrease) in quantity leads to a decrease (increase) in price. This phenomenon seems to be an anomaly as an effect of prices while it could be explained as an outcome of rational behavior; for instance, it could emerge when the demand is expanded by attracting new customers with lower willingness-to-pay [32].

The increased exposures to TV advertising lag behind the increased purchases for seven of the 13 products that executed TV advertising in the observed period. This may suggest that TV advertising by firms is a reaction to an increase in demand, not as an upfront investment or, in

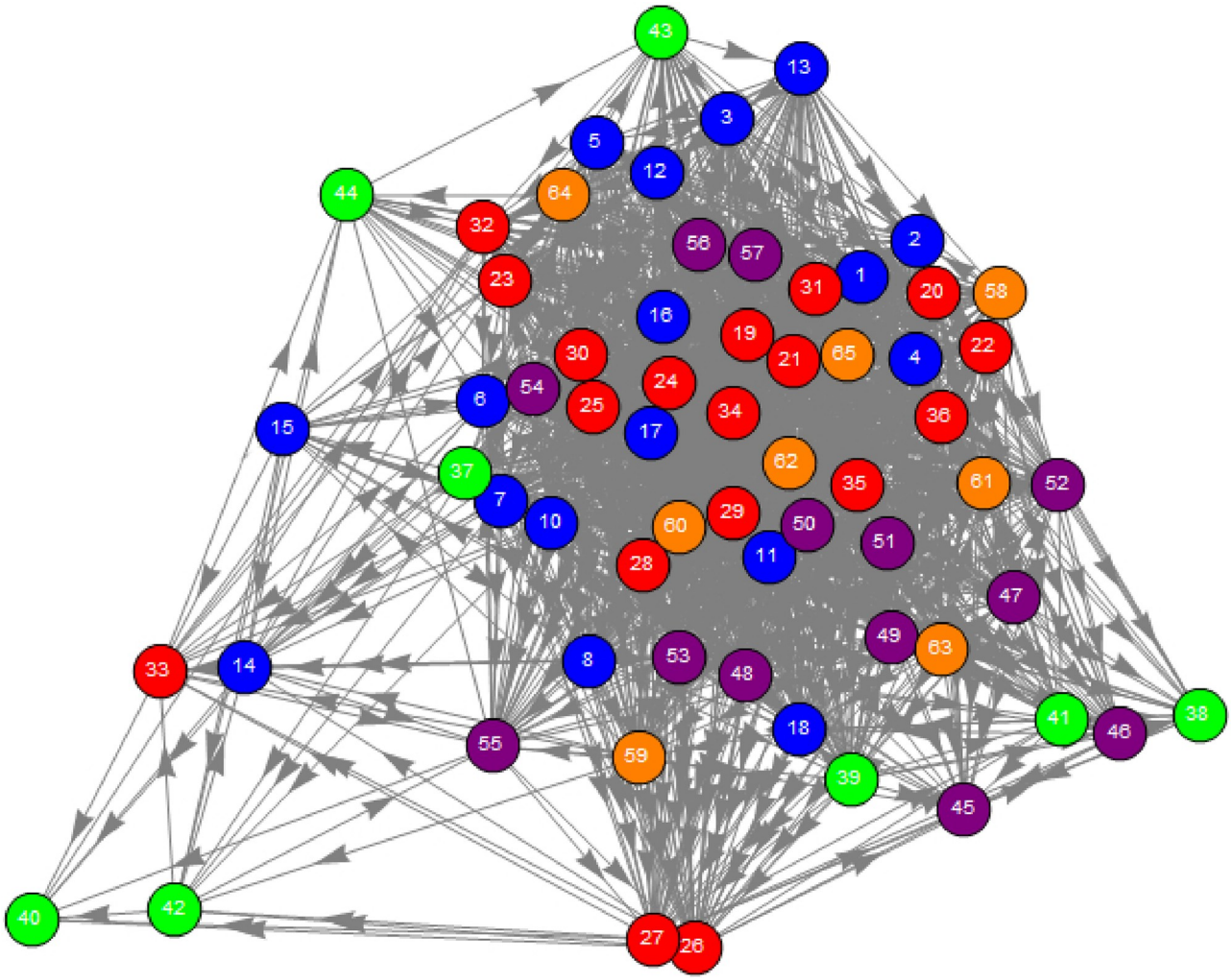

Fig 7. The synchronization network's graph layout. The vertical position of each node corresponds to its Hodge potential as a constraint in the force-directed algorithm of the graph layout. Upstream (leading) nodes are located toward the top while downstream (lagging) nodes are toward the bottom. The node no.9 is not drawn as it has only one link and is not relevant to this visualization.

https://doi.org/10.1371/journal.pone.0245531.g007

the latter case, after a significant time lag. The time sequence of web/mobile site visits or searches is less consistent across products than price or TV advertising. A possible reason is that the exposures to these marketing instruments is less controllable for firms, reflecting the idiosyncratic nature of individual products. In that sense, this inconsistency shows the advantage of our approach that analyzes the observed data purely empirically without any strong assumptions.

Second, we can compare the Hodge potentials horizontally between products, which indicates synchronization (comoving almost simultaneously) of marketing instruments between different products of a firm or even between firms. For instance, as Fig 8 shows, a price cut and sales increase for Products 1, 2 and 3, which are the main products of Firm 1, tend to be synchronized. That is, Firm 1 might coordinate price promotion consistently among their own products compared to rivals. These variables seem to be synchronized also between Firms 1 and 3, suggesting that these firms are mutually competing more intensively. As the potentials

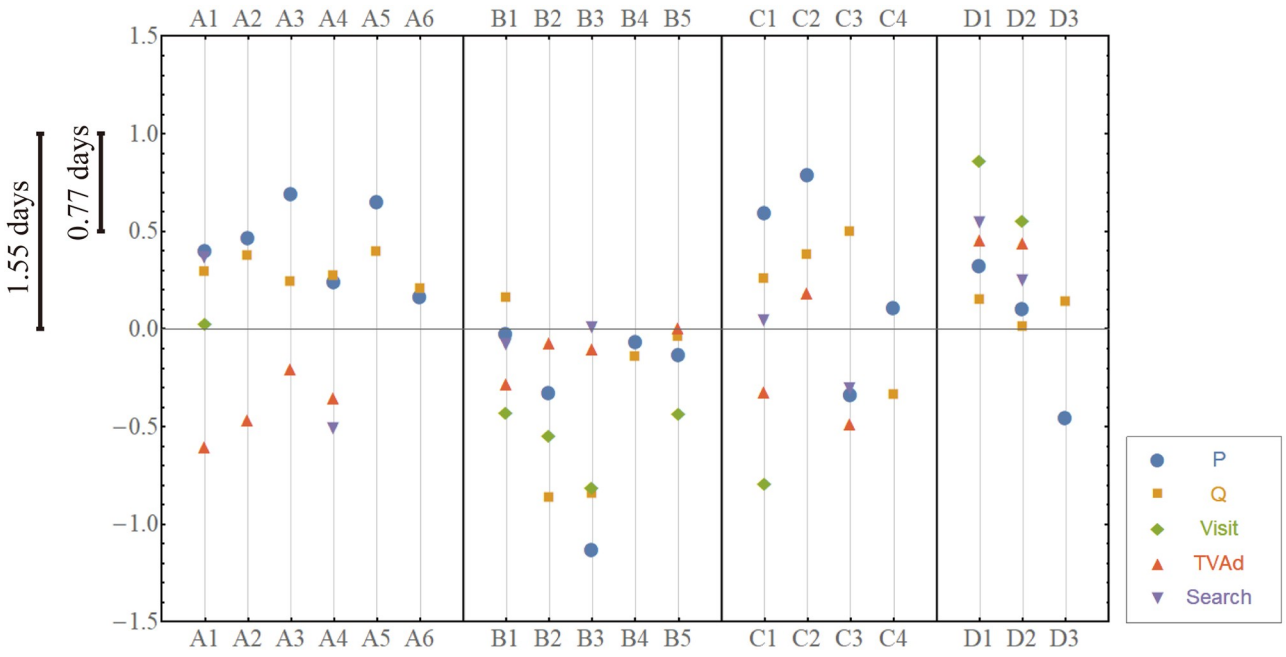

Fig 8. The fundamental sequence of exposures to marketing instruments and purchases for each product. The vertical axis represents the Hodge potential, whose value is larger as it leads than others. Horizontally, products are arrayed in an arbitrary order. If the positions of two time series are closer, they are comoving almost simultaneously with each other.

https://doi.org/10.1371/journal.pone.0245531.g008

show that these firms tend to change prices before sales, their main weapon for competitive reaction is price promotion.

It is noteworthy that the potentials for purchase quantity are relatively concentrated within the narrow band for most products, implying that beer consumption is highly synchronized as a whole. The reason is easily explained by the established knowledge that typical beer consumption increases during higher temperatures or special occasions such as weekends or holidays. A more interesting finding is the existence of a few products (Products 8 and 9 of Firm 2) outside the band. These products are interpreted as satisfying some niche demand in the market. Firm 2 seems to be differentiated since its pricing behavior is not necessarily synchronized with Firms 1 and 3 as a whole. Another prominent feature of this firm is that customers visited its web/mobile site more frequently while they seldom visited the sites of Firms 1 and 3 (or there may not be a competitor). Customers may visit Firm 2's site after making a purchase or exposure to TV advertising. On the other hand, customers seem to visit the site in advance. Such differences might be due to variations in marketing strategies.

Customer profile

We found in the Method Section that there are two significant eigenmodes in the aggregate behavior of customers; the remaining $N - 2$ modes can be discarded as "noise." We are interested to see how the *individual* behavior of customers can be represented in terms of these two significant eigenmodes. Such representation can provide deeper insight into how the two significant eigenmodes can be interpreted by examining individual customer's profiles such as their gender, age, income, other attributes, and their preferences for specific products.

Let us denote individual customer's time series by $x_{p,\alpha}(t)$ ($p = 1, \ldots, P$) where the index p denotes individual customers, and P is the total number of customers in our data; $P = 1,738$. $\alpha = 1, \ldots, N$ is the same index as used in the Method Section with $N = 65$.

We first complexify $x_{p,\alpha}(t)$ into complex time series, denoted by $z_{p,\alpha}(t)$, and standardize (subtract mean and normalize by standard deviation) it precisely in the same way as we did in the Method Section. Thus, we have

$$\hat{z}_{p,\alpha}(t) = \frac{z_{p,\alpha}(t) - \langle z_{p,\alpha} \rangle}{\sigma_{p,\alpha}}. \tag{22}$$

If $x_{p,\alpha}(t)$ is identically equal to zero during all of time t for some α (e.g., a customer was not exposed to any advertising), we use the convention that $\hat{z}_{p,\alpha}(t) = 0$.

Then, we project the time series to a space spanned by two significant eigenvectors (we term it *customer space* hereafter); that is,

$$a_{n,p}(t) = \sum_{\alpha=1}^{N} (\mathbf{e}_n)_\alpha^* \hat{z}_{p,\alpha}(t), \tag{23}$$

for the two significant modes $n = 1, 2$ where $(\mathbf{e}_n)_\alpha$ is the α-th component of the eigenmode $\mathbf{e}_n$. To locate each customer in a space spanned by the two eigenmodes, we calculate the temporal mean of the squared magnitude of the projected time series $a_{n,p}(t)$, namely,

$$X_{n,p} = \frac{1}{T} \sum_t |a_{n,p}(t)|^2, \tag{24}$$

which gives us two-dimensional coordinates for each customer p.

Fig 9 shows the resulting spatial representation of Eq (24) for all the P customers. Recalling (6) in the Method Section, each coordinate's value $X_{n,p}$ can be compared with the eigenvalues λ_1 and λ_2, which are numerically given by Eq (8). We observe that there are customers whose positions are consistent with Eq (8) in the sense that $X_{1,p}/X_{2,p} \sim \lambda_1/\lambda_2$. There are, however,

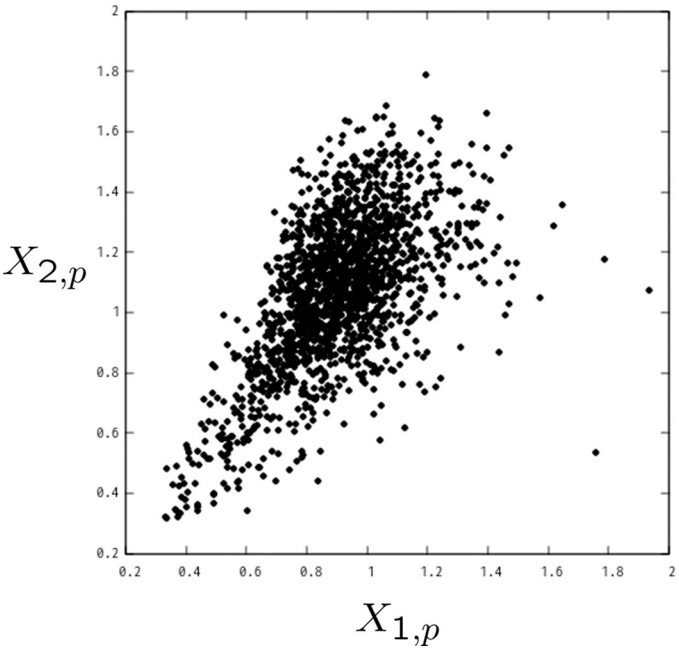

Fig 9. Individual customer's projected representation for the two significant modes. Each location is calculated by the temporal mean of the squared magnitude of the projected time series (see Eq (24)).

https://doi.org/10.1371/journal.pone.0245531.g009

Table 4. Regression of coordinates of customer space to customer profiles.

	Model 1.1			Model 1.2			Model 2.1			Model 2.2		
Criterion Variable:	**coord. for 1st eigen mode: $X_{1,p}$**						**coord. for 2nd eigen mode: $X_{2,p}$**					
Explanatory Variables:	coef.	s.e.		coef.	s.e.		coef.	s.e.		coef.	s.e.	
Intercept	.9050	.0045	aa	.9050	.0044	aa	1.0720	.0058	aa	1.0720	.0058	aa
Age	.0164	.0048	aa	.0159	.0047	aa	.0347	.0061	aa	.0352	.0061	aa
Gender	.0033	.0059		.0040	.0059		.0036	.0077		.0036	.0077	
Marrital Status	.0013	.0050		.0007	.0049		.0030	.0064		.0018	.0064	
Personal Income	.0071	.0065		.0084	.0065		−.0099	.0084		−.0092	.0084	
Houshold Income	−.0011	.0055		−.0034	.0055		−.0067	.0071		−.0078	.0072	
Total Purchase (freq.)	.0351	.0055	aa	.0331	.0056	aa	.0257	.0070	aa	.0273	.0073	aa
Total Purchase (mℓ)	.0139	.0055	b	—	—		.0165	.0071	b	—	—	
Purchase—A1				.0091	.0047	c				−.0007	.0060	c
A2				.0028	.0045					.0082	.0058	
A3				.0035	.0045					.0066	.0058	
A4				.0031	.0047					.0086	.0061	
A5				.0059	.0045					.0107	.0058	
A6				−.0005	.0047					.0007	.0060	
B1				.0044	.0046					.0008	.0059	
B2				.0046	.0045					−.0013	.0058	
B3				−.0058	.0045					−.0048	.0059	
B4				−.0002	.0045					.0070	.0058	
B5				.0117	.0047	b				.0092	.0061	
C1				−.0073	.0046					−.0033	.0060	
C2				−.0006	.0045					−.0037	.0059	
C3				.0035	.0046					.0162	.0059	aa
C4				.0067	.0045					.0073	.0059	
D1				.0168	.0046	aa				.0031	.0059	
D2				−.0014	.0048					−.0048	.0062	
D3				.0170	.0045	aa				.0104	.0058	
R^2	.0650			.0870			.0484			.0870		
Adjusted R^2	.0612			.0742			.0445			.0742		

(aa: $p < .001$; a: $p < .01$; b: $p < .05$; c: $p < .10$).

https://doi.org/10.1371/journal.pone.0245531.t004

more diversified customers in the two-dimensional space. Such diversification tells us that the location of each customer might be related to the heterogeneity in customer behavior.

To assess how the customer space is associated with each customer's profile, we conduct regression analysis where either of the coordinates in the customer space, $X_{1,p}$ or $X_{2,p}$, is used as a criterion variable, and multiple variables representing customer profiles are used as explanatory variables, which are all available in our data set. First, we select age (nine-point scale from age 20 to 24 to age 60 or older), gender (0 for male, 1 for female), marital status (0 for unmarried, 1 for the married), personal income (nine-point scale), and household income (five-point scale) by preliminary analysis. Second, to capture each customer's beer preference, total purchase frequency and quantity (ml) over all products are added to the predictors. As the results of Model 1.1 and 2.1 of Table 4 show, the estimated coefficients are significant only for age (0.1% significant), total purchase frequency (0.1% significant), and quantity (5% significant) for both $X_{1,p}$ or $X_{2,p}$. The values of R^2 indicate that most of the variations are not explained by the above predictors.

Alternatively, we replace total purchase quantity with each product's purchase quantities to capture individual differences in product-level preference. The results are presented for Model 1.2 and 2.2 in Table 4. Compared to the above models, the coefficients for age and total purchase frequency are consistently significant while the adjusted R^2s are slightly increased, implying separating total purchase quantity into purchase quantity for products may marginally improve the model fit. The coefficients of a few products are significant from the 0.1% to 10% level for $X_{1,p}$ and $X_{2,p}$. Thus, locations in the customer space could be explained to some extent by age, purchase frequency at the category level, and purchases of some remarkable products. However, it should be noted that most of the variations in the customer space remain unexplained. In other words, the individual locations in a customer space might reflect an infinite number of factors, only some of which could be measured via customer surveys or purchase history tracking. Hence, our proposed projection method contributes to the evaluation of individual deviations from a representative customer choice process.

Conclusion

The proliferation of marketing instruments and competitive products are rendering the consumer choice process increasingly complex to define. Existing methods used for this purpose ignore the existence of competitive products or underestimate the range of competition, even in cases where most customers consider multiple alternative products by searching or shopping. One of the main reasons for this is the limited ability of the existing methods, such as multivariate time series analyses, to handle the complexity caused by multiple products competing with multiple instruments. The market with 18 products and 5 variables (4 marketing instruments and one purchase quantity), analyzed in this study, is not tractable without strong assumptions to drastically reduce parameters. Our proposed CHPCA overcomes this limitation without any strong assumptions. Furthermore, CHPCA's supplemental methods, synchronization network and Hodge decomposition, can be used to summarize and visualize the results to be more interpretable.

This study shows that a set of our proposed methods could be used to effectively understand the consumer choice process embedded in enormously high-dimensional time-series data. The application of our method to the beer market in Japan derives some interesting findings. First, for most products on the market, the increase (decrease) in a product's price leads to the decrease (increase) in its purchase quantity (as the standard economic theory predicts) with a few exceptional cases. Second, the exposure to TV advertising increases later than the purchase quantity in many cases. Simply put, consumers notice a price change in a store, buy a product, and are then exposed to TV ads later. Third, the timing of consumers' visits to product web/mobile sites or the usage of search engines varies across products. Fourth, we find synchronization across products, in particular within each firm, rather than across firms. These findings imply that individual firms are heterogeneous, each adopting a distinctive coherent marketing strategy.

The fourth point has an important implication for economic policy making. Synchronization of marketing strategy between firms indicates that their behavior could be competitive if prices are decreasing but be collusive if prices are increasing. The latter case should attract a strong interest of anti-trust agencies. Our result might deny this possibility, while suggesting another difficulty in economic policy making. If corporate behaviors are heterogeneous than expected, policymakers must allow for such heterogeneity in evaluating the effectiveness of planned policies in advance. Firms should not be treated as an aggregation of representative agents.

For managers in firms, the above-mentioned findings may be instructive to improve their marketing practices. If TV advertising reaches customers later than their purchase, the timing of advertisement should be reconsidered, since consumers cannot choose the timing of exposure to it. On the other hand, if the ad campaign intends to reinforce customer retention, the marketing practice could be successful. Our method reveals which products could be real rivals, without any prejudice, by showing synchronization of marketing instruments between competitive products. This information can be used by marketers to investigate the dynamics of competition or substitution for their product.

Researches in marketing science have traditionally measured brand loyalty by incorporating own-product inertia [33] and the long-term effect of advertising by incorporating ad stock into the model [34]. Our method could implicitly capture the effect of brand loyalty that is embedded in the comovement of marketing instruments. How to explicitly quantify the magnitude of loyalty or inertia may be a possible challenge for us. We have already attempted incorporating ad stock with exponentially-distributed weights and alternative parameters. As the result was not sensitive to such modification, we tentatively conclude that accounting for the long-term effects of advertising in our method is not a priority.

In our opinion, it is desirable to develop further extension of our method, aiming to offer some numerical indications for policymakers and marketers, to enable better actions, by proceeding to sensitivity analysis/simulations, using the synchronization structure discussed in this paper. A fluctuation-dissipation approach [35] may be useful, assuming that the impact of external stimuli does not change the correlation structure, but simply excites some of the structure. In other words, this approach deals with small perturbations on the existing structure, which is, in general, true when promoting specific products or regulating specific firm's behavior. Research in this direction, therefore, would be fruitful.

Supporting information

S1 File. Appendix.
(PDF)

Acknowledgments

This study is conducted as a part of the Project "Macro-Economy under COVID-19 influence: Data-intensive analysis and the road to recovery" undertaken at the Research Institute of Economy, Trade and Industry (RIETI). The authors are grateful for helpful comments and suggestions by Discussion Paper seminar participants at RIETI. The authors appreciate the support by INTAGE Inc. to make the data available. The authors would also like to thank participants of the conferences of Computational Social Science Japan, Japan Institute of Marketing Science, and Japan Society for Evolutionary Economics for their helpful discussions and comments. We would like to thank Editage (www.editage.com) for English language editing.

Author Contributions

Conceptualization: Makoto Mizuno, Hideaki Aoyama, Yoshi Fujiwara.

Data curation: Makoto Mizuno.

Formal analysis: Hideaki Aoyama, Yoshi Fujiwara.

Funding acquisition: Hideaki Aoyama, Yoshi Fujiwara.

Investigation: Makoto Mizuno, Hideaki Aoyama, Yoshi Fujiwara.

Methodology: Makoto Mizuno, Hideaki Aoyama, Yoshi Fujiwara.

Project administration: Makoto Mizuno, Hideaki Aoyama, Yoshi Fujiwara.

Resources: Makoto Mizuno, Hideaki Aoyama, Yoshi Fujiwara.

Software: Hideaki Aoyama, Yoshi Fujiwara.

Supervision: Makoto Mizuno, Hideaki Aoyama, Yoshi Fujiwara.

Validation: Makoto Mizuno, Hideaki Aoyama, Yoshi Fujiwara.

Visualization: Makoto Mizuno, Hideaki Aoyama, Yoshi Fujiwara.

Writing – original draft: Makoto Mizuno, Hideaki Aoyama, Yoshi Fujiwara.

Writing – review & editing: Makoto Mizuno, Hideaki Aoyama, Yoshi Fujiwara.

References

1. Edelman DC, Singer M. Competing on customer journeys. Harvard Business Review. 2015; 93(11):88–100.
2. Lemon KN, Verhoef PC. Understanding Customer Experience Throughout the Customer Journey. Journal of Marketing. 2016; 80(11):69–96. https://doi.org/10.1509/jm.15.0420
3. Anderl E, Becker I, von Wangenheim F, Schumann JH. Mapping the Customer Journey: Lessons Learned from Graph-based Online Attribution Modeling. International Journal of Research in Marketing. 2016; 33(3):457–474. https://doi.org/10.1016/j.ijresmar.2016.03.001
4. De Haan E, Kannan PK, Verhoef PC, Wiesel T. The Role of Mobile Devices in the Online Customer Journey; 2015.
5. Pauwels K. How Retailer and Competitor Decisions Drive The Long-Term Effectiveness of Manufacturer Promotions for Fast Moving Consumer Goods. Journal of Retailing. 2007; 83(3):297–308. https://doi.org/10.1016/j.jretai.2006.03.001
6. Ataman MB, Van Heerde HJ, Mela CF. The long-term effect of marketing strategy on brand sales. Journal of Marketing Research. 2010; 47(5):866–882. https://doi.org/10.1509/jmkr.47.5.866
7. Sriram S, Balachander S, Kalwani MU. Monitoring the Dynamics of Brand Equity Using Store-Level Data. Journal of Marketing. 2007; 71(4):61–78. https://doi.org/10.1509/jmkg.71.2.061
8. Kolsarici C, Vakratsas D. Correcting for Misspecification in Parameter Dynamics to Improve Forecast Accuracy with Adaptively Estimated Models. Management Science. 2016; 61(10):2495–2513. https://doi.org/10.1287/mnsc.2014.2027
9. Steenkamp JBEM, Nijs VR, Hanssens DM, Dekimpe MG. Competitive Reactions to Advertising and Promotion Attacks. Marketing Science. 2005; 24(1):35–54. https://doi.org/10.1287/mksc.1040.0069
10. Li H. Accurate and Efficient Classification Based on Common Principal Components Analysis for Multivariate Time Series. Neurocomputing. 2016; 171:744–753. https://doi.org/10.1016/j.neucom.2015.07.010
11. Aoyama H, Fujiwara Y, Ikeda Y, Iyetomi H, Souma W. Econophysics and Companies: Statistical Life and Death in Complex Business Networks. Cambridge University Press; 2010.
12. Aoyama H, Fujiwara Y, Ikeda Y, Iyetomi H, Yoshikawa H. Macro-Econophysics, New Studies on Economic Networks and Synchronization. Cambridge University Press; 2017.
13. Kichikawa Y, Iyetomi H, Aoyama H, Fujiwara Y, Yoshikawa H. Interindustry Linkages of Prices: Analysis of Japan's Deflation. PLoS ONE. 2020; 15(2):e0228026. https://doi.org/10.1371/journal.pone.0228026 PMID: 32053604
14. Vodenska I, Aoyama H, Fujiwara Y, Iyetomi H, Arai Y. Interdependencies and causalities in coupled financial networks. PloS one. 2016; 11(3):e0150994. https://doi.org/10.1371/journal.pone.0150994 PMID: 26977806
15. Jiang X, Lim LH, Yao Y, Ye Y. Statistical Ranking and Combinatorial Hodge Theory. Mathematical Programming. 2011; 127(1):203–244. https://doi.org/10.1007/s10107-010-0419-x
16. Hartmann W, Nair HS, Narayanan S. Identifying Causal Marketing Mix Effects Using a Regression Discontinuity Design. Marketing Science. 2011; 30(6):1079–1097. https://doi.org/10.1287/mksc.1110.0670

17. Mizuno M, Hoshino T. Assessing the Short-term Causal Effect of TV Advertising via the Propensity Score Method; 2006.

18. Rasmusson EM, Arkin PA, Chen WY, Jalickee JB. Biennial variations in surface temperature over the United States as revealed by singular decomposition. Mon Wea Rev. 1981; 109:587–598. https://doi.org/10.1175/1520-0493(1981)109%3C0587:BVISTO%3E2.0.CO;2

19. Barnett TP. Interaction of the monsoon and Pacific trade wind system at interannual time scales part I: The equatorial zone. Mon Wea Rev. 1983; 111:756–773. https://doi.org/10.1175/1520-0493(1983)111%3C0756:IOTMAP%3E2.0.CO;2

20. Horel JD. Complex principal component analysis: Theory and examples. J Appl Meteor. 1984; 23:1660–1673. https://doi.org/10.1175/1520-0450(1984)023%3C1660:CPCATA%3E2.0.CO;2

21. Stein K, Timmermann A, Schneider N. Phase synchronization of the El Niño-Southern oscillation with the annual cycle. Phys Rev Lett. 2011; 107:128501. https://doi.org/10.1103/PhysRevLett.107.128501 PMID: 22026806

22. Hannachi A, Jolliffe IT, Stephenson DB. Empirical orthogonal functions and related techniques in atmospheric science: A review. International Journal of Climatology. 2007; 27(9):1119–1152. https://doi.org/10.1002/joc.1499

23. Hilbert D. Grundzge einer allgemeinen theorie der linearen integralgleichungen. Druck und Verlag con B. G. Teubner; 1912.

24. Gabor D. Theory of communication. J Inst Electr Eng–Part III, Radio Commun Eng. 1946; 93:429–457.

25. Granger CWJ, Hatanaka M. Spectral analysis of economic time series. Princeton Univ. Press.; 1964.

26. Bendat JS, Piersol AG. Random Data: Analysis and Measurement Procedures. Wiley Series in Probability and Statistics. Wiley; 2011. Available from: http://books.google.co.jp/books?id=iu7pq6_vo3QC.

27. Feldman M. Hilbert transform in vibration analysis. Mechanical systems and signal processing. 2011; 25(3):735–802. https://doi.org/10.1016/j.ymssp.2010.07.018

28. Ikeda Y, Aoyama H, Yoshikawa H. Synchronization and the coupled ocsillator model in international business cycles. RIETI discussion paper. 2013;13-E-086.

29. Ikeda Y, Aoyama H, Iyetomi H, Yoshikawa H. Direct evidence for synchronization in Japanese business cycles. Evolutionary and Institutional Economics Review. 2013; 10(2):315–327. https://doi.org/10.14441/eier.A2013016

30. Arai Y, Iyetomi H. Complex principal component analysis of dynamic correlations in financial markets. Intelligent Decision Technologies, Frontiers in Artificial Intelligence and Applications. 2013; 255:111–119.

31. Iyetomi H, Aoyama H, Fujiwara Y, Souma W, Vodenska I, Yoshikawa H. Relationship between Macroeconomic indicators and economic cycles in U.S. Scientific Reports. 2020; 10(1):1–12. https://doi.org/10.1038/s41598-020-70100-3

32. Kwon M, Erdem T, Ishihara M. Counter-Cyclical Price Promotion: Capturing Seasonal Category Expansion Under Endogenous Consumption; 2018.

33. Guadagni PM, Little JD. A Logit Model of Brand Choice Calibrated on Scanner Data. Marketing Science. 1983; 2(3):203–238. https://doi.org/10.1287/mksc.2.3.203

34. Nerlove M, Arrow KJ. Optimal Advertising Policy under Dynamic Conditions. Economica. 1962; 29(114):129–142. https://doi.org/10.2307/2551549

35. Iyetomi H, Nakayama Y, Aoyama H, Fujiwara Y, Ikeda Y, Souma W. Fluctuation-dissipation theory of input-output interindustrial relations. Physical Review E. 2011; 83(1):016103. https://doi.org/10.1103/PhysRevE.83.016103 PMID: 21405740