RIETI Discussion Paper Series 20-E-070

Constructing the Customer Journey Map of Competitive Brands: A Complex Time-series Analysis

MIZUNO, Makoto
Meiji University

AOYAMA, Hideaki
RIETI

FUJIWARA, Yoshi
University of Hyogo

RIETI
Research Institute of Economy, Trade & Industry, IAA
The Research Institute of Economy, Trade and Industry
https://www.rieti.go.jp/en/

RIETI Discussion Paper Series 20-E-070

August 2020

Constructing the Customer Journey Map of Competitive Brands: A Complex Time-series Analysis *

Makoto MIZUNO

Meiji University

Hideaki AOYAMA

Research Institute of Economy, Trade and Industry

RIKEN iTHEMS

Yoshi FUJIWARA

University of Hyogo

Abstract

In today's rapidly evolving consumer markets, obtaining a quantitative grasp of the customer journey (the sequence of touch points where customers and brands meet, which is important for marketing strategy) requires analysis of extremely high-dimensional data. Existing studies ignore the effects of touch points of multiple brands that are mutually competitive. We propose to apply a novel method called complex Hilbert principal component analysis (CHPCA) to allow unbiased, model-free analysis, and construct a synchronization network using Hodge decomposition. We apply this method to Japanese beer market data and show that it is suitable for the construction of the customer journey map both within-brand and across brands, the latter reflecting competition among firms. Furthermore, we capture 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, the limitations, and further development of the proposed method.

Keywords: Hilbert transformation, Hodge decomposition, co-movement, beer market, single-source data, customer heterogeneity, marketing strategy

JEL classification: M31, M37, C32

*This study is conducted as a part of the Project "Dynamics of Economy and Finance from the Economic Network Point of View" undertaken at the Research Institute of Economy, Trade and Industry (RIETI). The author is grateful for helpful comments and suggestions by Discussion Paper seminar participants at RIETI.

1. Introduction

In rapidly evolving, highly competitive consumer market environments, marketers are faced with a countless number of factors affecting business. These factors often veer away from conventional rules of thumb or theory and can be mutually interacting. Marketers and researchers face the challenge of detecting the meaningful relationships between such factors and predicting the possible performance of brands. This challenge represents a significant marketing mix problem caused by the proliferation of viable media and competing brands. Traditionally, consumers' consideration of a consumer goods purchase began with exposure to TV advertising. However, penetration via a variety of media and channels has increased the number of "touch points" between customers and firms (Winer 2001). As a result, the proliferation of touch points has intensified competition across brands, creating an increasingly complex marketing mix.

Recently, to untangle the growing complexity of the customer decision process, practitioners have increasingly begun to refer to the customer journey framework (Edelman and Singer 2015, Lemon and Verhoef 2016). The customer journey represents a transition of the customer state or a sequence of touch points that finally converges into conversion (in many cases, a purchase or sale). Several quantitative models have been proposed to untangle this complexity: Anderl et al. (2016) applied a Markov graphical model for individual customers' history of contacts with touch points, and De Haan et al. (2015) applied an advanced regression to the aggregated data. Nevertheless, these methods fail to cover all relevant information because they focus on the customer journey within a set of touch points of the focal brand. Consequently, they ignore the effects of the touch points of multiple brands that are mutually competitive. To investigate how customers choose a certain brand, a customer journey map should be extended to include active competitors' touch points.

The first difficulty in extending the customer journey map lies in collecting the data on competitors' touch points. Fortunately, particularly for consumer packaged goods, syndicated data are often available that include information on the touch points of multiple competitive brands. In this study, we use the data offered by a leading marketing research company in Japan, INTAGE Inc. These data are a compilation of touch point information or marketing variables based on consumer exposure to TV advertising, visits to brand websites through search engine queries, and purchase actions with retailers.

The second difficulty in the mapping process lies in identifying the effects of touch points of multiple brands on their overall performance and the mutual relationships between those brands. For this purpose, the most suitable method used in marketing science may be multivariate time-series analyses, which have been intensively applied to assess the long-term effectiveness of the marketing mix. This method includes a variety of models such as the vector auto-regressive model (Pauwels 2007), dynamic linear model (Ataman and Mela 2010), varying parameter model (Sriram et al. 2007), and the Kalman filter (Kolsarici and Vakratsas 2016).

However, even these methods face challenges in handling competitive customer journeys because of the limited number of parameters tractable for the model. Extant multivariate time-series models are applied to cases with only three to five competitive brands and a few marketing variables, although most consumer goods markets are composed of many more brands and touch points. For simplicity, we first consider the within-brand effects of each brand. Even in this case, a number of possible effects need to be measured, which are classified as follows:

(1) Own-brand effect: A brand's activity at each touch point could affect its own performance, which is of primary importance to marketers.

(2) Own-brand feedback: A brand's past performance could affect its activity at each touch point, for instance, through budget constraints based on past performance,

(3) Own-brand synergy: A brand's activity could affect its another activity, which is likely to occur if positive (or negative) synergy leads to joint (or disjointed) implementation of these activities.

(4) Own-brand inertia: A brand's activity or performance might depend on its past state; for instance, brand loyalty (positive inertia) or variety-seeking (negative inertia). (Guadagni and Little 1983).

For a market with m brands, each with n touch points and one performance measure, a simple linear model would require the estimation of at least $m \times n$ parameters for (1) and (2) each, $m \times (m-1) \times n$ parameters for (3) and $(m+1) \times n$ parameters for (4) (in total, $(m+1)^2 \times n$ parameters). If $m=10$ and $n=5$, which seems to be a realistic setting, at least 360 parameters should be estimated. To measure the synergistic effects between activities more directly, some interaction terms would be required(Naik and Ramon 2003, Naik and Peters 2009), further increasing the number of parameters.

Additionally, marketers are interested in how competitors interrupt their own customers' journeys and the effect of this on a brand's performance. Thus, the following effects should be considered:

(1) Cross-brand effect: A brand's activity at each touch point could affect the performance of its competitors. The most popular notion regarding this effect is the influence of cross-elasticity of prices on other brands' sales.

(2) Cross-brand feedback: A brand's past performance could affect its competitors' activities at touch points since competitors monitor each other's performance to plan strategic actions.

(3) Cross-brand reaction: A brand's activity at a touch point could increase (or decrease) a competitor's' retaliatory activities (or accommodation),

(4) Cross-brand dependence: A brand's past performance could directly affect the performance of its competitors, whose effect could be negative (or positive) if the brands are substitutional (or complementary),

Incorporating cross-brand relationships into the model would increase the number of parameters disproportionately. To measure both own- and cross-brand effects via a linear model, we estimate $((m+1)\times n)^2$ parameters (for the above numerical example, $(6\times 10)^2 = 3,600$ parameters in total). Adding time-lag or nonlinear effects increases the number further, exceeding the limitation of parameters tractable for ordinary multivariate time-series analyses. To avoid such an explosive increase in parameters, researchers have often imposed strong restrictions on variables and their relationships based on their experience or research tradition. Otherwise, researchers employ exploratory analyses to reduce the dimensionality of given multivariate data, such as principal component analysis (PCA). However, as discussed later, ordinary PCA often fails to treat time-series data with time-lag effects in a proper way.

Alternatively, we propose a novel method called complex Hilbert principal component analysis (CHPCA) (Aoyama et al. 2010, 2017) to depict a customer journey map with own and competitors' touch points. CHPCA was developed originally in econophysics as an extension of PCA to uncover temporal comovements among variables observed in the macro economy (Kichikawa et al. 2020) or foreign exchange markets (Vodenska et al. 2016). CHPCA can handle enormously high-dimensional time-series data without any strong assumptions concerning the phenomenon of interest. This method provides information on almost all own- and cross-brand effects, except own-brand inertia and cross-brand effects. Although this method is advantageous in capturing a number of effects among multiple brands and touch points, can it capture other aspects that the existing models satisfy? Ataman and Mela (2010) evaluate alternative multivariate time-series models to determine whether they satisfy the following conditions:

- Endogeneity: Recently, this issue has been increasingly emphasized because it is widely recognized that many marketing actions are not independent of future performance. Suppose customers who are more likely to purchase are targeted in advertising campaigns. With no consideration for endogeneity in advertising, the effects of endogeneity on a purchase would be overestimated.
- Performance feedback: As already discussed in relation to own-brand feedback, consideration for performance feedback effects may lead to a more accurate assessment of the customer journey.
- Competitive reactions: This concept is discussed above as cross-brand reaction and has been studied by a variety of data and analytical methods (Steenkamp et al. 2005).

CHPCA satisfies the first condition (endogeneity) since the model treats all variables as comoving variables. Additionally, the model satisfies the second condition (feedback from past outcomes), except for the cross-brand condition and the third condition (competitive reactions) 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. If using random rotation simulation (RRS), significant eigenmodes (principal components in PCA) can be selected automatically in a theoretically justifiable manner. 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. Finally, this method provides a skeleton of a customer journey map using aggregate marketing data, which is available for consumer packaged goods markets. The model also provides information on customer profile heterogeneity.

In addition, 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 causality 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 (Hartmann et al. 2011) or a propensity score method (Mizuno and Hoshino 2006), 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 detect the customer journey in section 3 and the procedure for depicting the map 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.

2. Data

2.1. Data Collection

In this paper, 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 brands 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.

https://www.overleaf.com/project/5ecc5fdae0d84e0001f5bc3f

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.

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

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

The potential heterogeneity among customers is represented as the location of a few-dimensional space (see section 5). The purchase data are documented at the store-keeping unit (SKU) level while the advertising and other communication variables are documented at the brand level. In general, one brand is composed of multiple SKUs. When merging the two types of data, the purchase data are summed across SKUs by their corresponding brand.

There are 163 beer brands from 14 firms in the original data. We selected the top 18 brands according to the total quantity in the data during the stated period: Fig. 1 is the rank-size plot of some of the top brands. The top 18 brands selected for the current analysis are those beyond the thin vertical line, Total Quantity $> 1 \times 10^4$. As is apparent in this plot, these brands form a distinctive top group with a large gap between this group and the followers. 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 brands cover 64.9% of all the sales. Fig. 2 shows the daily total quantities for both all brands (in light gray) and the selected 18 brands (in dark gray). 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.[1]

The brands are shown with codes, the first letter of which corresponds to the firm, and the second letter (digit) corresponds to the brand: For example, the code “A1” means that the product is from firm ‘A’ and the brand ‘1’. With five types of data for each brand as listed in Table 1, we have $18 \times 5 = 90$ time series altogether. However as no communication

[1] We have data beyond this day for another several months. However, 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).

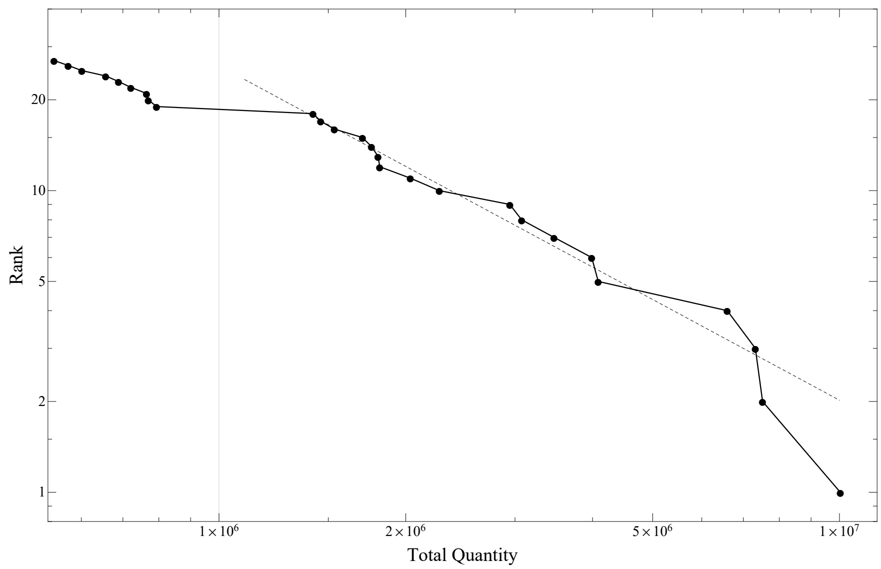

Figure 1 **Rank-size plot of the top selling brands.**

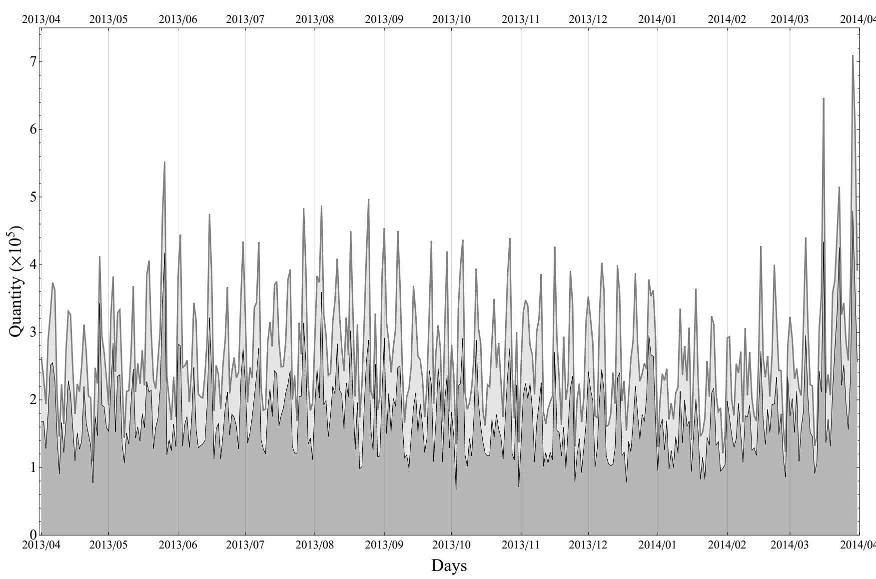

Figure 2 **Daily total quantities. Light gray: all brands, Dark gray: Selected 18 brands. Apparent periodic peaks correspond to Sundays.**

activity was observed for some brands, 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.

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

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

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)	—	—	—

Table 3 Descriptive statistics for the top-selling 18 brands: means and standard deviations (in parentheses). The symbol "—" corresponds to discarded data because of sparsity (see text).

2.2. 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. We then use the standard method of subtracting the mean and dividing it by the standard deviation. Let us denote the resulting time series by $x_\alpha(t_i)$ where $\alpha = 1, \cdots, N(=65)$ is the label for the time series, and $t_i = 1, \cdots, 365$ denotes the number of days. 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.

We depict, as a sample of $x_\alpha(t_i)$, the five types of time series for the top brand "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 brand visits via web site or mobile device (Visit) have tranquility with sudden and short activities. The frequency of TV advertising (TVAd) exposure has weak periodic behavior presumably due to the TV advertising activities of the brand's firm and

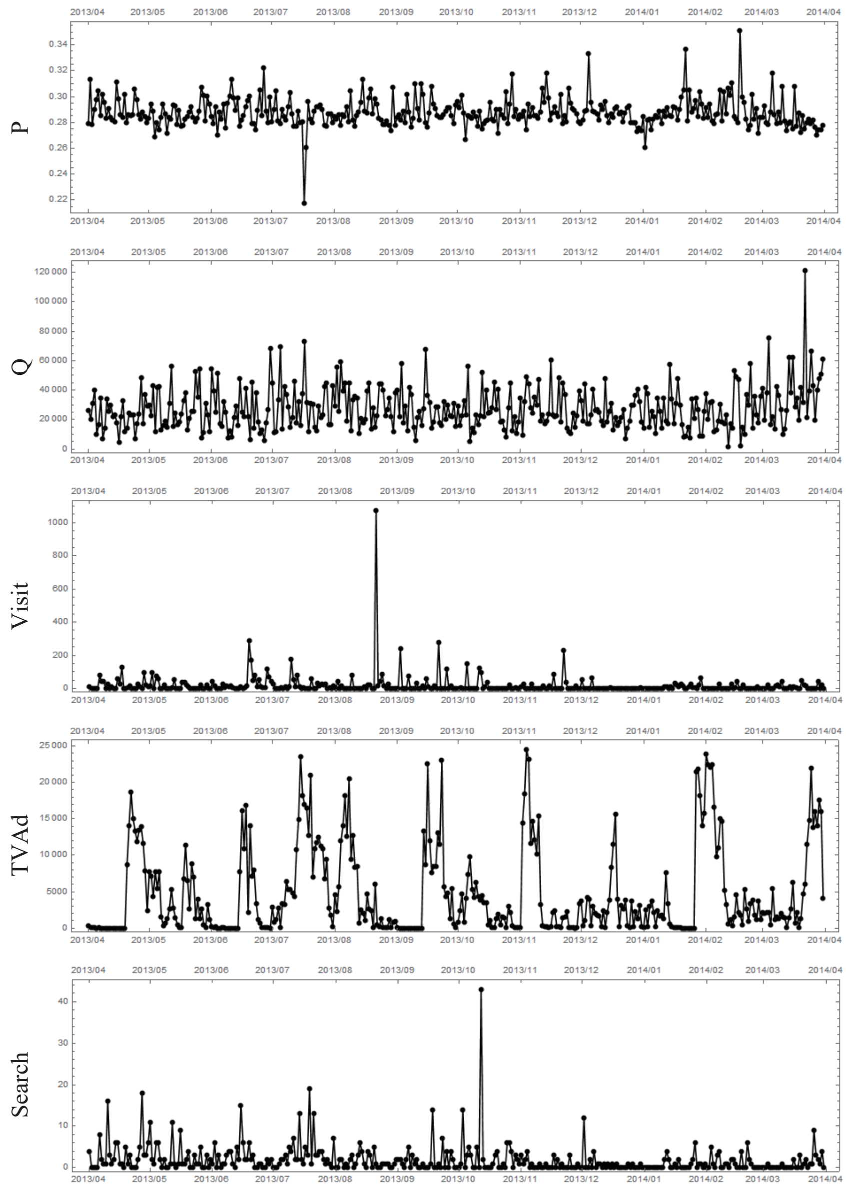

Figure 3 Sample of time-series $x_i(t)$ for the top brand "D2" for the five kinds of time-series, namely, P, Q, Visit, TVAd, and search from top to bottom.

corresponding exposure to customers. The frequency of brand searches via PC or mobile device (Search) have continuous activities with bursts.

3. Complex Hilbert Principal Component Analysis (CHPCA)

3.1. 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 paper, we are interested in inter correlations in the time series. To identify the structure and dynamics of the customer journey, 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 Rasmusson et al. (1981), Barnett (1983), Horel (1984), Stein et al. (2011), and Hannachi et al. (2007) using the Hilbert transformation developed in Hilbert (1912), Gabor (1946), Granger and Hatanaka (1964), Bendat and Piersol (2011), Feldman (2011), among others. The approach has been successfully applied in several areas of natural science and economics (Ikeda et al. (2013b), Ikeda et al. (2013a), Kichikawa et al. (2020), Vodenska et al. (2016)). We further introduced improvements on CHPCA by Aoyama et al. (2017).

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 $\cos\omega t \to -\sin\omega t$ and $\sin\omega t \to \cos\omega t$. Therefore, the complexification converts $\cos\omega t$ to $e^{-i\omega t}$ and $\sin\omega t$ to $i\,e^{-i\omega t}$; clockwise rotation on its complex plane. Furthermore, the Hilbert transformation converts

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

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;[2]

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

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:

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

[2] Hereafter, $\cdot^*$ denotes the complex conjugate of $\cdot$.

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 Arai and Iyetomi (2013). 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) comparing each distribution of the eigenvalues and the actual eigenvalue from the largest in descending order: identifying the eigenmodes whose eigenvalue is larger than that of the one obtained in the step (3) as significant modes. This methodology overcomes the shortcoming of the RRS 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 journey in our data.

3.2. Results

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

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

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 CHPCA identifies the eigenmode more easily than PCA. This is natural since PCA misses movements with lead/lag.

The result of the RRS analysis of 10^4 times the RRS 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.

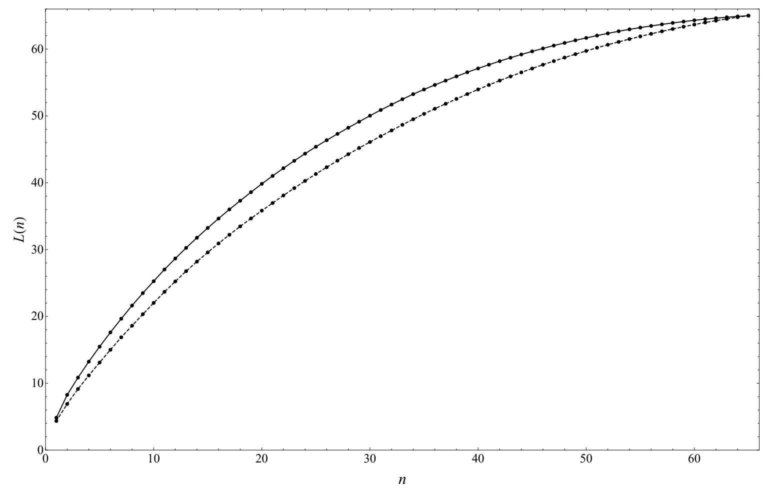

Figure 4 Cumulative eigenvalue $L(n)$ defined in Eq.(7).

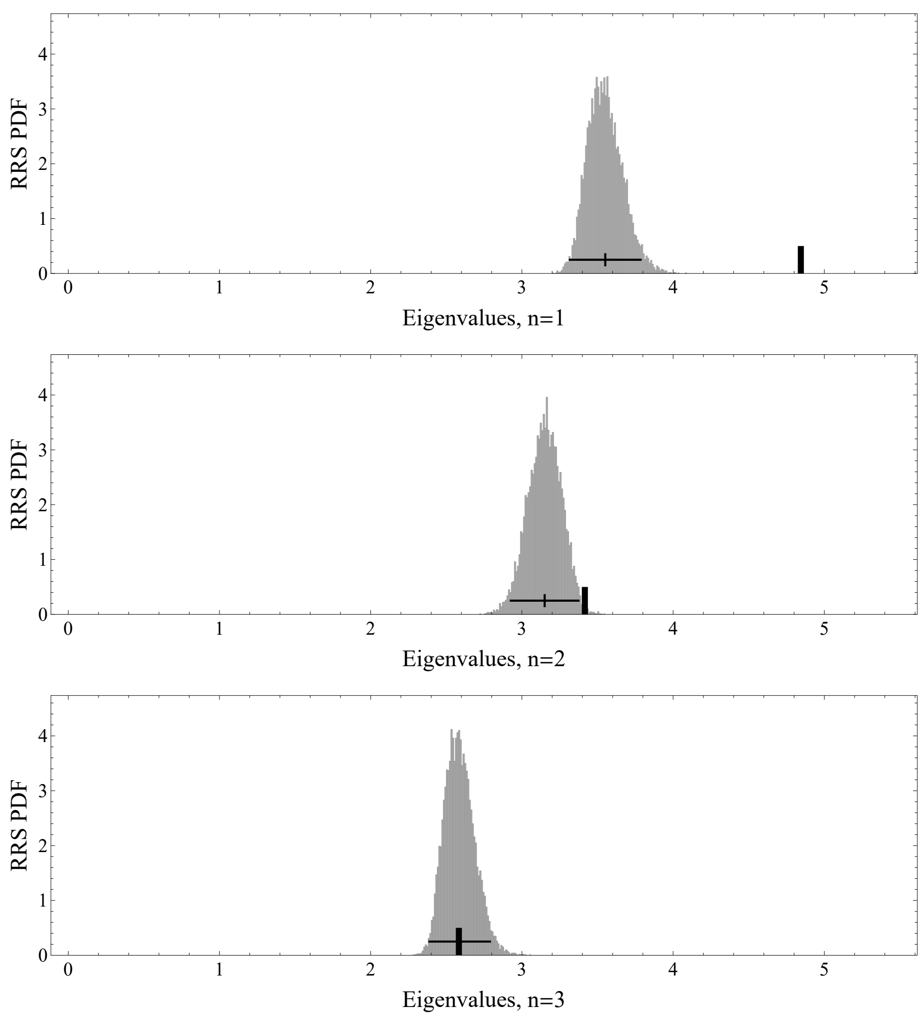

Figure 5 The eigenvalues (thick ticks) with the corresponding RRS eigenvalue distributions (shaded bell-shapes) and their 2σ ranges (horizontal bars) for eigenvalues $n = 1, 2, 3$. The eigenvalues #1 and #2 are above the RRS 2σ range and, therefore, are significant. The eigenvalues #3 and below are not.

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.

The top and the second eigenvector components are shown in Fig. 6, where each component is shown by a marker specified by the brand code at its top and the style shown in the legend. 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 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 touch points across brands 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.

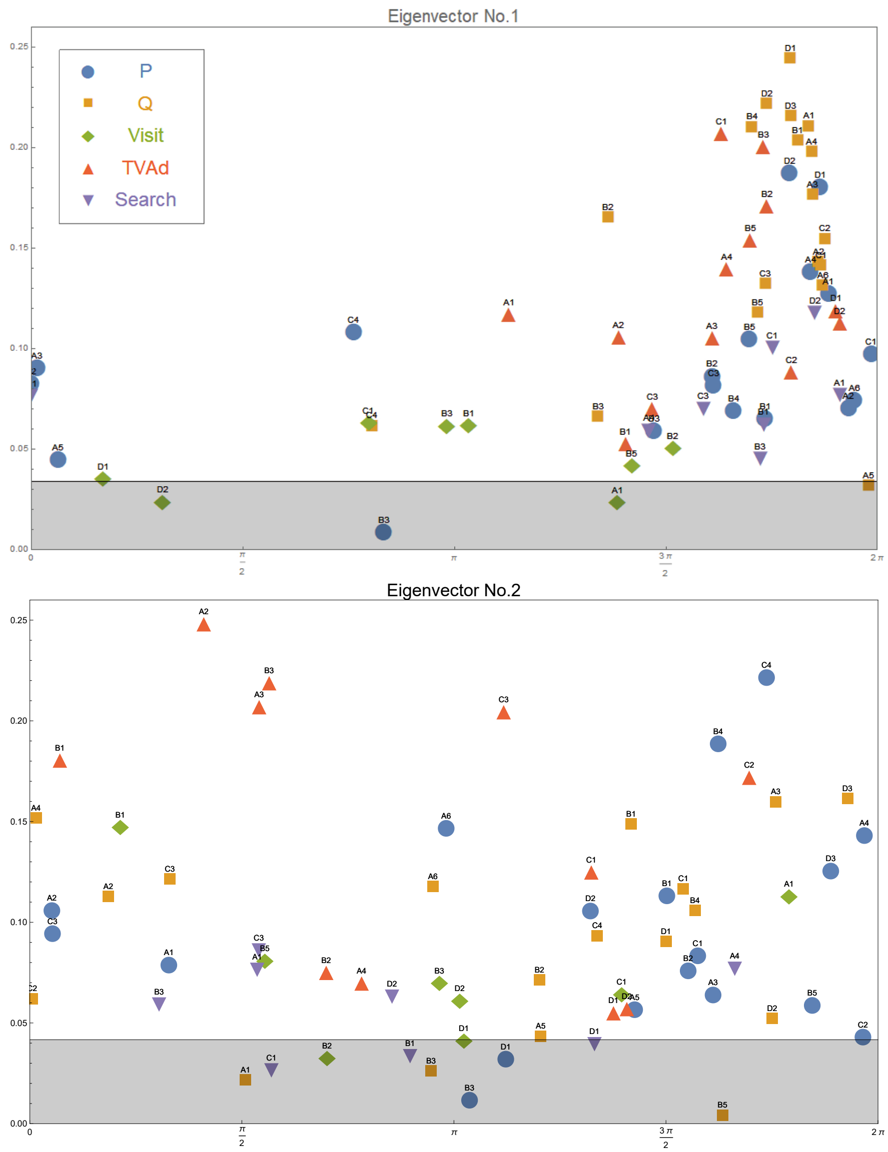

Figure 6 The components of the first (upper) and the second (lower) eigenvectors.

4. Hodge Decomposition and Synchronization Network

4.1. Method

The complex correlation coefficient, $C_{\alpha\beta}$, represents how strongly a pair of α and β are correlated possibly with lead and lag. The strength of the correlation is given by the magnitude

$$\rho_{\alpha\beta} := |C_{\alpha\beta}| \ , \tag{9}$$

and the lead and lag can be measured by the phase

$$\theta_{\alpha\beta} := \arg C_{\alpha\beta} \ . \tag{10}$$

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 $C_{\alpha\beta}$, 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, $C_{\beta\alpha} = C^*_{\alpha\beta}$, 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 Jiang et al. (2011) for example), which is called *Hodge decomposition*. Denote the adjacency matrix of the binary and weighted network by

$$A_{\alpha\beta} = \begin{cases} 1 & \text{if there is a directed edge from } \alpha \text{ to } \beta, \\ 0 & \text{otherwise,} \end{cases} \tag{11}$$

and

$$B_{\alpha\beta} = \begin{cases} f_{\alpha\beta} & \text{if there is a directed edge with a flow,} \\ 0 & \text{otherwise,} \end{cases} \tag{12}$$

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

$$F_{\alpha\beta} = B_{\alpha\beta} - B_{\beta\alpha} \ . \tag{13}$$

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

$$W_{\alpha\beta} = A_{\alpha\beta} + A_{\beta\alpha} \ . \tag{14}$$

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

Hodge decomposition is given by

$$F_{\alpha\beta} = W_{\alpha\beta} \left(\phi_\alpha - \phi_\beta \right) + F_{\alpha\beta}^{(\mathrm{loop})} \ , \tag{15}$$

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

$$\sum_\beta F_{\alpha\beta}^{(\mathrm{loop})} = 0 \tag{16}$$

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

Rewriting Eq. (16), we have for each $\alpha = 1, \cdots, N$,

$$\sum_\beta L_{\alpha\beta}\, \phi_\beta = \sum_\beta F_{\alpha\beta} \ , \tag{17}$$

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

$$L_{\alpha\beta} = \delta_{\alpha\beta} \sum_\gamma W_{\alpha\gamma} - W_{\alpha\beta} \ , \tag{18}$$

where $\delta_{\alpha\beta} = 1$ if $\alpha = \beta$ and $\delta_{\alpha\beta} = 0$ otherwise. Given $F_{\alpha\beta}$ and $W_{\alpha\beta}$, Eqs. (17) are simultaneous linear equations to determine the Hodge potential ϕ_α of all the nodes α.

Note that simultaneous linear equations (17) are not independent of each other. In fact, the summation over α of (17) 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 (Iyetomi et al. 2020). In the following, we use the convention that the mean is zero:

$$\sum_{\alpha} \phi_\alpha = 0 \ . \tag{19}$$

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.

4.2. 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 journey map covering all competitive brands, Hodge potentials are used to constitute the fundamental time sequence of the exposures to touch points (TV Ad, web/mobile site visit, search, price, and quantity) for each brand. 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 brands are arrayed in an arbitrary order. For instance, for Brand 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 touch points within each brand. For more than half of the brands (11 out of 18), a decrease (increase) in price leads to an increase (decrease)

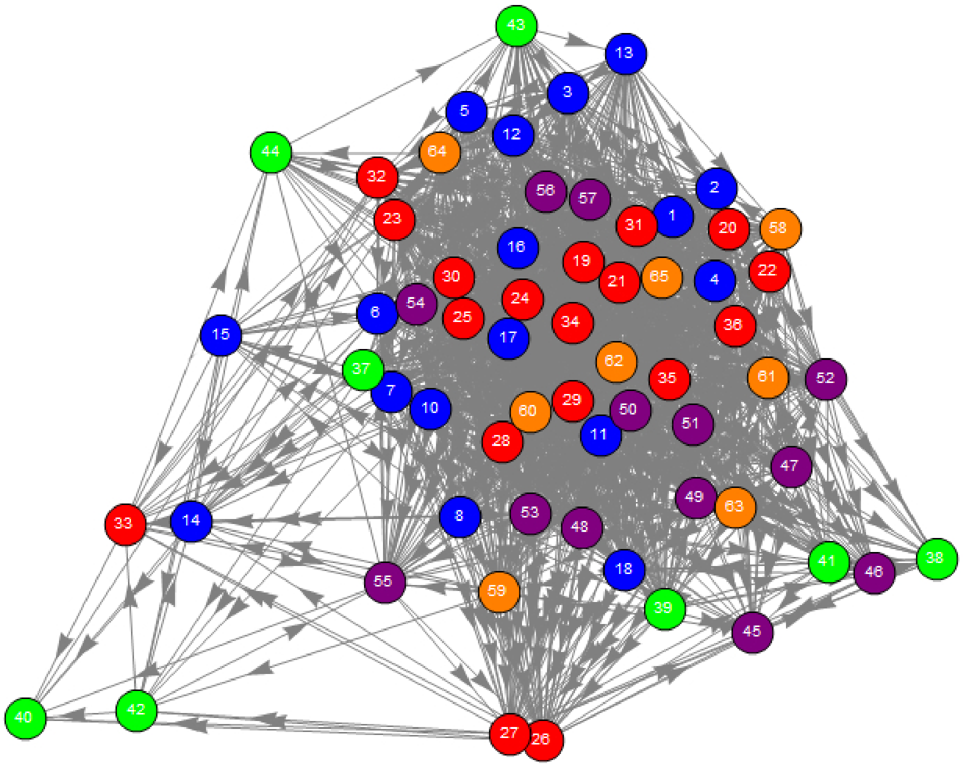

Figure 7 **The synchronization network's graph layout. The vertical position of each node corresponds to its Hodge potential as a constraint in the force-directed algorithm of the graph layout. Upstream (leading) nodes are located toward the top while downstream (lagging) nodes are toward the bottom.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

in quantity to a certain extent. For some brands, a change in price occurs almost simultaneously with a change in quantity. These two cases are consistent with standard economic theory. On the other hand, for Brands 15 and 18, an increase (decrease) in quantity leads to a decrease (increase) in price. This phenomenon seems to be an anomaly as a pricing effect 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 (Kwon et al. 2018).

The increased exposures to TV advertising lag behind the increased purchases for seven of the 13 brands 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 the latter case, after a significant time lag. The time sequence of web/mobile site visits or searches is less consistent across brands than price or TV advertising. A possible reason is that visiting these touch points is less controllable for firms, reflecting the idiosyncratic nature of individual brands. In that sense, this inconsistency

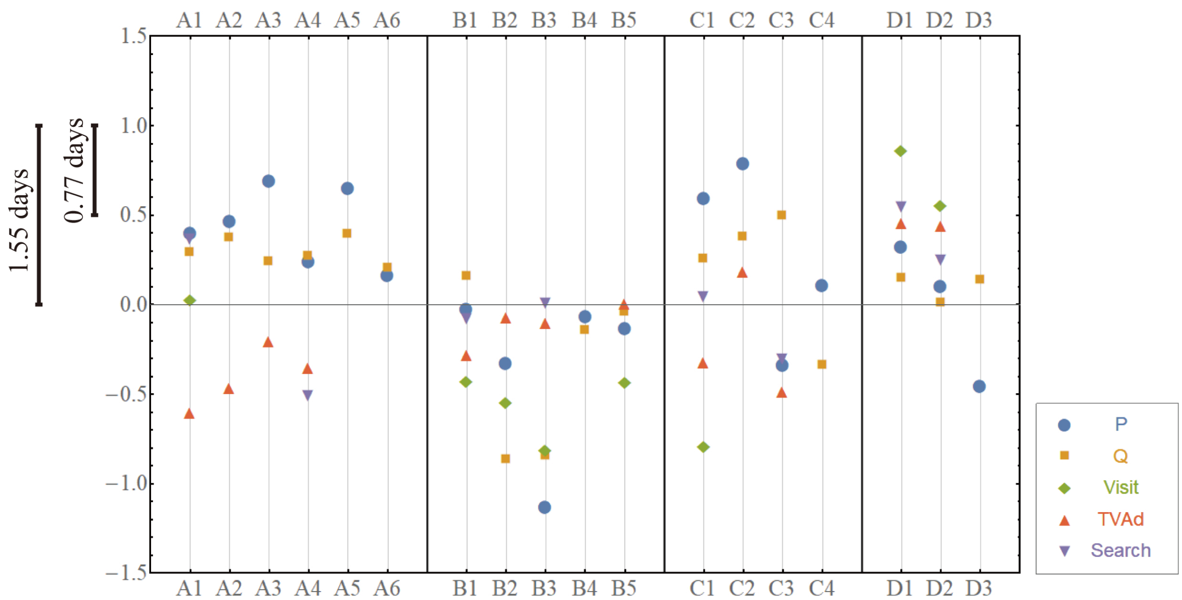

Figure 8 The fundamental sequence of touch points, price information, and purchases for each brand. The vertical axis represents the Hodge potential, whose value is larger as it leads than others. Horizontally, brands are arrayed in an arbitrary order. If the positions of two time series are closer, they are comoving almost simultaneously with each other.

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 brands, which indicates synchronization (comoving almost simultaneously) of touch points between different brands of a firm or even between firms. For instance, as Fig.8 shows, a price cut and sales increase for Brands 1, 2 and 3, which are the main brands of Firm 1, tend to be synchronized. That is, Firm 1 might coordinate price promotion consistently among their own brands 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 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 brands, 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 brands (Brands 8 and 9 of Firm 2) outside the band. These brands 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.

5. Customer Profile

We found in Sec. 3 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 brands.

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 Sec. 3 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 Sec. 3. Thus, we have

$$\hat{z}_{p,\alpha}(t) = \frac{z_{p,\alpha}(t) - \langle z_{p,\alpha}\rangle}{\sigma_{p,\alpha}}. \tag{20}$$

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,

$$a_{n,p}(t) = \sum_{\alpha=1}^{N} (\mathbf{e}_n)^*_\alpha \, \hat{z}_{p,\alpha}(t), \tag{21}$$

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,

$$X_{n,p} = \frac{1}{T}\sum_t |a_{n,p}(t)|^2, \tag{22}$$

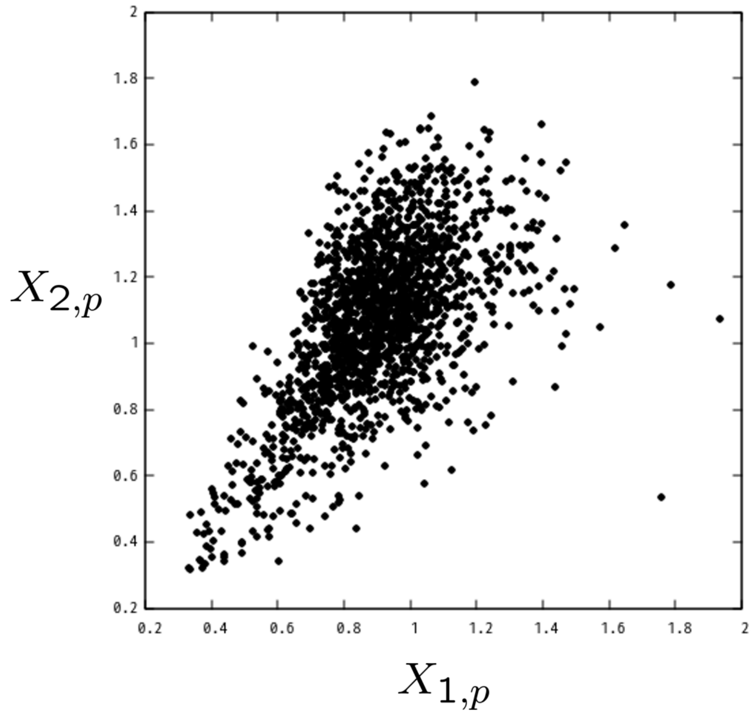

Figure 9 **Individual customer's projected representation for the two significant modes. See Eq.** (22) **for the representation.**

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

Fig. 9 shows the resulting spatial representation of Eq. (22) for all the P customers. Recalling (6) in Sec. 3, 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, 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 brands are

	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	—	—	
Brand 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		

Table 4 Regression of coordinates of customer space to customer profiles. (aa: $p < .001$; a: $p < .01$; b: $p < .05$; c: $p < .10$)

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 brand's purchase quantities to capture individual differences in brand-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 brands may marginally improve the model fit. The coefficients of a few brands 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 brands. 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 journey.

6. Conclusion

The expansion of digital media and distribution channels that serve customers is rendering the so-called customer journey across possible touch points increasingly complex to define. Most methods used for this purpose are problematic and ignore the existence of competing brands, even in cases where most customers consider multiple alternative brands by searching or shopping. One of the main reasons for the weak methods is the difficulty in collecting data on rivals. Another difficulty is applying the existing methods, such as multivariate time series analysis, as the number of brands and touch points increases. The number of parameters is evident for a case with 18 brands and 5 touch points. However, without strong assumptions to reduce parameters drastically, applying such methods is difficult. 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 depict customer journey maps embedded in enormously high-dimensional time-series data. From the data with 18 brands and 5 variables, we detect principal sequences of exposures to touch points by individual brands and their interactions without any assumptions or prior knowledge. The application of our method to the beer market in Japan derives some interesting findings. First, for most brands on the market, a change in price leads to a change in purchase quantity followed by a change in exposure to TV advertising. Simply put, customers notice a price change in a store, buy a product, and are then exposed to TV ads later. The timing of visiting other touch points (brand web/mobile sites, search engines) is not consistent between brands. Second, we find synchronization across brands, in particular within each firm, rather than across firms. It suggests that individual firms are heterogeneous, each adopting a distinctive coherent marketing strategy.

The second point has an important implication for economic policy. 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 it suggests 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.

For marketers, the above-mentioned information is instructive to improve their marketing strategies. If TV advertising reaches customers later than their purchase, the timing of ad insertion should be reconsidered. On the other hand, if the ad campaign intends to reinforce customer loyalty, the marketing strategy could be successful. Our method reveals which brands could be real rivals without any prejudice. With this information, marketers can investigate the dynamics of competition or substitution for their brand.

Some marketing researchers might be concerned with the way our method deals with brand loyalty or own-brand inertia (Guadagni and Little 1983) and the long-term effect of advertising or the ad stock (Nerlove and Arrow 1962). Both are popular topics in marketing modeling. Regarding brand loyalty, the effect of brand loyalty is implicitly embedded in the comovement of touch points. How to explicitly quantify the effect may be a possible challenge for us. Regarding ad stock, we have already incorporated the ad stock with exponentially-distributed weights with some alternative parameters. As the result was not sensitive to such modification, we tentatively conclude that accounting for advertising long-term effects in our method is not prioritized.

From our viewpoint, a more serious limitation of our method is that it offers some critical information for policymakers or marketers but not exact numerical indications to enable better marketing actions. As an extension of our method, therefore, it is important to proceed to sensitivity analysis/simulations using the synchronization structure discussed in this paper. A fluctuation-dissipation approach (Iyetomi et al. (2011)) 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 brands. Research in this direction, therefore, would be fruitful.

Finally, we confirm that our method would not necessarily outperform existing methods such as multivariate time-series analysis in every aspect. If the number of brands and variables is limited to a certain range, and if sufficiently reliable knowledge is available for selecting parameters, such existing methods may be highly productive. Additionally, other methods based on statistics and machine learning already prepare a number of powerful tools for a wider range of problems. In rapidly evolving markets, however, marketers must grasp the critical relationships of a number of time series without enough prior knowledge, and from scratch. In this situation, we believe our method is useful and should be added to the business analytics toolbox.

Acknowledgments

This study is an outcome of the Project "Dynamics of Economy and Finance from the Economic Network Point of View" at the Research Institute of Economy, Trade and Industry (RIETI). This work is partially supported by Grant-in-Aid for Scientific Research (KAKENHI) of JSPS Grant Numbers 17H02041, 20H02391 and by MEXT as Exploratory Challenges on Post-K Computer (Studies of Multilevel Spatiotemporal Simulation of Socioeconomic Phenomena: Macroeconomic Simulations). 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.

References

Anderl E, Becker I, von Wangenheim F, Schumann JH (2016) Mapping the customer journey: Lessons learned from graph-based online attribution modeling. *International Journal of Research in Marketing* 33(3):457–474.

Aoyama H, Fujiwara Y, Ikeda Y, Iyetomi H, Souma W (2010) *Econophysics and Companies: Statistical Life and Death in Complex Business Networks* (Cambridge University Press).

Aoyama H, Fujiwara Y, Ikeda Y, Iyetomi H, Yoshikawa H (2017) *Macro-Econophysics, New Studies on Economic Networks and Synchronization* (Cambridge University Press).

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

Ataman VHHJ M B, Mela CF (2010) The long-term effect of marketing strategy on brand sales. *Journal of Marketing Research* 47(5):866–882.

Barnett TP (1983) Interaction of the monsoon and pacific trade wind system at interannual time scales part i: The equatorial zone. *Mon. Wea. Rev.* 111:756–773.

Bendat J, Piersol A (2011) *Random Data: Analysis and Measurement Procedures*. Wiley Series in Probability and Statistics (Wiley), ISBN 9781118030684, URL `http://books.google.co.jp/books?id=iu7pq6\_vo3QC`.

De Haan E, Kannan PK, Verhoef PC, Wiesel T (2015) The role of mobile devices in the online customer journey. MSI Working Paper Series No. 15-124, Cambridge MA, Marketing Science Institute.

Edelman DC, Singer M (2015) Competing on customer journeys. *Harvard Business Review* 93(11):88–100.

Feldman M (2011) Hilbert transform in vibration analysis. *Mechanical systems and signal processing* 25(3):735–802.

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

Granger CWJ, Hatanaka M (1964) *Spectral analysis of economic time series.* (Princeton Univ. Press.).

Guadagni PM, Little JD (1983) A logit model of brand choice calibrated on scanner data. *Marketing Science* 2(3):203–238.

Hannachi A, Jolliffe IT, Stephenson DB (2007) Empirical orthogonal functions and related techniques in atmospheric science: A review. *International Journal of Climatology* 27(9):1119–1152, ISSN 1097-0088, URL `http://dx.doi.org/10.1002/joc.1499`.

Hartmann W, Nair HS, Narayanan S (2011) Identifying causal marketing mix effects using a regression discontinuity design. *Marketing Science* 30(6):1079–1097.

Hilbert D (1912) *Grundzğe einer allgemeinen theorie der linearen integralgleichungen* (Druck und Verlag con B. G. Teubner).

Horel JD (1984) Complex principal component analysis: Theory and examples. *J. Appl. Meteor.* 23:1660–1673.

Ikeda Y, Aoyama H, Iyetomi H, Yoshikawa H (2013a) Direct evidence for synchronization in japanese business cycles. *Evolutionary and Institutional Economics Review* 10(2):315–327.

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

Iyetomi H, Aoyama H, Fujiwara Y, Souma W, Vodenska I, Yoshikawa H (2020) Relationship between macroeconomic indicators and economic cycles in U.S. *to appear* .

Iyetomi H, Nakayama Y, Aoyama H, Fujiwara Y, Ikeda Y, Souma W (2011) Fluctuation-dissipation theory of input-output interindustrial relations. *Physical Review E* 83(1):016103.

Jiang X, Lim LH, Yao Y, Ye Y (2011) Statistical ranking and combinatorial hodge theory. *Mathematical Programming* 127(1):203–244.

Kichikawa Y, Iyetomi H, Aoyama H, Fujiwara Y, Yoshikawa H (2020) Interindustry linkages of prices: Analysis of japan's deflation. *PLoS ONE* 15(2):e0228026.

Kolsarici C, Vakratsas D (2016) Correcting for misspecification in parameter dynamics to improve forecast accuracy with adaptively estimated models. *Management Science* 61(10):2495–2513.

Kwon M, Erdem T, Ishihara M (2018) Counter-cyclical price promotion: Capturing seasonal category expansion under endogenous consumption. SSRN 3138863.

Lemon KN, Verhoef PC (2016) Understanding customer experience throughout the customer journey. *Journal of Marketing* 80(11):69–96.

Mizuno M, Hoshino T (2006) Assessing the short-term causal effect of tv advertising via the propensity score method. University of Tsukuba, Department of Social Systems and Management Discussion Paper Series No. 1162.

Naik PA, Peters K (2009) A hierarchical marketing communications model of online and offline media synergies. *Journal of Interactive Marketing* 23(4):288–299.

Naik PA, Ramon K (2003) Understanding the impact of synergy in multimedia communications. *Journal of Marketing Research* 40(11):375—-388.

Nerlove M, Arrow KJ (1962) Optimal advertising policy under dynamic conditions. *Economica* 29(114):129–142.

Pauwels K (2007) How retailer and competitor decisions drive the long-term effectiveness of manufacturer promotions for fast moving consumer goods. *Journal of Retailing* 83(3):297–308.

Rasmusson EM, Arkin PA, Chen WY, Jalickee JB (1981) Biennial variations in surface temperature over the united states as revealed by singular decomposition. *Mon. Wea. Rev.* 109:587–598.

Sriram S, Balachander S, Kalwani MU (2007) Monitoring the dynamics of brand equity using store-level data. *Journal of Marketing* 71(4):61–78.

Steenkamp JBEM, Nijs VR, Hanssens DM, Dekimpe MG (2005) Competitive reactions to advertising and promotion attacks. *Marketing Science* 24(1):35–54.

Stein K, Timmermann A, Schneider N (2011) Phase synchronization of the el niño-southern oscillation with the annual cycle. *Phys. Rev. Lett.* 107:128501, URL `http://dx.doi.org/10.1103/PhysRevLett.107.128501`.

Vodenska I, Aoyama H, Fujiwara Y, Iyetomi H, Arai Y (2016) Interdependencies and causalities in coupled financial networks. *PloS one* 11(3):e0150994.

Winer R (2001) A framework for customer relationship management. *California Management Review* 43(4):89–105.

Appendix to "Constructing the Customer Journey Map of Competitive Brands: A Complex Time-series Analysis "

Makoto Mizuno[1], Hideaki Aoyama[2] and Yoshi Fujiwara[3]

[1]School of Commerce, Meiji University, Tokyo 101-8301, Japan

[2] Research Institute of Economy, Trade and Industry, Tokyo 100-8901, Japan,

and RIKEN iTHEMS, Wako, Saitama 351-0198, Japan

[3]Graduate School of Simulation Studies, University of Hyogo, Kobe 650-0047, Japan.

APPENDIX: The eigenmodes 1 and 2: detail

We list the components of the first eigenmode with absolute value above the 2σ range in Fig. 5 and Fig. 6 and, similarly, for the second eigenmode in Fig. 7 and Fig. 8.

Brand	Variable	Phase	Abs.
A1	TVAd	2.22	0.21
A1	WebV	2.57	0.07
A1	MobV	2.85	0.10
A1	Q	4.91	0.12
A1	P	5.09	0.08
A2	WebV	2.90	0.14
A2	TVAd	2.93	0.09
A2	MobV	3.08	0.08
A2	P	5.57	0.07
A3	MobV	2.59	0.18
A3	WebV	2.8	0.20
A3	TVAd	2.93	0.08
A3	Q	4.83	0.09
A3	P	5.32	0.08
A4	TVAd	0.44	0.08
A4	Q	5.29	0.07
A6	P	3.88	0.10
A6	Q	5.49	0.11
A7	P	4.85	0.11
A7	Q	5.99	0.14
B1	WebV	1.82	0.07
B1	TVAd	2.12	0.08
B1	Q	5.41	0.09
B1	P	5.83	0.07
B2	Q	3.90	0.08
B2	TVAd	4.81	0.08
B3	WebS	3.27	0.07
B3	TVAd	5.39	0.14
B4	Q	4.91	0.09

Table 5 List of components in the first eigenmode with absolute value above the 2σ range (to be continued to Fig. 6)

Brabd	Variable	Phase	Abs.
B5	WebV	3.52	0.08
B5	TVAd	4.18	0.15
B5	P	4.21	0.08
B5	Q	5.03	0.07
C1	WebV	3.62	0.07
C1	TVAd	4.57	0.11
C1	Q	5.33	0.08
C2	MobS	1.86	0.10
C2	MobV	1.86	0.10
C2	Q	5.63	0.11
C2	P	6.26	0.10
C3	TVAd	5.50	0.12
C3	Q	5.89	0.25
C3	P	6.16	0.27
C4	Q	0.96	0.11
C4	WebS	2.89	0.07
C4	WebV	3.37	0.09
C4	TVAd	4.68	0.11
C4	P	6.12	0.11
C5	Q	2.88	0.22
C5	P	6.01	0.33
D1	Q	5.24	0.09
D1	P	5.47	0.11
D1	TVAd	5.84	0.15
D1	WebS	6.07	0.07
D2	MobV	3.05	0.06
D2	WebV	3.54	0.07
D2	WebS	5.04	0.08
D2	P	5.52	0.09
D2	Q	5.70	0.09
D3	TVAd	3.87	0.19
D3	Q	4.77	0.07

Table 6 -continued from Fig. 5

Brand	Variable	Phase	Abs.
A1	TVAd	5.49	0.10
A1	Q	5.70	0.17
A1	P	5.84	0.09
A2	TVAd	4.52	0.09
A2	Q	5.73	0.11
A3	MobS	4.77	0.07
A3	TVAd	5.2	0.08
A3	Q	5.54	0.18
A3	P	5.92	0.08
A3	MobV	6.06	0.07
A4	WebS	4.09	0.08
A4	TVAd	4.76	0.13
A4	WebV	5.24	0.12
A4	Q	5.80	0.20
A4	P	6.11	0.14
A6	P	0.73	0.06
A7	Q	4.92	0.12
B1	P	5.19	0.10
B1	WebS	5.28	0.07
B1	Q	5.36	0.18
B1	TVAd	5.53	0.08
B2	Q	4.06	0.16
B2	WebS	5.02	0.07
B2	P	5.27	0.09
B2	TVAd	5.64	0.11
B3	WebV	2.91	0.08
B3	TVAd	5.21	0.09
B4	Q	5.15	0.19
B4	P	5.46	0.12

Table 7 **List of components in the second eigenmode with absolute value above the 2σ range (to be continued to Fig. 8)**

Brand	Variable	Phase	Abs.
B5	Q	5.07	0.11
B5	P	5.36	0.10
B5	TVAd	5.63	0.10
C1	WebV	2.02	0.07
C1	TVAd	5.08	0.19
C1	WebS	5.25	0.07
C1	Q	5.96	0.12
C1	P	6.28	0.08
C2	TVAd	5.22	0.12
C2	P	5.37	0.08
C2	Q	5.49	0.13
C3	Q	3.18	0.12
C3	P	3.38	0.18
C3	TVAd	3.69	0.12
C3	WebV	5.2	0.08
C4	TVAd	4.47	0.09
C4	WebV	5.14	0.07
C4	Q	5.34	0.22
C5	P	3.10	0.18
C5	Q	5.99	0.23
D1	Q	5.52	0.19
D1	WebS	5.74	0.07
D1	P	5.81	0.14
D2	WebS	5.34	0.08
D2	Q	5.34	0.18
D2	P	5.39	0.12
D2	TVAd	5.58	0.12
D3	TVAd	0.29	0.12
D3	P	5.14	0.08
D3	Q	5.81	0.21

Table 8 **-continued from Fig. 7**