A Compositional Data Analysis of Market Share Dynamics*

Yoshiyuki Arata†† Tamotsu Onozaki‡

April 5, 2017

Abstract: Market share is an important variable for understanding how competition works in a market. However, we cannot resort to the conventional multivariate statistics in analyzing market share dynamics because market shares are subject to the constraint that their total sum must be 1. This constrained nature of market share has been ignored in the literature and led to misleading results. The present paper sheds new light on this issue by applying compositional data analysis to market share data on Japanese manufacturing firms. Our findings are as follows: First, there are several industries showing the variation of changes in market share becomes large as market concentration becomes higher. Firms in these industries face an even more intense struggle for market share in spite of high concentration. Second, in the industries examined, shares of the top 2 or 3 firms fluctuate in the same direction and their changes are compensated by changes in the shares lower-ranked firms. Thus, market competition is characterized by competition between the subgroup of the top 2 or 3 firms and lower-ranked firms rather than competition between the top 2 firms.

Keywords: Market share dynamics; Market competition; Compositional data analysis

JEL Classification Numbers: L22; L11; D43

---

*This research was conducted as a part of the Project ‘Sustainable Growth and Macroeconomic Policy’ undertaken at Research Institute of Economy, Trade and Industry (RIETI). This work was partly supported by JSPS KAKENHI (Grant Number 23330086).
†Corresponding author.
††Research Institute of Economy, Trade and Industry (RIETI), 1-3-1 Kasumigaseki, Chiyoda-ku, Tokyo, 100-8901 Japan. E-mail: arata-yoshiyuki@rieti.go.jp.
‡Department of Economics, Rissho University, 4-2-16 Osaki, Shinagawa-ku, Tokyo, 141-8602 Japan. E-mail: onozaki@ris.ac.jp.
1 Introduction

Inter-firm competition is at the heart of our modern economy: Issues like innovation, regulations, and economic growth cannot be discussed without market competition. Also in many theoretical models, competition among firms has been taken for granted. However, an empirical investigation of market competition is not as simple as it might seem. Indeed, how to empirically explore and characterize competition in real market situations is still a challenging issue and no consensus about the methodology has yet emerged (see, e.g., Boone et al. (2007)). In this regard, market share is an important variable for understanding how competition works in a market. This paper examines market share dynamics of Japanese manufacturing firms and sheds new light on this issue. In particular, by treating a change in market share as a random variable and applying a newly introduced statistical method called compositional data analysis (CDA), we find statistical regularities of market share dynamics which have not been addressed in the existing literature.

There has been a strand of literature on market share dynamics closely related to the aim of this paper. The early literature includes Stigler (1964), Shepherd (1970), Heggestad and Rhoades (1976), Caves and Porter (1978), Baldwin and Gorecki (1994), Geroski and Toker (1996), and Davies and Geroski (1997), in which the variability of market share changes (sometimes called market mobility) is viewed as a measure of competition and its relation with the market concentration has been discussed.\(^1\) While their findings are limited owing to a lack of detailed data, recent studies have used comprehensive data and provided new insights from a different perspective. For example, Sutton (2007a) uses a data set on Japanese manufacturing firms and finds that the correlation coefficient between market shares of the top 2 firms is close to 0.\(^2\) Based on this finding, he argues that shocks to firms are independent in many markets, on the contrary to the conventional argument of competition.

This impressive finding, however, has a severe drawback. Sutton (2007a) considers market shares as a random variable in the \(D\)-dimensional real space \(\mathbb{R}^D\) and ignores an important constraint: \(\sum_{i=1}^{D} x_i = 1\), where \(x_i\) denotes market share of firm \(i\). A constellation of market shares of \(D\) firms

---

\(^1\)There seem to be two different views on market competition: static and dynamic views. For example, Hayek (1948) argues that the notion of competition in traditional microeconomics based on a static view differs from what is used in economic actuality and it fails to describe the dynamical property of competition (see Yanagita and Onozaki (2010) for a detailed discussion). We focus on dynamic aspects of market competition in this paper.

should not be viewed as a point moving around in the $D$-dimensional real space $\mathbb{R}^D$ but as a point traveling on the $(D-1)$-dimensional hyper-plane, $\sum_{i=1}^{D} x_i = 1$. In order for graphical understanding of this constrained nature, a 3-dimensional case (i.e., $D = 3$) is depicted in Figure 1, where sample points are plotted not in $\mathbb{R}^D$ but on the triangle representing the plane $x_1 + x_2 + x_3 = 1$. By this constraint, statistical analysis based on, for example, correlation coefficients is severely distorted: An increase in a component must be compensated by decreases in other components, leading to a negative bias of the correlation coefficient.\(^3\)

![Figure 1: 3-dimensional case.](image-url)

For an illustration of how this constraint causes problems in the analysis of market share dynamics, suppose that there are three firms in a market, each of which has an equal market share. Let $X_i$ and $x_i$, $i = 1, 2, 3$ be the sales and the shares of the three firms, respectively (i.e., $x_1 = x_2 = x_3 = 1/3$). We assume that the sales of each firm grow independently and the growth rate follows a normal distribution: The sales of a firm in the next year are given by $\varepsilon_i X_i$ and $\varepsilon_i$ is drawn from $\mathcal{N}(1, \sigma_i^2)$. If our focus is on the relation between firms 1 and 2, the correlation coefficient

\[^3\text{Formally, this bias is described as follows. From the constraint and its expectation, we have}
\]

\[x_1 - E[x_1] + x_2 - E[x_2] + ... + x_D - E[x_D] = 0.\]

By multiplying both sides of this equation by $x_1 - E[x_1]$ and taking expectation, we obtain

\[
\begin{align*}
\text{Var}(x_1) + \text{Cov}(x_1, x_2) + ... + \text{Cov}(x_1, x_D) &= 0 \\
\text{Cov}(x_1, x_2) + ... + \text{Cov}(x_1, x_D) &= -\text{Var}(x_1) (< 0),
\end{align*}
\]

where $\text{Var}$ and $\text{Cov}$ denote the variance and covariance, respectively.
between \( \varepsilon_1 \) and \( \varepsilon_2 \) gives us an answer. Since the growth rates, \( \varepsilon_1 \) and \( \varepsilon_2 \), are drawn independently, the sample correlation coefficient is close to 0 (see Figure 2, in which \( \text{Corr}(\varepsilon_1, \varepsilon_2) = -0.000790 \)). Given the sales of the three firms and their growth rates, growth rate of market shares can be obtained: For each \( i \), the growth rate is given by \( \varepsilon_i^* := \frac{\varepsilon_i X_i}{\sum_{i=1,2,3} \varepsilon_i X_i} = \frac{3\varepsilon_i}{\sum_{i=1,2,3} \varepsilon_i} \). As in the case of sales, we can obtain the correlation coefficients of the growth rates of market shares. The samples of \( \varepsilon_1^* \) and \( \varepsilon_2^* \) with different values of \( \sigma_3 \) are plotted in Figure 3. This suggests that the correlation coefficients between firms 1 and 2 depend significantly on \( \sigma_3 \): If \( \sigma_3 \) is tiny, the changes of market shares are due mainly to the fluctuations of the sales of firms 1 and 2 and, therefore, the negative bias noted above is dominant in the correlation coefficient, \( \text{Corr}(\varepsilon_1^*, \varepsilon_2^*) = -0.798 \) in Figure 3(a). In contrast, the fluctuation of \( \varepsilon_3 \) is large, the market shares of firms 1 and 2 increases and decreases together to compensate the fluctuation of the share of firm 3. Because of this effect, the correlation coefficient becomes positive, \( \text{Corr}(\varepsilon_1^*, \varepsilon_2^*) = 0.580 \), in Figure 3(c). Note that the underlying relation between firms 1 and 2 is exactly the same as in Figure 2, that is, independence. Since this bias is caused by the constraint itself, it is seriously misleading to apply correlation analysis to market share data. The point is that given the constraint, the bias depends on the behavior of firm 3 and we do not know the reference state corresponding to the zero correlation defined on \( \mathbb{R}^2 \). Thus, even if our focus is on the relation between firms 1 and 2, it is impossible to infer implications from the correlation coefficient of market shares of the two firms.

Figure 2: Scatter plot of sales growth, \( \varepsilon_1 \) and \( \varepsilon_2 \). In this simulation, \( \sigma_1 = \sigma_2 = 0.1 \) and the number of samples is 500. The sample correlation coefficient is \(-0.000790\).
Figure 3: Scatter plot of market share growth, $\varepsilon_1^*$ and $\varepsilon_2^*$. The sample correlation coefficients are given by (a) $-0.798$ (b) $-0.526$ and (c) $0.580$, respectively. Note that the sales of the two firms $\varepsilon_1$ and $\varepsilon_2$ used to obtain $\varepsilon_1^*$ and $\varepsilon_2^*$ are exactly the same as the ones in Figure 2. The only difference in $\sigma_3$ yields differences in these panels.

One might say that this difficulty can be avoided by using data free from such a constraint as $\sum_{i=1}^{n} x_i = 1$. Indeed, Coad and Teruel (2012) follow this strategy and inspect firm growth measured by employees, sales, and value added instead of market share. They find the uncorrelated growth rates of rival firms, consistent with Sutton’s finding. However, another problem arises concerning this type of analysis. Let us suppose a market whose size fluctuates due to demand shocks. Market expansion and contraction may lead to the comovement of sales of firms and a positive correlation between sales, but this positive correlation cannot be interpreted as evidence of a complementary relation between rival firms. In other words, even a fiercely competitive relationship can be positively biased due to the fluctuation of market size. Thus, we should exclude the bias caused by fluctuations in market size, but this is not a simple task. One might introduce normalization by dividing individual firm’s sale by their total.\footnote{This procedure is essentially equivalent to subtracting growth rate of total sales from that of individual sales.} Note that normalized sales are nothing but market shares and the same difficulty mentioned above confronts us: An increase in normalized sales of a firm implies decreases in normalized sales of other firms, that is, a negative bias.

This difficulty should not be viewed as just a technical one: It is related to which aspect of market competition is under consideration. The fact that an increase in a firm’s share must be compensated by decreases in others’ means that market share has information on the relative scale rather than the absolute scale. For illustration, suppose that market share of a firm is determined
by the quality of its product, that is, a firm whose product is of the highest quality takes the largest share, and each firm is engaged in improving its quality to take market share away from competitors. In this situation, improvement of the quality does not necessarily imply an increase in market share if its competitors make more rapid progress. The quality improvement of a firm must be compared with the others': Namely, the relative quality does matter for market share. Because of this relative nature of market share, it is meaningless to consider that the shares of all firms simultaneously increase.

In this paper, we focus on this point: How does the relative market position change? Our focus is not on whether the average quality (or sales) of their products improves or deteriorates but on how a firm cuts off and outperforms its rivals. Since the conventional statistical methods are not suited for such relative variables, we need an alternative method to analyze market share. For this purpose, we use the recently developed statistical approach called CDA, which enables us to obtain implications of market share dynamics without the bias mentioned above. To the best of our knowledge, this paper is the first application of CDA to the comprehensive market share data.

Applying CDA to changes in market shares, we explore the following two points: (1) the variability in the market shares of the top 2 firms and its relation with market concentration and (2) the structure of competition among the top 5, 6, and 7 firms. For (1), we find that there are several industries showing the variance of the changes in the market shares becomes larger as market concentration becomes higher. High market concentration does not mean that firms in a market get rid of competition for market share but rather even more fierce competition. For (2), we find a particular structure of competition among the top 5, 6, and 7 firms in the industries examined: The shares of the top 2 or 3 firms increases and decreases together and their changes are compensated by those of lower-ranked firms. It suggests that market competition is characterized by competition between the subgroup of the top 2 or 3 firms and the lower-ranked firms rather than competition between the top 2 firms.

The rest of paper is organized as follows. Section 2 introduces CDA and shows its applicability to the analysis of market share dynamics. Section 3 applies CDA to market share data of Japanese manufacturing firms. In particular, Section 3.2 analyzes the variability in the market shares of the top 2 firms and its relation with market concentration. Section 3.3 analyzes the structure

5For the explanation of CDA, see the next section.
of competition among the top 5, 6, and 7 firms. Section 4 concludes this paper. Appendix A summarizes the method of outlier detection employed in our analysis.

2 Compositional Data Analysis (CDA)

CDA is a rapid growing field in statistics and explicitly takes into account the fact that the components sum up to unity: \( x_1 + x_2 + \ldots + x_D = 1 \). The difficulty concerning correlation discussed above is called spurious correlation, which is firstly pointed out by Pearson (1897). The spurious correlation is by no means pathological in real applications. In the field of geology, which is one of the main application fields of CDA, a series of papers (e.g., Chayes (1960)) have confirmed that the spurious correlation is widespread in the literature. Given the importance of compositional data and the obvious constraint, it is surprising that problems related to the spurious correlation have remained unnoticed in other fields including economics.\(^6\) The spurious correlation becomes serious especially in the analysis of market share. Suppose that our interest lies in whether the properties of the dynamics of market share depend on market concentration, which is one of the fundamental issues in the early literature (see the Introduction). However, it is problematic to compare the correlation coefficients with different concentrations because the bias by the constrained nature of market share also depends on the concentration. There is no easy way to distinguish correlation representing the underlying economic relation from the bias.

Related to the spurious correlation, the identification of the boundary of relevant markets is another problem which makes an analysis based on the conventional correlation questionable. While the boundary of a market has been explicitly given in most of the theoretical studies and its identification has been viewed as a technical one, this problem turns out to be serious to empirical researchers.\(^7\) Moreover, it is in practice unavoidable that some firms are missing, which means that the boundary of a market is misspecified. Any reliable analysis of market share should be robust to the misspecification of a boundary, but the conventional correlation does not satisfy this property. In contrast, CDA overcomes these difficulties in a consistent manner.

---

\(^6\)One exception in the economic literature is a series of studies by Fry et al. (1996, 2000), in which they apply CDA to budget share models of households’ expenditure. However, to our knowledge, no study has applied CDA to market share data in the existing literature on industrial organization.

\(^7\)See, e.g., Kaplow (2015). It is common that the boundary of a market is defined in terms of competition, that is, firms are in a market if they compete against each other. However, competition is a concept difficult to define and sometimes depends on the boundary itself.
In the 1980s, a series of papers in the statistical literature have tackled the difficulties of compositional data and these efforts have culminated in the seminal work by Aitchison (1986), who develops an axiomatic approach satisfying a set of fundamental principles. Among them, a principle called subcomposition coherence in CDA literature is worth mentioning in our analysis. A subcomposition is defined to be a subset of components; for example, if there are $D$ firms in a market and we have $D$ shares of firms, $x_1, x_2, \ldots, x_D$, $\sum_{i=1}^{D} x_i = 1$, the shares of two firms, $x_1' := cx_1, x_2' := cx_2, x_1' + x_2' = 1$, where a constant $c$ is introduced so that the sum is unity, is a subcomposition of the full composition. The subcomposition coherence means that results obtained from the subcomposition are coherent with those obtained from the full composition. The conventional correlation does not satisfy this principle whereas CDA does, which is one of our motivations to use CDA. In the 2000s, the approach has been further elaborated and generalized by several statisticians (for reviews, see Pawlowsky-Glahn and Buccianti (2011), Pawlowsky-Glahn et al. (2015)). Following this line of literature, we apply CDA to market share dynamics in this paper.

Let us begin with notations. We define a sample space called simplex as follows:

$$S^D := \left\{ \mathbf{x} = (x_1, x_2, \ldots, x_D) : x_i > 0 (i = 1, 2, \ldots, D), \sum_{i=1}^{D} x_i = 1 \right\}$$

As noted above, the difficulties related to compositional data arise from the fact that the structure of $S^D$ is different from that of the real space $\mathbb{R}^D$. For example, the simplex is not a vector space with $+$ and $\cdot$: $\exists \mathbf{x}, \mathbf{y} \in S^D$ and $a \in \mathbb{R}$ such that $a \mathbf{x} + \mathbf{y} \notin S^D$. It means that we cannot discuss a linear combination such as linear regression and principal component analysis because a linear combination may not be an element in $S^D$. Namely, the operations, $+$ and $\cdot$, are not suited for $S^D$. What are operations in $S^D$ playing the role of $+$ and $\cdot$ in $\mathbb{R}^D$? These operations called perturbation (denoted by $\oplus$) and powering ($\odot$) are defined as follows:

$$\mathbf{x} \oplus \mathbf{y} := \mathcal{C}(x_1 y_1, x_2 y_2, \ldots, x_D y_D), \quad \alpha \odot \mathbf{x} := \mathcal{C}(x_1^\alpha, x_2^\alpha, \ldots, x_D^\alpha),$$

$$\mathcal{C}\mathbf{x} := \left( \frac{x_1}{\sum_{i=1}^{D} x_i}, \frac{x_2}{\sum_{i=1}^{D} x_i}, \ldots, \frac{x_D}{\sum_{i=1}^{D} x_i} \right),$$

where the operation $\mathcal{C}$ is called closure.
It should be noted that the two operations $\oplus$ and $\odot$ have economic meaning, especially in the context of firm growth models. Suppose that the firm growth process follows Gibrat’s law, that is, the sales of firm $i$, $s_{i,t}$, grow proportionally to its previous sales:\footnote{In the literature on firm growth, it is well-known that Gibrat’s law provides a good fit to empirical data, especially for large firms. See, e.g., Coad (2009).}

$$s_{i,t} = \varepsilon_{i,t} s_{i,t-1}, \quad (1)$$

where $\varepsilon_{i,t}$ is a growth shock independent from its previous sale. Expressing sales of firms in terms of market share (i.e., $x_t := \mathcal{C}(s_t)$), equation (1) is written as follows:

$$x_t = x_{t-1} \oplus \varepsilon_t.$$

Thus, the shares $x_t$ can be seen as the sum of the previous shares and a growth shock with the operation $\oplus$. As in the same manner, the difference of shares between successive years can be defined as follows: $x_t \ominus x_{t-1} := x_t \oplus (-1) \odot x_{t-1} = \varepsilon_t.$\footnote{In a different strand of literature, a replicator model is used to describe the path of firm growth (see, e.g., Mazzucato (2000)):}

$$\frac{ds_{i,t}}{dt} = \lambda_i s_{i,t}, \quad s_t = s_0 \cdot \exp(\lambda t),$$

where $\lambda = \{\lambda_1, \lambda_2, \ldots, \lambda_D\}$ is a constant vector representing the competitiveness of firms. The equation above can be written in terms of $\oplus$ and $\odot$ as follows:

$$x_t = x_0 \oplus t \odot \exp(\lambda).$$

Thus, the replicator model can be seen as a straight line in $\mathcal{S}^D$ with angle $\exp(\lambda)$.\footnote{In the literature on firm growth, it is well-known that Gibrat’s law provides a good fit to empirical data, especially for large firms. See, e.g., Coad (2009).}
\[
v = \text{clr}(x) := \log \left[ \frac{x_1}{g_m(x)}, \frac{x_2}{g_m(x)}, \ldots, \frac{x_D}{g_m(x)} \right], \quad g_m(x) = \left( \prod_{i=1}^{D} x_i \right)
\]
with inverse,

\[
x = \text{clr}^{-1}(v) := C \exp(v).
\]

The clr transformation has several useful properties; for example, it preserves the structure given by perturbation and powering, that is,

\[
\text{clr}(\alpha \odot x \oplus y) = \alpha \text{clr}(x) + \text{clr}(y).
\]

Furthermore, the Aitchison inner product and distance can be simplified by the clr transformation:

\[
\langle x, y \rangle_A = \langle \text{clr}(x), \text{clr}(y) \rangle, \quad d_A(x, y) = d(\text{clr}(x), \text{clr}(y)) = \sqrt{\sum_{i=1}^{D} (\text{clr}_i(x) - \text{clr}_i(y))^2},
\]

where \(\langle \cdot, \cdot \rangle\) and \(d(\cdot, \cdot)\) are the usual inner product and Euclidean distance in \(\mathbb{R}^D\), respectively. Thus, by the clr transformation, \(\oplus, \odot, \langle \cdot, \cdot \rangle_A, \text{ and } d_A(\cdot, \cdot)\) in \(S^D\) correspond to \(+, \cdot, \langle \cdot, \cdot \rangle, \text{ and } d(\cdot, \cdot)\) in \(\mathbb{R}^D\), providing a Euclidean structure to \(S^D\). This means that we are able to deal with elements in \(S^D\) as if they are variables in \(\mathbb{R}^D\) with the usual operations. However, it should be noted that \(\text{clr}(x)\) has a new constraint, \(\sum_{i=1}^{D} \text{clr}_i(x) = 0\), that is, the transformed data are collinear. To overcome this disadvantage, Egozcue et al. (2003) introduce the isometric logratio (ilr) transformation.

The ilr transformation is essentially equivalent to choosing an orthonormal basis on the hyperplane \(H := \{v \in \mathbb{R}^D : \sum_{i=1}^{D} v_i = 0\}\) by, for example, the Gram-Schmidt algorithm. Formally, this is defined as follows: let \(\{e_1, e_2, \ldots, e_{D-1}\}\) be an orthonormal basis of \(S^D\), i.e., \(\langle e_i, e_j \rangle_A = \delta_{ij}\). For a fixed orthonomal basis, the ilr transformation is given as follows:

\[
x^* = \text{ilr}(x) := (\langle x, e_1 \rangle_A, \langle x, e_2 \rangle_A, \ldots, \langle x, e_{D-1} \rangle_A)
\]

\[
x = \text{ilr}^{-1}(x^*) := \oplus_{i=1}^{D-1} x_i^* \odot e_i.
\]

The ilr transformation gives the coordinates of \(x\) represented in \(\mathbb{R}^{D-1}\). Analogous to the clr
transformation, the ilr transformation satisfies the following relations:

\[ \text{ilr}(\alpha \odot x \oplus y) = \alpha \text{ilr}(x) + \text{ilr}(y). \]

\[ \langle x, y \rangle_\alpha = \langle \text{ilr}(x), \text{ilr}(y) \rangle, \quad d_\alpha(x, y) = d(\text{ilr}(x), \text{ilr}(y)) \]

Note that \( x \in S^D \) is transformed into \( x^* \in \mathbb{R}^{D-1} \) by ilr and \( x^* \) has no additional restriction. \( x^* \) is a variable in \( \mathbb{R}^{D-1} \) and therefore the conventional multivariate statistics in \( \mathbb{R}^{D-1} \) can be directly applied to \( x^* \). In short, our strategy consists of the following steps (see Table 1):\(^1\)

1. Variables in \( S^D \), that is, market share in our analysis, are transformed into \( \mathbb{R}^{D-1} \) by ilr.
2. Multivariate statistical analysis in \( \mathbb{R}^{D-1} \) (e.g., principal component analysis (PCA) and cluster analysis) are carried out on the transformed variables.
3. The results are inversely transformed to the original space \( S^D \) by ilr\(^{-1}\).

\[ S^D \xrightarrow{\text{ilr}} \mathbb{R}^{D-1} \xrightarrow{\text{ilr}^{-1}} S^D \]

<table>
<thead>
<tr>
<th>Compositional data</th>
<th>Multivariate statistical analysis</th>
<th>Interpretation</th>
</tr>
</thead>
</table>

Table 1: Working on coordinates.

In the next section, market share dynamics is examined based on this strategy.

3 Market Share Dynamics

3.1 Data

Our dataset consists of annual observations of market shares of manufacturing firms in Japan over the period of 1980–2009. The source of our data is *Market Share in Japan*, published by Yano Research Institute Ltd.\(^1\) The classification corresponds roughly to 6-digit commodity classification for the Census of Manufactures in Japan, in which manufacturing goods are classified into 2,363 markets.\(^2\) This source is unique in that it provides information for each market: Namely, we obtain

\(^1\)This strategy is called the principle of working on coordinates in the CDA literature. See, e.g., Mateu-Figueras et al. (2011).

\(^2\)This data source is the same one used in Sutton (2007a) and Kato and Honjo (2006).

\(^3\)Hereafter, we call 6-digit classification *markets* (e.g., heavy bearing rings) and 3-digit classification (e.g., iron & steel) *industries*.
market composition and the names of firms for each market. Although databases used in previous works (e.g., Coad and Teruel (2012)) have detailed information on firms’ attributes, they provide no detailed information on the goods that each firm supplies. It is common to classify firms to one sector according to their main activity. However, not a few firms, especially large firms, supply more than one product. In contrast, our database focus on markets rather than individual firms, and therefore a firm appears in multiple markets in our dataset.

The choice of markets examined in our analysis is based on two criteria: the length of the time series and the number of firms in a market. Since we focus on markets existing over a long period rather than emerging or disappearing markets, we choose markets with more than 25-annual observations over the period of 1980–2009. The industries and the number of markets examined in our analysis are given in the following sections.

In our analysis, we examine the growth of market share defined by $\varepsilon_t := x_t \otimes x_{t-1}$. To catch an initial glimpse of the data, we first display $\varepsilon_t$ by ternary diagrams. As noted before, compositional data in $S^3$ represented in the three-dimensional space are contained in a planar triangle. The ternary diagram is the two-dimensional projection of this triangle viewed along the direction of the arrow in Figure 1. In Figure 4, the pooled growth of market share in the food and pharmaceutical industries are shown. It suggests that while both sample points are scattered around the center of the triangle, the dispersion for the pharmaceutical industry is much larger; namely, market mobility of the pharmaceutical industry is higher than that of the food industry. We further analyze these data by CDA methods developed in the previous section.

3.2 Dependence of market growth on its concentration

Market share, especially market concentration, has been viewed as an important variable in the context of market competition. Monopoly, oligopoly, and perfect competition are concepts closely related to the state of market competition.\footnote{Even when we have a year’s observation of a market, the number of firms in data can be insufficient because smaller firms are often agglomerated into others. Samples with smaller number of firms (\(<\ 2\) in Section 3.2 and \(<\ 5, 6,\ and\ 7\) in Section 3.3) are removed.} While these concepts play a fundamental role in the\footnote{The center of the triangle (denoted by $\lambda_3$) is defined as $\lambda_3 := (1/3, 1/3, 1/3)$ and the center in $S^D$ in general is defined as $\lambda_D := (1/D, 1/D, 1/D)$. If $\varepsilon_t = \lambda_D$, the market composition remains the same, that is, $x_t = x_{t-1} \otimes \lambda_D = x_{t-1}$.} competition becomes fierce as the number of firms becomes large and the market shares of each firm become tiny. On the other hand, fierce competition may force inefficient firms out of the market and only efficient firms survive, resulting in an increase in concentration. Namely, a small
IO literature, they are based on a static description of the market. Dynamic properties are another important issue: How does market concentration influence market share dynamics? Is the struggle for market share exacerbated or alleviated as the concentration becomes higher? In this section, we answer these questions by investigating whether the distribution of ε of the top 2 firms depends on market concentration. Complementary to the previous works, this analysis shows another aspect of market competition.

For this purpose, we need to define the distribution of ε := (ε₁, ε₂) and other probability distributions such as normal distributions for statistical tests, which becomes possible by CDA: Once variables in SD are transformed, we can apply the usual tools (e.g., normal distributions in R^{D−1} and variance tests) to them. Thus, we first transform ε in S² into a variable in R by ilr: ilr(ε) = \frac{1}{\sqrt{2}} \log(\frac{ε₁}{ε₂}).\(^{16}\) Figure 5 shows the histogram of ilr(ε), clearly suggesting the significant departure from normality: It has excessive kurtosis and fatter tail than that of normal distribution.\(^{17}\) In number of firms in a market may be a consequence of reallocation. For example, Boone et al. (2007) empirically show that the effect of reallocation is dominant in many markets, that is, fierce competition leads to higher concentration.

\(^{16}\)This ilr representation is obtained as follows. First, we transform ε by the clr transformation: clr(ε) = (\log \frac{ε₁}{g_m(ε)}, \log \frac{ε₂}{g_m(ε)}) with \log \frac{ε₁}{g_m(ε)} + \log \frac{ε₂}{g_m(ε)} = 1. Second, we choose an orthonormal basis in this space. Since the dimension of this space is 1, the choice of an orthonormal basis is either \frac{1}{\sqrt{2}}(1, 1) or \frac{1}{\sqrt{2}}(−1, 1) in R². The former is chosen. This choice does not affect any conclusions of our analysis. Finally, taking an inner product with this basis, we obtain the ilr representation of ε, ilr(ε) = \frac{1}{\sqrt{2}} \log(\frac{ε₁}{ε₂}).

\(^{17}\)For lack of space, a histogram only for general machinery is shown in Figure 5 and others are omitted. Histograms for other industries show a very similar shape.
other words, compared with normal distribution, market shares remain almost the same in most of the cases (excessive kurtosis), but we sometimes observe drastic changes of market positions (fatter tail). This finding is closely related to the literature on the distribution of firm growth rates. In this literature, it is empirically known that the distribution of firm growth rates follows a Laplace distribution, which is similar to the one shown in Figure 5, rather than a normal distribution.\textsuperscript{18} However, since the cross-sectional distribution of firm growth rates is considered in this literature, the relationship with rivals is not taken into account. There is a possibility that shocks to a whole market such as demand shocks are dominant and the sales of the two firms increase and decrease together with the market position unchanged. Figure 5 shows that this is not the case: As in the distribution of firm growth rates, the growth of market share is also characterized by excessive kurtosis and fatter tail. Thus, it suggests that the Laplace shape is due mainly to the intrinsic feature of the growth pattern of individual firms rather than that of shocks to a whole market.

\begin{figure}
\centering
\includegraphics[width=0.5\textwidth]{figure5}
\caption{Histogram of ilr(ε) for general machinery.}
\end{figure}

Next, we discuss the variance of ε. The variance of ε can be defined based on the Aitchison

Thus, this is the variance of the transformed data in $\mathbb{R}^{D-1}$. Descriptive statistics of $\text{ilr}(\varepsilon) = \frac{1}{\sqrt{2}} \log(\frac{\xi_1}{\xi_2})$ are given in Table 2. It shows that the mean is quite close to 0, which corresponds to the neutral element $\lambda_2 := (\frac{1}{2}, \frac{1}{2})$ in $S^D$. Note that the shock $\varepsilon = \lambda_2$ does not change the market position in the previous year. Thus, Table 2 implies that there is no particular tendency in favor of or against the market leader.

<table>
<thead>
<tr>
<th>Industry</th>
<th># of markets</th>
<th># of obs.</th>
<th>Mean of $\varepsilon$ (1st &amp; 2nd)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Iron &amp; Steel</td>
<td>23</td>
<td>581</td>
<td>.503, .497</td>
</tr>
<tr>
<td>General Machinery</td>
<td>132</td>
<td>3605</td>
<td>.501, .499</td>
</tr>
<tr>
<td>Transportation</td>
<td>10</td>
<td>249</td>
<td>.504, .496</td>
</tr>
<tr>
<td>Precision Machinery</td>
<td>25</td>
<td>658</td>
<td>.500, .500</td>
</tr>
<tr>
<td>Electrical Machinery</td>
<td>35</td>
<td>833</td>
<td>.499, .501</td>
</tr>
<tr>
<td>Chemical</td>
<td>20</td>
<td>501</td>
<td>.499, .502</td>
</tr>
<tr>
<td>Food</td>
<td>37</td>
<td>880</td>
<td>.498, .502</td>
</tr>
<tr>
<td>Paper</td>
<td>11</td>
<td>283</td>
<td>.499, .501</td>
</tr>
<tr>
<td>Pharmaceutical</td>
<td>44</td>
<td>1102</td>
<td>.501, .499</td>
</tr>
<tr>
<td>Cosmetics</td>
<td>19</td>
<td>515</td>
<td>.498, .502</td>
</tr>
</tbody>
</table>

Table 2: Descriptive statistics. The first column shows the industries (3-digit classification) considered in our analysis. The second column is the number of markets (6-digit classification). The third column represents the number of sample observations. The fourth and fifth columns show the mean of $\varepsilon$ in $S^2$.

In order to examine the dependence of the variance on its concentration, we decompose our sample data of $\varepsilon$ for each industry by the median of its concentration. Here, the concentration is defined to be the sum of shares of the two firms. Obtaining the two subsets with higher and lower concentrations, we compare the two variances. The results are given in Table 3. We employ two

\[^{19}\text{For later purpose, the variance is defined in a more general form. For } D = 2, \text{ the first line of the equation above becomes } \text{Var} [\varepsilon] = \frac{1}{2} \text{Var} \left[ \log \frac{\xi_1}{\xi_2} \right].\]

15
Figure 6: Boxplots of samples. The subset with lower (higher) concentration is denoted by 1 (2). For lack of space, only four industries are shown.
<table>
<thead>
<tr>
<th>Industry</th>
<th>Ratio of var.</th>
<th>Robust</th>
<th>F-test</th>
<th>Levene</th>
<th>F-K</th>
<th>Median(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Iron &amp; Steel</td>
<td>.678</td>
<td>.620</td>
<td>.00108</td>
<td>.0221</td>
<td>.0170</td>
<td>54.8</td>
</tr>
<tr>
<td>General Machinery</td>
<td>1.141</td>
<td>1.412</td>
<td>.00635</td>
<td>.0177</td>
<td>.00435</td>
<td>44.2</td>
</tr>
<tr>
<td>Transportation Equip.</td>
<td>.579</td>
<td>.527</td>
<td>.00256</td>
<td>.0247</td>
<td>.0282</td>
<td>69.9</td>
</tr>
<tr>
<td>Precision Machinery</td>
<td>.667</td>
<td>.875</td>
<td>.000299</td>
<td>.0952</td>
<td>.251</td>
<td>62.1</td>
</tr>
<tr>
<td>Electrical Machinery</td>
<td>.831</td>
<td>.579</td>
<td>.0550</td>
<td>.0223</td>
<td>.00344</td>
<td>40.2</td>
</tr>
<tr>
<td>Chemical</td>
<td>.327</td>
<td>.191</td>
<td>&lt; 2.2e-16</td>
<td>4.40e-08</td>
<td>5.63e-10</td>
<td>57.5</td>
</tr>
<tr>
<td>Food</td>
<td>.695</td>
<td>.578</td>
<td>.000202</td>
<td>.0151</td>
<td>.0180</td>
<td>51.6</td>
</tr>
<tr>
<td>Paper</td>
<td>.475</td>
<td>.329</td>
<td>1.67e-05</td>
<td>.00221</td>
<td>.00138</td>
<td>47.4</td>
</tr>
<tr>
<td>Pharmaceutical</td>
<td>1.474</td>
<td>2.88</td>
<td>7.46e-06</td>
<td>3.35e-05</td>
<td>7.86e-07</td>
<td>36.8</td>
</tr>
<tr>
<td>Cosmetics</td>
<td>1.402</td>
<td>2.24</td>
<td>.00862</td>
<td>.0928</td>
<td>.122</td>
<td>63.4</td>
</tr>
</tbody>
</table>

Table 3: Ratio of the two variances and tests for equality. The second column shows the ratio of variances (variance for low concentration divided by variance of high concentration) based on the classical method. The third column shows the ratio of variances based on a robust method. The fourth column is the p-value derived from the F-test of equality of the two variances. The fifth (sixth) column is the p-value derived from Levene’s (Fligner-Killeen) test. The seventh column is the median of market concentration at which our samples are decomposed.

For tests of equality of variances, we first use the usual F-test. Since the F-test is sensitive to the normality assumption and the distribution of $\text{ilr}(\varepsilon)$ does not seem to follow a normal distribution as shown in Figure 5, we also employ Levene’s and Fligner-Killeen tests. Roughly speaking, the results in Table 3 shows that these industries can be classified into three groups:

- **Group 1:** Iron & Steel, Transportation Equipment, Chemical, Food, and Paper. The variance of the growth $\varepsilon$ becomes larger when market concentration is high.
- **Group 2:** General Machinery, Precision Machinery, and Electrical Machinery. The variance is less dependent on market concentration.\(^{21}\)
- **Group 3:** Pharmaceutical and Cosmetics. The variance becomes larger when market concentration is low.

In addition, boxplots of $\frac{1}{\sqrt{2}} \log\left( \frac{\hat{\varepsilon}_1}{\hat{\varepsilon}_2} \right)$ in Figure 6 show a consistent picture with results in Table 3.

\(^{20}\)Here, the robust variance estimate means an estimate based on the median absolute deviation. See, e.g., Filzmoser and Hron (2008).

\(^{21}\)While the tests for general machinery show the statistically significant difference of the two variances by the large number of observations, the point estimates show that the difference is relatively small. Thus, general machinery is classified into group 2. Precision machinery is classified into group 2 because its robust estimates of the ratio (.875) is relatively close to 1.
The reason why the variance for the pharmaceutical industry is negatively correlated with market concentration can be interpreted as follows: In this industry, patents play an important role in the business and become high entry barriers for new entrants, leading to high concentration in a market (see e.g., Bergman and Rudholm (2003)). However, once patents expire, the barriers are removed and the existing firms have to compete against new entrants. The high variability of $\varepsilon$ with low concentration reflects this fierce competition.

However, this scenario cannot be generalized to other manufacturing industries. Figure 3 shows that the variance of industries within group 1 becomes larger when its concentration is high. Namely, even if the concentration becomes high, firms in these industries are unable to get rid of a struggle for market share: Rather, they face even more intense competition for market share. Regarding to this point, the analysis by Boone et al. (2007) may give us a clue to understanding our results. Introducing a new index of competition called profit elasticity, they focus on the reallocation effect of competition: Fierce competition forces inefficient firms out of its market and leads to an increase in market concentration. They empirically show that the reallocation effect dominant in many markets. Taking into account this point, an increase in the variability of $\varepsilon$ can be seen as a consequence of this effect. The top firms are in a fierce competition with each other and the changes of market share are volatile, preventing new comers from entering market. While it might be debatable to argue that the degree of competition is measured solely by the variance of $\varepsilon$, its dependence on market concentration and differences across industries reflect one of the aspects of market competition.

### 3.3 Who competes against whom?

We have so far focused the growth of market share of the top 2 firms. In this section, we examine market share dynamics including other firms in a market (i.e., firms ranked 3, 4, and 5) and then explore the structure of $\varepsilon := (\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4, \varepsilon_5)$. For example, consider the following question: Is an increase in a firm’s share compensated by a decrease in the share of another particular firm? If two particular firms compete against each other, the changes of market shares would happen within the two firms keeping other firms’ shares unchanged. In many theoretical models (e.g., see Sutton (2007b)), rankings represent firms’ superiority (e.g., productivity) but no particular relationship between firms has been addressed. To explore this question empirically, we perform compositional
PCA and cluster analysis to $\varepsilon$ of the top 5 firms. Since both methods require a stable covariance matrix, we assume that the covariance matrix of $\varepsilon$ does not depend on the previous shares $x$. Thus, we focus on industries in group 2 based on the results in the previous section: general machinery, precision machinery, electrical machinery, and food industries.$^{22}$

Descriptive statistics are given in Table 4. As in the case of $D = 2$, the mean of $\varepsilon$ is very close to the neutral element $\lambda_5 := (1/5, 1/5, 1/5, 1/5, 1/5)$. It means that, on average, there is no tendency in favor of or against the market leader. This point can be confirmed by the ternary diagrams given in Figure 7, which shows that sample points scatter around the center corresponding to $\lambda_5$.

<table>
<thead>
<tr>
<th>Ind.</th>
<th># of markets</th>
<th># of obs</th>
<th>Mean, 1st</th>
<th>2nd</th>
<th>3rd</th>
<th>4th</th>
<th>5th</th>
</tr>
</thead>
<tbody>
<tr>
<td>General Machinery</td>
<td>105</td>
<td>1885</td>
<td>.201</td>
<td>.200</td>
<td>.199</td>
<td>.200</td>
<td>.200</td>
</tr>
<tr>
<td>Precision Machinery</td>
<td>13</td>
<td>102</td>
<td>.200</td>
<td>.199</td>
<td>.201</td>
<td>.201</td>
<td>.199</td>
</tr>
<tr>
<td>Electrical Machinery</td>
<td>32</td>
<td>670</td>
<td>.199</td>
<td>.199</td>
<td>.200</td>
<td>.201</td>
<td>.201</td>
</tr>
<tr>
<td>Food</td>
<td>27</td>
<td>259</td>
<td>.199</td>
<td>.200</td>
<td>.199</td>
<td>.199</td>
<td>.203</td>
</tr>
</tbody>
</table>

Table 4: Descriptive statistics. For explanation, see the caption in Table 2.

Next, we perform compositional PCA, which is done as follows: We first transform $\varepsilon$ in $S^5$ to ilr($\varepsilon$) in $\mathbb{R}^4$. We then apply PCA to ilr($\varepsilon$): We estimate the location vector and covariance matrix of ilr($\varepsilon$), and apply singular value decomposition.$^{23}$ In this analysis, we focus on two principal components (PCs). Note that, since the ilr transformation means the dimensional reduction of the data by one ($5 \rightarrow 4$), the ilr coordinates has no direct interpretation relating to the original components. Following the convention in the CDA literature, the PCs are inversely transformed to the clr representation for interpretation.

Results are shown in Figure 8. Arrows in these figures represent clr coordinates of the two PCs.$^{24}$ Links between arrow heads provide information on the variability of the corresponding log-ratio, i.e., the length of a link (i.e., $|X_iX_j|^2$ in Figure 8 is proportional to $\text{Var}\left[\log \frac{X_i}{X_j}\right]$). Figure 8 clearly shows that links between $\varepsilon_1$ and $\varepsilon_2$ (and $\varepsilon_3$) are short, indicating that the shares of firms ranked 1 and 2 (and 3) move up and down in the same direction rather than the opposite direction. Namely, a struggle for market share among the top 2 firms is weak; rather, the subgroup of the

$^{22}$The food industry is added to the industries being examined in this section because the ratio of the two variances (.695) are closest to 1 in group 1.

$^{23}$We use a robust method developed by Filzmoser et al. (2009).

$^{24}$If the clr coordinates of the two PCs are represented as $(a_1, a_2, \ldots, a_D)$ and $(b_1, b_2, \ldots, b_D)$, the coordinate of the head of arrow $X1$ in Figure 8 is $(a_1, b_1)$. 

19
Figure 7: Ternary diagrams.
top 2 (or 3) firms competes against the lower-ranked firms (4th or 5th) for market share. The same picture can be observed by cluster analysis shown in Figure 9. Here, the distance between two components is measured by $\text{var} \log(\frac{E_i}{E_j})$, based on which the components are clustered. Figure 9 shows that the top 2 or 3 firms are clustered as close components and distant from the lower-ranked firms.

Interestingly, this property can be observed even when we increase the number of firms considered. In Figures 10 and 11, we perform the same analysis with additional firms. Figure 10 shows the compositional PCA and cluster analysis to $\epsilon := (\epsilon_1, \epsilon_2, \ldots, \epsilon_6)$ of the top 6 firms for general machinery and electric machinery industries. Figure 11 is the results of the top 7 firms for the same industries. Both figures show that the shares of the top 3 (or 4) firms are likely to move up and down in the same direction. As in the case of 5 firms, the subgroup of the top firms competes against the lower-ranked firms (6th or 7th) for market share.

As noted above, any particular relation or structure between firms has not been addressed in the standard game-theoretic modeling. In these models, the competitive relation between the market leader and the 2nd firm is essentially the same as the one between the market leader and a lower ranked firm. In this sense, these models are symmetric with respect to ranking. However, our finding suggests that competitive relations in a market have a particular structure. The reason for this structure is not given in this paper and needs to be addressed in future research.

4 Conclusions

Competition among firms has been taken for granted in theoretical models while empirical analysis on this issue is by no means an easy task. Market share and its dynamics are an important clue to explore inter-firm competition, but the constrained nature of market share as compositional data has impeded us from using conventional multivariate statistics. To overcome this difficulty, this paper applies the recently developed statistical method called CDA to market share data and explores market share dynamics. To the best of our knowledge, this is the first application of CDA in this literature. We have shown that the space structure introduced by CDA has a natural interpretation in the context of firm growth models, which justifies our usage of CDA for the

\footnote{The reason why these two industries are that the number of samples for these industries are large.}
Figure 8: Compositional principal component analysis. The proportions of variance explained by the two PCs are (a) 65.5%, (b) 71.2%, (c) 69.2%, and (d) 68.1%.
Figure 9: Cluster analysis.
Figure 10: Compositional PCA and cluster analysis. The number of firms is six. The proportions of variance explained by the two PCs are (a.1) 58.0%, (b.1) 66.0%.
(a.1) Compositional PCA. General Machinery.  
(b.1) Compositional PCA. Electric Machinery.  

(a.2) Cluster analysis. General Machinery.  
(b.2) Cluster analysis. Electric Machinery.  

Figure 11: Compositional PCA and cluster analysis. The number of firms is seven. The proportions of variance explained by the two PCs are (a.1) 44.4%, (b.1) 52.4%.
analysis of market share dynamics.

We have examined the dependence of changes in the market share on market concentration. We have found that there are several industries showing the variation of market share becomes larger as market concentration is high. High market concentration does not mean that firms get rid of competition for market share but rather even more fierce one. The analysis of the top 5, 6, and 7 firms shows that there is a particular structure in competition: Market competition is characterized by competition between the subgroup of the top firms and the lower-ranked firms rather than competition between the top 2 firms.

Our analysis has shown the applicability of CDA in the literature on industrial organization. Since it enables us to apply the conventional statistical tools to market share without inconsistency, it broadens the scope of the analysis of market share. Since the use of CDA is extremely rare in the economic literature, there is plenty of room for exploitation.

Appendix

A Outliers

For detecting outliers in our samples, we follow the approach developed by Filzmoser and Hron (2008). In this approach, the Mahalanobis distance (MD) based on the Minimum Covariance Determinant (MCD) estimates for location and covariance matrix are used as a criteria for outliers. Since the squared MD follows the \( \chi^2_{D-1} \) distribution under the normality assumption of samples, the .975 quantile of \( \sqrt{\chi^2_{D-1}} \) is used as the cut-off value in the literature.

We compute the MD values for every sample and plot them in Figure 12. The solid line refers to the cut-off values corresponding to the .975 quantile. It should be noted, however, that our samples do not seem to follow normal distribution (see Figure 5) and therefore samples above the line may be due to the departure from normality. If so, removing all samples above the cut-off value would be too excessive. In our analysis, we decide to only remove extreme samples which are visually isolated from the main cloud of samples. For the analysis in Section 3.2 and Section 3.3, the cut-off values are MD = 7 and MD = 20, respectively. Although the choice of the cut-off value by visual inspection is debatable, we have confirmed that our conclusion does not significantly depend on
the choice of the cut-off values.

![Graphs showing outlier detection](image)

(a) Iron & Steel for $D = 2$.  
(b) General Machinery for $D = 5$.

Figure 12: Outlier detection.

References


Mateu-Figueras, G., Pawlowsky-Glahn, V., Egozcue, J. J., 2011. The principle of working on co-
ordinates. In: Pawlowsky-Glahn, V., Buccianti, A. (Eds.), Compositional data analysis: Theory
and applications. John Wiley & Sons, Chichester, pp. 31–42.

concentration and instability. Edward Elgar Publishing.

John Wiley & Sons.

positional data. John Wiley & Sons.

Pearson, K., 1897. Mathematical Contributions to the Theory of Evolution. On a Form of Spurious
Correlation Which May Arise When Indices Are Used in the Measurement of Organs. Proceedings


Sutton, J., 2007a. Market Share Dynamics and the “Persistence of Leadership” Debate. The Amer-


Yanagita, T., Onozaki, T., 2010. Dynamics of market structure driven by the degree of consumer’s