RIETI Discussion Paper Series 17-E-076

A Compositional Data Analysis of Market Share Dynamics

ARATA Yoshiyuki
RIETI

ONOZAKI Tamotsu
Rissho University

RIETI
Research Institute of Economy, Trade & Industry, IAA

The Research Institute of Economy, Trade and Industry
http://www.rieti.go.jp/en/

RIETI Discussion Paper Series 17-E-076
May 2017

A Compositional Data Analysis of Market Share Dynamics[1]

ARATA Yoshiyuki[2]
Research Institute of Economy, Trade and Industry
ONOZAKI Tamotsu
Department of Economics, Rissho University

Abstract

The existing literature has shown that there are several statistical regularities in industrial dynamics, which are an important clue to understanding the underlying mechanism. This paper focuses on market share changes and shows that its distribution has some remarkable properties. Because of the constrained nature of market share, namely, the sum must be unity, this paper applies the recently developed method called compositional data analysis (CDA) to market share data. We find the distribution does not follow a Gaussian but instead a tent-shaped distribution with a fatter tail, which is closely related with the findings of firm growth rate distribution. With some exceptions, this statistical feature can be observed across different sectors. Furthermore, this property can be observed when we focus on the relation between the top subgroup and lower-ranked firms. This distribution shape implies that market share growth cannot be described by an accumulation of small shocks. Rather, lumpy jumps that transform the market structure are crucial in market share dynamics. Put differently, radical change in market structure is a relatively frequent phenomenon. Such implications based on statistical properties of observed data help us further investigate industrial dynamics theoretically.

Keywords: Market share dynamics, Compositional data analysis, Subbotin family
JEL classification: L22; L11; D43

[1]This research was conducted as a part of the “Sustainable Growth and Macroeconomic Policy” project undertaken at the Research Institute of Economy, Trade and Industry (RIETI). This work was partly supported by JSPS KAKENHI (Onozaki: Grant Number 23330086).
[2]Corresponding author, E-mail: arata-yoshiyuki@rieti.go.jp

1 Introduction

During the past few decades, a series of studies have revealed a number of remarkable statistical regularities in industrial dynamics: the positive skewness and fat-tailedness of firm size distributions (e.g., Axtell (2001), Gabaix (2009)), Laplace shape of firm growth rate distributions (e.g., Stanley et al. (1996), Bottazzi et al. (2002, 2007, 2011), Bottazzi and Secchi (2006)) and productivity dispersion (e.g., Dosi et al. (2016); for review, see, e.g., Dosi et al. (2017)). Surprisingly, it has been shown that these *stylized facts* are quite robust and hold across sectors and countries as well as time periods, highlighting universal properties of industrial dynamics. The importance of these statistical regularities cannot be overstated because they give us an important clue to understanding of underlying mechanism. The aim of this paper is to make a contribution to this literature by presenting another new empirical regularity in market share dynamics. Applying a newly introduced statistical method called *compositional data analysis* (CDA) to Japanese manufacturing firms, we find remarkable distributional features in market share dynamics which have not been addressed in the existing literature.

There has been a strand of literature empirically analyzing market share dynamics (e.g., Geroski and Toker (1996), Davies and Geroski (1997), Mazzucato (2000, 2002), Mazzucato and Parris (2015), and Sutton (2007)), which gives us insight into how market structure evolves over time.[1] This type of analysis, however, is prone to suffer from difficulties caused by the constraint: the sum adds up to unity, $\sum_{i=1}^{D} x_i = 1$, where x_i denotes market share of firm i. In the previous studies, this constraint has not been explicitly taken into account and the conventional multivariate statistical methods developed for the D-dimensional real space $\mathbb{R}^D$ has been applied to data. However, as explained in the following, this procedure leads to biased results. A constellation of market shares of D firms should be viewed as a point on the $(D-1)$-dimensional hyper-plane given by $\sum_{i=1}^{D} x_i = 1$ rather than $\mathbb{R}^D$. 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 $x_1 + x_2 + x_3 = 1$.

For an illustration of how this constraint causes difficulties, let us consider correlation analysis as in Sutton (2007), where he examines the correlation coefficient between market share changes

[1] Among them, Sutton (2007) is closest to the aim of this paper. Sutton (2007) uses a large and disaggregated data set on Japanese manufacturing firms and tries to find new statistical regularities that holds universally.

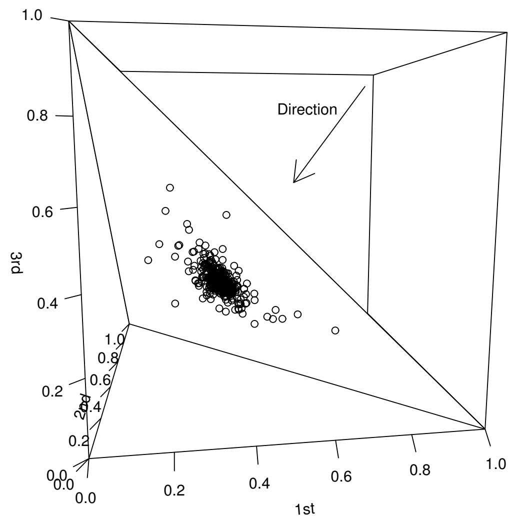


Figure 1: 3-dimensional case.

of the top 2 firms, finding that it is close to 0. Suppose that there are three firms in a market, each of which has an equal market share, and our focus is to examine the relation between firms 1 and 2. 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 ε_i is drawn from $\mathcal{N}(1, \sigma_i^2)$. Since the growth rates, ε_1 and ε_2, are drawn independently, the sample correlation coefficient is close to 0 (see Figure 2, in which $\mathrm{Corr}(\varepsilon_1, \varepsilon_2) = -.000790$). Thus, the sample correlation coefficient gives us an insight into the underlying mechanism.

How about market share? Given the sales of the three firms and their growth rates, the growth rate of market shares, ε_i^*, can be obtained as follows: for each i, $\varepsilon_i^* := \frac{\varepsilon_i X_i}{\sum_{j=1,2,3} \varepsilon_j X_j} / x_i = \frac{3\varepsilon_i}{\sum_{j=1,2,3} \varepsilon_j}$ because of the equal market shares. The sample plots of ε_1^* and ε_2^* with different values of σ_3 and correlation coefficients are given in Figure 3. Panel (a) and (b) in Figure 3 suggests that when σ_3 is not large, that is, the variance of ε_3 is not large, the sample correlation coefficient shows a negative value even though the growth rates of *sales* are independent with each other as shown Figure 2. This is due to a negative bias by the constraint of market share because an increase in a firm's share must be offset by decreases in other firms' shares.[2]

[2]Formally, this bias is described as follows. From the constraint and its expectation, we have

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

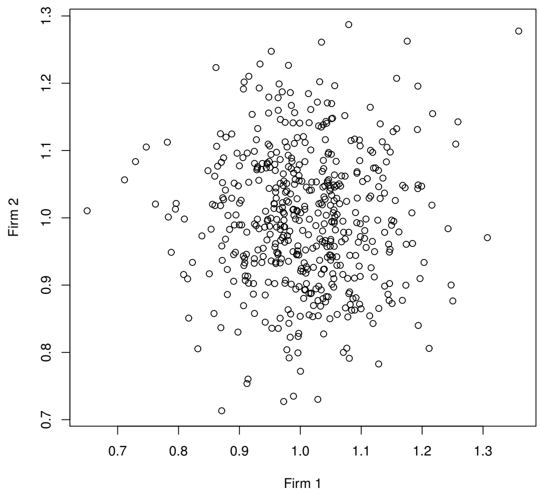


Figure 2: Scatter plot of sales growth, ε_1 and ε_2. In this simulation, $\sigma_1 = \sigma_2 = .1$ and the number of samples is 500. The sample correlation coefficient is $-.000790$.

In contrast, when σ_3 is large, the market shares of firms 1 and 2 have to increase and decrease together to offset the fluctuation of the firm 3 market share. Because of this effect, the sample correlation coefficient becomes positive, $\mathrm{Corr}(\varepsilon_1^*, \varepsilon_2^*) = .580$, as shown 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 these positive and negative values of sample correlation coefficients are caused by the constraint, it is seriously misleading to interpret them as an evidence of some economic mechanism. The point is that the bias depends on the behavior of firm 3 even if our focus is on the relation between firms 1 and 2. Since the correlation coefficient is biased to the unknown extent, it is impossible to obtain implications about firms' relation from the correlation analysis.[3] Given

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

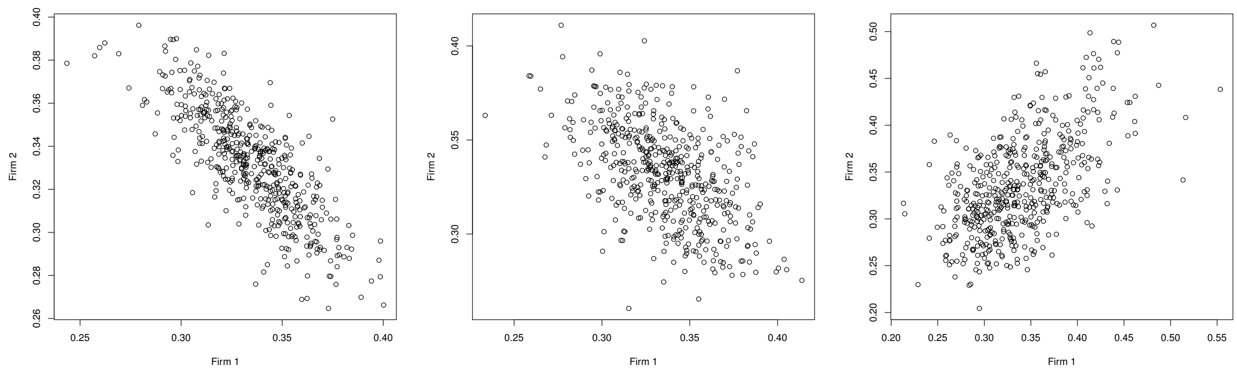
$$\begin{aligned} \mathrm{Var}(x_1) + \mathrm{Cov}(x_1, x_2) + \ldots + \mathrm{Cov}(x_1, x_D) &= 0 \\ \mathrm{Cov}(x_1, x_2) + \ldots + \mathrm{Cov}(x_1, x_D) &= -\mathrm{Var}(x_1) \ (< 0), \end{aligned}$$

where Var and Cov denote the variance and covariance, respectively.

[3] 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.

The point in our analysis is that we focus on the *relative* market position: Market share captures how a firm's position changes *compared with its rivals.*

the fundamental role of correlation coefficient in statistical analysis, this suggests that we need an alternative approach to uncover empirical regularities in market share dynamics.



(a) $\sigma_3 = .01$. (b) $\sigma_3 = .1$. (c) $\sigma_3 = .4$.

Figure 3: Scatter plot of market share growth, ε_1^* and ε_2^*. The sample correlation coefficients are given by (a) $-.798$ (b) $-.526$ and (c) $.580$, respectively. Note that the sales of the two firms ε_1 and ε_2 used to obtain ε_1^* and ε_2^* are exactly the same as the ones in Figure 2. The only difference in σ_3 yields differences in these panels.

To overcome these difficulties, we introduce CDA in this paper, which enables us to obtain implications of market share dynamics without the bias mentioned above. In particular, by defining new operators, CDA enables us to use statistical concepts such as mean, variance, distribution and the central limit theorem on the space of market shares. We apply CDA to market share data on Japanese manufacturing firms and explore statistical properties of market share dynamics. To the best of our knowledge, this paper is the first application of CDA to the comprehensive market share data.

Our analysis shows that the distribution of market share change displays a remarkable feature: The distribution does not follow a Gaussian but a Laplace-like tent-shaped distribution with a fatter tail. This distribution is closely related with the findings of firm growth rate distribution. This shape of the distribution implies that market share change cannot be described by an accumulation of small shocks. Rather, lumpy jumps than completely transform the market structure are crucial in market share dynamics. Namely, such a radical change in market structure is relatively frequent. Interestingly, with some exceptions, this statistical feature can be observed across different sectors. Furthermore, this property can be observed when we focus on the relation between the top subgroup and lower-ranked firms. Therefore, our analysis shows that this statistical property captures an essential feature of market share dynamics.

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 top 2 firms and explore its distributional properties. Section 3.3 extend our analysis to the multivariate case (the top 5, 6, and 7 firms) and analyze its marginal distribution. 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: $x_1 + x_2 + ... + x_D = 1$. 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.[4] 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

[4] 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.

researchers.[5] 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.

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, $x_1, x_2, ..., x_D$, $\sum_{i=1}^{D} x_i = 1$, 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:

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

As noted above, the difficulties related to compositional data arise from the fact that the structure of $\mathcal{S}^D$ is different from that of the real sapce $\mathbb{R}^D$. For example, the simplex is not a vector space with $+$ and $\cdot$: $\exists \mathbf{x}, \mathbf{y} \in \mathcal{S}^D$ and $a \in \mathbb{R}$ such that $a\mathbf{x} + \mathbf{y} \notin \mathcal{S}^D$. 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 $\mathcal{S}^D$. Namely, the operations, $+$ and $\cdot$, are not suited for $\mathcal{S}^D$. What are operations in $\mathcal{S}^D$ playing the role of $+$ and $\cdot$ in $\mathbb{R}^D$? These operations called *perturbation*

[5] 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.

(denoted by $\oplus$) and *powering* ($\odot$) are defined as follows:

$$\mathbf{x} \oplus \mathbf{y} := \mathcal{C}(x_1y_1, x_2y_2, ..., x_Dy_D), \quad \alpha \odot \mathbf{x} := \mathcal{C}(x_1^\alpha, x_2^\alpha, ..., x_D^\alpha),$$

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

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:[6]

$$s_{i,t} = \varepsilon_{i,t} s_{i,t-1}, \tag{1}$$

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

$$\mathbf{x}_t = \mathbf{x}_{t-1} \oplus \boldsymbol{\varepsilon}_\mathbf{t}.$$

Thus, the shares $\mathbf{x}_t$ 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: $\mathbf{x}_t \ominus \mathbf{x}_{t-1} := \mathbf{x}_t \oplus (-1) \odot \mathbf{x}_{t-1} = \boldsymbol{\varepsilon}_t$.[7]

As in the conventional linear regression and principal component analysis, the orthogonality needs to be defined in $\mathcal{S}^D$ for further analysis. We introduce Aitchison inner product in $\mathcal{S}^D$ as follows:

$$\langle \mathbf{x}, \mathbf{y} \rangle_A := \frac{1}{D} \sum_{i<j} \log \frac{x_i}{x_j} \log \frac{y_i}{y_j}$$

[6]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).

[7]In a different strand of literature, a replicator model is used to describe the path of firm growth (see, e.g., Mazzucato (2000), See ICC Dosi et al 2017 footnote 4):

$$\frac{ds_{i,t}}{dt} = \lambda_i s_{i,t}, \quad \mathbf{s}_t = \mathbf{s}_0 \cdot \exp(\boldsymbol{\lambda} t),$$

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

$$\mathbf{x}_t = \mathbf{x}_0 \oplus t \odot \exp(\boldsymbol{\lambda}).$$

Thus, the replicator model can be seen as a *straight line* in $\mathcal{S}^D$ with angle $\exp(\boldsymbol{\lambda})$.

The induced distance is given by

$$d_A(\mathbf{x},\mathbf{y}) := \|\mathbf{x} \ominus \mathbf{y}\|_A = \sqrt{\frac{1}{D}\sum_{i<j}\Big(\log\frac{x_i}{x_j} - \log\frac{y_i}{y_j}\Big)^2}$$

The meanings of the inner product and distance become clear by considering a transformation from $\mathcal{S}^D$ to $\mathbb{R}^D$. First, we define the *centered log-ratio* (clr) transformation:

$$\mathbf{v} = \text{clr}(\mathbf{x}) := \log\Big[\frac{x_1}{g_m(\mathbf{x})}, \frac{x_2}{g_m(\mathbf{x})}, ..., \frac{x_D}{g_m(\mathbf{x})}\Big], \quad g_m(\mathbf{x}) = \Big(\prod_{i=1}^{D} x_i\Big)$$

with inverse,

$$\mathbf{x} = \text{clr}^{-1}(\mathbf{v}) := \mathcal{C}\exp(\mathbf{v}).$$

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

$$\text{clr}((\alpha \odot \mathbf{x}) \oplus \mathbf{y}) = \alpha\text{clr}(\mathbf{x}) + \text{clr}(\mathbf{y}).$$

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

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

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, $\oplus$, $\odot$, $\langle\cdot,\cdot\rangle_A$, and $d_A(\cdot,\cdot)$ in $\mathcal{S}^D$ correspond to $+$, $\cdot$, $\langle\cdot,\cdot\rangle$, and $d(\cdot,\cdot)$ in $\mathbb{R}^D$, providing a Euclidean structure to $\mathcal{S}^D$. This means that we are able to deal with elements in $\mathcal{S}^D$ as if they are variables in $\mathbb{R}^D$ with the usual operations. However, it should be noted that $\text{clr}(\mathbf{x})$ has a new constraint, $\sum_{i=1}^{D}\text{clr}_i(\mathbf{x}) = 0$, 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 $H := \{\mathbf{v} \in \mathbb{R}^D : \sum_{i=1}^{D} v_i = 0\}$ by, for example, the Gram-Schmidt algorithm. Formally, this is defined as follows: let $\{\mathbf{e}_1, \mathbf{e}_2, ..., \mathbf{e}_{D-1}\}$ be an orthonormal basis of $\mathcal{S}^D$, i.e., $\langle \mathbf{e}_i, \mathbf{e}_j\rangle_A = \delta_{ij}$. For a fixed orthonomal basis, the ilr transformation is given as follows:

$$\mathbf{x}^* = \text{ilr}(\mathbf{x}) := (\langle \mathbf{x}, \mathbf{e}_1 \rangle_A, \langle \mathbf{x}, \mathbf{e}_2 \rangle_A, ..., \langle \mathbf{x}, \mathbf{e}_{D-1} \rangle_A)$$

$$\mathbf{x} = \text{ilr}^{-1}(\mathbf{x}^*) := \oplus_{i=1}^{D-1} x_j^* \odot \mathbf{e}_i.$$

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

$$\text{ilr}((\alpha \odot \mathbf{x}) \oplus \mathbf{y}) = \alpha \text{ilr}(\mathbf{x}) + \text{ilr}(\mathbf{y}).$$

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

Note that $\mathbf{x} \in \mathcal{S}^D$ is transformed into $\mathbf{x}^* \in \mathbb{R}^{D-1}$ by ilr and $\mathbf{x}^*$ has no additional restriction. $\mathbf{x}^*$ is a variable in $\mathbb{R}^{D-1}$ and therefore the conventional multivariate statistics in $\mathbb{R}^{D-1}$ can be directly applied to $\mathbf{x}^*$. In short, our strategy consists of the following steps (see Table 1):[8]

1. Variables in $\mathcal{S}^D$, 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 $\mathcal{S}^D$ by ilr^{-1}.

$\mathcal{S}^D$	$\overset{\text{ilr}}{\Longrightarrow}$	$\mathbb{R}^{D-1}$	$\overset{\text{ilr}^{-1}}{\Longrightarrow}$	$\mathcal{S}^D$
Compositional data		Multivariate statistical analysis		Interpretation

Table 1: Working on coordinates.

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

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

3 Market Share Dynamics

3.1 Data

Our dataset consists of annual observations of market shares of Japanese manufacturing firms over the period of 1980–2009. The source of our data is *Market Share in Japan*, published by Yano Research Institute Ltd.[9] 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.[10] This source is unique in that the unit of analysis is market: we obtain market composition and the names of firms for each market. While databases used in previous works (e.g., Coad and Teruel (2012)) have detailed information on firms' attributes, firms are classified to a single 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 firms supplying more than one product appear across 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 restrict our attention to markets with more than 25-annual observations over the period of 1980–2009.The sectors and the number of markets examined in our analysis are given in the following sections.

3.2 Distribution of market share growth

In this section, we focus on the top 2 firms and examine market share growth defined by $\boldsymbol{\varepsilon}_t := \mathbf{x}_t \ominus \mathbf{x}_{t-1}$. As we have discussed above, we can define the distribution of market share growth by CDA. Based on this method, we explore its distributional properties in the following analysis.

We first transform $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2)$ in $\mathcal{S}^2$ into a one-dimensional variable in $\mathbb{R}$ by the ilr transformation: $\mathrm{ilr}(\boldsymbol{\varepsilon}) = \frac{1}{\sqrt{2}} \log(\frac{\varepsilon_1}{\varepsilon_2})$.[11] Figure 4(a) shows the kernel density estimate of pooled market

[9] This data source is the same one used in Sutton (2007) and Kato and Honjo (2006).

[10] Hereafter, we call 6-digit classification *markets* (e.g., heavy bearing rings) and 3-digit classification (e.g., iron & steel) *sectors*.

[11] The explicit form of the ilr representation is obtained as follows. First, we transform $\boldsymbol{\varepsilon}$ by the clr transformation: $\mathrm{clr}(\boldsymbol{\varepsilon}) = (\log \frac{\varepsilon_1}{g_m(\boldsymbol{\varepsilon})}, \log \frac{\varepsilon_2}{g_m(\boldsymbol{\varepsilon})})$ with $\log \frac{\varepsilon_1}{g_m(\boldsymbol{\varepsilon})} + \log \frac{\varepsilon_2}{g_m(\boldsymbol{\varepsilon})} = 0$. 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 $\mathbb{R}^2$. The former is chosen here. Finally, taking an inner product with this basis, we obtain the ilr representation of $\boldsymbol{\varepsilon}$, $\mathrm{ilr}(\boldsymbol{\varepsilon}) = \frac{1}{\sqrt{2}} \log(\frac{\varepsilon_1}{\varepsilon_2})$.

share growth, $\text{ilr}(\boldsymbol{\varepsilon})$, over 1980–2009, aggregated across over all the sectors. Descriptive statistics is given in Table 2. The first to be noticed is that the mean of $\text{ilr}(\boldsymbol{\varepsilon})$ is very close to 0, which corresponds to the neutral element $\boldsymbol{\lambda_2} = (\frac{1}{2}, \frac{1}{2})$ in $\mathcal{S}^D$. Since $\boldsymbol{\varepsilon} = \boldsymbol{\lambda_2}$ does not change its relative position in the next year, that is, $\mathbf{x} \oplus \boldsymbol{\lambda_2} = \mathbf{x}$, market share growth is on average like a fair coin tossing: The chances of taking market share away from its rival are fiftyfifty. Regarding distributional properties, Figure 4(a) clearly suggests the significant departure from normality:[12] Rather, the distribution is leptokurtic (tent-shape) and has fatter tail than that of Gaussian distribution. Namely, compared with a Gaussian distribution, we often observe drastic change in market structure with non-negligible probability. Moreover, this shape of the distribution is stable over time: Figure 4(b) plots kernel density estimates of $\text{ilr}(\boldsymbol{\varepsilon})$ in different years, showing similar tent-shaped distributions.

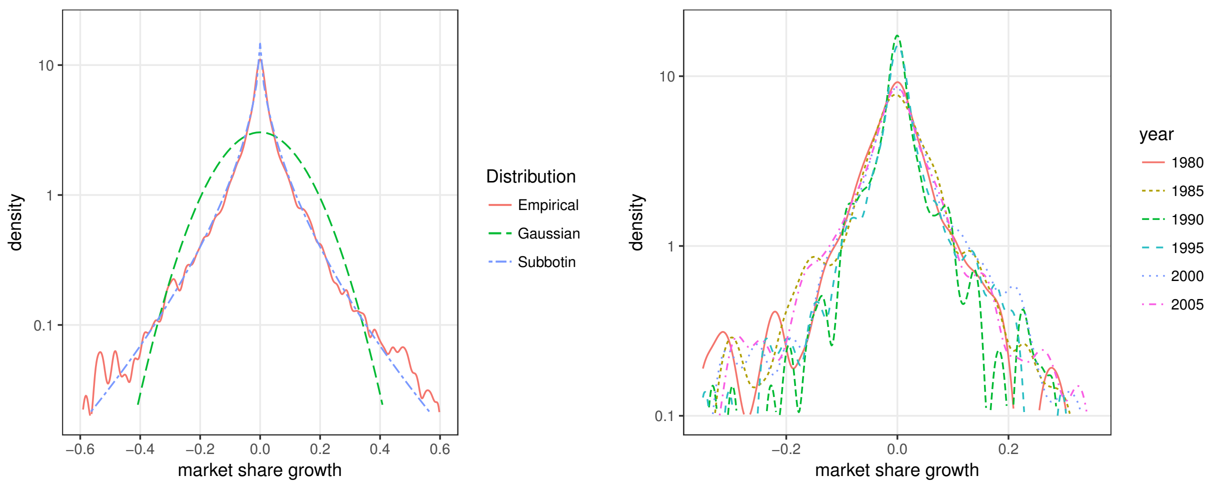
This finding is closely related to *stylized facts* of the distribution of firm growth rates. In the existing literature, it is empirically known that the distribution of firm growth rates does not follow a Gaussian but a Laplace distribution, which is similar to the one shown in Figure 4.[13] Interestingly, this statistical feature is observed at a disaggregated level and the shape of distributions across different sectors shows a surprising degree of homogeneity. This remarkable regularity implies that firm growth cannot be described by an accumulation of small independent shocks: if firm growth is a consequence of many small shocks, the distribution of firm growth rate would be Gaussian by the central limit theorem, which contradicts the stylized fact. Rather, the Laplace shape and the fatter tail implies that firm growth is characterized by lumpy *jumps*: Namely, drastic change in market structure is not rare but relatively common.

While the relationship with rivals is not taken into account in the literature on the distribution of firm growth rates, Figure 4 captures the statistical properties of change in relative market position. The same argument as in firm growth rates applies to the market share growth: market share dynamics cannot be characterized by gradual and smooth change. Rather, significant episodes of complete transformation of the market structure are relatively frequent.

To further characterize the shape of the empirical distributions, we consider a family of distributions called *Subbotin* distributions, which are used to describe firm growth rate distribution

[12] The null hypothesis of normality is rejected at 1 percent significance level by Anderson-Darling normality test.

[13] See a series of papers by G. Bottazzi and his coauthor (e.g., Bottazzi et al. (2002, 2007, 2011)). For theoretical explanations of the Laplace distribution, see Bottazzi and Secchi (2006) and Arata (2014).



(a) (b)

Figure 4: Kernel density estimation and fitted density functions. In (a), market share growth over the period 1980–2009 is pooled. "Empirical" refers to the kernel density estimation with Gaussian kernel. Bandwidth is chosen following the method in Scott (1992), using factor 1.06. "Gaussian"("Subbotin") refers to the Gaussian (Subbotin) fit. In (b), kernel density estimations in 5 different years are plotted. Estimation method is the same as in (a).

(see, e.g., Bottazzi and Secchi (2006)). The probability density function of Subbotin distributions is given as follows:

$$p(x) := \frac{1}{2ab^{\frac{1}{b}}\Gamma(1+1/b)} \exp\left(\frac{-(|x-\mu|^b)}{ba^b}\right), \tag{2}$$

where μ is a location parameter, a is a scale parameter, and b represents the shape of the distribution. Subbotin family includes Gaussian ($b = 2$) and Laplace distributions ($b = 1$) as special cases. Thus, a smaller value of b, especially $b < 2$, indicates the fatness of the tail. Maximum likelihood estimates (MLE) of a and b are reported in Table 2. This shows that the parameter b is not only lower than 2 (Gaussian case) but lower than 1 (Laplace case).[14] The tail of $\mathrm{ilr}(\varepsilon)$ is significantly fatter than that of a Laplace distribution. As noted above, leptokurticity and fat-tailedness are remarkable characteristics of the distribution of market share growth.

Next, we analyze this shape of the distribution at more disaggregated level, that is, each of the 10 3-digit sectors are considered separately. In Figure 5, all the 30 years of market share growth are pooled together under the stationary assumption. MLEs for each sector are given in Table 3.

[14] The null hypothesis of $b = 1$ is rejected for all cases at 1 percent significant level by the loglikelihood ratio test.

	# obs.	mean	s.d.	a	b
Pooled	9273	-0.00189	0.222	0.0448 (0.00076)	0.5507 (0.0112)
1980	296	-0.00232	0.109	0.0464 (0.00385)	0.6986 (0.06988)
1985	290	-0.01407	0.149	0.0524 (0.00433)	0.7737 (0.08236)
1990	308	-1e-05	0.089	0.0253 (0.00245)	0.4455 (0.04333)
1995	311	0.01124	0.128	0.0314 (0.00283)	0.5294 (0.05017)
2000	320	0.00631	0.093	0.0498 (0.00399)	0.6957 (0.0683)
2005	325	0.00563	0.114	0.0493 (0.00401)	0.673 (0.06651)

Table 2: Descriptive statistics and MLE. The 5th and 6th column shows the MLEs of the parameter a and b (standard error in parenthesis).

Figure 5 and Table 3 show that the distribution in all sectors except Transportation Equipment is tent-shaped as in the aggregate case: the parameter b is close to or smaller than 1. While we can observe some heterogeneity across sectors, especially Gaussian shape in Transportation Equipment, we can say that a tent-shaped distribution can be observed at a disaggregated level and the tent-shape observed at the aggregated level is not a mere statistical effect of aggregation. In most of the sectors, drastic and radical change in market structure is a frequent phenomenon.

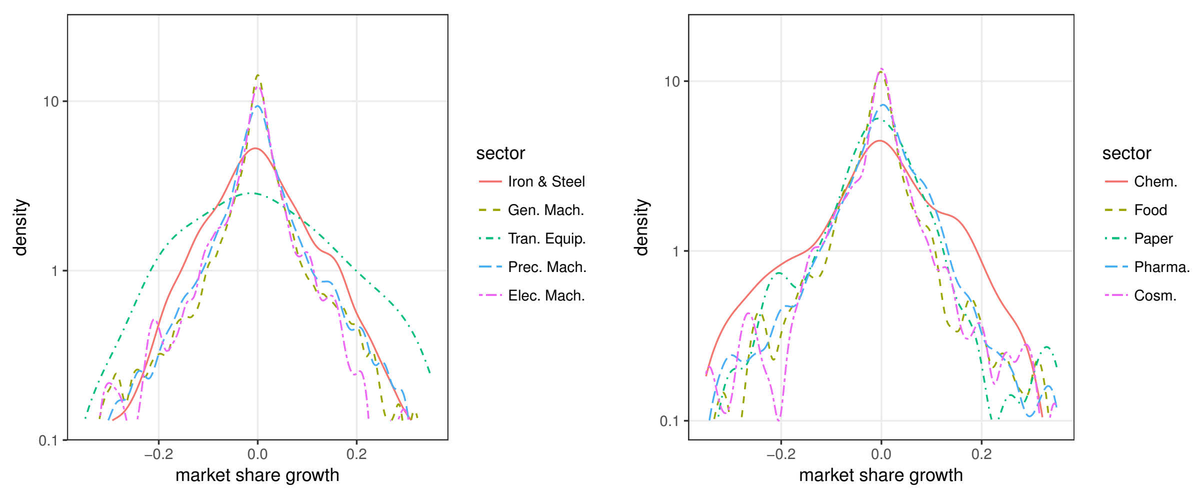


Figure 5: Kernel density estimations. Market share growth over the period 1980–2009 are pooled under the assumption that the distributions are stationary.

Finally, we use the variance of ε as an index of market mobility and examine its relation

	# obs.	mean	s.d.	a	b
Iron Steel	581	0.00131	0.215	0.079 (0.00423)	1.1672 (0.1036)
Gen. Mach.	3605	-0.00137	0.18	0.0368 (0.00101)	0.5069 (0.01469)
Tran. Equip.	264	-0.00235	0.164	0.1429 (0.01066)	2.1911 (0.37765)
Prec. Mach.	658	-0.00112	0.222	0.0465 (0.00277)	0.6225 (0.04468)
Elec. Mach.	883	-0.00222	0.165	0.0389 (0.00204)	0.5662 (0.03327)
Chem.	501	-0.00114	0.508	0.096 (0.00631)	1.1132 (0.12801)
Food	880	-0.00979	0.228	0.0414 (0.0022)	0.5772 (0.03448)
Paper	283	0.00405	0.188	0.0643 (0.0052)	0.8731 (0.09771)
Pharma.	1102	-0.00115	0.202	0.0594 (0.00256)	0.7867 (0.04605)
Cosm.	515	-0.00374	0.202	0.0407 (0.00291)	0.5255 (0.04033)

Table 3: Descriptive statistics and MLE. Thus, this is the variance of the transformed data in $\mathbb{R}^{D-1}$. The 5th and 6th column shows the MLEs of the parameter a and b (standard error in parenthesis).

with market concentration.[15] In a strand of literature focusing on evolutionary aspects of market structure (e.g., a series of studies by Mazzucato), the variability of market share (i.e., market mobility) and market concentration are viewed as important indexes characterizing the evolutionary stage of a market.[16] In particular, in order to describe market mobility, several indexes have been used in the literature (e.g., market share instability in Mazzucato (2002)) but suffer from the bias caused by the constrained nature of market share. In contrast, the variance of $\boldsymbol{\varepsilon}$ captures the variability of market share by definition and is free from the bias by virtue of CDA.

In order to examine the dependence of the variance on its concentration, we decompose our sample data of $\boldsymbol{\varepsilon}$ 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 4. Roughly speaking,

[15] The variance of $\boldsymbol{\varepsilon}$ can be defined based on the Aitchison distance:

$$\begin{aligned}\mathrm{Var}[\boldsymbol{\varepsilon}] &:= \frac{1}{D}\sum_{i<j}\mathrm{Var}\Big[\log\frac{\varepsilon_i}{\varepsilon_j}\Big]\\ &= \sum_{i=1}^{D}\mathrm{Var}[\mathrm{clr}_i(\boldsymbol{\varepsilon})]\\ &= \sum_{i=1}^{D-1}\mathrm{Var}[\mathrm{ilr}_i(\boldsymbol{\varepsilon})]\end{aligned}$$

Thus, this is the variance of the transformed data $\mathrm{ilr}(\boldsymbol{\varepsilon})$ in $\mathbb{R}^{D-1}$. For later purpose, the variance is defined in a more general form. For $D = 2$, the first line of the equation above becomes $\mathrm{Var}[\boldsymbol{\varepsilon}] = \frac{1}{2}\mathrm{Var}\Big[\log\frac{\varepsilon_1}{\varepsilon_2}\Big]$.

[16] For example, in an early stage, there are many firms in a market, that is, low market concentration and market share is instable (e.g., high entry-exit rate). As it become mature, market competition forces inefficient firms out of the market (high market concentration) and their market shares become stable.

the results in Table 4 shows that these sectors can be classified into three groups:

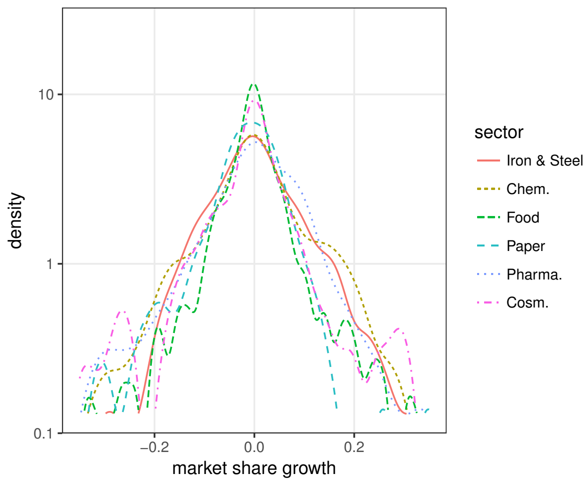
- Group 1: Iron & Steel, Transportation Equipment, Chemical, Food, and Paper. The variance of the growth ε becomes larger when market concentration is high.
- Group 2: General Machinery, Precision Machinery, and Electrical Machinery. The variance is less dependent on market concentration.[17]
- Group 3: Pharmaceutical and Cosmetics. The variance becomes larger when market concentration is low.

While these two market indexes are important to describe the status of market, we cannot find a simple relationship between these two indexes in our analysis. The inter-sectoral heterogeneity of the relation between is quite large: Group 1 shows a positive relationship and Group 3 shows a negative relationship. In contrast, as shown in Figure 6, in which kernel density estimation of the two subgroups are plotted, we can observe tent-shaped distributions in both subgroups (i.e., high and low concentrations), similar to the ones in Figure 5. In this sense, the tent-shaped distribution is a rather robust feature of market share dynamics. Specifically, Figure 6(c) shows the kernel density estimates of sectors in Group 2 and suggests that the shape of the distribution is insensitive with respect to market concentration. Focusing on the sectors in Group 2 and extending our analysis to multivariate cases, we further explore the distributional properties of market share growth.

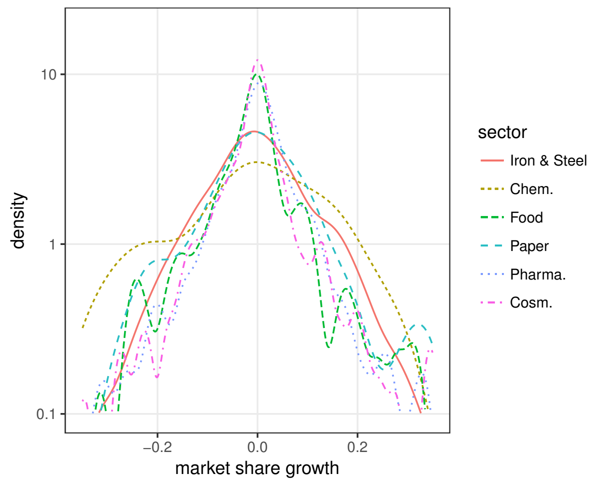
3.3 Who competes against whom?

We have so far analyzed the univariate case (i.e., market share growth of the top 2 firms). In this section, we extend our analysis to a multivariate case, that is, market share dynamics of the top 5, 6, and 7, and then explore distributional properties of $\boldsymbol{\varepsilon}_t := \mathbf{x}_t \ominus \mathbf{x}_{t-1}$. By increasing the dimension, we can consider more complex relation among firms. For example, consider the following question: Is an increase in a firm's share offset by a decrease in the share of another *particular* firm? If two particular firms compete for market share against each other, market share change would occur within the two firms keeping other firms' shares unchanged. To explore such relation among

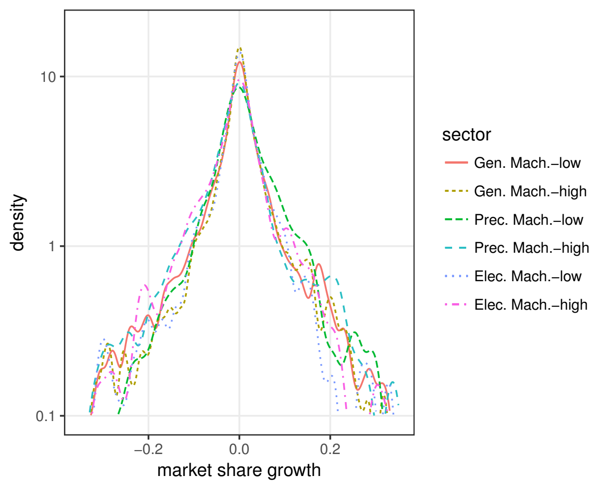
[17] While the tests for general machinery show the statistically significant difference of the two variances by the large number of observations, the point estimates show that the difference is relatively small. Thus, general machinery is classified into group 2. Precision machinery is classified into group 2 because its robust estimates of the ratio (.875) is relatively close to 1.



(a)



(b)



(c)

Figure 6: Kernel density estimations. In (a), kernel density estimations of market share growth with low concentration in Group 1 and 3 are plotted. In (b) kernel density estimations of market share growth with high concentration in Group 1 and 3 are plotted. In (c), kernel density estimations of market share growth of both subgroups in Group 2 are plotted. Estimation method is the same as in Figure 4.

Industry	Ratio of var.	Robust	F-test	Levene	F-K	Median(%)
Iron & Steel	.678	.620	.00108	.0221	.0170	54.8
General Machinery	1.141	1.412	.00635	.0177	.00435	44.2
Transportation Equip.	.579	.527	.00256	.0247	.0282	69.9
Precision Machinery	.667	.875	.000299	.0952	.251	62.1
Electrical Machinery	.831	.579	.0550	.0223	.00344	40.2
Chemical	.327	.191	< 2.2e-16	4.40e-08	5.63e-10	57.5
Food	.695	.578	.000202	.0151	.0180	51.6
Paper	.475	.329	1.67e-05	.00221	.00138	47.4
Pharmaceutical	1.474	2.88	7.46e-06	3.35e-05	7.86e-07	36.8
Cosmetics	1.402	2.24	.00862	.0928	.122	63.4

Table 4: Ratio of the two variances and tests for equality. The second column shows the ratio of variances (variance for low concentration markets divided by variance for high concentration markets) 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.

firms, we first perform compositional PCA and cluster analysis to $\boldsymbol{\varepsilon}$ of the top 5 firms. Based on these results, we examine the marginal distribution of $\boldsymbol{\varepsilon}$. Since these methods implicitly assume homogeneity of covariance structure, we focus on sectors in group 2 based on the results in the previous section: general machinery, precision machinery, and electrical machinery.[18] Descriptive statistics are given in Table 5. As in the previous section, the mean of $\boldsymbol{\varepsilon}$ is very close to the neutral element $\boldsymbol{\lambda}_D := (\frac{1}{D}, \frac{1}{D}, ..., \frac{1}{D})$. On average, the relative market position has no information about market share growth in the next year. In other words, in terms of market share, a market leader has no advantages/disadvantages.

Next, we perform compositional PCA, which is done as follows: We first transform $\boldsymbol{\varepsilon}$ in $\mathcal{S}^5$ to $\text{ilr}(\boldsymbol{\varepsilon})$ in $\mathbb{R}^4$. Next, we apply PCA to $\text{ilr}(\boldsymbol{\varepsilon})$, that is, we estimate the location vector and covariance matrix of $\text{ilr}(\boldsymbol{\varepsilon})$. Finally, we apply singular value decomposition.[19] In this analysis, we focus on two principal components (PCs). Following the convention in the CDA literature, the PCs are inversely transformed to the clr representation for interpretation.

Results are shown in Figure 7. Arrows in these figures represent clr coordinates of the two PCs.[20] Figure 7 clearly shows that links between ε_1 and ε_2 (and ε_3) are short, indicating that the

[18]Another reason of this choice is the number of observations in each sector.

[19]We use a robust method developed by Filzmoser et al. (2009).

[20]If the clr coordinates of the two PCs are represented as $(a_1, a_2, ..., a_D)$ and $(b_1, b_2, ..., b_D)$, the coordinate of the

	# markets	# obs.	Mean, 1st	2nd	3rd	4th	5th	6th	7th
Top 5 firms									
Gen. Mach.		1843	0.201	0.200	0.199	0.200	0.200		
Pre. Mach.		102	0.200	0.199	0.201	0.201	0.199		
Elec. Mach.		656	0.199	0.199	0.200	0.201	0.201		
Top 6 firms									
Gen. Mach.		1052	0.167	0.166	0.165	0.167	0.166	0.170	
Elec. Mach.		431	0.165	0.166	0.167	0.167	0.168	0.166	
Top 7 firms									
Gen. Mach.		483	0.143	0.143	0.141	0.143	0.142	0.145	0.144
Elec. Mach.		212	0.141	0.142	0.143	0.143	0.147	0.141	0.143

Table 5: Descriptive statistics. The first column the number of markets. The second column the number of pooled market share growth. The rest of columns refers to the sample mean represented in $\mathcal{S}^D$.

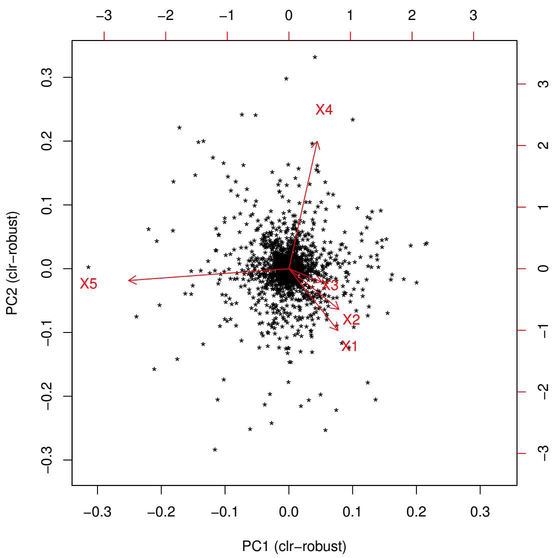
shares of firms ranked 1 and 2 (and 3) move up and down together. Namely, when the largest firm succeeds to increase its market share, the market share of the second largest firm is likely to increase and their increases are offset by decrease of the market shares of lower-ranked firms. Put differently, the subgroup of the 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 8.[21] Figure 8 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 9 and 10, we perform the same analysis with additional firms. Figure 9 shows the compositional PCA and cluster analysis to $\boldsymbol{\varepsilon} := (\varepsilon_1, \varepsilon_2, ..., \varepsilon_6)$ of the top 6 firms for general machinery and electric machinery.Figure 10 is the results of the top 7 firms for the same sectors. Both figures show that the shares of the top 3 (or 4) firms are likely to move up and down together. 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.

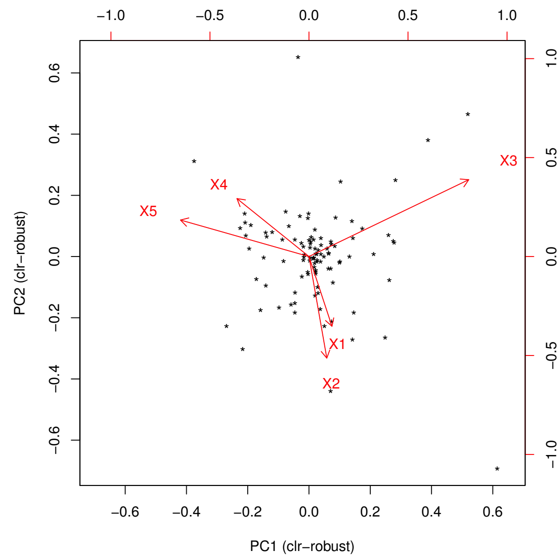
Next, we examine the distribution of $\boldsymbol{\varepsilon}$ based on this finding. While the distribution of $\boldsymbol{\varepsilon}$ in $\mathcal{S}^D$ can be, in principle, expressed by the corresponding distribution in $\mathbb{R}^{D-1}$, its density estimation becomes practically difficult as the number of dimension increases: Especially when the number of

head of arrow $X1$ in Figure 7 is (a_1, b_1).

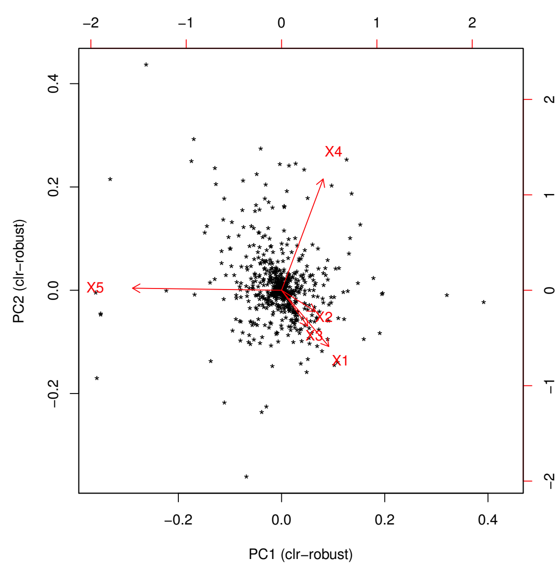
[21] Here, the distance between two components is measured by $\operatorname{var}\log(\frac{\varepsilon_i}{\varepsilon_j})$, based on which the components are clustered. For detail, see ...



(a) General Machinery.

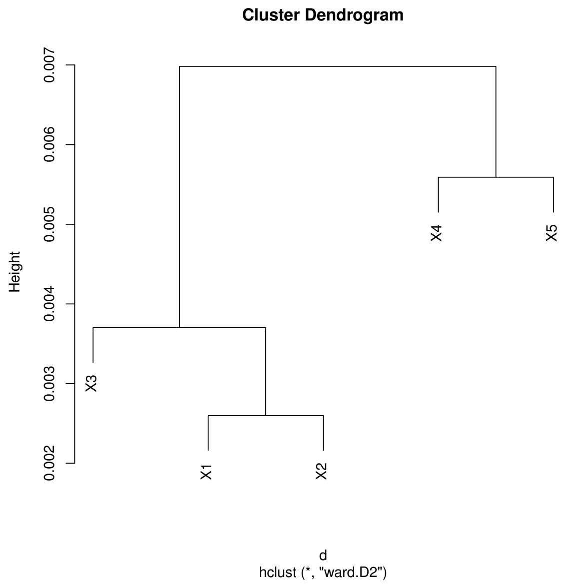


(b) Precision Machinery.

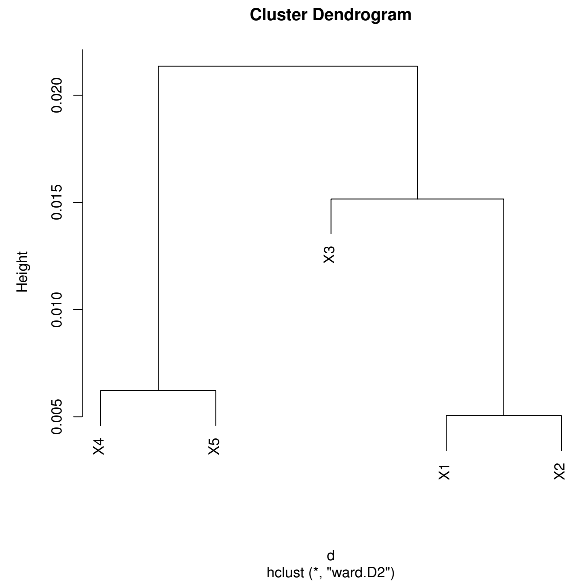


(c) Electrical Machinery.

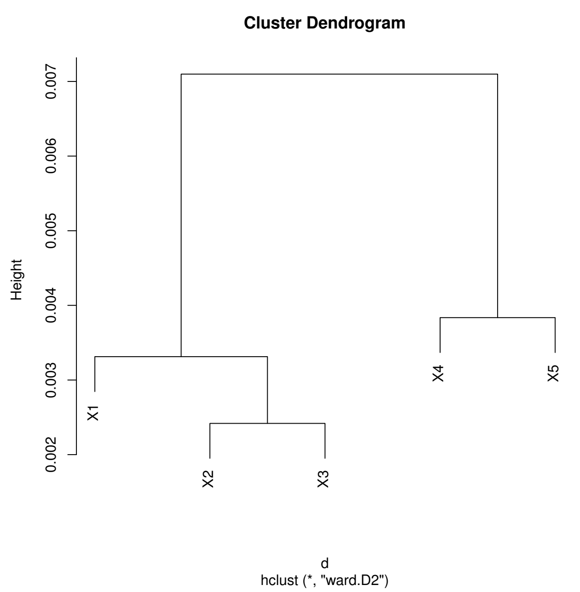
Figure 7: Compositional principal component analysis. The proportions of variance explained by the two PCs are (a) 65.5%, (b) 71.2%, (c) 69.2%, and



(a) General Machinery.

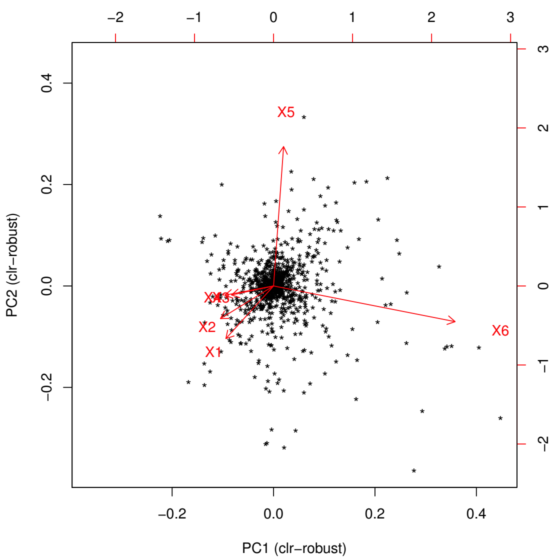


(b) Precision Machinery.

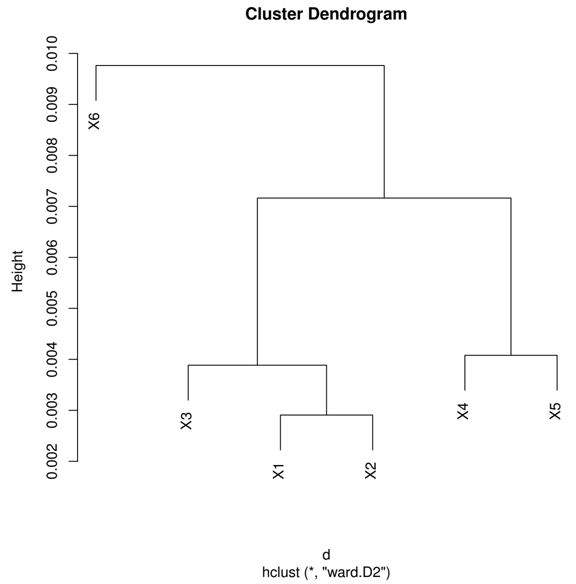


(c) Electrical Machinery.

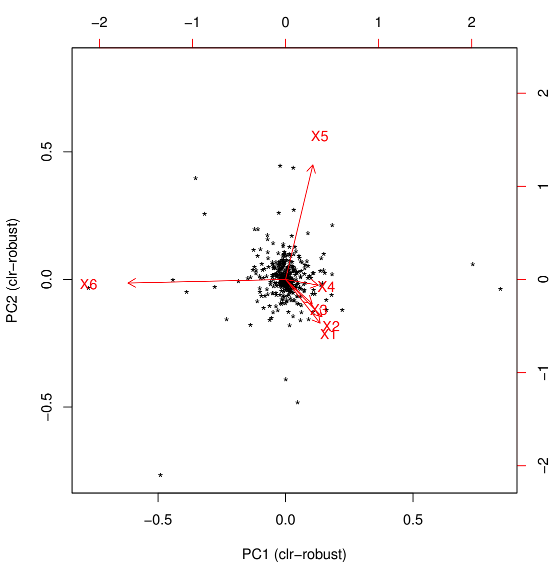
Figure 8: Cluster analysis.



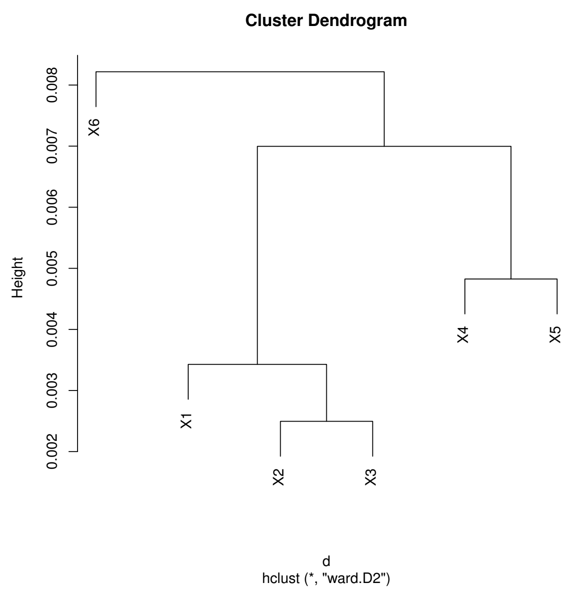
(a.1) Compositional PCA. Genaral Machinery.



(a.2) Cluster analysis. Genaral Machinery.

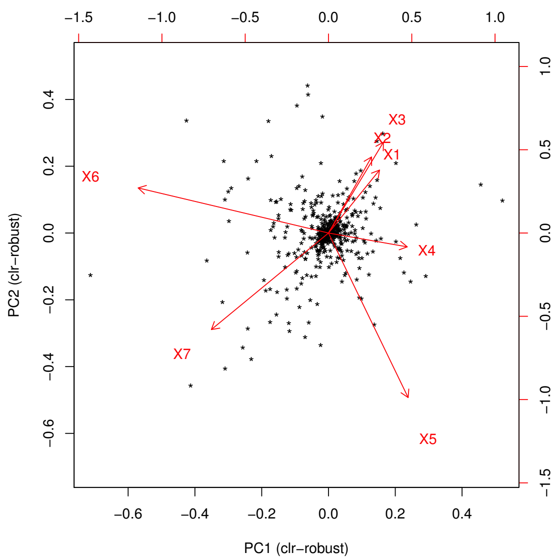


(b.1) Compositional PCA. Electric Machinery.

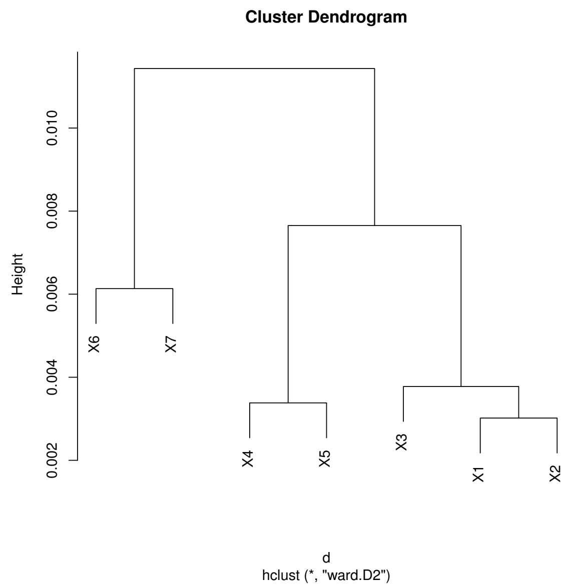


(b.2) Cluster analysis. Electric Machinery.

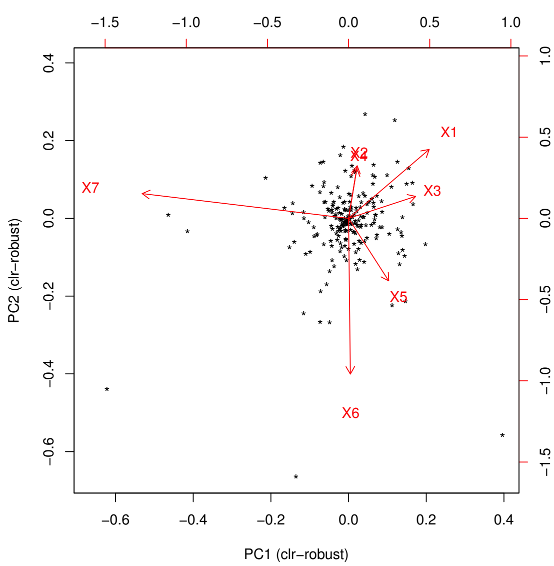
Figure 9: Compositional PCA and cluster analysis. The number of firms is six. The proportions of variance explained by the two PCs are (a.1) 58.0%, (b.1) 66.0%.



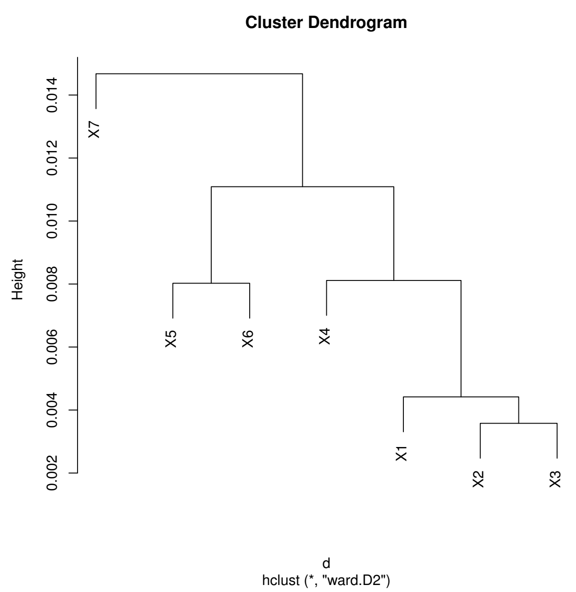
(a.1) Compositional PCA. Genaral Machinery.



(a.2) Cluster analysis. Genaral Machinery.



(b.1) Compositional PCA. Electric Machinery.



(b.2) Cluster analysis. Electric Machinery.

Figure 10: 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%.

observations is not so large, the density estimation becomes unreliable. Thus, in the remainder of this section, we focus on marginal distributions. As in the usual case in multivariate analysis, there are an infinite number of ways of choosing an orthonormal basis represented in $\mathbb{R}^D$. Taking into account the finding above, we consider an orthonormal basis $\{e_i\}_{1 \le i \le D-1}$ whose two elements are given as follows:

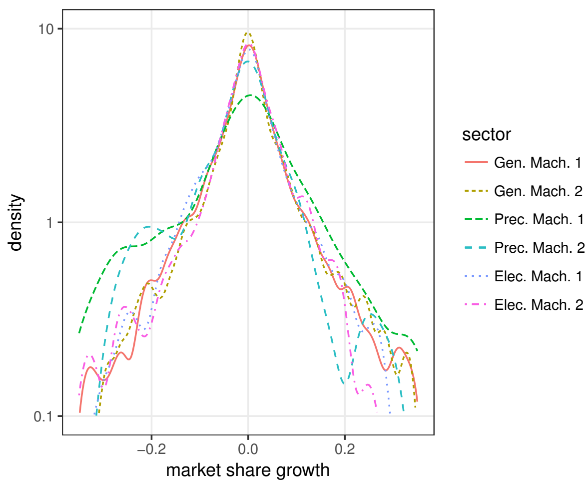
$$e_1 = \sqrt{\frac{D-1}{D}}\Big(\overbrace{\frac{1}{D-1}, \frac{1}{D-1}, ..., \frac{1}{D-1}}^{D-1}, -1\Big), \quad e_2 = \sqrt{\frac{D-2}{D-1}}\Big(\overbrace{\frac{1}{D-2}, \frac{1}{D-2}, ..., \frac{1}{D-2}}^{D-2}, -1, 0\Big)$$

Its coordinate represented by this basis is $y_1 = \sqrt{\frac{D-1}{D}} \log \frac{(\varepsilon_1 \varepsilon_2 ... \varepsilon_{D-1})^{\frac{1}{D-1}}}{\varepsilon_D}$ and $y_2 = \sqrt{\frac{D-2}{D-1}} \log \frac{(\varepsilon_1 \varepsilon_2 ... \varepsilon_{D-2})^{\frac{1}{D-2}}}{\varepsilon_{D-1}}$. Since these coordinate is the logratio of the geometric mean of the top subgroup to a lower-rankd firm, y_1 and y_2 capture the most variable part of $\boldsymbol{\varepsilon}$. The distributional properties of y_1 and y_2 are shown in Figure 11, where we can observe a tent-shaped distribution similar to the ones found in the previous section. MLEs of the parameters of Subbotin family are reported in Table 6, which suggests that the parameter b is close to or smaller than 1. This implies that market share changes between the top subgroup and lower-ranked firms are characterized as in the case of the top 2 firms discussed in the previous section: Episodes of lumpy *jumps* are relatively frequent.

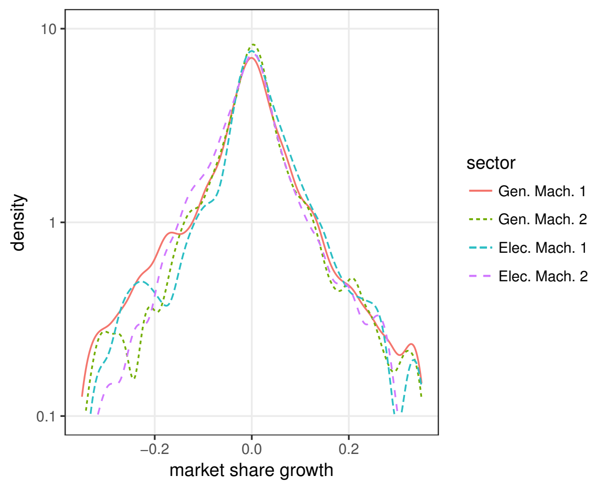
4 Conclusions

As previous studies have shown, statistical regularities are an important clue to further understanding of industrial dynamics. Market share representing relative market position is also an important variable 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 method called CDA to market share data and explores its statistical properties. 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 analysis of market share dynamics.

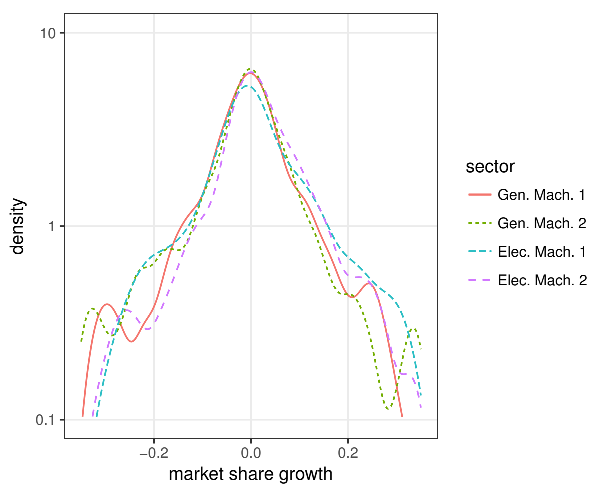
We have found that the distribution of market share growth displays a remarkable feature: the distribution does not follow a Gaussian but a tent-shaped distribution with a fatter tail, which is



(a) The top 5 firms.



(b) The top 6 firms



(c) The top 7 firms.

Figure 11: Marginal distributions. "Gen. Mach. 1" stands for the marginal distribution of y_1 in General machinery.

	obs.	mean	s.d.	a	b
Top 5 firms					
Gen. Mach.(4,1)	1843	-0.00083	0.148	0.0551 (0.00196)	0.6766 (0.03107)
Gen. Mach.(3,1)	1843	0.00125	0.137	0.0517 (0.00181)	0.6483 (0.02738)
Prec. Mach.(4,1)	102	0.00743	0.209	0.0896 (0.01277)	0.8971 (0.19305)
Prec. Mach.(3,1)	102	-0.00187	0.166	0.0658 (0.009)	0.8562 (0.16032)
Elec. Mach.(4,1)	656	-0.00728	0.135	0.0568 (0.00319)	0.7667 (0.05805)
Elec. Mach.(3,1)	656	-0.00816	0.115	0.0517 (0.0029)	0.7248 (0.05139)
Top 6 firms					
Gen. Mach.(5,1)	1052	-0.02173	0.188	0.0647 (0.00297)	0.7254 (0.04385)
Gen. Mach.(4,1)	1052	0.00322	0.146	0.0539 (0.00253)	0.6585 (0.0386)
Elec. Mach.(5,1)	429	0.00345	0.149	0.0571 (0.00401)	0.7105 (0.06245)
Elec. Mach.(4,1)	429	-0.01017	0.133	0.0583 (0.00399)	0.7957 (0.07458)
Top 7 firms					
Gen. Mach.(6,1)	483	-0.00934	0.157	0.0677 (0.00428)	0.8463 (0.07408)
Gen. Mach.(5,1)	483	-0.01512	0.18	0.0671 (0.00441)	0.7578 (0.06535)
Elec. Mach.(6,1)	212	-0.00039	0.17	0.0807 (0.00749)	0.9534 (0.13309)
Elec. Mach.(5,1)	212	0.01665	0.202	0.0674 (0.00633)	0.8697 (0.11321)

Table 6: MLEs of marginal distributions. For explanation, see Table 2.

closely related with the findings of firm growth rate distribution. This shape of the distribution implies that market share growth cannot be described by an accumulation of small shocks: Rather, lumpy jumps are crucial in market share dynamics. Put differently, it suggests that radical change in market structure is relatively frequent. Interestingly, with some exceptions, this statistical feature can be observed across different sectors.

We extend our analysis to the multivariate case. The analysis of the top 5, 6, and 7 firms shows that there is a particular relation among firms: The main part of the total variation of market share growth is explained by one between the subgroup of the top firms and the lower-ranked firms. Seeing the marginal distribution describing this relation, we have found that a tent-shaped distribution similar to the case of the top 2 firms emerges. As in the analysis of the top 2 firms, market share change between the top subgroup and lower-ranked firms is also characterized by a lumpy type behavior.

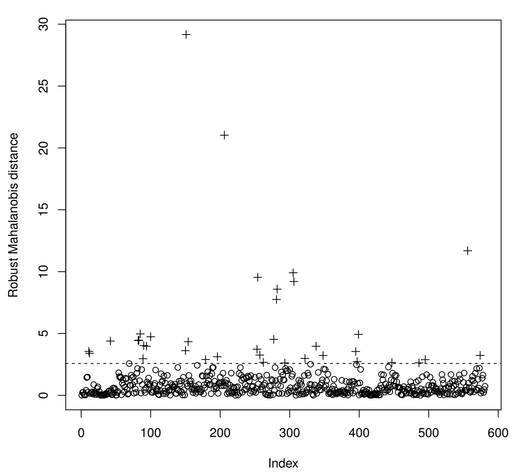
Our analysis implies that the distribution of market share growth has a remarkable feature and a drastic transformation of market structure is rather frequent. Such implications based on statistical properties of observed data help us further investigate industrial dynamics theoretically. This paper has added a new finding to this literature.

Appendix

A Outliers

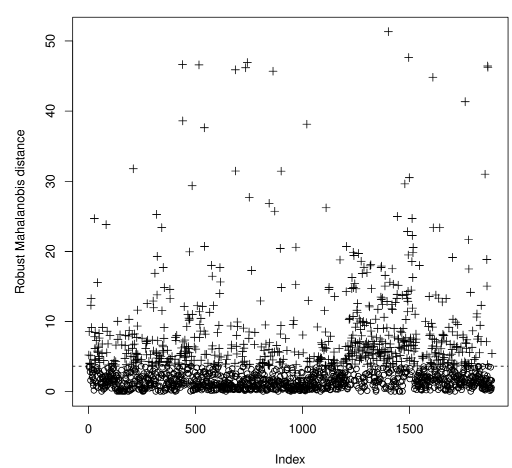
For detecting outliers in our samples, we follow the approach developed by Filzmoser and Hron (2008). In this approach, the *Mahalanobis distance* (MD) based on the *Minimum Covariance Determinant* (MCD) estimates for location and covariance matrix are used as a criteria for outliers. Since the squared MD follows the χ^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 **??**) 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.



(a)

Iron & Steel for $D = 2$.



(b)

General Machinery for $D = 5$.

Figure 12: Outlier detection.

References

Aitchison, J., 1986. The Statistical Analysis of Compositional Data. Monographs on Statistics and Applied Probability. Chapman and Hall Ltd.

Arata, Y., 2014. Are there any empirical regularities of firms' growth patterns? The importance of large jumps. RIETI Discussion Paper 14-E-033.

Axtell, R. L., 2001. Zipf distribution of u.s. firm sizes. Science 293 (5536), 1818–1820.

Bottazzi, G., Cefis, E., Dosi, G., 2002. Corporate growth and industrial structures: Some evidence from the Italian manufacturing industry. Industrial and Corporate Change 11 (4), 705–723.

Bottazzi, G., Cefis, E., Dosi, G., Secchi, A., 2007. Invariances and Diversities in the Patterns of Industrial Evolution: Some Evidence from Italian Manufacturing Industries. Small Business Economics 29 (1-2), 137–159.

Bottazzi, G., Coad, A., Jacoby, N., Secchi, A., 2011. Corporate growth and industrial dynamics: Evidence from French manufacturing. Applied Economics 43 (1), 103–116.

Bottazzi, G., Secchi, A., 2006. Explaining the distribution of firm growth rates. The RAND Journal of Economics 37 (2), 235–256.

Chayes, F., 1960. On correlation between variables of constant sum. Journal of Geophysical Research 65 (12), 4185–4193.

Coad, A., 2009. The growth of firms; A survey of theories and empirical evidence. Cheltenham: Edward Elgar Publishing.

Coad, A., Teruel, M., 2012. Inter-firm rivalry and firm growth: Is there any evidence of direct competition between firms? Industrial and Corporate Change.

Davies, S. W., Geroski, P. A., 1997. Changes in concentration, turbulence, and the dynamics of market shares. Review of Economics and Statistics 79 (3), 383–391.

Dosi, G., Grazzi, M., Marengo, L., Settepanella, S., 2016. Production theory: Accounting for firm heterogeneity and technical change. The Journal of Industrial Economics 64 (4), 875–907.

Dosi, G., Pereira, M. C., Virgillito, M. E., 2017. The footprint of evolutionary processes of learning and selection upon the statistical properties of industrial dynamics. Industrial and Corporate Change 26 (2), 187–210.

Egozcue, J. J., Pawlowsky-Glahn, V., Mateu-Figueras, G., Barceló-Vidal, C., 2003. Isometric logratio transformations for compositional data analysis. Mathematical Geology 35 (3), 279–300.

Filzmoser, P., Hron, K., 2008. Outlier detection for compositional data using robust methods. Mathematical Geosciences 40 (3), 233–248.

Filzmoser, P., Hron, K., Reimann, C., 2009. Principal component analysis for compositional data with outliers. Environmetrics 20 (6), 621–632.

Fry, J. M., Fry, T. R., McLaren, K. R., 1996. The stochastic specification of demand share equations: Restricting budget shares to the unit simplex. Journal of Econometrics 73 (2), 377–385.

Fry, J. M., Fry, T. R., McLaren, K. R., 2000. Compositional data analysis and zeros in micro data. Applied Economics 32 (8), 953–959.

Gabaix, X., 2009. Power laws in economics and finance. Annual Review of Economics 1, 255–294.

Geroski, P., Toker, S., 1996. The turnover of market leaders in UK manufacturing industry, 1979-86. International Journal of Industrial Organization 14 (2), 141–158.

Kaplow, L., 2015. Market definition, market power. International Journal of Industrial Organization 43, 148–161.

Kato, M., Honjo, Y., 2006. Market share instability and the dynamics of competition: A panel data analysis of Japanese manufacturing industries. Review of Industrial Organization 28 (2), 165–182.

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

Mazzucato, M., 2000. Firm size, innovation, and market structure: The evolution of industry concentration and instability. Edward Elgar Publishing.

Mazzucato, M., 2002. The pc industry: new economy or early life-cycle? Review of Economic Dynamics 5 (2), 318–345.

Mazzucato, M., Parris, S., 2015. High-growth firms in changing competitive environments: the us pharmaceutical industry (1963 to 2002). Small Business Economics 44 (1), 145–170.

Pawlowsky-Glahn, V., Buccianti, A., 2011. Compositional Data Analysis: Theory and Applications. John Wiley & Sons.

Pawlowsky-Glahn, V., Egozcue, J. J., Tolosana-Delgado, R., 2015. Modeling and analysis of compositional 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 of the Royal Society of London LX, 489–502.

Scott, D. W., 1992. Multivariate density estimation: theory, practice, and visualization. John Wiley & Sons.

Stanley, M. H., Amaral, L. A., Buldyrev, S. V., Havlin, S., et al., 1996. Scaling behaviour in the growth of companies. Nature 379 (6568), 804–806.

Sutton, J., 2007. Market Share Dynamics and the “Persistence of Leadership” Debate. The American Economic Review 97 (1), 222–241.