

PAPER • OPEN ACCESS

Time evolution of companies towards a stable scaling curve obtained from flow diagrams in three-dimensional phase space

To cite this article: Yuh Kobayashi *et al* 2019 *New J. Phys.* **21** 043038

View the [article online](#) for updates and enhancements.

**PAPER**

Time evolution of companies towards a stable scaling curve obtained from flow diagrams in three-dimensional phase space

OPEN ACCESS**RECEIVED**

23 January 2019

REVISED

27 February 2019

ACCEPTED FOR PUBLICATION

8 March 2019

PUBLISHED

23 April 2019

Original content from this work may be used under the terms of the [Creative Commons Attribution 3.0 licence](https://creativecommons.org/licenses/by/4.0/).

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

Yuh Kobayashi¹, Hideki Takayasu^{2,3}, Shlomo Havlin^{3,4} and Misako Takayasu^{1,3} ¹ Department of Mathematical and Computing Science, School of Computing, Tokyo Institute of Technology, Yokohama 226-8502, Japan² Sony Computer Science Laboratories, Tokyo 141-0022, Japan³ Institute of Innovative Research, Tokyo Institute of Technology, Yokohama 226-8502, Japan⁴ Department of Physics, Bar-Ilan University, Ramat-Gan 52900, IsraelE-mail: takayasu.m.aa@m.titech.ac.jp**Keywords:** flow diagram, complex systems, scaling relation, growth rate, attractor, business firm, econophysics**Abstract**

Although scaling relations of power law forms have been revealed in a wide variety of complex systems, the origin of scaling relations in real systems remains mostly unsolved. Based on a long tradition in physics, we here explore the phase space dynamics of business companies, whose different measures of size have been found to exhibit multiple scaling relations. Using a large scale dataset of Japanese companies covering over two million companies for over two decades (1994–2015), we compile the data of three measures of size, namely, annual sales, number of employee and number of trading partners. Tracking the historical time evolution of companies, we first show that there exists a stable region that attracts most of the data points in the long time scale, although the company dynamics in the three-dimensional phase space looks random in a shorter time scale. We then elaborate on ‘evolving flow diagrams’ of the averaged motion in the 3D space. Remarkably, the flow diagrams indicate that a 3D curve which represents the scaling relations between the three size measures can be regarded as an approximate attractor of the averaged flows. The results could serve for better modeling and understanding of companies dynamics and our method can be applied to other dynamics such as social and biological phenomena.

1. Introduction

Scaling relations and scaling laws with non-trivial exponents have been central issues in physics of complex phenomena. Since the non-trivial scaling relation around a critical point was first discovered over a century ago [1], the scaling concept had guided the exploration into critical phenomena [2–7]. It was subsequently applied to soft matter physics and led to the discovery of universal features of polymers [8, 9] and colloidal aggregates [10]. Meanwhile, scaling relations had also been found for complex systems such as biodiversity in ecological communities [11, 12], animal bodies and cells [13–16], human cities [17–19] and business companies [20–25]. Recent physical and mathematical studies have attempted to explain scaling relations in such systems and provided plausible theories for animal bodies or cells [26, 27] and for human cities [28, 29]. All these studies have employed statistical mechanical approach, where macroscopic laws and patterns are explained by assuming approximate laws and dynamics at more fundamental levels.

However, there remains a significant challenge in statistical physics approach to complex phenomena, because the dynamics at fundamental levels has not been sufficiently established for biological, social or economic systems. We assume that this lack of established laws at microscopic levels has led to a plethora of models for such systems. On business companies, for example, models have highlighted several mechanisms or factors as diverse as hierarchical organization [30], stochasticity in competition [31, 32], financial [33] or hiring/firing [34] behaviors, preferential attachment of company units [35], social networks [36] and multiple independent components in companies [35, 37–40]. Although they often agree with the data at the level of distributions calculated from a snapshot of the systems, these models lack validation by direct observation at

microscopic levels, or fail when their microscopic dynamics is explored [41]. The situation is similar in biodiversity research: two classes of models with radically different assumptions on dynamics at microscopic levels both can explain the rank-size distribution of abundance [42, 43] and spatial distributions [44, 45] of natural species.

During the development of statistical mechanics, there emerged an alternative approach, in which one explores laws and dynamics only in terms of the ‘macroscopic’ variables that allow direct observation. More specifically, the evolution of complex systems in the phase space of ‘macroscopic’ variables is studied. This approach dates back at least to Lorenz [46, 47], who is also famous for the pioneering work [48] in chaos theory. He proposed the analogue method: trajectories of systems with unknown dynamics could be predicted by collecting the data from the past with configurations similar to the present. By applying this method to the world trade, Cristelli *et al* [49] recently demonstrated that countries and territories were in either turbulent or laminar zones in the phase space of their GDP and ‘fitness’ measure in the trading network. This suggests that the methodology in this line could be vital to studies of complex phenomena. We consider this approach to be complementary to the statistical mechanical approach. Although observed scaling properties in complex systems may be explained by the statistical physics models enumerated above, explanation of such properties by the dynamics at macroscopic levels would be valuable: the characterized dynamics could be used to justify or falsify the possible theories on dynamics at microscopic levels. Nevertheless, such an approach to the scaling relations has not been tested yet to the best of our knowledge.

Here we empirically observe how companies evolve into the previously known scaling relations in a three-dimensional phase space. Instead of the statistical mechanical approach, we introduce the macro-variables approach. To this aim, we utilize a comprehensive dataset of approximately 2 million Japanese companies accumulated over more than 20 years since 1994. We compile the data of three measures of size from the dataset: (i) monetary size estimated typically by the annual sales, (ii) labor size measured by the number of employees, and (iii) the size of transactional activity which can be characterized by the number of companies with which the focal company has a direct trading relationship. These measures of size were previously studied in the context of scaling relations with scale-invariant fluctuations [24]. Based on this huge dataset, several pieces of evidence are presented which indicate that companies tend to evolve into a single three-dimensional ‘scaling curve’, on which companies are ‘stable’ in the sense that their average velocity is close to zero.

The rest of the paper is organized as follows. First, we briefly discuss our compilation of the data. Second, we show that the previously reported multiple two-dimensional scaling relations can be represented by a curve in the three-dimensional phase space. We then inspect the time evolution of companies testing and disproving the hypothesis that it is an anisotropic random walk with constant diffusion coefficients. We further derive evolving flow diagrams from the data to demonstrate that companies move towards the scaling curve on average. Lastly, we discuss the implications of our results and future research directions.

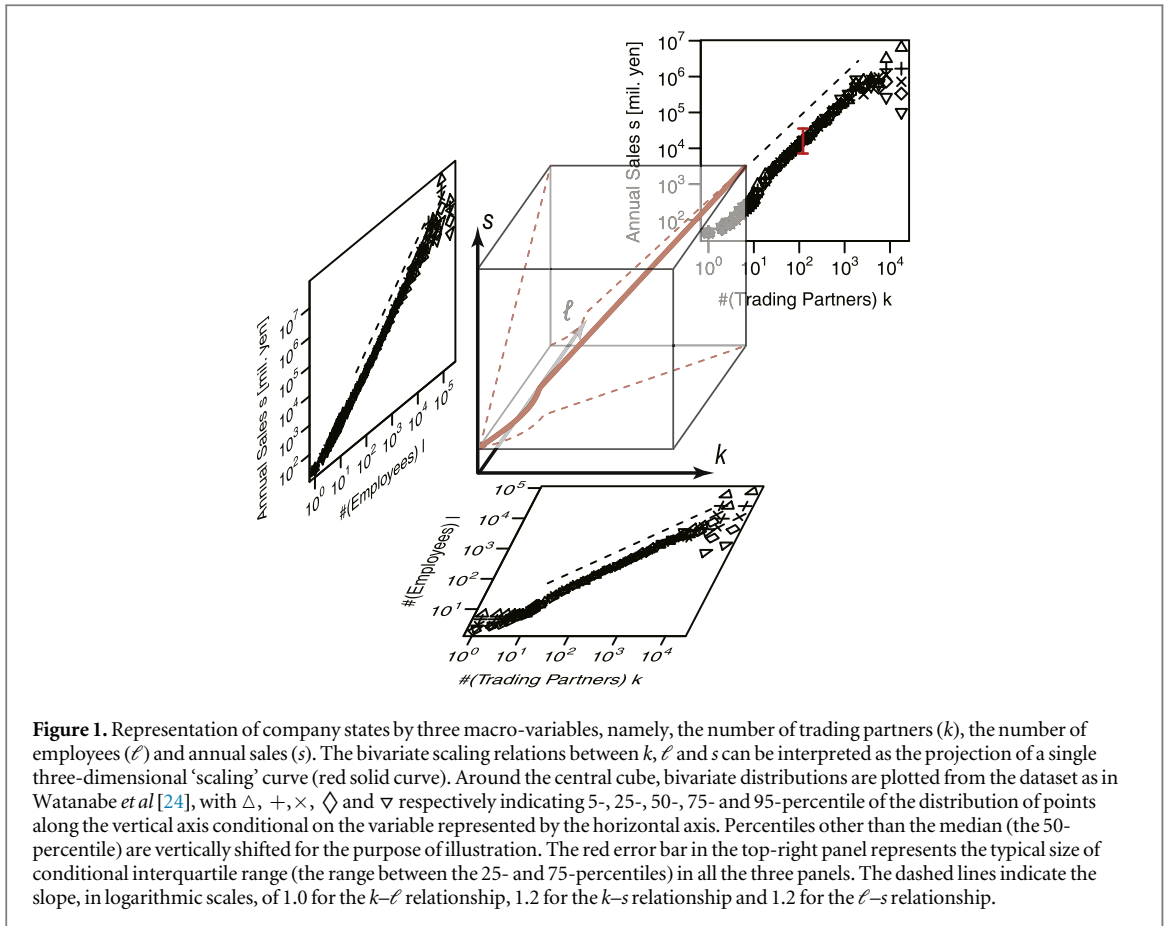
2. Data compilation

We compile the data utilized here from an exhaustive dataset by a major credit reporter that summarizes the description of Japanese companies in the period of 1994–2015 (COSMOS 2 by Teikoku Databank, Ltd.), available to us in January, 2017. The dataset contains total of 2.415×10^6 companies (1.263×10^6 in yearly average). We filtered out a small fraction of financial companies and governmental organizations whose sales are defined very differently from other ordinary companies. We also removed a very small number of sales data which were recorded more than 8 years after the publication of the financial statement. Additionally, we excluded the sales data where the end of fiscal year changes, because some of them are not likely to be annual sales. Therefore, our final dataset, which consists of totally 2.395×10^6 companies (1.247×10^6 in yearly average), primarily concerns manufacturing, construction or wholesale companies. Here, we have the data of annual sales, s , and the number of employees, ℓ .

We also construct a trading network for every year from the list of trading relationships between companies, to obtain the degree k (i.e. the number of direct trading partners) of every company. The network includes 3.051×10^6 trading links per year on average. This enables us to consider the companies’ size in terms of transactional activity in the network of trading partnership [24]. We note that data of the trading network is almost monotonically growing with years, from 1.826×10^6 trading links in 1994 to 3.873×10^6 in 2015.

3. Three-dimensional scaling curve

We first present a unified picture for the previously reported triple scaling relations between pairs of variables in the three-dimensional phase space of (k, ℓ, s) . Watanabe *et al* [24] demonstrated the triple scaling relations between the number of trading partners, k , the number of employees, ℓ , and the annual sales, s , for middle- and



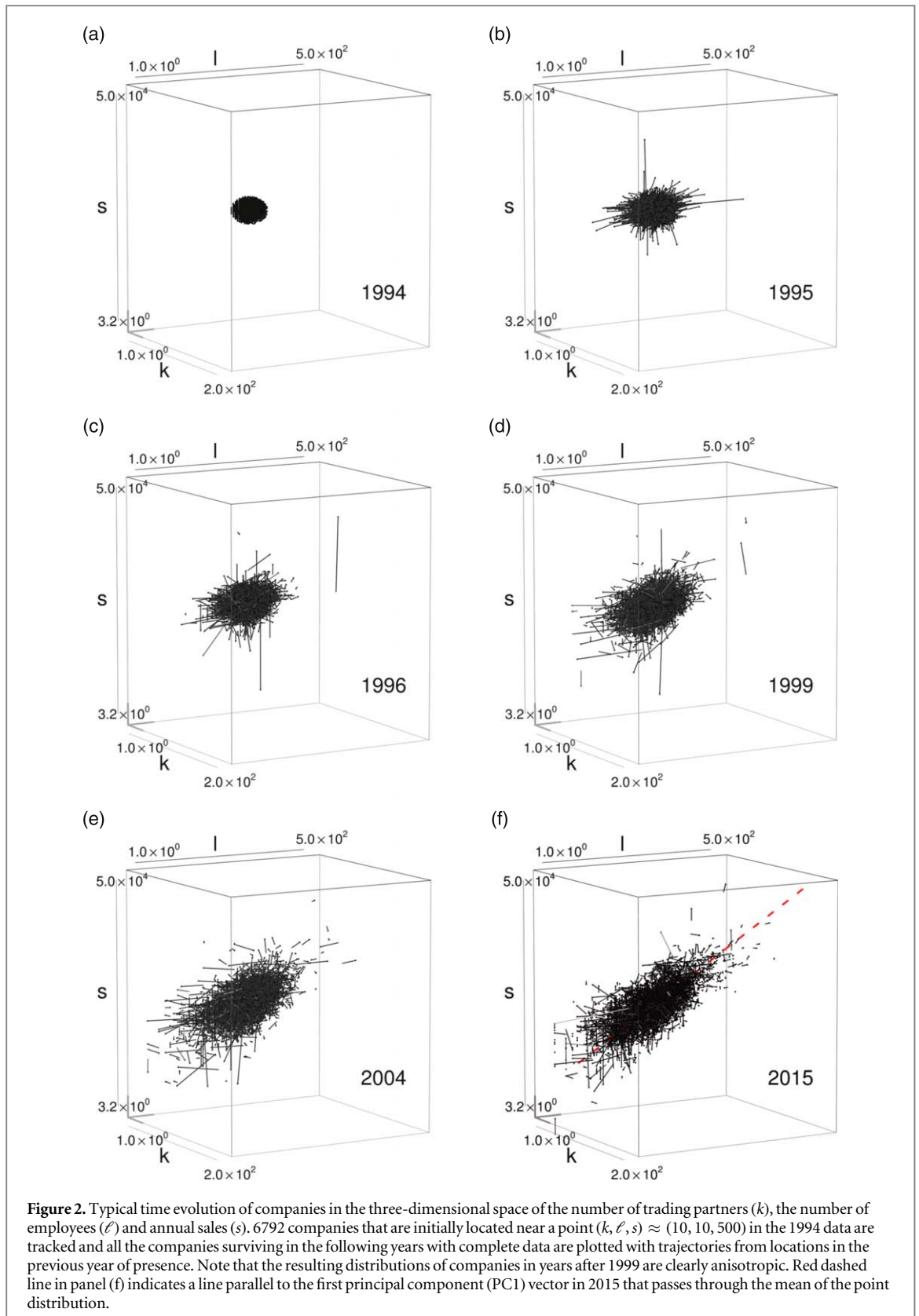
large-sized companies: $\ell \propto k^{1.0}$, $s \propto k^{1.3}$ and $s \propto \ell^{1.3}$. These relations were defined in terms of conditional medians. For example, the equation, $s \propto \ell^{1.3}$, means that the median s for companies with ℓ employees increases as $\ell^{1.3}$ when ℓ is increased. Furthermore, fluctuations around the conditional medians were found to be scale-invariant: the probability of having, for example, double or half the conditional median sales is independent of k and ℓ . In other words, the distributions of $\ell/k^{1.0}$, $s/k^{1.3}$ and $s/\ell^{1.3}$ are independent of the size of companies, and these variables are, in this sense, intensive. Such scaling relations fit our dataset well for all 22 years with slightly different scaling exponents of 1.2 for k - s and ℓ - s scaling (figure 1). This means that when the three-dimensional distribution of companies is projected to a two-dimensional plane by neglecting one of the variables, companies are densely located on a line defined by the scaling relations. This highly suggests that the three-dimensional distribution of companies can be represented by a single curve in the three-dimensional phase space (red line in figure 1), near which the data points are densely located. We hereafter refer to this hypothetical curve as the ‘scaling curve’.

While the scaling curve is linear in logarithmic scales for middle- and large-sized companies, this does not hold for small-sized companies. Indeed, in the two-dimensional plots of conditional quantiles (figure 1), the scaling exponents are evidently different at the zone of small-sized companies and that of middle- or large-sized ones. Therefore, we hereafter refer to these two different regimes of scaling as the small-company and large-company regimes, considering them separately.

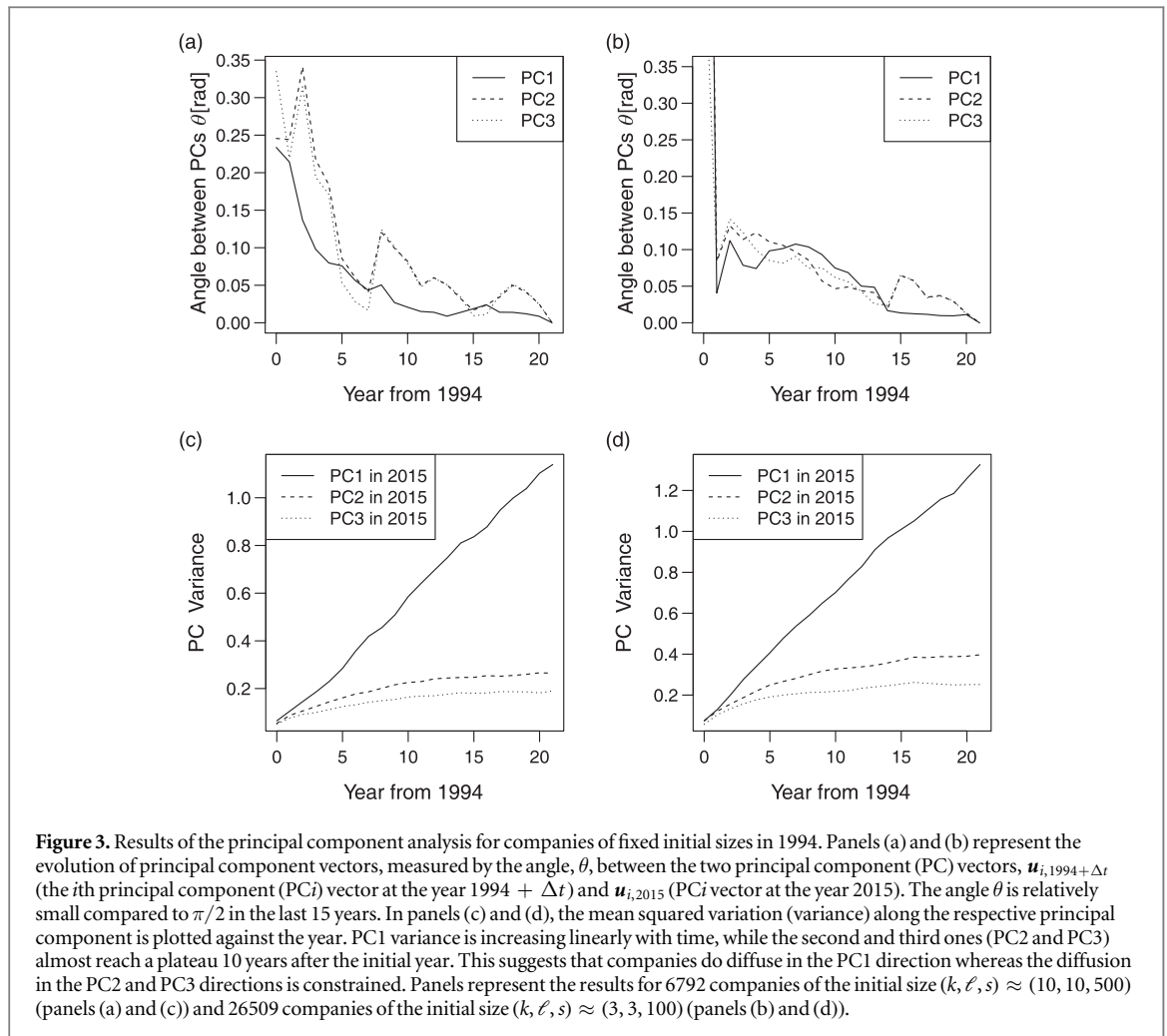
4. Alignment of companies towards a curve

Here we first visually inspect the data to obtain an intuitive view of the three-dimensional dynamics of companies. As in the previous section, we regard a company as a point in the three-dimensional space of the log-transformed measures of size (K, L, S) \equiv ($\log k, \log \ell$ and $\log s$). Historical changes of size measures of companies, initially near a point in the space, are demonstrated in figure 2. Historical changes of total of 6792 companies near $(k, \ell, s) = (10, 10, 500)$ are tracked to show the representative evolution in the phase space (see figure 2). It is evident from the figure that companies are not evolving isotropically in this 3D representation. Rather, there seems to exist a curve along which companies are most likely to move.

Applying principal component analysis (PCA) on the data can further reveal whether the evolving dynamics of companies can be modeled as an anisotropic random walk with constant diffusion coefficients. If we assume



the dynamics as direction-dependent yet time- or location-independent diffusion from a specific point, two consequences follow. First, the characteristic direction along which the points are most widely distributed does not change with time. Second, the three-dimensional shape of point distributions at different times are similar. Both conclusions can be conveniently tested with PCA. Indeed, the first principal component (PC1) resulting from PCA specifies the characteristic direction along which the data are most widely distributed, whereas the second principal component (PC2) indicates the second characteristic direction of the distribution after it is



projected into the plane perpendicular to PC1. The procedure is repeated until all the variances are accounted. Note that the principal components (PCs) are perpendicular to each other. Stability of characteristic directions can thus be evaluated by the stability of PCs, while similarity of distributions can also be tested by the stability of ratios between the variances along the PCs.

To further characterize the apparent anisotropic moves, we apply PCA on two different groups of companies starting in 1994 from different points in the three-dimensional space, $(k, \ell, s) \approx \mathbf{x}_1 = (10, 10, 500)$ and $(k, \ell, s) \approx \mathbf{x}_2 = (3, 3, 100)$. We let $\mathbf{u}_{i,t}^{(1)}$ and $\mathbf{u}_{i,t}^{(2)}$ denote the unit PC i vectors in the year t for companies starting from \mathbf{x}_1 and \mathbf{x}_2 , respectively. Two results emerge from the analysis. First, the PC unit vectors turn to be remarkably stable across years (see figure 3). We measure the differences between the principal component vectors in different years by the angles between them, $\theta(\mathbf{u}_{i,t}, \mathbf{u}_{i,t'}) \equiv \arccos \mathbf{u}_{i,t} \cdot \mathbf{u}_{i,t'}$. Overall, the differences from the final year, $\theta(\mathbf{u}_{i,1994+\Delta t}, \mathbf{u}_{i,2015})$, are very small ($\theta \ll \pi/2$) and become almost negligible 5 years after the initial year, as seen in figures 3(a) and (b). Therefore, we conclude that the characteristic directions do not change over time. Second, however, there is a clear sign of deviation from a random walk. The variance along PC2 and PC3 almost reaches a plateau 10 years after the initial year, while variance along PC1 increases linearly (figures 3(c) and (d)), which is enough to reject the random walk hypothesis. This indicates that fluctuations along PC1 of the distribution may be modeled as a random walk, but fluctuations perpendicular to PC1 reach a steady state. It thus suggests that there exists a Langevin-like ‘field’ that drives back the randomly deviating companies into a line parallel to PC1, as seen in figure 2(f) with the red dashed central line.

5. Evolving flow diagrams

Considering the results in the previous section, it is natural to hypothesize that companies that are distant from the scaling curve have the tendency of flow towards this curve. To support this hypothesis, we elaborate on what we refer to as the ‘evolving flow diagrams’ by plotting the estimated vector field of the annual growth in the three

dimensional space of k , ℓ and s . This is inspired by previous work on the prediction of countries' economic growth [49], where the authors advocate the applicability of analogue methods [46, 47] to economic systems.

We first estimate the average growth rates of size measures, k , ℓ and s , at a specific point (k_0, ℓ_0, s_0) in the three-dimensional phase space. We do this by collecting the data sufficiently near the point and computing the arithmetic mean of log-transformed growth rates. Let us define $d_c^{\log}(k_0, \ell_0, s_0)$ as the Euclidean distance of company c from a fixed point (k_0, ℓ_0, s_0) in the previous year $t - 1$ in the logarithmically scaled space:

$$d^{\log} = \sqrt{(d_c^{\log(k)})^2 + (d_c^{\log(\ell)})^2 + (d_c^{\log(s)})^2},$$

where

$$\begin{aligned} d_c^{\log(k)} &= \log[k_c(t-1)/k_0], \\ d_c^{\log(\ell)} &= \log[\ell_c(t-1)/\ell_0], \\ d_c^{\log(s)} &= \log[s_c(t-1)/s_0]. \end{aligned}$$

We sample the companies within the Euclidean radius $d_{\log} < \log[10^{1/8}]$ and eventually get the Nadaraya–Watson estimate with the kernel function of rectangular pulse [50]. Nevertheless, when the resulting sample size is less than 200, the threshold of d_{\log} is enlarged until the sample number reaches or exceeds 200 to suppress the variability of estimates, therefore employing the 200-nearest neighbor method. Thus, the final estimation of mean velocity $\tilde{V}(k, \ell, s)$ in the log-scaled space is obtained from the equation

$$\tilde{V}(k, \ell, s) = \begin{cases} \frac{\#(C_{t-1}(d_{\text{thr}}))^{-1} \sum_{c \in C_{t-1}(d_{\text{thr}})} V_{c,t}}{\#(C_{t-1}(d_{\text{thr}})) \geq 200}, \\ \frac{\#(C_{t-1}(r_{200}))^{-1} \sum_{c \in C_{t-1}(r_{200})} V_{c,t}}{\text{(otherwise)}} \end{cases},$$

where

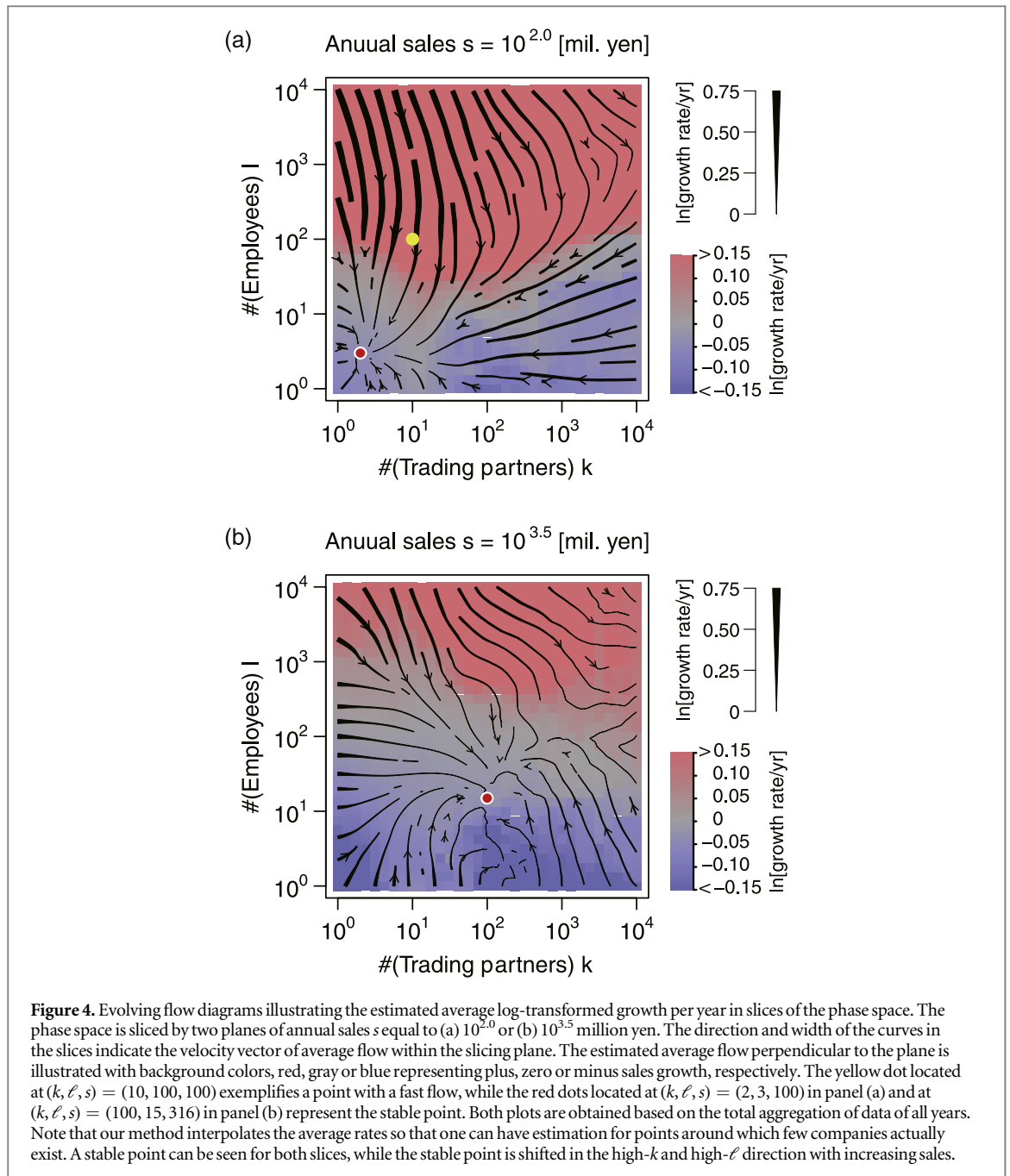
$$\begin{aligned} V_{c,t} &= \begin{pmatrix} \log[k_c(t)/k_c(t-1)] \\ \log[\ell_c(t)/\ell_c(t-1)] \\ \log[s_c(t)/s_c(t-1)] \end{pmatrix}, \\ C_t(d; k, \ell, s) &= \{c \mid d_c^{\log}(k, \ell, s) \leq d\}, \\ r_N &= \min \{r \in \mathbb{R} \mid \#(C_{t-1}(r)) \geq N\}, \\ d_{\text{thr}} &= \log[10^{1/8}], \end{aligned}$$

and $\#(S)$ represents the number of elements in a set S . Taking the logarithm of growth rates, we can limit their possible ranges of several orders of magnitude within those expected from exponential distributions. This makes the arithmetic mean a more robust estimator of the typical value. Note that the arithmetic mean of log-transformed growth rates is equal to the logarithm of geometric mean of growth rates.

We then render two-dimensional diagrams that represent slices of the three-dimensional space, exploiting the method developed in the field of computer graphics [51]. In short, the method follows two steps: (i) placing some points randomly on the plane and drawing streamlines that pass through them, and (ii) repeatedly comparing the original image with a randomly modified one and selecting the one in which the streamlines are placed more homogeneously. Random modifications include inserting, deleting, lengthening, shortening and parallel moving of a streamline.

We show some slices of the vector space in figure 4. Two slices of constant sales ($s = 10^{2.0}$ and $s = 10^{3.5}$ million yen, respectively) are shown in the figure. The streamlines with arrows represent the average movement of companies parallel to the slice, while background colors indicate the average flow orthogonal to the slicing plane. For example, one can see the average time evolution of a company of the size $(k, \ell, s) = (10, 100, 100)$ from figure 4. The point $(k, \ell) = (10, 100)$ can be specified on the slice of $s = 100$ (see the yellow dot in figure 4(a)). The red background around the dot indicates that the company grew in sales by a factor of over 1.15 on average in a year, while the employees was declining rapidly by a factor of 1.6 and almost no change is expected for the number of trading partners. Note that the data in all years of 1994–2015 are joined together to have the maximal sample size.

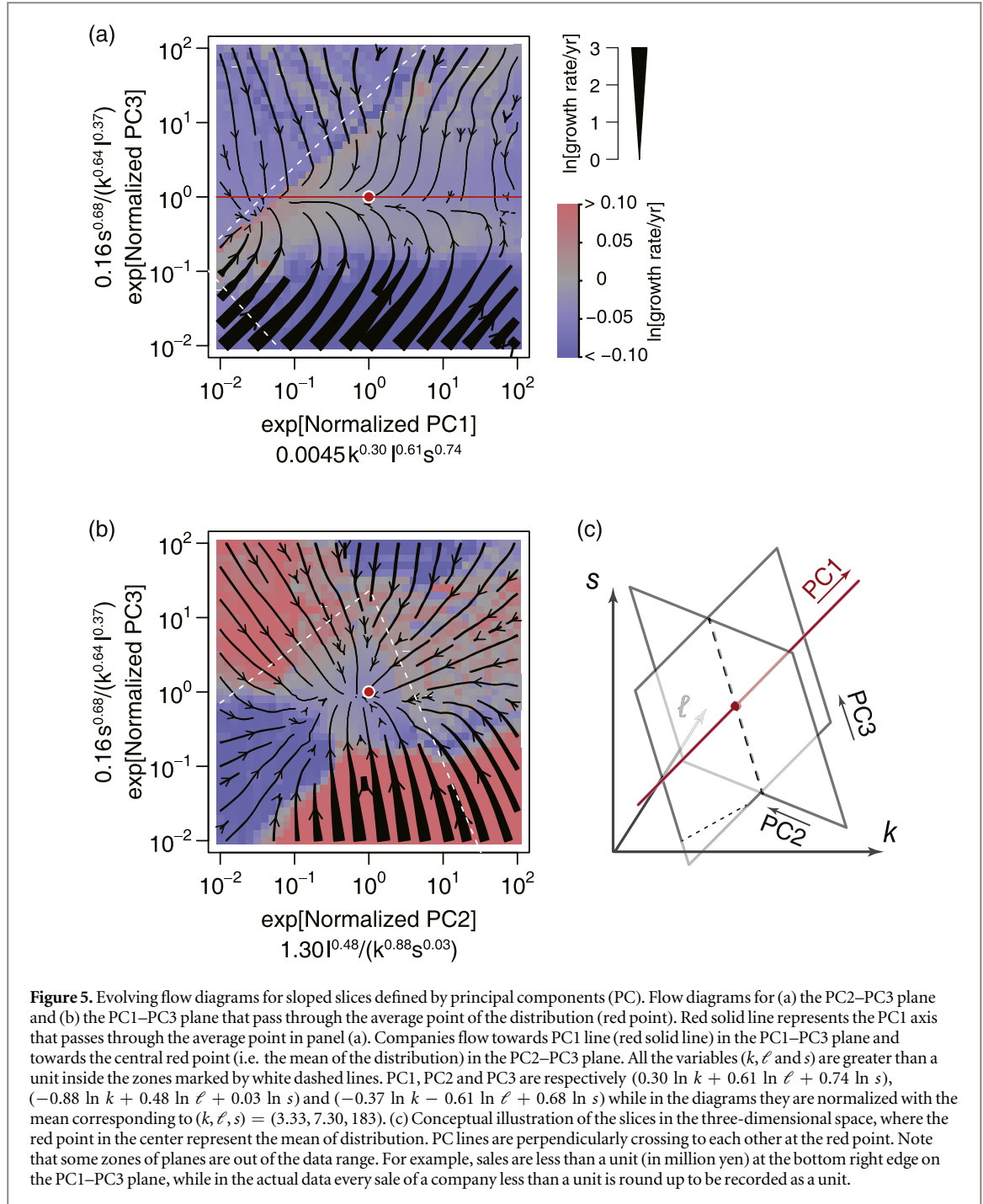
Clearest in figure 4 is the mean flows in ℓ (the number of employees) and k (the number of trading partners) towards a point of stability. It can be seen in figure 4 that the point of stability is located near $(k, \ell) = (2, 3)$ in the slice of $s = 10^{2.0}$, and near $(k, \ell) = (100, 15)$ in the slice of $s = 10^{3.5}$, respectively (see the red dots in both panels). Deviations from the point are compensated by the moving parallel to the slice, i.e. their simultaneous changes in the trading (k) and employment (ℓ). Therefore, this point marks the steady state of the companies having a constant sale.



It is also seen from comparison of figures 4(a) and (b) that the point of stability shifts in the direction of a large number of trading partners (k) and employees (ℓ) as the fixed value of sale (s) is set to be higher. This means that the points of stability for slices of constant sales constitute a curve in the three-dimensional space, which changes from low values of k , ℓ and s to higher values. We hereafter refer to this curve defined from the mean flows as the stability curve.

We further notice in figure 4 that those points on the stability curve can be characterized by relatively very slow absolute changes in sales: average flows around the points of stability are close to zero in comparison to other zones of the space. Therefore, the stability curve can be regarded as an attractor of the mean flow of companies in the phase space, since the average flows on the curve is almost zero in any direction.

Although visually impressive, the results should be interpreted with some care. First, the estimates are most accurate in zones of small sizes (bottom-left in figure 4) and less accurate in other zones, because of poor statistics due to less numbers of sample companies. In particular, the number of samples is fewest at the edges of bottom-right and top-left, where very few companies exist. When the growth rates are estimated for such points without any company around, these estimates are based on data points on the nearest edge of the distribution. Furthermore, the estimates are biased when there is a gradient of data density: the center of the distribution then has more weight than peripheral regions, so that the effects of moving outward from the center of the



distribution on growth rate are always underestimated. Finally, the continuous increase of trading link data (almost two-fold in the two decades) is likely to add positive bias in the estimate of k flow (the change in number of trading partners), turning the direction of average flows rightward in figure 4. This might be the cause of flows in top-right of figure 4(b) that never reach the stability points but continuously go in the right. However, our main conclusion remains unaffected, as the flows towards the steady state cannot be explained merely by these biases and errors.

Considering the results presented in figures 3 and 4 together, it is natural to hypothesize that the stability curve is almost identical to the scaling curve as shown in figure 2(f). We support this by plotting the evolving flow diagrams for sloped slices parallel to the PC vectors (figure 5(a)). PCA is here newly applied to all the data joined together from all years, yielding the PC vectors,

$$\mathbf{u}_1^{\text{all}} = (0.30\mathbf{u}_K + 0.61\mathbf{u}_L + 0.74\mathbf{u}_S), \quad (1a)$$

$$\mathbf{u}_2^{\text{all}} = (-0.88\mathbf{u}_K + 0.48\mathbf{u}_L - 0.03\mathbf{u}_S), \quad (1b)$$

$$\mathbf{u}_3^{\text{all}} = (-0.37\mathbf{u}_K - 0.64\mathbf{u}_L + 0.68\mathbf{u}_S), \quad (1c)$$

where \mathbf{u}_K , \mathbf{u}_L and \mathbf{u}_S are the unit vectors in the direction of $\log k$, $\log \ell$ and $\log s$ axes, respectively. The differences between these PC vectors and those calculated above for the two groups of companies are small ($\theta(\mathbf{u}_i^{\text{all}}, \mathbf{u}_{i,2015}) < 0.15$). Note that PC1, i.e. $\mathbf{u}_1^{\text{all}} \cdot \mathbf{x}$, where $\mathbf{x} \equiv (k, \ell, s)^T$, indicates the overall size of the company, while PC2, i.e. $\mathbf{u}_2^{\text{all}} \cdot \mathbf{x}$, indicates the ratio of the employee number to the number of trading partners. PC3, i.e. $\mathbf{u}_3^{\text{all}} \cdot \mathbf{x}$, can be regarded as the sales scaled against k (the number of trading partners) and ℓ (the number of employees), because PC3 scales $\log [k^{-0.37}\ell^{-0.64}s^{0.68}] = \log [(s/k^{0.55}\ell^{0.94})^{0.68}]$. Note that the PC1 value is extensive in the sense that it scales with the company size while the PC2 and PC3 values are intensive just as $\ell/k^{1.0}$, $s/k^{1.3}$ and $s/\ell^{1.3}$ discussed in section 3.

As seen in the diagram, it supports our hypothesis that companies are on average flowing towards the PC1 line that pass through the mean of the distribution (marked by the red dot in figures 5(a) and (b)). On the other hand, companies are almost ‘neutrally’ stable on the PC1 line: after a small perturbation, a company on the line tends back towards the line, but a displacement parallel to the PC1 line would not be recovered, due to the fact that the flow along the PC1 line is close to zero near the line (see figure 5(b)).

Some non-ideal dynamics of the system is evident in figure 5. First, one would notice the small flow on the PC1 line in the declining direction. This is consistent with the previous studies that have reported negative sales growth of companies on average [52], considering that a large proportion of the companies are located near the PC1 line by definition. Second, the blue-colored area in figure 5(a) is far larger than the red-colored area, meaning that PC2 is generally highly likely to decrease. Since the PC2 value (i.e. $-0.88 \ln k + 0.48 \ln \ell - 0.03 \ln s$) is determined mainly by the number of trading partners (k), this suggests that the number of trading partners (k) of a company tends to increase constantly. This is particularly well shown in figure 4(b): high- k and high- ℓ companies seem to tend towards higher- k states rather than the stable state. We attribute the increase of k to the constantly growing amount of data itself, since adding new companies to the database necessarily leads to increasing the trading relations previously not recognized in the data. Nevertheless, the PC1 line remains an approximate attractor in the dynamical system of the averaged motion, in the sense that companies flow primarily towards the line and the speed of flow is considerably lower on the line. We conclude that the PC1 line, which we consider to be a part of the scaling curve, approximately agrees well with the stability curve in the small-company regime. Note that we obtain similar results for the large-company regime as presented in [appendix](#).

6. Dissipation rates

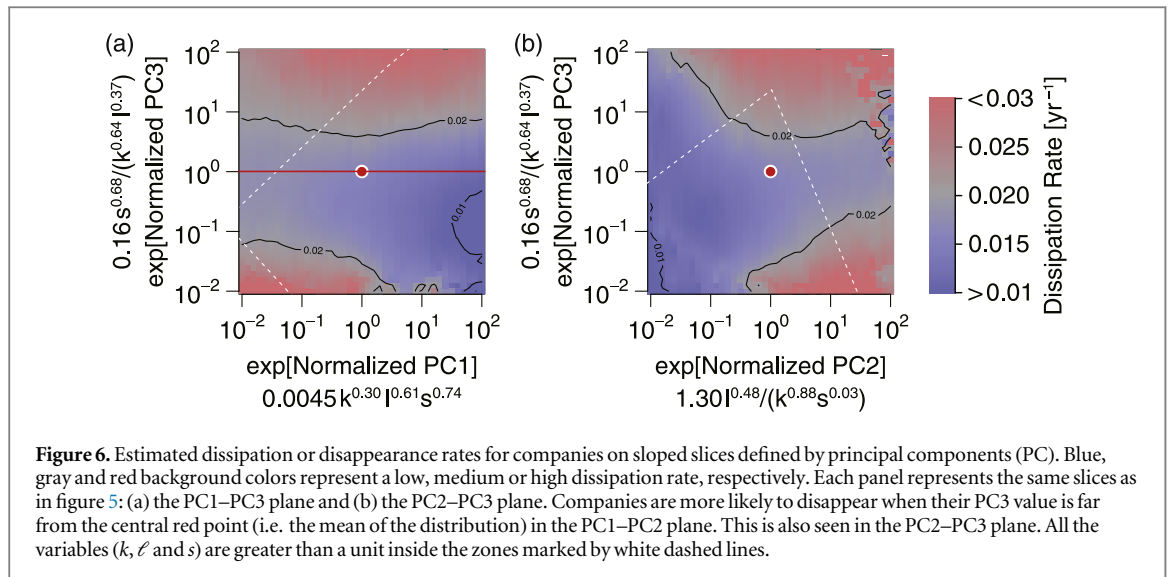
Apart from the average flow that we described above, we can also regard the difference in the disappearance or dissipation rate as the origin of a stable distribution of companies in the phase space. Recall that a random walk process has no steady state. In order for a stochastic process to have a steady state, diffusion has to be compensated by some countereffects. In our case, we suppose that if companies disappear more rapidly at the zones far from the scaling curve compared to the zones near it, this difference in the dissipation rate could keep the distribution near the scaling curve against the diffusive effects that lead the system to non-stationary dynamics.

In order to explore our hypothesis, we employ a method slightly modified from the one in the previous section. First, a company is determined to be nonexistent in a year t if there is no record for the company’s existence either in the year t or in the following year $t + 1$. We then compute the estimated dissipation rate at each point, $\tilde{\mu}(k, \ell, s)$, from the data of all years as follows:

$$\tilde{\mu}(k, \ell, s) = \begin{cases} \#(C_{t-1}(d_{\text{thr}}))^{-1} \sum_{c \in C_{t-1}(d_{\text{thr}})} \delta_{c,t} \\ \quad (\#(C_{t-1}(d_{\text{thr}})) \geq 1000) \\ \#(C_{t-1}(r_{1000}))^{-1} \sum_{c \in C_{t-1}(r_{1000})} \delta_{c,t} \\ \quad (\text{otherwise}) \end{cases},$$

where $\delta_{c,t}$ is 1 (if the company c does not exist in year t) or 0 (otherwise) and d_{thr} is here set to be $\log[10^{1/4}]$.

We plot two slices of the scalar field of dissipation rates, $\tilde{\mu}(k, \ell, s)$, into respective heatmaps (figure 6). The slices are set in the same way as in figure 5. Figure 6(a) clearly shows that $\tilde{\mu}$ is higher at zones far from the PC1 line compared to zones near the line. It is also evident that the rate of dissipation is lower at the zones of larger PC1, which is in line with an earlier report that such a rate decreases with company size [53]. The difference in dissipation rate in the PC3 direction also conforms to the expectation in figure 6(b), although $\tilde{\mu}$ is almost monotonically increasing with the PC2 value against our hypothesis. Note that a low value of PC2 ($\approx \log [k^{0.88}/\ell^{0.48}]$) indicates that the company has more trading partners compared to the average one of those with the same number of employees, therefore suggesting superiority, rather than inferiority, of the company. Note that in similar plots for the large-company regime, however, the tendency of higher dissipation rate at zones far from the scaling curve is more evident ([appendix](#)). In conclusion, the difference in the dissipation also



plays a role in stabilizing the distribution of companies by keeping them near the scaling curve against random diffusion.

7. Discussion and conclusion

We have analyzed the dynamics of companies, highlighting the general tendency of companies towards a steady state. As we have seen in figure 5, there exist tendencies in the time evolution leading companies to a ‘stability curve’ in the three-dimensional space. For small-size companies, the curve is parallel to the PC1 vector of the company distribution (figure 5(a)). Note that the PC1 vector does not agree with the previously reported scaling exponents, especially with regard to the number of trading partners, k . As the companies are primarily scattered along the first principal component (PC1) direction by definition, equation (1a) indicates that $\ln \ell$ increases by 0.61 on average when $\ln k$ increases by 0.30, thus suggesting the square relationship between k and ℓ ($\ell \propto k^2$). On the other hand, the relation between k and ℓ has been found to be almost linear ($\ell \propto k$) in the previous study [24]. This is probably because of the different data that has been analyzed. In our analysis, the vast majority of companies are the small companies, while the small-size companies have been excluded from the analysis in the previous study. Indeed, when companies of $k \lesssim 10$ that belong to the small-company regime are excluded (see appendix), the PCA recovers the previously reported scaling relations (see figure A1) and the stability curve is again parallel to the PC1 vector of the truncated company distribution (see figure A2). Therefore, we conclude that the stability curve can be approximated by a single scaling curve (composed of two functions for different regimes) over the whole range.

We are inclined to interpret our results of flows towards the scaling curve as an emergent property. If the scaling relations and the bounded dynamics were the result of a rigid regulation, there could be a clear-cut threshold over which companies are not allowed to exist. On the contrary, the distribution of companies in the three-dimensional space is rather smooth, as revealed in the previous study [24]. Moreover, the dissipation rate is higher at zones far from the scaling curve compared to zones near the curve. This suggests that there are difficulties in staying at the zones distant from the scaling curve, and that the moves towards the scaling curve are adaptive and spontaneous.

It remains an open question what are the origins of the two different regimes of scaling. One possible factor is the discretized values of k and ℓ that might strongly affect an analysis particularly when they are less than 10. Nevertheless, different scaling exponents might also be relevant to differences in actual dynamics of companies. For example, the number of companies having 5–10 trading partners are more than expected from the power-law tail of the distribution in our dataset (as seen in [54]). This might suggest that there is a ‘barrier’ that prevents those in the small-company regime from entering the large-company regime. Existence of such a barrier would be suggestive of different mechanisms actually at work in different regimes.

While all kinds of companies are collectively used in our current analysis, finer-grained analyses are possible in the framework we have presented here. For example, the dynamical similarity between companies of different industrial sectors can be evaluated. Since survival rates of companies are known to be surprisingly similar across different sectors [55], it seems reasonable to hypothesize that the dynamics in the phase space is also similar.

Another possible direction is to study the dynamics for companies of different ages. Despite much consideration devoted to the topic, empirical research on whether companies behave differently depending on their age is still lacking.

It is important to note that there are always temporal fluctuations in companies' size measures [20, 30, 35], which are presumably a dominant factor of company dynamics (diffusion) especially around the stability curve. We speculate that the diffusion effects of the stochastic growth rates are in equilibrium with the average flow discussed above. This might lead to the universal fat-tailed distribution of companies around the scaling relations [24] through a process similar to a random multiplicative process [56]. However, the connection between the common scaling function and the stochastic dynamics is yet to be established.

Our results might support the claim [49] that the analogue method is beneficial in addressing the problem of dynamics which cannot be solved by the standard regression analysis, or more generally, by the statistical analysis of a snapshot of a system. As shown in figure 5(b), low-PC3 companies subsequently tend to increase their PC1 value, while high-PC3 (and high-PC2) ones are likely to decrease it slightly. Therefore, we could expect that the 'natural' size of companies would be larger for low PC3 compared to high PC3, even if their PC1 values are the same. This point cannot be revealed only by the PCA.

Finding the novel stylized fact of agreement between the scaling and stability curves, we believe that our results could be also beneficial to future modeling and theoretical considerations. Although researchers have formulated numerous models [30–41], we are not aware of any work that could be used to explain the agreement between the scaling and stability curves demonstrated here. Thus, we suggest that future theoretical studies should incorporate this phenomenological multi-variate evolution of companies.

We expect that similar results would emerge when applying our method to a different set of companies in other areas or countries. Interestingly, different scaling exponents for the $\ell - s$ relation have been found in datasets from different countries and areas. While sales (s) increase superlinearly with employees (ℓ) in the Japanese data, the relation is almost linear in the dataset from the United States of America and sublinear in the one from the People's Republic of China [25]. It would be a promising research direction to carry out comparative studies across different countries and areas. Furthermore, analogous results might emerge when one examine diverse natural, social or economic systems such as city growth [57] or the species-area relationship in ecological systems [12].

Acknowledgments

We thank Orr Levy for discussions. The authors appreciate Teikoku Databank, Ltd., Center for TDB Advanced Data Analysis and Modeling in Tokyo Institute of Technology for providing both the data and financial support. This work is partially supported by the Grant-in-Aid for Scientific Research (B), Grant Number 26310207, by JST, Strategic International Collaborative Research Program (SICORP) on the topic of 'ICT for a Resilient Society' by Japan and Israel, and by MEXT as 'Exploratory Challenges on Post-K computer (Study on multilayered multiscale spacetime simulations for social and economical phenomena)'.

MT directed the project. YK, HT and SH developed the data analysis. YK analyzed the data and generated the diagrams. All authors contributed in writing the paper.

Appendix. Scaling and stability curve in large-company regime

A.1. PCA with unbiased thresholding

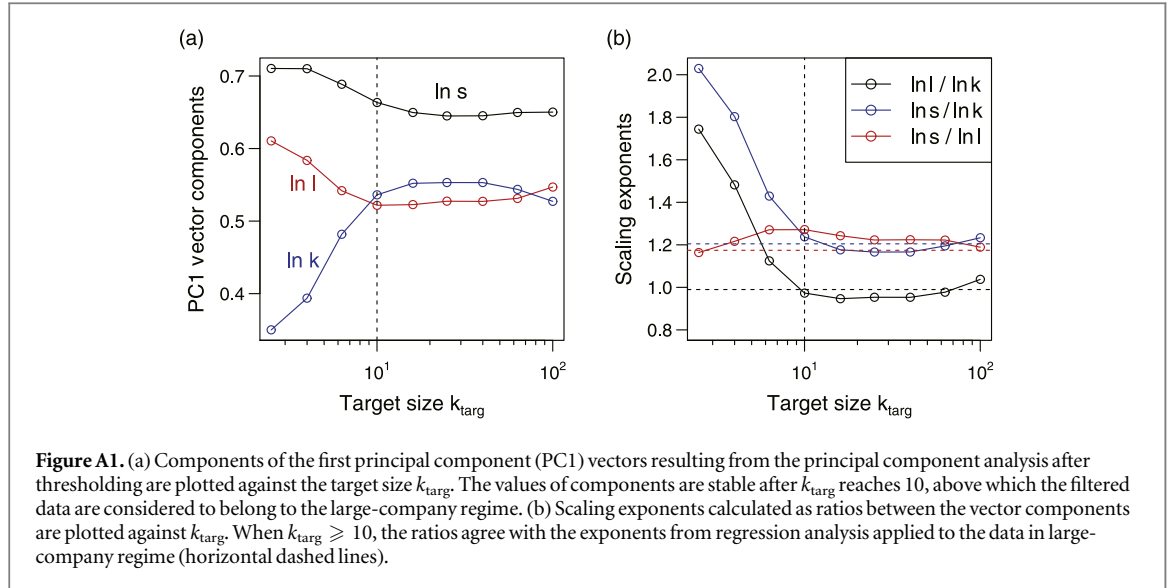
Here we discuss the method to obtain the threshold that discriminates between the large- and small-company regimes, so that we can filter out the data points in the small-company regime. We first define a property of data filtering (i.e. self-consistency) in relation to the principal component vectors for the filtered data. We then present an algorithm to search for filtering criteria that have the property. This algorithm gives a series of possible threshold planes as a function of a scalar parameter of company size. In ideal cases, one can find a single 'true' threshold of company size, above which the candidate threshold planes are all placed in parallel to each other.

We define a 'self-consistency' for a data filtering,

$$\sum_{i=1}^N a_i x_i \geq 1,$$

to be a filtering such that the PC1 vector for the filtered data is perpendicular to the line or (hyper-)plane,

$$\sum_{i=1}^N a_i x_i = 1. \quad (\text{A.1})$$



We hereafter refer to the threshold plane (A.1) of a self-consistent criterion as a self-consistent threshold plane. The definition is inspired by the notion of self-consistency for a principal curve along which the data are aligned [58, 59]. In our data, the principal curve presumably amounts to the scaling curve (red bold curve in figure 1).

Self-consistent threshold planes are all parallel to each other in certain ideal cases. Suppose that the original dataset can be modeled as

$$\mathbf{x}_i = \mathbf{v}_0 + \rho_i \mathbf{u}_0 + \mathbf{e}_i, \quad (\text{A.2})$$

where \mathbf{x}_i denotes the vector representing the i th data, \mathbf{v}_0 a vector representing the origin, ρ_i i.i.d. random scalar variables, \mathbf{u}_0 a vector representing the characteristic direction and \mathbf{e}_i i.i.d. random vectors that are perpendicular to \mathbf{u}_0 . Note that multivariate Gaussian distributions with a non-identity covariance matrix are included in this class of multivariate distributions. Given that the variance of \mathbf{e}_i along any direction is smaller than the variance of ρ_i , the PC1 vector should be almost parallel to \mathbf{u}_0 [58]. In this case, removing the data below a threshold plane perpendicular to the original PC1 vector (i.e. accepting the data with ρ_i larger than a threshold) does not affect the location of the PC1 line, as long as the variance of ρ_i in the new dataset is larger than those of \mathbf{e}_i . As a result, self-consistent threshold planes are always almost perpendicular to the PC1 vector for the original dataset, and thus almost parallel to each other.

We can determine if the data conform to the model (A.2) based on whether the self-consistent threshold planes are placed in parallel. Assuming that the multivariate company distribution in the large-company regime can be described by the model (A.2) (i.e. ‘straightly’ distributed as in figure 1), the self-consistent threshold planes should be all parallel to each other above a ‘true’ threshold.

A series of self-consistent threshold planes for our data is determined by minimizing the difference between the initial threshold plane and a plane that is perpendicular to the PC1 line for the filtered data. In more detail, our procedure is as follows. First, we set an arbitrary threshold plane and define a filtering criterion,

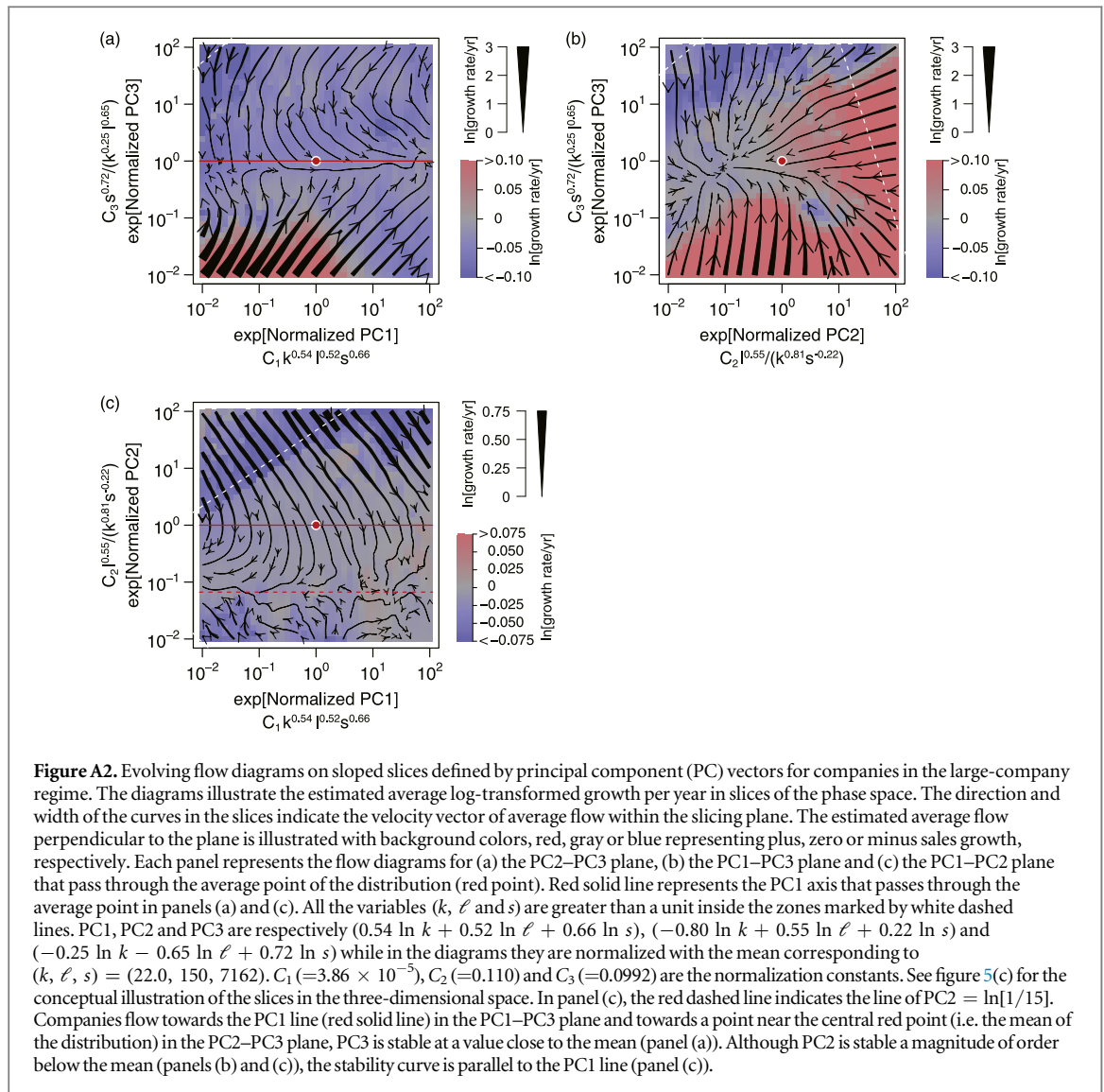
$$a_1^{(0)} \log k + a_2^{(0)} \log \ell + a_3^{(0)} \log s \geq 1.$$

Next, we apply PCA to the filtered data and obtain a PC1 line, which passes through the mean of the filtered data distribution. Then, we can determine the plane,

$$a_1^{(1)} \log k + a_2^{(1)} \log \ell + a_3^{(1)} \log s = 1,$$

that is perpendicular to the PC1 line and intersects the PC1 line at a number of trading partners, k_{targ} . We hereafter refer to this parameter, k_{targ} , as the target size of a self-consistent threshold plane. Then, the sum of squared errors, $\sum_{i=1}^3 (a_i^{(1)} - a_i^{(0)})^2$, is minimized until it almost reaches zero. The method thus determines a self-consistent threshold plane given the single parameter of company size. Data for all years (1994–2015) are used as the input to the procedure. The method is tried for values of k_{targ} that are evenly spaced in logarithmic scale. We use the Nelder–Mead method [60] as implemented in R (ver.3.5.1) for minimization. The final sum of squared errors after the minimization is less than 10^{-10} for all the values of k_{targ} plotted in figure A1.

The principal component vectors for the self-consistently filtered datasets are plotted against the target size, k_{targ} , in figure A1(a). The vector components largely fluctuates at $k_{\text{targ}} < 10$, but are remarkably stable at $k_{\text{targ}} \geq 10$. In this zone, the scaling exponents for pairs of variables also approximately agree with the results from the standard regression analysis, when the data of the explanatory variable less than a threshold (set to be

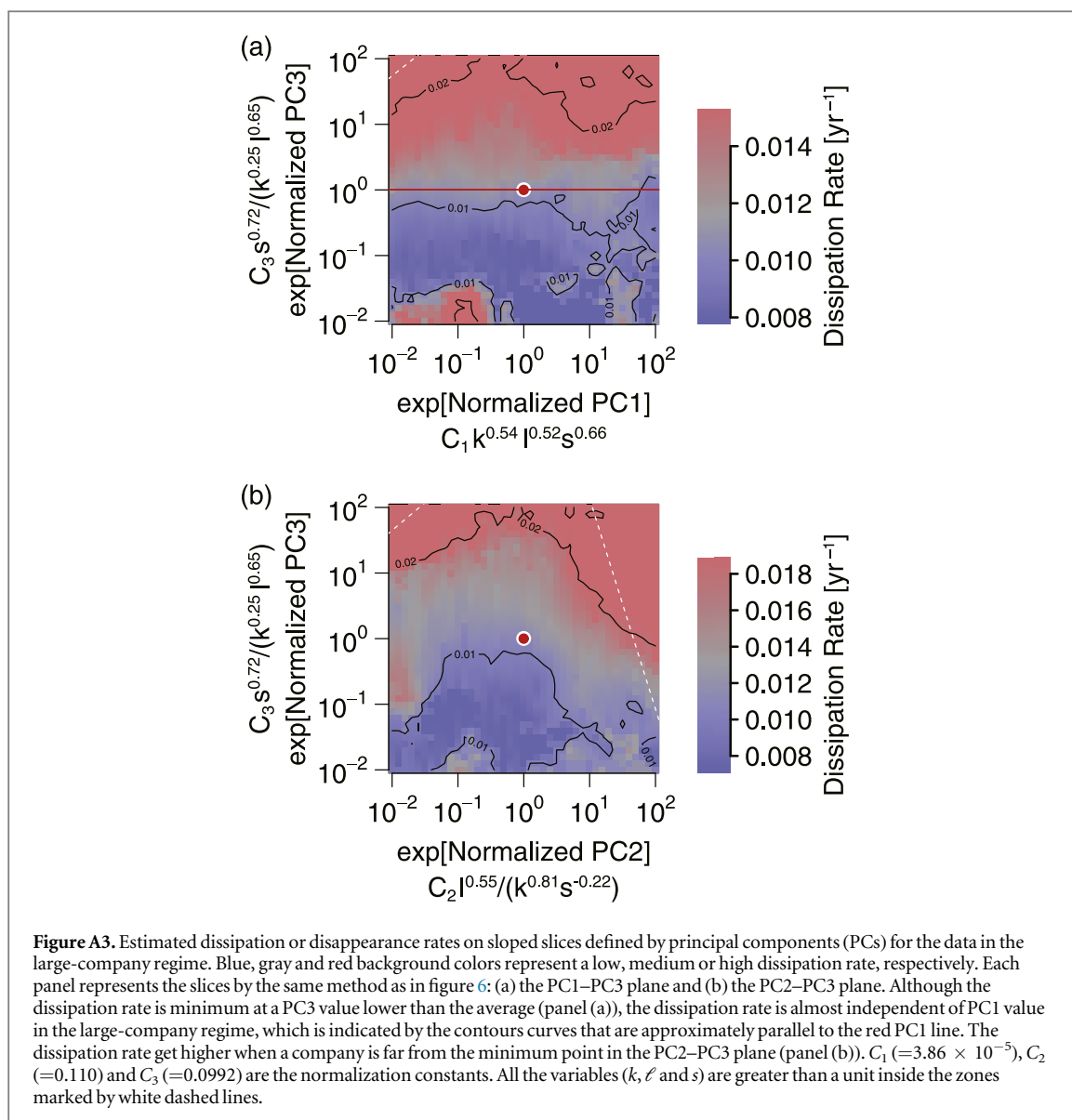


100 here) are excluded from the regression. The regression analysis is used here to quantify the previous examination of scaling exponents [24]. The horizontal dashed lines in figure A1 (b) indicate the scaling exponents from the standard regression analysis of the pairs of variables, $(\log \ell, \log k)$, $(\log s, \log k)$ and $(\log s, \log \ell)$, with the former being explained and the latter being explanatory. These results highly suggest that the self-consistent threshold plane for $k_{\text{targ}} = 10$ best discriminates between the small- and large-company regimes.

A.2. Evolving flow and dissipation rates in large-company regime

We here discuss the evolving flow diagrams and the heatmaps of dissipation rates for the large-company regime. To obtain the figures, we apply the same methods as in the main text, except using the PCA results after the self-consistent data filtering of $k_{\text{targ}} = 10$.

As seen in each panel of figure A2, the tendency towards a stable state is also evident in the large-company regime. The stability curve is parallel to the scaling curve (the PC1 line marked by a red line in figures A2(a) and (c)), although the location of the stability curve, especially regarding the PC2 value, differs from the scaling curve. Indeed, as shown in figure A2(b), companies are stable at a PC2 value considerably lower than the average (central red dot). In figure A2(c), the scaling PC1 line (red solid line) is located differently from the stability line (red dashed line). We consider this to be the consequence of continuously growing amount of the trading network data. Due to this increase, a company without any real change in k typically experience an increase of the trading partners in the dataset. As the normalized PC2 value (i.e. $-0.80 \ln k + 0.55 \ln \ell + 0.22 \ln s + \ln 0.110$) is highly affected by the number of trading partners (k), increasing k would decrease the PC2 value without any real change, pushing the stability curve towards a low-PC2 location. We expect that the position of the scaling and the stability curves would be identical if the amount of data were stable.



Likewise, the estimated dissipation rates ($\tilde{\mu}$) of companies in the large-company regime are consistent with our conclusion on the smaller-company regime. The rate, $\tilde{\mu}$, is typically higher in the zones far from the scaling line compared to those near the scaling line, as illustrated in figure A3. In particular, the dissipation rate gets higher as the PC1 value gets higher (figures A3(a) and (b)) or the PC2 value get more distant from the average (marked by red dot in the center of figure A3(b)). This might help to constrain the multivariate company distribution around the PC1 line. Although the minimum of $\tilde{\mu}$ is not located in the exactly same place as the scaling line, the contours of constant $\tilde{\mu}$ are almost parallel to the scaling line (the horizontal red line in figure A3(a)). This indicates that $\tilde{\mu}$ is almost independent of the PC1 value, which is consistent with the previous observation that the survival rate of publicly traded North American companies are almost independent of their age and business sector [55]. Note that companies with the lowest dissipation rate are located in the zone of a high average sales growth, as shown in figures A2(a) and (b). We consider that this high growth prevents companies from disappearing.

ORCID iDs

Misako Takayasu  <https://orcid.org/0000-0002-2467-614X>

References

- [1] van der Waals JD 1894 Thermodynamische Theorie der Kapillarität unter Voraussetzung stetiger Dichteänderung *Z. Phys. Chem.* **13U** 657–725
- [2] Guggenheim EA 1945 The principle of corresponding states *J. Chem. Phys.* **13** 253–61

- [3] Widom B 1965 Equation of state in the neighborhood of the critical point *J. Chem. Phys.* **43** 3898–905
- [4] Patashinskii A Z and Pokrovskii V L 1966 Behavior of ordered systems near the transition point *Sov. Phys. JETP* **23** 292–7
- [5] Kadanoff L P 1966 Scaling laws for Ising models near T_c *Physics* **2** 263–72
- [6] Fisher M E 1967 The theory of equilibrium critical phenomena *Rep. Prog. Phys.* **30** 615–730
- [7] Eugene Stanley H 1971 *Phase Transitions and Critical Phenomena* (Oxford: Oxford University Press)
- [8] de Gennes P G 1972 Exponents for the excluded volume problem as derived by the Wilson method *Phys. Lett. A* **38** 339–40
- [9] de Gennes P-G 1979 *Scaling Concepts in Polymer Physics* (London: Cornell University Press)
- [10] Vicsek T 1992 *Fractal Growth Phenomena* 2nd edn (Singapore: World Scientific)
- [11] Arrhenius O 1921 Species and area *J. Ecol.* **9** 95–9
- [12] Rosenzweig M L 1995 *Species Diversity in Space and Time* (Cambridge: Cambridge University Press)
- [13] Kleiber M 1947 Body size and metabolic rate *Physiol. Rev.* **27** 511–41
- [14] Schmidt-Nielsen K 1984 *Scaling: Why is Animal Size So Important?* (Cambridge: Cambridge University Press)
- [15] Martin A P and Palumbi S R 1993 Body size, metabolic rate, generation time, and the molecular clock *Proc. Natl Acad. Sci.* **90** 4087–91
- [16] Savage V M, Gillooly J F, Woodruff W H, West G B, Allen A P, Enquist B J and Brown J H 2004 The predominance of quarter-power scaling in biology *Funct. Ecol.* **18** 257–82
- [17] Bettencourt L M A, Lobo J, Helbing D, Kuhnert C and West G B 2007 Growth, innovation, scaling, and the pace of life in cities *Proc. Natl Acad. Sci.* **104** 7301–6
- [18] Samaniego H and Moses M E 2008 Cities as organisms: allometric scaling of urban road networks *J. Transp. Land Use* **1** 21–39
- [19] Li X, Wang X, Zhang J and Wu L 2015 Allometric scaling, size distribution and pattern formation of natural cities *Palgrave Commun.* **1** 15017
- [20] Stanley M H R, Amaral L A N, Buldyrev S V, Havlin S, Leschhorn H, Maass P, Salinger M A and Stanley H E 1996 Scaling behaviour in the growth of companies *Nature* **379** 804–6
- [21] Amaral L A N, Buldyrev S V, Havlin S, Maass P, Salinger M A, Stanley H E and Stanley M H R 1997 Scaling behavior in economics: the problem of quantifying company growth *Physica A* **244** 1–24
- [22] De Fabritiis G, Pammolli F and Riccaboni M 2003 On size and growth of business firms *Physica A* **324** 38–44
- [23] Saito Y U, Watanabe T and Iwamura M 2007 Do larger firms have more interfirm relationships? *Physica A* **383** 158–63
- [24] Watanabe H, Takayasu H and Takayasu M 2013 Relations between allometric scalings and fluctuations in complex systems: the case of Japanese firms *Physica A* **392** 741–56
- [25] West G B 2017 Toward a science of companies *Scale: The universal Laws of Growth, Innovation, Sustainability, and the Pace of Life in Organisms, Cities, Economies, and Companies* (New York: Penguin) ch 9
- [26] West G B, Brown J H and Enquist B J 1997 A general model for the origin of allometric scaling laws in biology *Science* **276** 122–6
- [27] West G B, Brown J H and Enquist B J 1999 The fourth dimension of life: fractal geometry and allometric scaling of organisms *Science* **284** 1677–9
- [28] Bettencourt L M A 2013 The origins of scaling in cities *Science* **340** 1438–41
- [29] Yakubo K, Saijo Y and Korošak D 2014 Superlinear and sublinear urban scaling in geographical networks modeling cities *Phys. Rev. E* **90** 022803
- [30] Buldyrev S V, Nunes Amaral L A, Havlin S, Leschhorn H, Maass P, Salinger M A, Eugene Stanley H and Stanley M H R 1997 Scaling behavior in economics: II. Modeling of company growth *J. Phys. I* **7** 635–50
- [31] Bottazzi G and Secchi A 2003 A stochastic model of firm growth *Physica A* **324** 213–9
- [32] Alfarano S, Milaković M, Irlé A and Kauschke J 2012 A statistical equilibrium model of competitive firms *J. Econ. Dyn. Control* **36** 136–49
- [33] Gallegati M, Guilioni G and Kichiji N 2003 Complex dynamics and financial fragility in an agent based model *Adv. Complex Syst.* **6** 267–82
- [34] Wright I 2005 The social architecture of capitalism *Physica A* **346** 589–620
- [35] Fu D, Pammolli F, Buldyrev S V, Riccaboni M, Matia K, Yamasaki K and Stanley H E 2005 The growth of business firms: theoretical framework and empirical evidence *Proc. Natl Acad. Sci.* **102** 18801–6
- [36] Mondani H, Holme P and Liljeros F 2014 Fat-tailed fluctuations in the size of organizations: the role of social influence *PLoS One* **9** e100527
- [37] Takayasu M, Watanabe H and Takayasu H 2014 Generalised central limit theorems for growth rate distribution of complex systems *J. Stat. Phys.* **155** 47–71
- [38] Malcai O, Biham O and Solomon S 1999 Power-law distributions and Lévy-stable intermittent fluctuations in stochastic systems of many autocatalytic elements *Phys. Rev. E* **60** 1299–303
- [39] Sutton J 2002 The variance of firm growth rates: the ‘scaling’ puzzle *Physica A* **312** 577–90
- [40] Wyart M and Bouchaud J P 2003 Statistical models for company growth *Physica A* **326** 241–55
- [41] Riccaboni M, Pammolli F, Buldyrev S V, Ponta L and Stanley H E 2008 The size variance relationship of business firm growth rates *Proc. Natl Acad. Sci.* **105** 19595–600
- [42] Hubbell S P 2001 *The Unified Neutral Theory of Biodiversity and Biogeography* (Princeton, NJ: Princeton University Press)
- [43] Tokita K and Yasutomi A 1999 Mass extinction in a dynamical system of evolution with variable dimension *Phys. Rev. E* **60** 842–7
- [44] Rosindell J and Cornell S J 2007 Species-area relationships from a spatially explicit neutral model in an infinite landscape *Ecol. Lett.* **10** 586–95
- [45] Filotas E, Grant M, Parrott L and Rikvold Per Arne P A 2010 Positive interactions and the emergence of community structure in metacommunities *J. Theor. Biol.* **266** 419–29
- [46] Lorenz E N 1969 Three approaches to atmospheric predictability *Bull. Am. Meteorol. Soc.* **50** 345–9
- [47] Lorenz E N 1969 Atmospheric predictability as revealed by naturally occurring analogues *J. Atmos. Sci.* **26** 636–46
- [48] Lorenz E N 1963 Deterministic nonperiodic flow *J. Atmos. Sci.* **20** 130–41
- [49] Cristelli M, Tacchella A and Pietronero L 2015 The heterogeneous dynamics of economic complexity *PLoS One* **10** e0117174
- [50] Hastie T, Tibshirani R and Friedman J 2009 Kernel smoothing methods *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* 2nd edn (New York: Springer) ch 6, pp 191–218
- [51] Turk G and Banks D 1996 Image-guided streamline placement *Proc. 23rd Annual Conf. on Computer Graphics and Interactive Techniques—SIGGRAPH ’96* (New York: ACM Press) pp 453–60
- [52] Fix B 2017 Energy and institution size *PLoS One* **12** e0171823

- [53] Ishikawa A, Fujimoto S, Mizuno T and Watanabe T 2017 Transition law of firms' activity and the deficit aspect of non-Gibrat's law *Proc. Asia-Pacific Econophysics Conf. 2016—Big Data Analysis and Modeling Toward Super Smart Society—(APEC-SSS2016)*; *J. Phys. Soc. Japan* **16** 011005
- [54] Goto H, Takayasu H and Takayasu M 2015 Empirical analysis of firm-dynamics on Japanese interfirm trade network *Proc. Int. Conf. on Social Modeling and Simulation, Plus Econophysics Colloquium 2014* ed H Takayasu (Cham: Springer) pp 195–204
- [55] Daepf M I G, Hamilton M J, West G B and Bettencourt L M A 2015 The mortality of companies *J. R. Soc. Interface* **12** 20150120
- [56] Takayasu H, Sato A-H and Takayasu M 1997 Stable infinite variance fluctuations in randomly amplified Langevin systems *Phys. Rev. Lett.* **79** 966–9
- [57] Batty M and Longley P 1994 *Fractal Cities: A Geometry of Form and Function* (New York: Academic)
- [58] Hastie T and Stuetzle W 1989 Principal curves *J. Am. Stat. Assoc.* **84** 502–16
- [59] Tarpey T and Flury B 1996 Self-consistency: A fundamental concept in statistics *Stat. Sci.* **11** 229–43
- [60] Nelder J A and Mead R 1965 A simplex method for function minimization *Comput. J.* **7** 308–13