Modeling Firm Birth and Death Dynamics using Survival Fractions and Age Distributions

Yipei Guo[1,2][0000−0002−4583−8865], Huynh Hoai Nguyen[1][0000−0002−8432−7435], and Feng Ling[1][0000−0001−6578−2558]

[1] Institute of High Performance Computing (IHPC), Agency for Science, Technology and Research (A*STAR), 1 Fusionopolis Way, Connexis (North Tower) #16-16, Singapore 138632, Republic of Singapore
[2] guoyp@ihpc.a-star.edu.sg

Abstract. The birth and death of firms govern firm population dynamics, and understanding these processes can guide urban planning and policy. Longitudinal data with full entry and exit records allow direct analysis of birth and death rates and how they depend on external factors and firm-level properties. However, real-world data are often incomplete, with missing records of extinct firms or exit dates. In such cases, it is unclear if and how we can extract information about the birth and death processes. By modeling how these processes shape firms' age distributions and survival fractions, we show how one can gain insights even from incomplete data. While age distributions are insufficient for inferring both processes, survival fractions reveal how death rates depend on firm age and sector size. Applying our approach to 14 major sectors in Singapore, we find that death rates decline with age and rise with sector size, with a multiplicative interaction between both effects. Assuming sigmoidal dependence on both factors, we infer sector-specific death models that accurately reproduce the data and enable reconstruction of key system features, e.g., sector size trajectories and birth-death rate correlations.

Keywords: birth-death dynamics · age distributions · survival fractions.

1 Introduction

The number of firms in a city evolves as new firms emerge and others close. These birth and death rates vary over time and across firm sectors [2, 3], and predicting sector size changes can inform urban planning and policy decisions such as space allocation across sectors or targeted incentives.

Understanding firm birth and death dynamics typically requires longitudinal data that track both entry and exit dates. Such data enable the extraction of annual births, closures, sector sizes, and the ages of surviving and exiting firms (Fig. 1a). This allows direct analysis of how births depend on market conditions (e.g., sector size) and how survival rates depend on firm-level properties (e.g., age, sector) and external factors. However, full entry and exit records are often unavailable—a limitation noted in prior studies [5, 6, 8]. For example, cross-sectional age distribution analyses often rely only on currently active firms [5,

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97635-3_24

6]. In Singapore, the Accounting and Corporate Regulatory Authority (ACRA) provides public data on all firms, including their sector, registration date, and current status (if alive) [1], but not exit dates of inactive firms.

When key information is missing, one potential approach to inferring birth and death dynamics is to use statistical quantities such as age distributions and survival fractions. Age distributions can be derived from registration dates of existing firms alone, while survival fractions for firms born in a given year can be obtained from registration dates and current status, without needing exit dates (Fig. 1a). While these quantities have been empirically studied in various countries [5, 6, 10, 13], the relationship between them and the underlying birth and death processes has been less explored. Hence, it is unclear to what extent they can be used to infer firm population dynamics.

By modeling how age and sector size dependencies in birth and death rates shape age distributions and survival fractions, we show how to infer firm population dynamics under 2 incomplete-data scenarios (Fig. 1a): (1) extinct firms are not recorded and (2) exit dates are missing. While the age distribution alone—available in both scenarios—is insufficient to recover birth and death dynamics, the survival fractions available in scenario 2 enable inference of how death rates vary with age and sector size. Applying our approach to 14 major sectors in Singapore, we find that death rates decline with age and rise with sector size, with multiplicative interaction between both effects. Assuming a sigmoidal dependence on both factors, we infer sector-specific death models that accurately reproduce empirical data and enable reconstruction of other system properties, e.g., sector size trajectories and correlations between birth and death rates.

2 Results

2.1 The model

Let $\mu(a, N)$ be the death rate (i.e., probability per unit time) of a firm of age a when N firms (of the same sector) are present. The survival fraction $f(a, T)$ is the probability that a firm survives from the time it was born $t = T - a$ to the current time T:

$$f(a, T) = e^{-\int_{T-a}^{T} \mu(s-(T-a), N(s)) ds}, \tag{1}$$

where $N(t) = \int_0^\infty \rho(a, t) da$, with $\rho(a, t)$ being the density of firms of age a.

Let $B(t)$ be the birth rate of new firms at time t. The density of firms of age a at time T is then the product of their birth rate at time $T - a$ (if $a < T$) or their initial density (if $a \geq T$), and the probability of surviving until T:

$$\rho(a, T) = \begin{cases} B(T-a) f(a, T) & \text{for } a < T \\ \rho(a-T, 0) e^{-\int_0^T \mu(s+a-T, N(s)) ds} & \text{for } a \geq T. \end{cases} \tag{2}$$

With the entry dates of all firms (Scenario 2, Fig. 1a), the birth rate trajectory $B(t)$ can be directly extracted. However, this is impossible with only the list of alive firms (Scenario 1, Fig. 1a). In this case, we model the birth dynamics by assuming that time variation in $B(t)$ arises solely from its dependence on $N(t)$: $B(t) = \beta(N(t))N(t)$, where $\beta(N)$ is the birth rate coefficient.

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97635-3_24

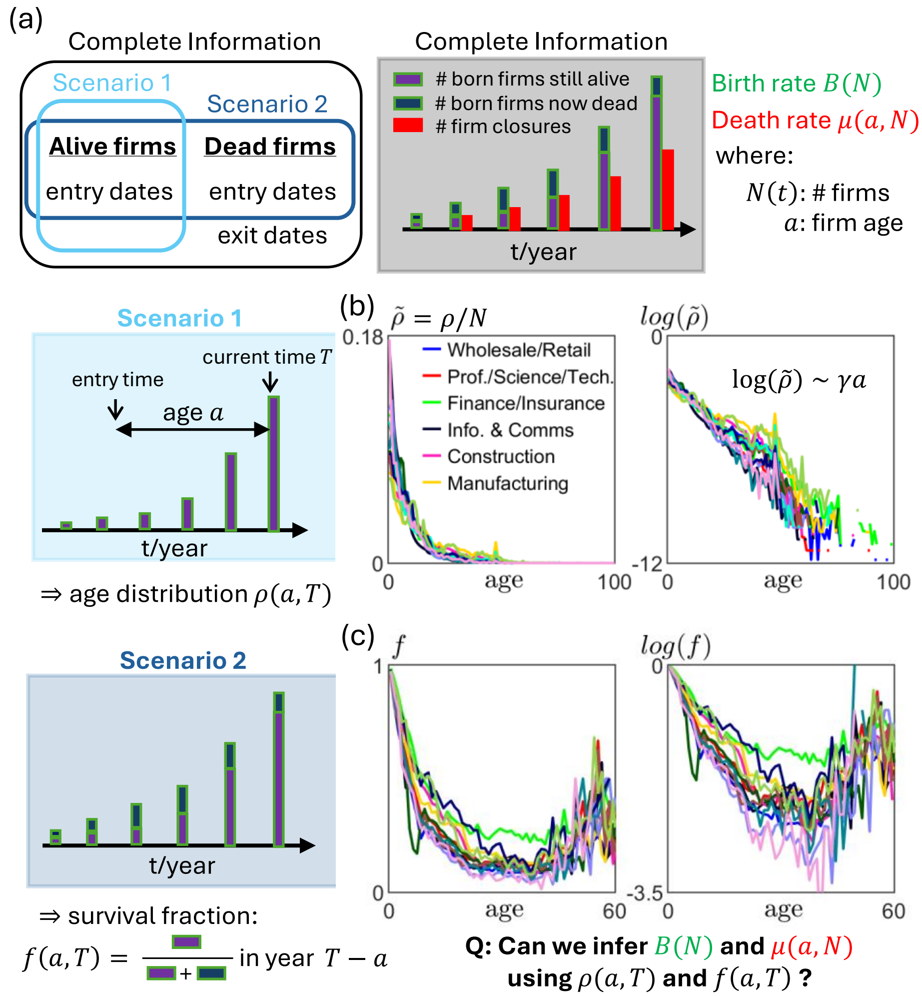

Fig. 1. Gaining insights into birth and death dynamics using incomplete data. (a) With complete information (entry and exit dates of all firms), one can directly extract birth rates $B(N)$ and death rates $\mu(N, a)$, where N is the number of firms and a is firm age. We consider two incomplete-data scenarios. In Scenario 1, only currently existing firms and their entry dates are available, from which the age distribution $\rho(a, T)$ can be computed. In Scenario 2, the dataset includes all firms and their alive status, but lacks exit dates; here, survival fractions $f(a, T)$ can be obtained. We examine whether birth and death models can be inferred in these cases. (b) Normalized age distribution $\tilde{\rho}(a) = \rho(a)/N$ (left) and $\log(\tilde{\rho}(a))$ (right) for each of the 14 largest firm sectors in Singapore. The approximate linear decay in $\log(\tilde{\rho}(a))$ implies that the age distribution approximately follows the exponential distribution. For clarity, only the 6 largest sectors are labeled. (c) The survival fraction $f(a)$ (left) and $\log f(a)$ of each firm sector. A non-linear $\log f(a)$ implies a non-constant death rate. We also ask how the observed non-monotonicity in $f(a)$ can arise.

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97635-3_24

2.2 Scenario 1: only age distributions are available

Prior studies in many countries have shown that firm age distributions often follow an exponential form [5, 6]. Using registration data for all active firms in Singapore (as of end-2022) and their sector classifications [1], we reconstruct $\rho(a)$ for the 14 largest sectors and find a similarly good approximation to the exponential distribution (Fig. 1b). This raises the question of how exponential age distributions arise. Proposed mechanisms include exponentially growing births with no deaths, or constant birth and death rates [5, 6]. A constant death rate with birth rates fluctuating around a constant mean can also yield exponential $\rho(a)$ [5]. However, these mechanisms may not be exhaustive. Notably, while deviations from exponential $\rho(a)$ was found to occur with age-dependent death rates under constant average birth rates [5], the effect of sector size-dependent death rates $\mu(N)$ remains unexplored. Here, we investigate N-dependent birth and death models (Eqs.2 and 1) and examine their resulting $\rho(a)$.

Constant birth rate. If the birth rate is constant $\left(\beta(N) = \frac{\beta_0}{N} \implies B(t) = \beta_0\right)$, and the death probability is age-independent ($\mu(a, N) = \mu(N)$), then from Eqs. 1- 2, $\frac{d \log \rho(a,T)}{da} = \frac{d \log f(a,T)}{da} = -\mu(N(T-a))$. Thus, if $\mu(N(T-a)) = \mu_0$, $\rho(a,T) \sim e^{-\mu_0 a}$, which is independent of β_0. This also implies that if births are approximately constant and $\rho(a)$ is exponential, then the death rate must also be approximately constant over time.

Constant birth rate coefficient. If $\beta(N) = \beta_0$, such that the overall birth rate $B(N) = \beta_0 N$, and $\mu(a, N) = \mu(N)$, then from Eqs. 1- 2, $\frac{d \log \rho(a,T)}{da} = -\frac{N'(T-a)}{N(T-a)} - \mu(N(T-a)) = -\beta_0$, where $N'(t) = \frac{dN(t)}{dt} = (\beta_0 - \mu(N))\, N$ is the dynamics of $N(t)$. Thus, $\rho(a,T) \sim e^{-\beta_0 a}$, regardless of whether $\mu(N)$ is constant or increases with the number of competitor firms.

These results show that while $\rho(a)$ constrains possible birth and death dynamics, multiple models can yield the same $\rho(a)$ shape. Distinguishing between them requires additional information. In particular, since both processes influence firm numbers, access to $N(t)$—e.g., via proxies like corporate tax records or revenue—can, together with $\rho(a)$, enable inference of both $B(N)$ and $\mu(N)$.

2.3 Scenario 2: survival fractions are also available

With all firms' entry dates and alive status, we can compute survival fractions $f(a)$ for each birth cohort, where a is the cohort age (Fig. 1b). Unlike the standard cohort-tracking approach (which gives survival fractions over time $f(t)$) [10, 13], our method does not need exit dates, which are missing in our dataset.

Across sectors, the $f(a)$ curves are convex and notably non-monotonic (Fig. 1c). While convexity has been observed in $f(t)$ [6, 10, 13], non-monotonicity is unique to our age-based $f(a)$, as $f(t)$ is always decreasing (by definition). This convexity and non-monotonicity in $f(a)$ are absent in $\rho(a)$ (Fig. 1b), and the shape difference between them implies that $B(t)$ is not a constant (Eq. 2).

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97635-3_24

Non-monotonic $f(a)$ suggests multiplicative interactions between the effects of age and sector size on death rate Since $f(a)$ depends solely on the death process, its shape can be used to infer properties of the death rate $\mu(a, N)$ (Eq. 1). In particular, a non-exponential $f(a)$, i.e., a non-linear $\log(f(a))$ curve (Fig. 1c, right), indicates that the death rate is not a constant.

To understand how a non-monotonic $f(a)$ can arise, we consider different forms of $\mu(a, N)$ and ask how they affect the sign of $\frac{d \log f(a,t)}{da}$ (using Eq. 1):

(i) **Effects of a and N are additive separable:** $\mu(a, N) = \mu_\alpha(a) + \mu_n(N)$. In this case, $\frac{d}{da}\mu(a, N) = \mu'_\alpha(a)$, and thus $\frac{d \log f(a,t)}{da} = -\mu_n(N(t-a)) - \mu_\alpha(a) < 0$ for all values of a, implying that $f(a)$ will always be monotonic. This includes cases where the death probability depends only on age ($\mu_n(N) = 0$) or sector size ($\mu_\alpha(a) = 0$). Hence, the non-monotonicity in the observed $f(a, t)$ implies an age- and N-dependent death probability.

(ii) **Effects of a and N are multiplicative separable:** $\mu(a, N) = \mu_\alpha(a)\mu_n(N)$. In this case, $\frac{d}{da}\mu(a, N) = \mu'_\alpha(a)\mu_n(N)$. Since $f(a, t)$ is non-monotonic in a if $\left.\frac{d \log f(a,t)}{da}\right|_{a \to t} > 0$, the condition for non-monotonicity is found to be: $\int_0^t \mu_\alpha(s)\mu'_n(N(s))N'(s)ds > \mu_\alpha(t)\mu_n(N(t))$. Since the number of firms grows over time ($N'(t) > 0$) and the death probability typically increases with competition ($\mu'_n(N) > 0$), this condition can be satisfied.

These results suggest a multiplicative interaction between age and sector size in determining firm death rates. We thus assume $\mu(a, N) = \mu_\alpha(a)\mu_n(N)$ and assess whether this model reproduces the observed survival and age distributions.

Hill functions for $\mu_\alpha(a)$ and $\mu_n(N)$ can reproduce data. We assume that death rates increase with sector size ($\mu'_n(N) > 0$) and decrease with age ($\mu'_\alpha(a) < 0$), consistent with higher exit rates for young firms [7, 11, 13]. To keep death probabilities bounded between 0 and 1, we use sigmoidal Hill functions:

$$\mu_\alpha(a) = \mu_{\alpha,ub} - (\mu_{\alpha,ub} - \mu_{\alpha,lb})\frac{a^{m_\alpha}}{a^{m_\alpha} + K_\alpha^{m_\alpha}} \tag{3}$$

$$\mu_n(N) = \mu_{n,lb} + (\mu_{n,ub} - \mu_{n,lb})\frac{N^{m_n}}{N^{m_n} + K_n^{m_n}}, \tag{4}$$

with $\mu_{\alpha,lb}$ $(\mu_{n,lb})/\mu_{\alpha,ub}$ $(\mu_{n,ub})$ being the lower/upper bounds; K_α (K_n) sets the midpoint; m_α (m_n) is the Hill coefficient controlling the transition steepness.

Since only the product $\mu_\alpha(a)\mu_n(N)$ affects dynamics, there are 7 effective parameters: $\mu_{\alpha,ub}\mu_{n,ub}$, $\tilde{\mu}_\alpha = \mu_{\alpha,lb}/\mu_{\alpha,ub}$, $\tilde{\mu}_n = \mu_{n,lb}/\mu_{n,ub}$, K_α, m_α, K_n, m_n. Given any parameter values and birth rate data, we solve for $N(t)$ and $f(a, T)$ using Eqs. 1-2. Starting from an initial guess, we iteratively adjust the model parameters to minimize the difference between predicted and observed survival fractions. Through inferring the best-fit parameters, we find that Hill functions for a decreasing $\mu_\alpha(a)$ and an increasing $\mu_n(N)$, combined multiplicative, can provide a good model for the death process (Fig. 2). By combining the predicted survival fractions with birth rate data, we also accurately recover the observed age distributions (Fig. 2c).

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97635-3_24

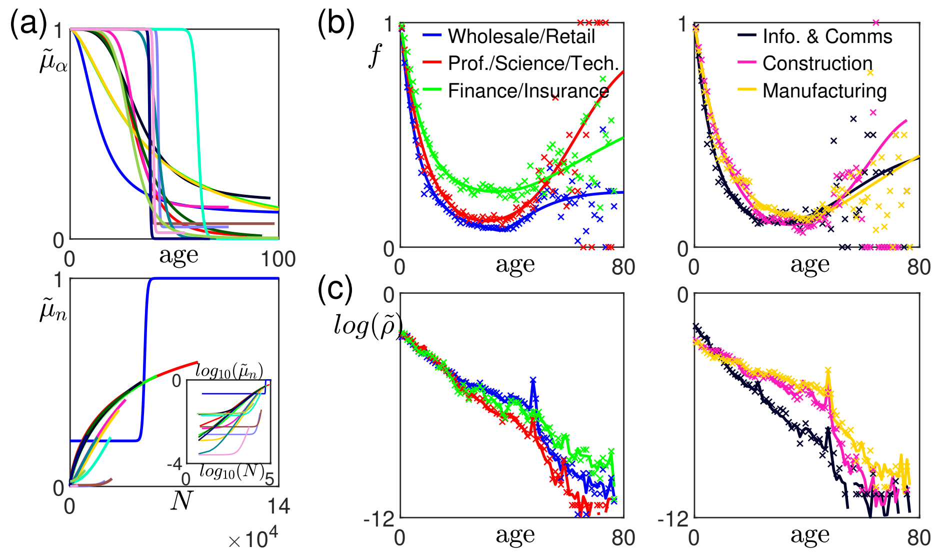

Fig. 2. Inferred death models accurately capture empirical data. (a) Inferred Hill functions for how death rates depend on age $\tilde{\mu}_\alpha(a)$ (Eq. 3) and sector size $\tilde{\mu}_n(N)$ (Eq. 4). (b) Inferred models (solid lines) give rise to $f(a)$ that agree well with the data ('x' markers). (c) Inferred models can recover observed normalized age distributions $\tilde{\rho}(a)$.

Inferred death models provide insights into other system properties With the inferred death models (Fig. 2a), we estimated the death rate of a newly born firm today, $\mu(0, N_f)$, where N_f is the current sector size. $\mu(0, N_f)$ falls between 0.08 and 0.15 and does not correlate with N_f (Fig. 3a). Such variation in survival rates across sectors could inform policies, e.g., subsidy allocation.

With both inferred models and birth data, we reconstructed sector sizes $N(t)$, revealing sharp growth transitions and shifts in the economy—while 'Wholesale/Retail Trade' sector has always been largest, 'Professional/Scientific/Technical Activities' only emerged as the second largest in the past 20 years (Fig. 3b). These $N(t)$ also allow model validation if other data sources become available.

We also recovered past age distributions $\rho(a, t)$ and used them to compute average death rates $\langle\mu(t)\rangle$. While $\langle\mu(t)\rangle$ rises over time for most sectors, the 'Wholesale/Retail Trade' sector has a stable death rate for many decades (Fig. 3c).

Using reconstructed $N(t)$ and birth data, we find an approximately linear birth model $B(N)$ across all sectors (Fig. 3d,e). Comparing the average birth coefficients $\beta = B/N$ with current death rates $\langle\mu\rangle$ reveals a positive correlation between them (Fig. 3f), consistent with studies in other countries [3, 4].

3 Discussion

Incomplete data is a common issue, with some detailed historical information lost for good. In such cases, inferring system dynamics from population-level

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97635-3_24

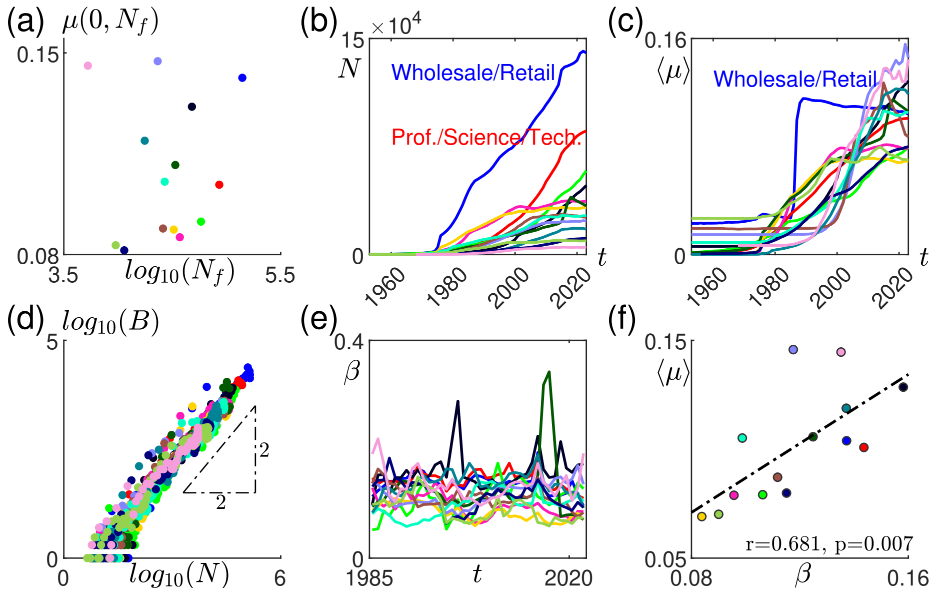

Fig. 3. Inferred death models reveal broader system features. (a) The current death rate of new firms, $\mu(0, N_f)$, shows no correlation with sector size N_f. (b) Reconstructed sector sizes $N(t)$ show rapid growth around 1980 and structural shifts— 'Wholesale/Retail Trade' stayed largest, while 'Professional/Scientific/Technical Activities' rose to second only in the last 20 years. (c) Average death rates $\langle\mu\rangle$ generally rose over time, except in 'Wholesale/Retail Trade', where rates stayed stable for 40 years. (d) Birth rates B scale approximately linearly with N across all sectors. (e) Birth rate coefficients $\beta = B/N$ are approximately constant in the last 40 years. (f) Sectors with higher average death rates tend have higher β (Pearson correlation coefficient $r = 0.681$, p-value $p = 0.007$).

statistics becomes invaluable. We show how firm age distributions and survival fractions can reveal underlying birth and death dynamics, even without data on extinct firms or exit dates. Though we focused on Singapore, our approach may apply to other countries and population types (e.g., ecological populations).

Age distributions alone constrain but do not fully determine underlying dynamics. In contrast, the convex, non-monotonic survival fraction curves suggest that death rates drop with age and rise with sector size, and both effects combine multiplicatively. Prior studies report differing age-dependence: death rates are constant in Japan but higher for young firms in the U.S. and many developing countries [9, 11]. Our findings offer a useful comparison and can guide incentive policies towards younger firms. Unlike age effects, sector size's influence on survival is less studied, and our results may prompt further exploration.

Other firm-level traits—e.g., size or profitability—can also affect survival [11]. While not explicitly modeled, some are implicitly captured (e.g., firm size may correlate with age). Our sector-level approach models average firm dynamics without tracking every micro-level detail. Nonetheless, these factors can lead to

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97635-3_24

heterogeneity in death rates and are critical for predicting individual firm survival. If such granular predictions are desired and the relevant data are available, our framework could be extended to include further dependencies and dynamics.

While we treated sectors independently and found that a common family of models captures their dynamics, cross-sector interactions exist [12]. Also, firm interactions depend on spatial proximity, with birth and death rates influenced by local competition or complementary businesses [14]. Incorporating such effects offers exciting directions for future work.

Acknowledgments. This work is supported by A*STAR Scholars' Development Fund (Y.G.) and Cities of Tomorrow Grant (H.H.N, F.L., project number CoT- H1-2025-3).

Disclosure of Interests. All authors declare that they have no competing interests.

References

1. List of firms in singapore, https://www.acra.gov.sg/about-bizfile/updates-and-announcements/acra-s-open-data-initiative [Last accessed: (2/29/25)]
2. Bartelsman, E., Scarpetta, S., Schivardi, F.: Comparative analysis of firm demographics and survival: evidence from micro-level sources in oecd countries. Industrial and corporate change **14**(3), 365–391 (2005)
3. Brixy, U., Grotz, R.: Regional patterns and determinants of birth and survival of new firms in western germany. Entrepreneurship and Regional Development **19**(4), 293–312 (2007)
4. Brown, J.P., Lambert, D.M., Florax, R.J.: The birth, death, and persistence of firms: Creative destruction and the spatial distribution of us manufacturing establishments, 2000–2006. Economic Geography **89**(3), 203–226 (2013)
5. Calvino, F., Giachini, D., Guerini, M.: The age distribution of business firms. Journal of Evolutionary Economics **32**(1), 205–245 (2022)
6. Coad, A.: Investigating the exponential age distribution of firms. Economics **4**(1), 20100017 (2010)
7. Coad, A.: Firm age: a survey. Journal of Evolutionary Economics **28**, 13–43 (2018)
8. Haviland, A., Savych, B.: A description and analysis of evolving data resources on small business (2007)
9. Ishikawa, A., Fujimoto, S., Mizuno, T., Watanabe, T.: The relation between firm age distributions and the decay rate of firm activities in the united states and japan. In: 2015 IEEE International Conference on Big Data (Big Data). pp. 2726–2731. IEEE (2015)
10. Mata, J., Portugal, P.: Life duration of new firms. The journal of industrial economics pp. 227–245 (1994)
11. McKenzie, D., Paffhausen, A.L.: Small firm death in developing countries. Review of economics and statistics **101**(4), 645–657 (2019)
12. O'Leary, D., Doran, J., Power, B.: The role of relatedness in firm interrelationships. Journal of Economic Studies **51**(9), 36–58 (2023)
13. Persson, H.: The survival and growth of new establishments in sweden, 1987-1995. Small Business Economics **23**, 423–440 (2004)
14. Rosenthal, S.S., Strange, W.C.: Evidence on the nature and sources of agglomeration economies. In: Handbook of regional and urban economics, vol. 4, pp. 2119–2171. Elsevier (2004)

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97635-3_24