



Deposited via The University of Leeds.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/135141/>

Version: Accepted Version

---

**Article:**

James, C, Azaele, S, Maritan, A et al. (2018) Zipf's and Taylor's laws. *Physical Review E*, 98 (3). 032408. ISSN: 2470-0045

<https://doi.org/10.1103/PhysRevE.98.032408>

---

© 2018 American Physical Society. This is an author produced version of a paper published in *Physical Review E*. Uploaded in accordance with the publisher's self-archiving policy.

**Reuse**

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

**Takedown**

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing [eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.

# Zipf's and Taylor's laws

Charlotte James

*Department of Engineering Mathematics, University of Bristol, Woodland Road, BS8 1UB, Bristol, UK*

Sandro Azaele

*School of Mathematics, University of Leeds, Leeds, UK.*

Amos Maritan

*Dipartimento di Fisica "G. Galilei", Università di Padova, INFN, via Marzolo 8, 35131 Padova, Italy.*

Filippo Simini\*

*Department of Engineering Mathematics, University of Bristol,  
Merchant Venturers Building, Woodland Road, BS8 1UB, Bristol, UK*

(Dated: August 24, 2018)

Zipf's law states that the frequency of an observation with a given value is inversely proportional to the square of that value; Taylor's law, instead, describes the scaling between fluctuations in the size of a population and its mean. Empirical evidence of the validity of these laws has been found in many and diverse domains. Despite the numerous models proposed to explain the presence of Zipf's law, there is no consensus on how it originates from a microscopic process of individuals dynamics without fine tuning. Here we show that Zipf's law and Taylor's law can emerge from a general class of stochastic processes at the individual level, which incorporate one of two features: environmental variability, i.e. fluctuations of parameters, or correlations, i.e. dependence between individuals. Under these assumptions, we show numerically and with theoretical arguments that the conditional variance of the population increments scales as the square of the population and that the corresponding stationary distribution of the processes follows Zipf's law.

Formally, a random variable follows Zipf's law if its probability density function is a power law with exponent  $-2$ . Evidence for Zipf's law has been found in many and diverse empirical domains, including systems where observations correspond to groups of individuals and the variable of interest is the group size, such as the number of employees in firms [1] or the distribution of family names [2]. In particular, one of the most documented empirical findings in human geography is Zipf's law for the distribution of city sizes: the probability to find a city with a given number of dwellers,  $n$ , is inversely proportional to the square of that number:  $P(n) \sim n^{-1-\gamma}$ , where  $\gamma \simeq 1$ . The exponent  $\gamma \simeq 1$  has been shown to apply to city sizes both globally and historically with surprisingly small deviations. Zipf's law also applies to the distribution of population of larger regions, for example countries in Europe [3] as well as cities and counties in the United States (see Fig.1a).

A number of general mechanisms exist to account for the emergence of Zipf's law in various systems [2, 6, 7]. A series of recent works [8, 9] has demonstrated that models with latent variables can lead to Zipf's law without fine tuning by mixing together narrow distributions with very different means. However, these works focus on static systems, without an explicit time dependence. Several models have been proposed to explain how Zipf's law can emerge as the stationary distribution

of dynamical processes for the sizes of groups of individuals. These models can be divided into two classes based on the scale considered: mesoscopic models at the scale of the groups (e.g., cities) and microscopic models at the scale of the individuals (e.g., dwellers). The mesoscopic models are stochastic processes describing the evolution of the group's population as a whole. Examples of mesoscopic models are the random multiplicative process [10], also called Gibrat's law [11] or proportionate random growth in the economics literature [12, 13], and those based on the interplay between intermittency and diffusion [14]. Mesoscopic models are able to explain the emergence of Zipf's law without the need to fine-tune their parameters to specific values [15], however they are coarse-grained descriptions of population dynamics and lack an explicit link to the underlying microscopic processes. Microscopic models provide a more fundamental description because they are stochastic processes describing the events experienced by an individual, namely births, deaths and migrations, that ultimately determine the change in the size of a population. Examples of microscopic models include Yule's and Simon's models based on the rich-get-richer mechanism [16, 17], cluster growth and aggregation [18, 19], preferential migration to large aggregates [20], and network growth with redirection [21]. However, microscopic models are only able to produce the power law exponent  $\gamma = 1$  for specific values of their parameters.

Here we present a class of microscopic stochastic processes that are able to reproduce Zipf's law with exponent  $\gamma = 1$  without fine tuning. These processes are charac-

---

\* f.simini@bristol.ac.uk

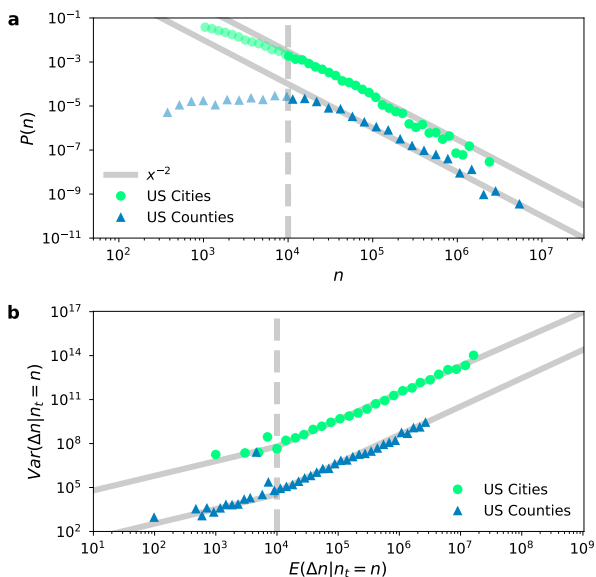


FIG. 1. a) Zipf's law, data. Probability density function  $P(n)$  (y-axis) vs size  $n$  (x-axis) for cities (circles) and counties (triangles) within the United States. The solid lines are guides for the eye corresponding to  $P(n) \propto n^{-2}$ . Data on the population of cities is obtained from the *Geonames* dataset [4]. County level data is obtained from the US Census Bureau [5]. The distribution for US cities has been shifted by a factor of 10 along the y-axis for clarity. b) Taylor's law, data. The variance of population in year  $t + 1$  conditioned to the population in year  $t$  (y-axis) vs the average population in year  $t + 1$  conditioned to the population in year  $t$  (x-axis) for cities (circles) and counties (triangles) in the United States during the period 1970 to 2010. The vertical dashed line denotes the cross-over city size  $n_c$  at which Taylor's exponent defined in Eq. 2 transitions from  $\alpha = 1/2$  to  $\alpha = 1$  (the best fit exponents in the regime  $n > n_c$  are  $\alpha = 0.92 \pm 0.01$  for cities and  $\alpha = 0.97 \pm 0.03$  for counties). As shown in panel a),  $n_c$  corresponds to the cross-over city size at which the distributions start following Zipf's law. The data for US cities have been shifted by a factor of 10 along the y-axis for clarity.

terised by an anomalous scaling of the fluctuations of the population increments, commonly known as Taylor's law in ecology [22–24]. In particular, we describe two general mechanisms to construct microscopic stochastic processes where the conditional variance of the population increments scale as the square of the population, and we demonstrate that the stationary distribution of such processes follows Zipf's law. Our processes are applicable to dynamical systems with an explicit time dependence where the stationary distribution of group sizes can be described by Zipf's law. In this respect, our derivation of Zipf's law differs from static models without an explicit time dependence and from models where groups can only grow [6, 8, 17, 25, 26].

We present our microscopic processes as models in which individuals belong to different local areas (groups or patches). To describe how the population in each group

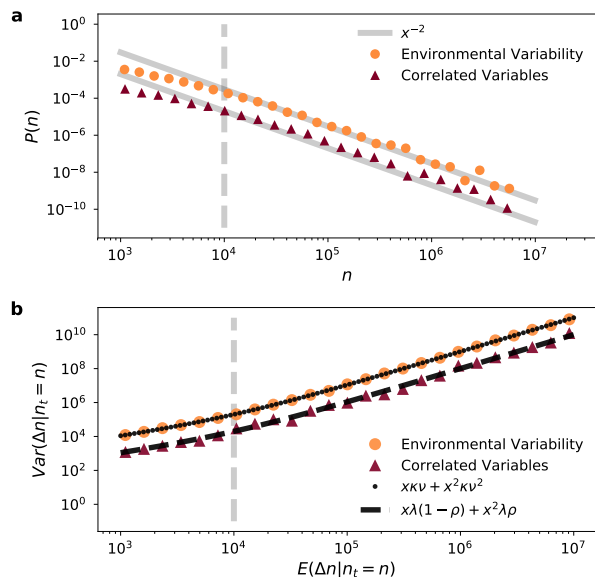


FIG. 2. a) Zipf's law, models. Stationary distributions of group sizes  $P(n)$  for modified branching processes with environmental variability (circles) and correlated individuals (triangles). See the main text for details on the numerical simulations and the parameter values. The distributions have been shifted by a factor of 10 along the y-axis for clarity. The vertical line denotes the cross-over city size  $n_c$  at which the distributions start following Zipf's law:  $n_c = 10^4$  for both models. b) Taylor's law, models. The variance of population change in time interval  $[t, t + 1]$  conditioned to the population at time step  $t$  (y-axis) vs the average population change in time interval  $[t, t + 1]$  conditioned to the population at time step  $t$  (x-axis) for branching processes with environmental variability (circles) and correlated variables (triangles). Simulations and parameter values are the same as in a). We observe a transition of Taylor's exponent from  $\alpha = 1/2$  for  $n < n_c$  to  $\alpha = 1$  for  $n > n_c$ . The black lines correspond to the analytical results of Eq. 3 and Eq. 5. Curves have been shifted by a factor of 10 along the y-axis for clarity.

evolves in time, we use a modified version of the Galton-Watson process. The original Galton-Watson process [27] is a discrete-time branching stochastic process describing the evolution of a population of  $n_t$  individuals at time  $t$  according to the equation

$$n_{t+1} = \sum_{i=1}^{n_t} x_i \quad (1)$$

where  $x_i$  are independent and identically distributed random variables over the integers, with finite mean and variance and with probability mass function  $P(x|\lambda)$ , where  $\lambda$  are parameters. If  $x_i = 0$ , individual  $i$  dies, if  $x_i = 1$ , they do nothing and if  $x_i = 2, 3, 4, \dots$  then they have 1, 2, 3, ... children. Extinction will occur with probability 1 if the average number of offspring per individual is less than or equal to one,  $\mathbb{E}(x) \leq 1$ . To avoid extinction and ensure the process has a stationary state we include

Distribution	$\gamma$	$D$	$p$
US Counties	0.86	0.04	< 0.01
US Cities	0.96	0.07	0.01
Environmental variability	0.91	0.02	< 0.01
Correlated individuals	1.00	0.01	< 0.01

TABLE I. Power law fit: exponent  $\gamma$  for the distribution of group sizes obtained using a Maximum likelihood estimate, Kolmogorov-Smirnov statistic  $D$  and  $p$ -value for the PDF distributions in Figs 1a and 2a. All numbers are rounded to two decimal places.

the boundary condition  $n_t \geq 1$  for all  $t$ , which accounts for immigration [28]. If  $\mathbb{E}(x) > 1$ , then the population will experience an exponential growth and there will be no stationary state. For  $\mathbb{E}(x) \leq 1$ , the stationary distribution cannot be described by Zipf's law; it is a power law with exponent  $-1$  ( $\gamma = 0$ ).

The conditional mean and variance of the population increments, defined as  $\mathbb{E}(\Delta n|n_t = n)$  and  $Var(\Delta n|n_t = n)$  respectively, where  $\Delta n = n_{t+1} - n_t$ , can be used to measure the scaling of the fluctuations of the population size. Given that the random variables  $x_i$  are independent and identically distributed, the conditional mean and variance of population increments both scale as  $n$ :  $\mathbb{E}(\Delta n|n) \sim n$  and  $Var(\Delta n|n) \sim n$ . This scaling behaviour can be summarised considering the relationship between these two quantities, i.e. Taylor's law [29]:

$$Var(\Delta n|n) \propto \mathbb{E}(\Delta n|n)^{2\alpha}. \quad (2)$$

The exponent  $\alpha$  often takes a value of either  $1/2$ , as in the original Galton-Watson process, or  $1$ , as in random multiplicative processes. As random multiplicative processes are known to produce Zipf's law, this suggests Zipf's law can be present with exponent  $\gamma = 1$  when Taylor's exponent is  $\alpha = 1$ . On the other hand we hypothesise that when Taylor's exponent is  $\alpha = 1/2$  Zipf's law will not be present. This suggests a connection between Taylor's law and Zipf's law, which we characterise with analytical arguments and numerical simulations.

Formalising this intuition, we propose two variations of the Galton-Watson process, namely processes with environmental variability and processes with correlated individuals, and show that for large populations they have exponents  $\gamma = 1$  and  $\alpha = 1$ .

We first present the case of processes with environmental variability. To this end, we consider a modified Galton-Watson process where the parameters  $\lambda$  of the distribution of the individuals are not constant values but random variables drawn from a distribution  $G(\lambda)$  at each time step. To be specific, we assume that the  $x_i$  are Poisson random variables with distribution  $P(x|\lambda) = \text{Pois}(\lambda) = e^{-\lambda}\lambda^x/x!$ .

In order to determine the fluctuations of the population increments, we compute the variance of the population change,  $\Delta n = n_{t+1} - n_t$ , conditioned to the

population at time  $t$  by applying the law of total variance [30],  $Var(\Delta n|n_t = n) = \mathbb{E}_\lambda[Var_{\Delta n|\lambda}(\Delta n)] + Var_\lambda(\mathbb{E}_{\Delta n|\lambda}[\Delta n])$ . Here  $\mathbb{E}_{\Delta n|\lambda}$  and  $Var_{\Delta n|\lambda}$  denote the mean and variance of  $\Delta n$  for a fixed  $\lambda$ , and  $\mathbb{E}_\lambda$  and  $Var_\lambda$  denote the mean and variance with respect to  $G(\lambda)$ . We obtain,

$$Var(\Delta n|n_t = n) = n \mathbb{E}_\lambda(\lambda) + n^2 Var_\lambda(\lambda). \quad (3)$$

Note that the fluctuations in the size of a population are proportional to  $n$  and follow Eq. 2 with exponent  $\alpha = 1/2$  for small populations,  $n \ll n_c \equiv \mathbb{E}_\lambda(\lambda)/Var_\lambda(\lambda)$ , whereas they scale as  $n^2$  for large populations  $n \gg n_c$ . The crossover population  $n_c$  marks the transition between these two scaling regimes. Empirical evidence of the presence of this crossover can be found analysing the fluctuations of the populations of cities and counties in the United States, where data shows a transition between Taylor exponents  $\alpha = 1/2$  and  $\alpha = 1$  around  $n_c \sim 10^4$  (Fig. 1b). Within our framework, this result means that on average the ratio between variance and mean of growth rates is very small, around  $n_c^{-1} = Var_\lambda(\lambda)/\mathbb{E}_\lambda(\lambda) \sim 10^{-4}$ . For example, if the distribution of  $\lambda$  is a Gamma distribution,  $G(\lambda) = \text{Gamma}(\lambda|\kappa, \nu)$ , then Eq. 3 becomes  $Var(\Delta n|n_t = n) = n \kappa \nu + n^2 \kappa \nu^2$  and  $n_c = \nu^{-1}$ . We verify the validity of this prediction with numerical simulations of Eq. 1 shown in Fig 2b, where the  $x_i$  are independent and identical Poisson random variables with distribution  $\text{Pois}(\lambda)$  and a new parameter  $\lambda$  is drawn at each discrete time step from a Gamma distribution with fixed parameters  $\kappa = 10^4$  and  $\nu = 10^{-4}$ .

Next we demonstrate that the stationary distribution of group sizes follows Zipf's law for large populations, as shown in Fig 2a and summarised in Table I. When the total population is sufficiently large, we expect that Eq. 1 can be approximated as (see martingale Central Limit Theorem [31])

$$\Delta n \equiv n_{t+1} - n_t \approx \hat{\mu}_n + \hat{\sigma}_n \xi(t), \quad (4)$$

where  $\Delta n(t)$  is a continuous random variable,  $\hat{\mu}_n = \mathbb{E}(\Delta n|n_t = n)$ ,  $\hat{\sigma}_n = \sqrt{Var(\Delta n|n_t = n)}$ , and  $\xi(t)$  is a zero-mean Gaussian white noise with autocorrelation  $\langle \xi(t)\xi(t') \rangle = 2\delta(t - t')$ . Figures 3a-b demonstrate that numerical simulations support the validity of the ansatz of Eq. 4. Using the law of total expectation and Eq. 3, we obtain  $\hat{\mu}_n = n(\mathbb{E}_\lambda(\lambda) - 1) \equiv n\hat{\mu}$  and  $\hat{\sigma}_n = \sqrt{n \mathbb{E}_\lambda(\lambda) + n^2 Var_\lambda(\lambda)} \equiv \sqrt{n\hat{\sigma}_1^2 + n^2\hat{\sigma}_2^2}$ . Using the formal substitutions  $t + 1 \rightarrow t + dt$ ,  $\hat{\mu} \rightarrow \mu dt$ ,  $\hat{\sigma}_1 \rightarrow \sigma_1 dt$  and  $\hat{\sigma}_2 \rightarrow \sigma_2 dt$  to first order in  $dt$ , we obtain the following stochastic differential equation:  $\dot{n}(t) = \mu n(t) + \sqrt{\sigma_1^2 n(t) + \sigma_2^2 n(t)^2} \xi(t)$ , with a reflecting boundary at  $n = 1$ . Hence, the proposed birth-death microscopic process can be well approximated by a mesoscopic proportionate random growth dynamics, when populations are large. In fact, for large populations,  $n \gg n_c = \sigma_1^2/\sigma_2^2$ , the above equation becomes a random multiplicative process with growth rate of mean  $\mu$  and variance  $\sigma_2^2$ . In this limit, we obtain a stationary

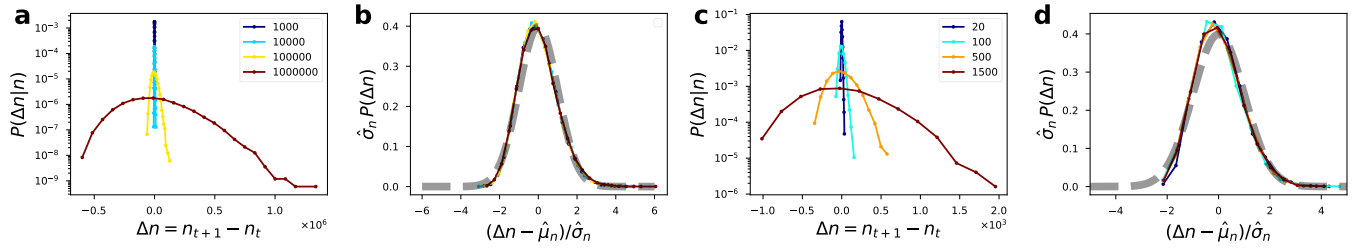


FIG. 3. Numerical simulations support the ansatz of Eq. 4 for processes with environmental variability (panels a,b), and with correlated individuals (panels c,d). a) Distribution of the fluctuations of the population increments in consecutive time steps,  $\Delta n = n_{t+1} - n_t$ , for a process with environmental variability where  $P(x|\lambda) = \text{Poiss}(\lambda)$  and  $G(\lambda) = \text{Gamma}(\lambda|\kappa, \nu)$  with  $\kappa = 19.8, \nu = 0.05$ . Different curves correspond to different values of populations  $n_t = n$ . b) The curves in panel a) collapse on the same distribution, a standard normal distribution (dashed line), when the population increments are shifted by removing the mean and rescaled by the square root of the variance derived in Eq. 3. c) Distribution of the fluctuations of the population increments in consecutive time steps for a process with correlated individuals, where the  $n$  Poisson random variables have covariance matrix  $\text{Cov}(x_i, x_j) = \delta_{ij}\lambda + (1 - \delta_{ij})\rho\lambda$  with  $\lambda = 0.999, \rho = 0.1$ . d) The curves in panel c) collapse on the standard normal distribution (dashed line) when the population increments are shifted by removing the mean and rescaled by the square root of the variance derived in Eq. 5.

distribution with a power-law tail  $P(n) \sim n^{-2 + \frac{\mu}{\sigma_2^2}}$  [32]. Notice that we get the exponent of Zipf's law,  $\gamma = 1$ , for  $|\mu| \ll \sigma_2^2$ . The parameter values used in the simulations of Fig. 2 are chosen to fit the models to city data, capturing both the Zipf exponent of the tail and the large value of  $n_c$ . Because of the specific choice of a Poisson distribution for  $P(x|\lambda)$ , this requires to use a set of parameter values that are very close to criticality (i.e.  $\lambda = 1$ ). However, it is important to emphasize that it is possible to find various distributions  $P(x|\lambda)$  and  $G(\lambda)$  that produce non-critical systems with stationary distributions with exponent  $\gamma = -2 + \frac{\mu}{\sigma_2^2}$  close to Zipf's law for any value of  $n_c = \sigma_1^2/\sigma_2^2$ . In general, for generic distributions  $P(x)$  and  $G(\lambda)$ , we have:

$$\begin{cases} \sigma_1^2 &= \mathbb{E}_\lambda (\text{Var}_{\Delta n|\lambda}(\Delta n)) \\ \sigma_2^2 &= \text{Var}_\lambda (\mathbb{E}_{\Delta n|\lambda}(\Delta n)) \\ \mu &= \mathbb{E}_\lambda (\mathbb{E}_{\Delta n|\lambda}(\Delta n)) - 1 \end{cases}$$

so it is possible to find many combinations of functions and parameter values such that  $|\mu| \ll \sigma_2^2 \ll \sigma_1^2$ , corresponding to Zipf's law and large  $n_c$ . For example, it might be possible to satisfy these conditions using a Negative Binomial distribution for  $P(x)$ , whose mean and variance can be independently set, and choosing an appropriate distribution over its parameters,  $G(\lambda)$ . The second case we present is the class of processes where individuals are correlated. To this end, we consider a modified Galton-Watson process with correlated individuals, where the joint probability  $P(x_1, \dots, x_n|\lambda)$  does not factorize into the product of the individual probabilities  $P(x_i|\lambda)$ . To be specific, we assume that the  $x$  random variables have a Poisson distribution with fixed parameter  $\lambda$  (the same for all individuals) and correlation matrix  $\rho_{ij} = \text{Cov}(x_i, x_j)/\lambda$ , where  $\text{Cov}(x_i, x_j) \equiv \mathbb{E}[(x_i - \lambda)(x_j - \lambda)]$ . When individuals are correlated the off-diagonal terms of the covariance matrix are non-zero, and in the simplest case that we consider, they are all

equal:  $\text{Cov}(x_i, x_j) = \delta_{ij}\lambda + (1 - \delta_{ij})\rho\lambda$ . The fluctuations of the population increments for the process with correlated individuals have the same scaling form of the fluctuations for the process with environmental variability. The conditional variance of  $\Delta n = n_{t+1} - n_t$  for a given population  $n_t = n$  is:

$$\text{Var}(\Delta n|n_t = n) = \sum_{i,j} \text{Cov}(x_i, x_j) = n\lambda(1 - \rho) + n^2\rho\lambda \quad (5)$$

Here the cross-over population between the regimes with Taylor exponents  $\alpha = 1/2$  and  $\alpha = 1$  is  $n_c = 1/\rho - 1$ . Within this framework, the cross-over population  $n_c \sim 10^4$  of United States cities and counties (Fig. 1b) corresponds to a very small value of correlation between people,  $\rho \sim 1/n_c \sim 10^{-4}$ . The physical meaning of this correlation can be interpreted considering that the inverse of the correlation matrix corresponds to the interaction between individuals [33, 34]. Since the off-diagonal elements of the inverse of the covariance matrix vanish in the small  $\rho$  limit we get that only a small amount of interaction is needed for large group sizes to follow Taylor's law with exponent  $\alpha = 1$ , yet in the absence of correlations this will not be present.

Using the ansatz of Eq. 4 and following the same steps taken for the case with environmental variability, it is possible to show that the process with correlated individuals has a stationary distribution of group sizes which for large populations,  $n > n_c$ , is a power-law,  $P(n) \sim n^{-2 + \frac{\lambda-1}{\rho\lambda}}$ , that follows Zipf's law when  $|\lambda - 1|/\lambda \ll \rho$ . Numerical simulations support the validity of the approximation of Eq. 4 on the scaling of the population increments (Figs. 3c-d and Fig. 2b) and the presence of Zipf's law in the distribution tail (Fig. 2a, Table I).

In the numerical simulations, we used two methods to generate correlated Poisson variables. In the first method, used in the simulations of Fig. 2, we consider

the modified Galton-Watson process  $n_{t+1} = \sum_{i=1}^{n_t} (x_i + z)$  where the  $x_i$  are independent and identical random variables with Poisson distribution  $\text{Pois}(\lambda(1 - \rho))$  and we introduce the random variable  $z$ , independent of the  $x_i$ , with Poisson distribution  $\text{Pois}(\rho\lambda)$ . One can show that for this process the conditional variance of  $\Delta n$  is identical to Eq. 5. In Fig. 2 we use  $\lambda = 1$  and  $\rho = 10^{-4}$ . In the second method, used in the simulations of Fig. 3c-d, we use a Gaussian copula model to link the Poisson marginals [35]. We verified with numerical simulations that for appropriate values of the multivariate Gaussian's parameters, this method generates Poisson variables with the desired values of  $\lambda$  and covariance matrix  $\text{Cov}(x_i, x_j) = C_{ij}$ ; the conditional variance of  $\Delta n$  is again identical to Eq. 5. Both methods give results compatible with the theoretical predictions.

To summarise our results, we have shown that envi-

ronmental variability and correlations between individuals are able to produce Zipf's law and Taylor's law [36]. The exponent of Taylor's law shifts from  $\alpha = 1/2$  for small populations to  $\alpha = 1$  for large populations. Also, when population is large Zipf's law emerges naturally without fine tuning whenever the growth rate's mean is much larger than the variance. The proposed framework can also naturally account for exponent values close but not identical to Zipf's law. This reveals a general connection between Zipf's law and Taylor's law in microscopic stochastic processes of population dynamics under realistic assumptions.

## ACKNOWLEDGMENTS

FS is supported by EPSRC First Grant EP/P012906/1.

- 
- [1] R. L. Axtell, *Science* **293**, 1818 (2001).  
 [2] M. E. Newman, *Contemporary physics* **46**, 323 (2005).  
 [3] K. Giesen and J. Südekum, *Journal of Economic Geography*, lbq019 (2010).  
 [4] B. Vatant and M. Wick, "Geonames ontology," (2012).  
 [5] U. C. Bureau, "County population totals," (1970-2010).  
 [6] D. Zanette and M. Montemurro, *Journal of quantitative Linguistics* **12**, 29 (2005).  
 [7] S. K. Baek, S. Bernhardtsson, and P. Minnhagen, *New Journal of Physics* **13**, 043004 (2011).  
 [8] L. Aitchison, N. Corradi, and P. E. Latham, *PLoS computational biology* **12**, e1005110 (2016).  
 [9] D. J. Schwab, I. Nemenman, and P. Mehta, *Physical review letters* **113**, 068102 (2014).  
 [10] S. Redner, *American Journal of Physics* **58**, 267 (1990).  
 [11] R. Gibrat, *Les inégalités économiques* (Recueil Sirey, 1931).  
 [12] D. Sornette, *Physical Review E* **57**, 4811 (1998).  
 [13] X. Gabaix, *The Quarterly Journal of Economics* **114**, 739 (1999).  
 [14] S. C. Manrubia and D. H. Zanette, *Physical Review E* **58**, 295 (1998).  
 [15] X. Gabaix, *Quarterly journal of Economics*, 739 (1999).  
 [16] G. U. Yule, *Philosophical Transactions of the Royal Society of London. Series B, Containing Papers of a Biological Character*, 21 (1925).  
 [17] H. A. Simon, *Biometrika*, 425 (1955).  
 [18] D. Rybski, A. G. C. Ros, and J. P. Kropp, *Physical Review E* **87**, 042114 (2013).  
 [19] H. A. Makse, J. S. Andrade, M. Batty, S. Havlin, H. E. Stanley, *et al.*, *Physical Review E* **58**, 7054 (1998).  
 [20] F. Leyvraz and S. Redner, *Physical review letters* **88**, 068301 (2002).  
 [21] G. F. Frasco, J. Sun, H. D. Rozenfeld, and D. Ben-Avraham, *Physical Review X* **4**, 011008 (2014).  
 [22] Z. Eisler, I. Bartos, and J. Kertész, *Advances in Physics* **57**, 89 (2008).  
 [23] M. A. De Menezes and A.-L. Barabási, *Physical review letters* **92**, 028701 (2004).  
 [24] A. Giometto, M. Formentin, A. Rinaldo, J. E. Cohen, and A. Maritan, *Proceedings of the National Academy of Sciences* **112**, 7755 (2015).  
 [25] F. Tria, V. Loreto, V. D. P. Servedio, and S. H. Strogatz, *Scientific reports* **4**, 5890 (2014).  
 [26] B. Corominas-Murtra, R. Hanel, and S. Thurner, *Proceedings of the National Academy of Sciences*, 201420946 (2015).  
 [27] T. E. Harris, *The theory of branching processes* (Courier Dover Publications, 2002).  
 [28] Modelling immigration using a boundary condition is a solution adopted in several models of population dynamics [37]. We verified that the specific implementation of the boundary (as a hard reflecting boundary or as a constant influx of individuals) and its value do not affect our results.  
 [29] Taylor's law originates from ecology, where it was originally defined as the scaling relationship between the variance and mean of populations in different areas [24, 38]. Since then, the name "Taylor's law" has been often used to denote the general scaling relationship between the variance and the mean of random variables in complex systems [22, 23]. In this case, the random variable of interest is the population increment in one time step,  $\Delta n$ , and we measure its mean and variance in patches of different sizes,  $n$ .  
 [30] N. A. Weiss, *A course in probability* (Addison-Wesley, 2006).  
 [31] A. N. Shiryaev, *Probability* (Springer New York, 1995).  
 [32] N. G. Van Kampen, *Stochastic processes in physics and chemistry*, Vol. 1 (Elsevier, 1992).  
 [33] S. Azaele, R. Muneeppeerakul, A. Rinaldo, and I. Rodriguez-Iturbe, *Journal of theoretical biology* **262**, 323 (2010).  
 [34] I. Volkov, J. R. Banavar, S. P. Hubbell, and A. Maritan, *Proceedings of the National Academy of Sciences* **106**, 13854 (2009).  
 [35] A. Barbiero and P. A. Ferrari, *Applied Stochastic Models in Business and Industry* **31**, 669 (2015).  
 [36] Indeed, numerical simulations show that the same results can be obtained when both mechanisms are present.

- [37] S. Azaele, S. Suweis, J. Grilli, I. Volkov, J. R. Banavar, and A. Maritan, *Reviews of Modern Physics* **88**, 035003 (2016).
- [38] L. Taylor, *Nature* **189**, 732 (1961).