GENUS, **LXIX** (No. 1), 25-60

JOEL E. COHEN* – MENG XU** – HELGE BRUNBORG***

Taylor's law applies to spatial variation in a human population

1. INTRODUCTION

The study of the spatial distribution of humans is a venerable and active field of demography (Duncan, 1957a, 1957b; Duncan, *et al*. 1963; Voss, 2007). Currently demographers often study changes in the human population density (number of people per unit of area) of different regions in terms of the births and deaths within each region and migration among regions (e.g., Rogers, 1995, 2008). Many ecologists share this perspective in studying the size or density of spatially separated populations of a single or of several nonhuman species (e.g., Gilpin and Hanski, 1991; Hanski, 1999; Dey and Joshi, 2006).

Another perspective, commoner in ecology than in human demography, focuses on the distribution of population size or density in an ensemble of populations. For example, Taylor (1961) and colleagues (Taylor, Woiwod and Perry, 1978, 1980; Taylor and Woiwod, 1980, 1982; Taylor, 1984; Perry and Taylor, 1985) observed that, in many species, the logarithm of the variance of the density (individuals per area or volume) of a set of comparable populations was an approximately linear function of the logarithm of the mean density: for some $a > 0$, log(variance of population density) = log(a) + $b \times$ log(mean population density). This relationship became known as Taylor's law (henceforth TL) or Taylor's power law of fluctuation scaling (Eisler *et al*., 2008). TL is one of the most widely verified empirical relationships in ecology.

Despite the abundance of demographic data on population density in different places and times, TL has hardly been tested in demography. This paper tests TL using high-quality human demographic data from Norway, 1978-2010, and proposes a simple model that accounts for some features of the results observed.

We know of only one prior test of TL using human demographic data. Taylor *et al*. (1978, p. 392, their Fig. 6a) plotted, on log-log coordinates, the spatial mean and spatial variance of human population using United States decennial census data from the "first census" (presumably of 1790) and 1900-1980. They

* The Rockefeller University and Columbia University, USA.
** The Rockefeller University, USA.
*** Statistics Norway.
Corresponding author: Joel E. Cohen; e-mail: cohen@rockefeller.edu.

did not specify whether the variable they used was population abundance or population density. They did not specify whether their spatial units were counties, states, or regions of several states. They gave no detailed information on how the population in 1980 was projected (Taylor *et al*., 1978, p. 393). They proposed no demographically interpretable model to explain their finding of a linear relationship between the log variance and the log mean. The limited scope and detail of this prior report leave ample room for testing and interpreting TL using human demographic data.

We focus on the parameter b of TL (the slope of log variance as a linear function of log mean population density) because b is independent of the unit used to measure population density. For example, if TL holds when population density is measured by people per km^2, then it will also hold with the same value of b when population density is measured by thousands of people per square mile. By contrast, the value of the parameter a depends on the units of measurement.

We take two approaches to estimating b. In the first, we fit a straight line to observations of the $\log_{10}$ mean and $\log_{10}$ variance of population density over different counties in a given year, for years 1978-2010. This approach measures and tests a spatial TL. Second, using each successive pair of years, we estimate the "local" slope as the ratio of the difference of $\log_{10}$ variance of population density from one year to the next divided by the difference of log10 mean of population density from one year to the next. It is intuitively clear that some average of the local slopes must approximate the slope of the spatial TL. The local slope gives temporally more refined information.

2. DATA, METHODS, AND THEORY

2.1 *Data*

Norway is divided geographically into five regions, which are further divided into 19 counties, one of which is the capital city, Oslo. The other 18 counties are further divided into about 430 municipalities. Norway's central statistical bureau Statistics Norway (SN) provided the population of every county yearly from 1978 to 2010 (33 years), based on the Central Population Register. SN also provided the land area of each county in 2006. We used the 2006 land area not covered by water.

With one exception, there were no changes in the boundaries of counties from 1978 to 2010. On January 1, 2002, 3,388 persons and 185 km2 of the municipality of Ølen were transferred from Hordaland county to Rogaland county. This change had no effect on the population density of 17 of the 19 counties, increased the population density of Hordaland by about 0.5% and reduced the population density of Rogaland by about 1.2%. These changes were comparable to the changes naturally occurring in these counties over one year (Table S1). We did not change the population densities of Hordaland and Rogaland for the years before 2002.

2.2 *Methods*

We let the number of counties be n and used the label $i = 1, \ldots, n$ to index the counties. When Oslo was included, $n = 19$. When Oslo was excluded, $n = 18$. We used the label $t = 1978, \ldots, 2010$ to index the years. Let $D_{i,t}$ be the population density (population size/area, number per km^2) of county i in year t. Let $w_{i,t}$ be a weight assigned to county i at time t such that $w_{i,t} > 0$, $w_{l,t} + \ldots + w_{n,t} = 1$, and let $\boldsymbol{w}$ be the vector of weights. We consider three sets of weights, specified below.

The weighted mean of population density over counties in year t is

$$M_{\boldsymbol{w}}(D_t) = \sum_{i=1}^{n} w_{i,t} D_{i,t} = \frac{\sum_{i=1}^{n} n w_{i,t} D_{i,t}}{n}.$$

The weighted variance of population density over counties in year t is

$$V_{\boldsymbol{w}}(D_t) = \sum_{i=1}^{n} w_{i,t} \left(D_{i,t} - M_{\boldsymbol{w}}(D_t)\right)^2 = \frac{\sum_{i=1}^{n} n^2 w_{i,t}^2 D_{i,t}^2}{n} - \left(M_{\boldsymbol{w}}(D_t)\right)^2$$
$$= \frac{\sum_{i=1}^{n} \left[n w_{i,t} D_{i,t} - M_{\boldsymbol{w}}(D_t)\right]^2}{n}.$$

This formula for the variance has n in the denominator on the right instead of n–1 because the counties were not sampled but were exhaustively enumerated.

The three sets of weights were equally weighted (denoted by Q), areally weighted (denoted by A), and population weighted (denoted by P). For Q, $w_{i,t} = 1/n$. For A, $w_{i,t} = a_i/a$ where a_i was the area of county i and a was the sum of the areas of the n counties considered. For P, $w_{i,t} = p_{i,t}/p_t$, where $p_{i,t}$ was the population of county i in year t and p_t was the sum of the populations of the n counties considered.

Least-squares linear and quadratic regressions of the dependent variable $\log_{10} V_w(D_t)$ as a function of the independent variable $\log_{10} M_w(D_t)$ were carried out using the JMP 9.0.1 platform called "Fit Y by X" (SAS Institute, 2010). All statistical significance tests used the critical value $P = 0.05$. All confidence intervals (CIs) were 95% CIs. Quadratic regression was used to test whether the dependent variable was a nonlinear function of the independent variable under each weighting method. When the evidence of nonlinearity was statistically significant, the relationship was either convex (positive coefficient of the quadratic term) or concave (negative coefficient of the quadratic term).

2.3 *Theory*

2.3.1 Taylor's law

We say that TL applies to population density $D(t)$ exactly for all times t if and only if there exist real constants $a > 0$ and b such that, for all t, $Var(D(t)) = a(E(D(t)))^b$. Equivalently, TL applies to$D(t)$ exactly for all t if and only if there exist constants $a > 0$ and b such that

$$\log Var\big(D(t)\big) - b \log E\big(D(t)\big) = \log a \ . \qquad [1]$$

The mean $E(D(t))$ and the variance $Var(D(t))$ refer to the mean and the variance over space at time t, not to a mean and variance over time t. Thus [1] specifies that $E(D(t))$ and $Var(D(t))$ satisfy exactly a spatial TL, not a temporal TL.

We say that TL applies to ($D(t)$ in the limit as t increases if there exist real constants $a > 0$ and b such that

$$\lim_{t\to\infty} [\log Var\big(D(t)\big) - b \log E\big(D(t)\big)] = \log a. \qquad [2]$$

If [2] holds, $E(D(t))$ and $Var(D(t))$ satisfy an asymptotic spatial TL, not a temporal TL. These definitions intentionally leave unspecified the base of the logarithms (e.g., e, 10, or 2) because TL is equally valid for logarithms to any base. For the following mathematical analysis, $\log = \log_e$ By contrast, we used $\log_{10}$ in the data analysis to make it easy to interpret the results. If the log mean and the log variance are differentiable functions of t, then taking the derivative with respect to t of both sides of [2] and switching the order of the derivative and the limit imply (Cohen, in press a)

$$\lim_{t\to\infty} [\frac{d \log Var\big(D(t)\big)}{dt} - b \frac{d \log E\big(D(t)\big)}{dt}] = 0, \qquad [3]$$

or equivalently

$$\lim_{t\to\infty} \left[\frac{\frac{dVar(D(t))}{dt}}{Var\big(D(t)\big)} - b \frac{\frac{dE(D(t))}{dt}}{E\big(D(t)\big)} \right] = 0. \qquad [4]$$

If $dE(D(t))/dt \neq 0$ (which means that the expected population density is growing or decreasing at t), define

$$b(t) = \frac{\left\{\frac{\frac{dVar(D(t))}{dt}}{Var(D(t))}\right\}}{\left\{\frac{\frac{dE(D(t))}{dt}}{E(D(t))}\right\}}. \qquad [5]$$

This $b(t)$ is the slope at a finite time t of log $Var(D(t))$ as a function of log $E(D(t))$. It will be called the local slope, to distinguish it from the large-time limit, defined next. If $\lim_{t\to\infty} b(t)$ exists and is a finite constant, that constant must be b in TL:

$$\lim_{t\to\infty}[b(t)] = b. \qquad [6]$$

In this case, TL holds in the limit of large time t. If $\lim_{t\to\infty} b(t)$ does not exist or is not a finite constant, TL does not hold in the limit as $t \to \infty$.

2.3.2 An exponential model

Suppose that the population density of each county changes exponentially (increasing or decreasing) at a fixed rate as time increases (Crow and Kimura, 1970, p. 10; Cohen, in press a). For $i = 1, \ldots, n$, county i has population density $D_i(0) > 0$ at $t = 0$, constant instantaneous rate of change r_i, and population size

$$D_i(t) = D_i(0)e^{r_i t} \qquad [7]$$

at every time t. Assume no two counties have the same constant instantaneous rate: $r_i \neq r_j$ if $i \neq j$ (this assumption held true of the estimated values of r_i in Norway). Some r_i might be positive, some negative, and at most one might be zero (since no two counties have the same exponential rate). Label the counties so that $+\infty > r_1 > \cdots > r_n > -\infty$. County i has a weight $w_{i,t}$, $0 < w_{i,t} < 1$, and $w_{i,t} + \cdots + w_{i,t} = 1$. Variability enters this deterministic model only through the parameter r_i, weight $w_{i,t}$ and population density $D_i(0)$ at time $t = 0$ of county i, $i = 1, \ldots, n$.

Under these assumptions, the weighted average (over the counties) of the population density at time t is

$$E(D(t)) = \sum_{i=1}^{n} w_{i,t} D_i(t) = \sum_{i=1}^{n} w_{i,t} D_i(0)e^{r_i t}. \qquad [8]$$

The weighted variance of population density among counties is

$$Var\big(D(t)\big) = E\big(D^2(t)\big) - \big[E\big(D(t)\big)\big]^2$$
$$= \textstyle\sum_{i=1}^{n} w_i\big(D_i(0)\big)^2 e^{2r_i t} - \big[E\big(D(t)\big)\big]^2. \qquad [9]$$

If all the weights $w_{i,t}$ are constant in time, then, in the limit of large time, the spatial TL [2] holds and the slope b is always 2 (Cohen, in press a):

$$\lim_{t\to\infty} b(t) = 2 = b. \qquad [10]$$

This limit $b = 2$ is independent of $D_i(0)$, r_i and w_i, $i = 1, \ldots, n$, when the r_i are all distinct and the $w_{i,t}$ and $D_i(0)$ are all positive and constant in time.

If counties are weighted by their population, i.e., if, for arbitrary positive constants a_i (here a_i is interpreted as the land area of county i),

$$w_i(t) = \frac{D_i(t)a_i}{\sum_{i=1}^{n} D_i(t)a_i},$$

then, from [5], $t \to \infty$,

$$b(t) \to 1 + \frac{r_2}{r_1} \qquad [11]$$

(proof in Appendix). We used the notation $w_{i,t}$ for data analysis and $w_i(t)$ in analysis of the model. This limit [11] is independent of $D_i(0)$ and a_i, $i = 1, \ldots, n$, when the r_i are all distinct and the a_i and $D_i(0)$ are all positive and constant in time. When all a_i are identical, w_i(t) is the weighting by population density.

The parameters of the exponential model for county i were estimated by fitting a least-squares line to $\log_{10}$ population density as the dependent variable and shifted calendar year (0 for 1978, 1 for 1979, …, and 32 for 2010) as the independent variable. Because the county population data described the population on January 1 of each year, we used both the data and the exponential model to calculate the mean $E(D(t))$ from [8] and the variance $Var(D(t))$ from [9] annually for t = 1978, …, 2010, and then computed a discrete analog $b^*(t)$ of the local slope for t = 1978, …, 2009:

$$b^*(t) = [\log_{10} Var\big(D(t+1)\big) - \log_{10} Var\big(D(t)\big)] / [\log_{10} E\big(D(t+1)\big) - \log_{10} E\big(D(t)\big)]. \qquad [12]$$

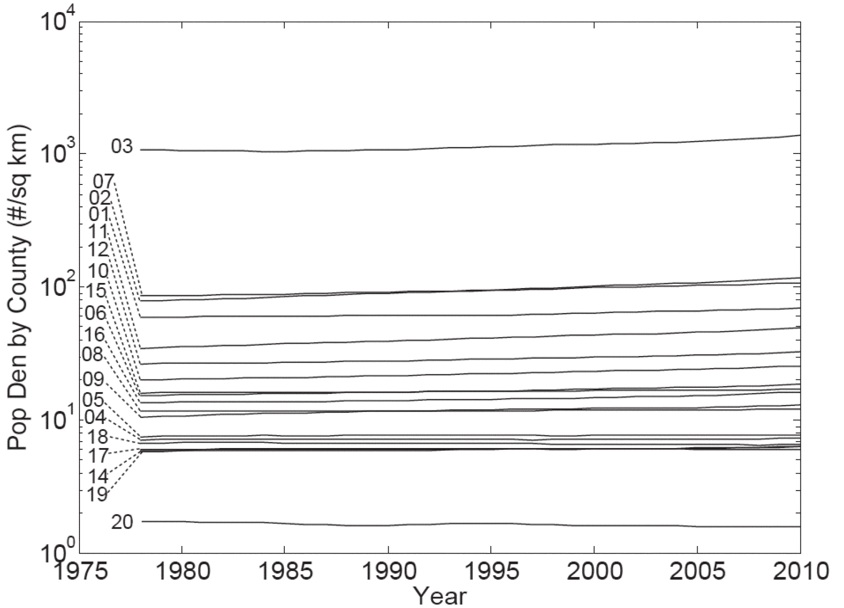
We compared this discrete local slope predicted by the exponential model with that from the data, using the same weighting for theory and data in each comparison. We also compared the relationship of log variance to log mean estimated from the data and the model under each of the three weightings.

3. RESULTS

3.1 *Population trajectories and exponential model*

From 1978 to 2010, the mean population densities of the counties varied by a factor of less than 26%, about 0.1 on a $\log_{10}$ scale. The population density of Akershus increased by nearly 49%, the population density of Finnmark decreased by nearly 8%, and some counties reversed the direction of demographic change once or repeatedly. Any reversal of the direction of change is qualitatively incompatible with the exponential model's assumption of a constant rate of growth or decline. However, the quantitative effect of such fluctuations in growth rate appeared to be small. On a log scale for population density and a linear scale for time, Figure 1 shows the population density of every county, including Oslo (03) at the top, from 1978 to 2010. The exponential model predicted that each county's trajectory should be a straight line with slope r_i for county i. On visual inspection, straight lines approximated each county's trajectory. When straight lines were fitted by least squares to the trajectories of all 19 counties, the linear regression fit between the independent variable t and the dependent variable $\log_e D_i(t)$ had R^2 from 0.4004 to 0.9964. Estimated slopes were all significantly non-zero. Two counties in the north, Nordland (18) and Finnmark (20), had negative slopes and 17 had positive slopes (Appendix Table S1).

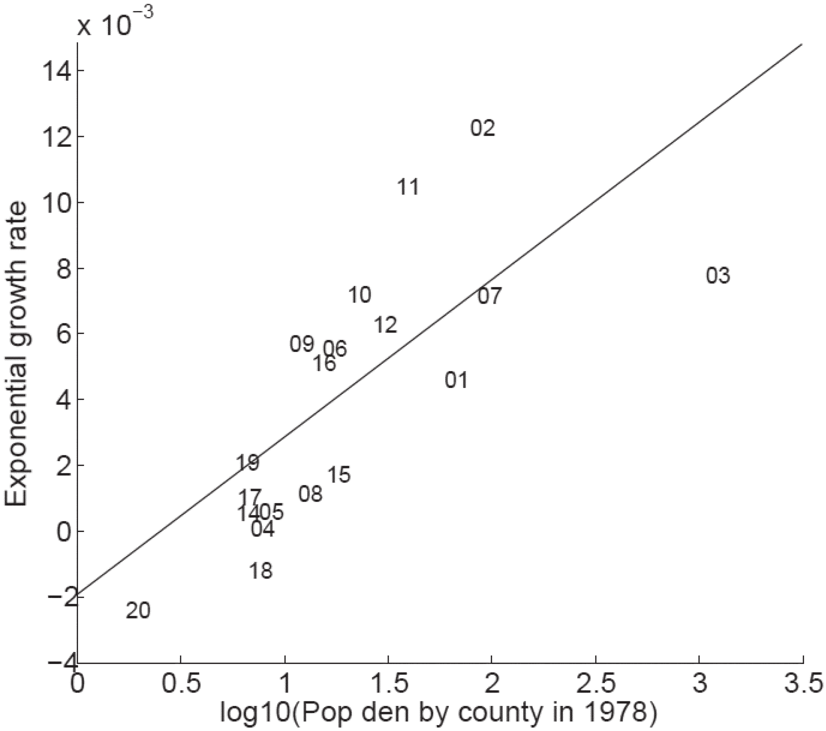
Figure 1 – *County population density, 1978-2010, Norway (logarithmic scale)*



Note: On these coordinates, exponential change would appear as a straight line. Counties are identified by their county number code (Appendix Table S1).

In general, the greater the initial population density $D_i(1978)$, the greater the growth rate r_i (Figure 2). The two counties with negative exponential parameters were initially among the least dense counties. Hence the denser the county initially, in general the greater the growth rate r_i. It is not surprising that, as the average population density rose in time, the variance of population density also increased.

Figure 2 – *Exponential rate of change against initial population density by county*

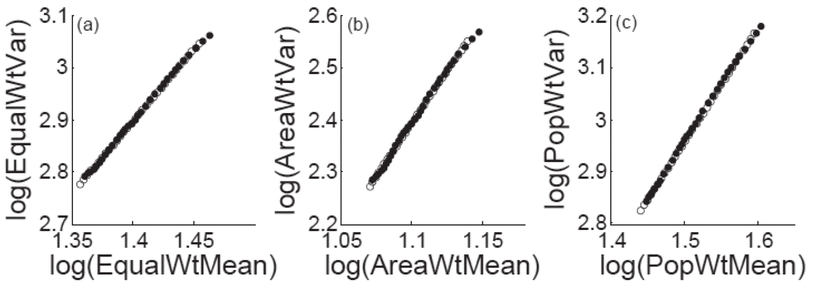


Note: Counties are identified by their county number code (Appendix Table S1). The solid line is the least-squares linear regression line: Exponential growth rate = –0.0019 (–0.0051, 0.0012) + 0.0048 (0.0025, 0.0071) × $\log_{10}$(Population density by county in 1978). 95% CI of the parameters are given after the corresponding point estimates. R^2 = 0.5407, adjusted R^2 = 0.5137 and root mean square error (RMSE) = 0.0028.

3.2 *Taylor's law excluding Oslo*

The means and variances of the population densities of Norway's 18 counties excluding Oslo agreed remarkably well with a spatial TL, whether the means and variances were equally weighted (Figure 3a), area weighted (Figure 3b), or population weighted (Figure 3c). The statistics of these regressions are summarized in Table 1. Despite the apparent linearity of the relationship, the evidence of concavity is statistically significant (e.g., equally weighted $P \approx 0.0004$, testing the null hypothesis of no quadratic term) for all three weightings but the effect is small enough to be almost invisible.

Figure 3 – *Observed variance of population density of 18 counties of Norway excluding Oslo (solid circles) as a function of observed mean of population density, on log_{10}-log_{10} coordinates, 1978-2010, using (a) equal weights, (b) areal weights, and (c) population weights, and predicted mean and variance (open circles) from an exponential model based on fitting a straight line to the log population density of each county separately*



Note: Data are shown in Figure 1 and Appendix Table S1. Points from the most recent years appear in the upper right corner. Early years are in the lower left. TL predicts a linear relationship. Statistics of each TL regression are summarized in Table 1.

For the entire period 1978-2010, for each weighting, the value of the spatial TL slope b was smaller when Oslo was included (Table 2) than when Oslo was excluded (Table 1). For no method of weighting was there overlap between the CIs obtained with and without Oslo. When Oslo was included in the analysis, for each weighted mean and variance, the estimate of the slope b using data from 1985 to 2010 was bigger, though not always significantly so, than the estimate of b using data from 1978 to 2010. This difference is due to the curvature in Figure 4 during 1978-1984.

When Oslo was included, the slope of spatial TL b was bigger than 2 when county population density was weighted equally or areally (Table 2). That $b > 2$ follows mathematically from the positive correlation (see Figure 2) between the exponential growth rate r_i and initial county population density $D_i(0)$ (proof in Appendix).

Table 1 – *Statistics of linear regressions of dependent variable* $\log_{10}$*(Variance of population density by county) on independent variable* $\log_{10}$*(Mean population density by county) excluding Oslo from the data and the model, 1978-2010*

Source	Weight	Slope			Intercept			Linear R^2	Quadratic pattern
		Estimate	95% CI		Estimate	95% CI			
			Lower bound	Upper bound		Lower bound	Upper bound		
Data	Equal	2.7667	2.7375	2.7959	-0.9772	-1.0182	-0.9361	0.9992	concave
	Area	3.9517	3.8696	4.0338	-1.9549	-2.0456	-1.8642	0.9968	concave
	Population	2.1916	2.1712	2.2120	-0.3321	-0.3630	-0.3011	0.9994	concave
Model	Equal	2.7772	2.7766	2.7777	-0.9920	-0.9928	-0.9911	1.0000	concave
	Area	3.9856	3.9737	3.9975	-1.9923	-2.0055	-1.9792	0.9999	concave
	Population	2.1995	2.1965	2.2025	-0.3460	-0.3506	-0.3415	1.0000	concave

Note: $R^2 > 0.99$ for any weighting method. Evidence for concavity in the data and the model is statistically significant ($P < 0.0004$) but visually small. The weights in the left column were used for the mean and variance. Using each weighting method, the 95% CI of the slope and the intercept from the model fell within the 95% CI of the corresponding parameters predicted from the data.

3.3 *Taylor's law including Oslo*

The capital Oslo had 10-12% of Norway's population and more people than any other county during 1978-2010 but only 0.14% of Norway's area. Its population density was above 1,000 people per km^2 and at least 10 times larger than that of any other county during this period. The population of Oslo decreased from 460,377 in 1978 to 447,257 in 1984 and increased 1985-2010, with accelerating growth in the last decade, largely because of immigration.

When Oslo was included in plots of log variance as a function of log mean population density, the 1978-1984 interval of Oslo's population decrease was visible as the small hook in the lower left of Figure 4a, where both the mean and the variance of population density declined. From 1985 to 2010, as Oslo's population grew, the mean and the variance of population density by county increased. In both periods, log variance was linearly related to log mean, but the slope was much higher in the earlier period than in the later. (Analysis of covariance (SAS Institute, 2010) rejected the null hypothesis of no effect of time period on the slope with $P < 0.001$.) When regressions were fitted separately to the equally weighted means and variances in 1978-84, the slope (± standard error, SE) was 4.9354 ± 0.2836, while for 1985-2010, the slope was 2.1835 ± 0.0111. During the earlier period, the relation of log variance to log mean was marginally concave, while during the later period the relation of log variance to log mean was convex (rejecting the null hypothesis of linearity with $P = 0.0194$). While the nonlinearity was statistically significant in the latter case, it was small quantitatively. Using areal weighting, the nonlinearity between log variance and log mean was significantly convex in 1978-1984 and not significant in 1985-2010. With population weights, the relation of log variance to log mean was significantly concave in 1978-1984 and significantly convex in 1985-2010. Again, in both periods, the nonlinearity could hardly be detected visually.

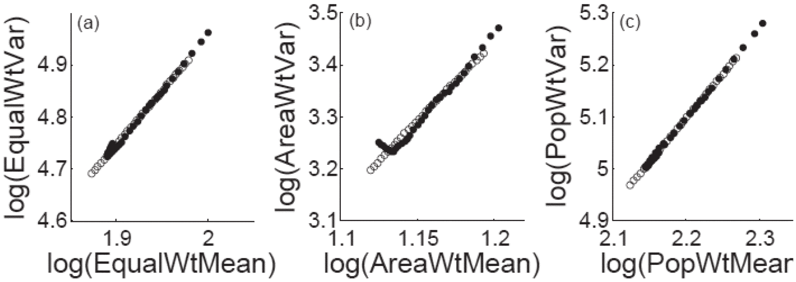
3.4 *Can the exponential model predict parameters of Taylor's law?*

We now test whether the exponential model can predict with useful accuracy the limiting slope b at large time and the local slope $b(t)$ in [5] at finite times.

3.4.1 Slope predicted for large time

In the exponential model, when the mean and variance are weighted by a constant, such as equally or by area, the local slope $b(t)$ converges at large time to 2 (Cohen, in press a, proves and explains this result). In the data, the equally weighted and areally weighted observed slopes for the counties of Norway were statistically significantly greater than 2, whether Oslo was excluded (Table 1) or included (Table 2). Hence the exponential model's limiting behavior predicted poorly the finite-time behavior of the data.

Figure 4 – *Observed variance of population density of 19 counties of Norway including Oslo (solid circles) as a function of observed mean of population density, on log_{10}-log_{10} coordinates, 1978-2010, using (a) equal weights, (b) areal weights, and (c) population weights, and predicted mean and variance (open circles) from an exponential model based on fitting a straight line to the log population density of each county separately*



Note: Points from the most recent years appear in the upper right corner. Early years are in the lower left. TL predicts a linear relationship, which was observed except for 1978-1984. Statistics of each TL regression are summarized in Table 2. Using areal weights (b) in the mean and variance changed the slope from positive to negative during 1978-1984. Using the population weighted mean and variance (c) made the rate of log variance decline with decreasing log mean in 1978-1984 ($b = 1.8628 \pm 0.0184$) similar to (though statistically significantly larger than, $P = 0.0224$) the rate that log variance rose with increasing log mean population density in 1985-2010 ($b = 1.7398 \pm 0.0051$), almost superimposing the two limbs of the curve.

In the exponential model, when the mean and variance are weighted by population size, the local slope $b(t)$ converges to $1 + r_2/r_1$, where r_1 is the largest exponential growth rate and r_2 is the second largest exponential growth rate (Appendix). This limit is always < 2, since $r_1 > r_2 > 0$. In the data, when Oslo was excluded (Table 1), the population-weighted slope for 1978-2010 was significantly > 2 statistically, in qualitative disagreement with the asymptotic theory. When Oslo was included (Table 2), the population-weighted slopes for 1978-2010, 1978-1984, and 1985-2010 were all significantly < 2 statistically, in qualitative agreement with the asymptotic theory.

In the latter case, it was worthwhile to test the asymptotic predicted slope quantitatively. The two counties in Norway with the first and second largest growth rates from 1978 to 2010 were Akershus and Rogaland with, respectively, $r_1 = 0.012293$ and $r_2 = 0.010512$. Therefore, as $t \to \infty$, using the population weighted mean and variance, the predicted slope in the limit was

$$b(t) \to 1 + \frac{r_2}{r_1} = 1 + \frac{0.010512}{0.012293} = 1.8551.$$

This value fell above the population-weighted 95% CIs for 1978-2010 and 1985-2010 of b in [2] (Table 2) when Oslo was included, and below the 95% CI when Oslo was excluded (Table 1). This value was bigger than the discrete local

Table 2 – *Statistics of linear regressions of dependent variable $\log_{10}$(Variance of population density by county) on independent variable $\log_{10}$(Mean population density by county) including Oslo in the data and the model, 1978-2010*

Source	Weighting	Interval	Slope			Intercept			Linear R^2	Quadratic pattern
			Estimate	95% CI		Estimate	95% CI			
				Lower bound	Upper bound		Lower bound	Upper bound		
Data	Equally	1978-2010	2.1108	2.0575	2.1640	0.7356	0.6331	0.8381	0.9953	convex
		1978-1984	4.9354	4.2064	5.6643	-4.6038	-5.9833	-3.2242	0.9838	concave (marginal)
		1985-2010	2.1835	2.1605	2.2065	0.5931	0.5486	0.6376	0.9994	convex
	Area	1978-2010	3.0985	2.8983	3.2988	-0.2691	-0.5006	-0.0377	0.9698	convex
		1978-1984	-2.0991	-2.5558	-1.6423	5.6114	5.0959	6.1269	0.9654	convex
		1985-2010	3.4543	3.3767	3.5318	-0.6864	-0.7766	-0.5963	0.9972	none
	Population	1978-2010	1.7183	1.7003	1.7363	1.3171	1.2774	1.3567	0.9992	convex
		1978-1984	1.8628	1.8155	1.9102	1.0088	0.9067	1.1109	0.9995	concave
		1985-2010	1.7398	1.7293	1.7502	1.2688	1.2457	1.2920	0.9998	convex
Model	Equally	1978-2010	2.0661	2.0682	2.0717	0.8219	0.8109	0.8176	1.0000	concave
		1978-1984	2.0848	2.0837	2.0860	0.7861	0.7839	0.7883	1.0000	concave
		1985-2010	2.0661	2.0645	2.0676	0.8219	0.8188	0.8249	1.0000	concave
	Area	1978-2010	3.0110	3.0266	3.0523	-0.1675	-0.2156	-0.1860	0.9999	concave
		1978-1984	3.1518	3.1428	3.1607	-0.3281	-0.3382	-0.3180	1.0000	concave
		1985-2010	3.0110	2.9997	3.0224	-0.1675	-0.1808	-0.1543	0.9999	concave
	Population	1978-2010	1.6949	1.6963	1.6985	1.3687	1.3606	1.3655	1.0000	concave
		1978-1984	1.7068	1.7061	1.7076	1.3429	1.3412	1.3445	1.0000	concave
		1985-2010	1.6949	1.6939	1.6959	1.3687	1.3665	1.3708	1.0000	concave

Note: Spatial Taylor's law describes $\log_{10}$ mean and $\log_{10}$ variance of population density in 19 counties of Norway (including Oslo) from the data and the model, for each of the three time intervals: 1978-2010, 1978-1984, and 1985-2010. The regression statistics shown are for linear models fitted by least squares to log variance as a linear function of log mean. For both the data and the model, under any weighting method, the 95% CI of slope for the earlier period (1978-1984) did not overlap the 95% CI of slope for the later period (1985-2010).

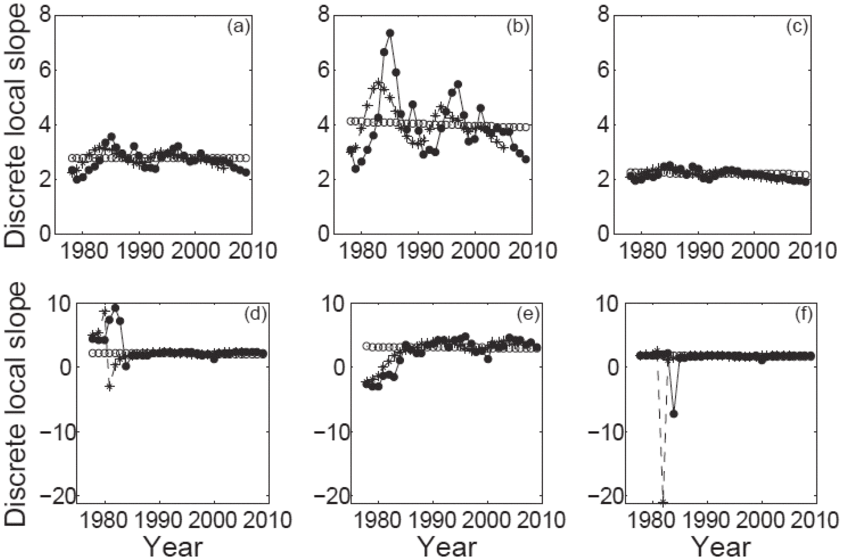
slope $b^*(2009) = 1.8290$ calculated from data according to [12] and fell above the corresponding 95% CI when Oslo was included (Table 3), and was smaller than $b^*(2009) = 1.9073$ and fell below the corresponding 95% CI when Oslo was excluded (Table 3).

In summary, the limiting slopes predicted by the exponential growth model, regardless of how the data were weighted, did not describe well the slopes in finite time for the counties of Norway, whether Oslo was included or not.

3.4.2 Slope predicted at finite times

For the years 1978, …, 2009, the discrete local slope [12] from the data using 1-year intervals (solid circles) and using 5-year intervals (asterisks) agreed reasonably well with that from the exponential model using 1-year intervals (open circles) whether Oslo was excluded (Figure 5a-c) or included (Figure 5d-f), using equal weighting (Figure 5a,d), areal weighting (Figure 5d,e), or population weighting (Figure 5c,f). When Oslo was excluded from the analysis, the discrete local slope from the data oscillated around that from the exponential model, where large oscillations occurred under areal weighting (Figure 5b). When Oslo was included in the analysis, around 1980, the discrete local slope deviated dramatically from that of the exponential model, under any weighting method.

Figure 5 – *Discrete local TL slope* [12] *from the data using 1-year intervals (solid circles) and 5-year intervals (asterisks), and from the exponential model using 1-year intervals (open circles)*



Note: Oslo was excluded in (a-c) and included in (d-f). Local slopes were calculated using equal weights (a, d), area weights (b, e) and population weights (c, f) respectively.

Table 3 – *Discrete local slopes b**(*t*) *from the data and from the model using* [12] *and 1-year intervals*

Source	Weighting	$b^*(t)$ excluding Oslo			$b^*(t)$ including Oslo		
		Mean	95% lower bound	95% upper bound	Mean	95% lower bound	95% upper bound
Data	Equal	2.6919	2.5595	2.8243	2.7884	2.1166	3.4602
	Area	3.9566	3.5449	4.3683	2.3004	1.4529	3.1479
	Population	2.1548	2.0953	2.2144	1.4406	0.8686	2.0126
Model	Equal	2.7771	2.7759	2.7783	2.0701	2.0663	2.0739
	Area	3.9876	3.9618	4.0134	3.0421	3.0142	3.0700
	Population	2.2000	2.1934	2.2065	1.6974	1.6950	1.6998

Note: The 95% CIs were calculated using normal theory. Under each weighting method, whether Oslo was excluded or included, the 95% CI from the model fell within that from the data, except when Oslo was included and equal weights were used. In this case, the 95% CI of the slope from the model fell below that from the data.

4. DISCUSSION

We report what we believe to be the first careful test of a spatial Taylor's law (TL) using high-quality data on human population density; the first comparative analysis (for any species or other data) of the consequences of different methods of weighting on the fit and parameters of TL; and the first empirical linkage of a simple model of exponential growth in population density with TL.

The principal finding is that a spatial Taylor's law (TL) described remarkably well the time course of spatial variation in the population density of the counties of Norway. The plots of log variance as a function of log mean in Figure 3, excluding Oslo, were as close to straight lines ($R^2 > 0.99$, Table 1) as any social scientific data we have seen. When Oslo was included, $R^2 > 0.96$ (Table 2).

The importance of this finding will depend in part on how well TL describes many other distributions of human population density (in e.g. the administrative subdivisions of many other countries, or the countries of the world by region or continent). If TL is successful, it could offer a new empirical regularity in human demography and a useful empirically tested baseline or standard against which to evaluate population projections at varied spatial scales. In this case, TL could be added to the ensemble of demographic techniques and models (like the exponential model and age-structured population models) shared by demographers of human and non-human populations. In any event, the introduction here of different weighting methods should be useful to future studies of TL in non-human demography and other fields of application.

Oslo is very different from the other counties. There is little empty land left in Oslo for housing and its density is roughly ten times that of the next densest county. Because of the exceptional situation of Oslo, we conducted two analyses, excluding Oslo (Figure 3) and including Oslo (Figure 4) as a county.

When Oslo was included, the relationship of log variance to log mean was non-linear for roughly a decade beginning in 1978, and linear regressions yielded different slopes for TL up to 1984 and after 1985, corresponding to the decline and growth of population in Oslo. When Oslo was excluded, the curvature disappeared and the linear relationship of log variance to log mean was remarkably precise throughout 1978-2010.

The hook in the lower left corners of Figure 4 deserves an interpretation. The short end of the hook arises from the years when the population of Oslo was declining (it fell from 1978 to 1984). Oslo's fall in population reduced the variance in population density among all counties, regardless of the method of weighting. However, from 1978 until 1992, the annual growth rate of Oslo's population increased because births and foreign immigration increased while net in-migration from the rest of Norway remained stable. Around 1970, Norway discovered oil in the North Sea. Within a few years the economy began to expand (Cappelen and Mjøset, 2012). The increasing migration to Oslo may have been associated with increasing employment opportunities resulting from the economic boom following the discovery of oil. From 1978 to 2010, the two counties with the largest rates of population growth (*r* in Table S1) were Akershus first and Rogaland second. Akershus shares the housing market and labor market of Oslo, and Rogaland on the North Sea was most directly affected by the discovery of oil. The population dynamics of the other counties were consistent before and since Oslo's demographic resurgence, judged by Figure 3, but all counties were positively affected by the strongly increasing immigration to Norway after 2004, when the European Union expanded to Eastern Europe. Norway, although not a member of the EU, is a member of the European Economic Area (EEA), which requires Norway to comply with EU regulations on the employment of labor from EU countries. This example suggests that non-demographic factors may influence the relationship between the spatial log variance and log mean of population density, but also that, in a stable environment (since the early 1980s), changes in the spatial distribution of population density can follow TL with remarkable precision.

In addition, the parameters of TL were sensitive to the weighting used in computing the mean, variance, and regressions. Although more than 1,000 papers have reportedly been published on TL (Eisler *et al*., 2008), we have not seen a prior examination of the influence on TL of different methods of weighting the data. For studies of land use, agriculture, forestry, and conservation, areal weighting may be appropriate because areal weighting considers the extent of land subject to each local human population density. On the other hand, for studies of urbanization, urban-rural dynamics, and migration, population weighting may be appropriate because population weighting considers the numbers of people who experience the local population density. The population-weighted mean population density is greater than or equal to the area-weighted mean population density, and the two are equal if and only if the population density of every county (or other unit) is equal (Cohen, in press b). Equal weighting gives precedence to political or administrative boundaries, rather than to the area of land or number of people affected by the local population density, and may be appropriate for political or administrative purposes, but should not be used automatically when the purpose of a study is ecological or sociological. When administrative boundaries reflect ethnic or linguistic boundaries, administrative units may be meaningful and natural for further analysis, but boundaries drawn along rivers may divide homogeneous watersheds and straight boundary lines drawn on maps may reflect no more than history and convenience.

The choice of any spatial scale of analysis is in part arbitrary. To evaluate TL at different spatial scales, we studied population density in Norway at a scale smaller than the county (namely, the municipality) and at a scale larger than the county (namely, the region). On each scale, we analyzed time series of population density, a spatial TL, and discrete local slopes, using data and the exponential model as on the county scale. In general, at the scales of the municipality and the region, the exponential model remained useful as a summary of the population trajectories, a spatial TL remained useful as a summary of the relation between the variance and mean of population density, and the local slope from the exponential model remained useful as an approximation to the local slope from the data, under all three methods of weighting, whether Oslo was included or excluded. We saw no clear relationship between the spatial scale of the unit of analysis (increasing from municipality to county to region) and the size of the regression slope or local slope, for any of the three weightings. However, when Oslo was excluded, the size of the regression slope or local slope was always smaller for municipalities than for counties by any weighting, in both the data and the exponential model. Equivalently, for a given proportional increase in mean population density, the increase in variance among municipalities was smaller than the increase in variance among counties, reflecting perhaps lesser demographic divergence of municipalities than of counties.

In a prior test of TL using demographic data, Taylor *et al*. (1978, p. 406, Appendix B) estimated a slope of $b = 2.04 \pm 0.01$ (estimate ± S.E.) for U.S.A. decennial census data and one projection, with no statistically significant evidence of curvature from a GLIM model. They stated no weighting method. Their slope fell just below our 95% CI for the equally weighted slope from 1978-2010 when Oslo was included (Table 2) and substantially below our 95% CI for the equally weighted slope from 1978-2010 when Oslo was excluded (Table 1). Nevertheless, their estimated slope and our equally weighted slopes all exceeded 2. Future analysis of U.S. census data at different scales of spatial aggregation (e.g., regions, states, counties, census tracts) could test the generality of these findings.

A practical application of TL would be to compare the parameters of TL estimated from historical data with the parameters of TL estimated from projections of future county populations. SN prepares such projections regularly under a range of assumptions about fertility, mortality, and migration, internal and external (Brunborg *et al*., 2012). It would be interesting to see whether these projections agree with TL as well as the historical data agree with TL and, if so, whether the parameters of one projection agree more closely with the historical parameters than do the parameters of another projection. In such a case, TL might help evaluate the relative plausibility of alternative projections. In such an application of TL, it would be essential to recognize modestly that external factors can change the direction of a trend, just as the discovery of North Sea oil was followed by a change in the slope of TL in the decade after 1978.

Acknowledgement

J.E.C. acknowledges with thanks the support of U.S. National Science Foundation grants EF-1038337 and DMS-1225529, and the assistance of Priscilla K. Rogerson. M.X. acknowledges with thanks the support of U.S. National Science Foundation grant EF-1038337.

References

BRUNBORG H., TEXMON I., TØNNESSEN M. (2012), “Population projections 2012-2100: Models and assumptions”, *Økonomiske analyser* 4/2012, 32-40: (in Norwegian), http://www.ssb.no/befolkning/artikler-og-publikasjoner/befolkningsframskrivning-2012-2100-modeller-og-forutsetninger; (in English), “Rapidly towards 6 million people”, Statistics Norway, http://www.ssb.no/english/subjects/02/03/folkfram_en/.

CAPPELEN Å., MJØSET L. (2012), “Can Norway be a role model for natural resource abundant countries?”, in FOSU A.K., *Development Success. Historical Accounts from More Advanced Countries*, Oxford University Press, Oxford, 44-72. http://www.wider.unu.edu/publications/working-papers/research-papers/2009/en_GB/rp2009-23/

COHEN J.E. (in press a), “Taylor’s power law of fluctuation scaling and the growth-rate theorem”, *Theoretical Population Biology* (accepted for publication 4 May 2013).

COHEN J.E. (in press b), “Cauchy inequalities for the spectral radius of products of diagonal and nonnegative matrices”, *Proceedings of the American Mathematical Society* (accepted for publication 18 December 2012).

DEY S., JOSHI A. (2006), "Stability via asynchrony in Drosophila metapopulations with low migration rates", *Science*, 312, 434-436. http://www.sciencemag.org/content/312/5772/434.full

DUNCAN O.D. (1957a), "The measurement of population distribution", *Population Studies*, 11, 27-45.

DUNCAN O.D. (1957b), "Population distribution and community structure", *Cold Spring Harbor Symposia on Quantitative Biology*, 22, 357-371.

EISLER Z., BARTOS I., KERTÉSZ J. (2008), "Fluctuation scaling in complex systems: Taylor's law and beyond", *Advances in Physics*, 57, 89-142. http://arxiv.org/abs/0708.2053

GILPIN M.E., HANSKI I. (eds.) (1991), *Metapopulation Dynamics: Empirical and Theoretical Investigations*, Academic Press, New York.

HANSKI I. (1999), *Metapopulation Ecology*, Oxford University Press, Oxford, New York.

PERRY J.N., TAYLOR L.R. (1985), "Ades: New ecological families of species-specific frequency distributions that describe repeated spatial samples with an intrinsic power-law variance-mean property", *Journal of Animal Ecology*, 54, 931-953. http://www.jstor.org/stable/4388

ROGERS A. (1995), *Multiregional Demography: Principles, Methods and Extensions*, John Wiley, New York, London.

ROGERS A. (2008), "Demographic modeling of the geography of migration and population: a multiregional perspective", *Geographical Analysis*, 40, 276-296.

SAS INSTITUTE INC. (2010), *Using JMP Version 9*, SAS Institute, Cary, NC.

TAYLOR L.R. (1961), "Aggregation, variance and the mean", *Nature*, 189, 732-735. DOI:10.1038/189732a0

TAYLOR L.R. (1984), "Assessing and interpreting the spatial distributions of insect populations", *Annual Review of Entomology*, 29, 321-357.

TAYLOR L.R., WOIWOD I.P. (1980), "Temporal stability as a density-dependent species characteristic", *Journal of Animal Ecology*, 49, 209-224.

TAYLOR L.R., WOIWOD I.P. (1982), "Comparative synoptic dynamics. I. Relationships between inter- and intra-specific spatial and temporal variance/mean population parameters", *Journal of Animal Ecology*, 51, 879-906.

TAYLOR L.R., WOIWOD I.P., PERRY J.N. (1978), "The density-dependence of spatial behaviour and the rarity of randomness", *Journal of Animal Ecology*, 47, 383-406, Stable URL: http://www.jstor.org/stable/3790.

TAYLOR L.R., WOIWOD I.P., PERRY J.N. (1980), "Variance and the large scale spatial stability of aphids, moths and birds", *Journal of Animal Ecology*, 49, 831-854.

VOSS P.R. (2007), "Demography as a Spatial Social Science", *Population Research and Policy Review*, 26, 457-476, DOI 10.1007/s11113-007-9047-4.

Appendix

This appendix proves two theoretical results and demonstrates two empirical findings. First, when counties are weighted by their population size or population density at time *t*, then the local slope *b*(*t*) of TL approaches $1 + r_2/r_1$ as $t\to\infty$, and approaches $1 + r_{n-1}/r_n$ as $t\to-\infty$, where $+\infty, > r_1 >\cdots> r_n > -\infty$, are the exponential rates of change of county population density. Second, positive correlation between the exponential rates and the initial county population density implies that $b(t) > 2$. Empirically, the exponential model and TL described well the variance and mean of the population density of municipalities (smaller than counties) and regions (larger than counties) separately.

A.1 *Theory*

A.1.1 Calculation of *b*(*t*) at finite time and in the limit of large time

An explicit formula for *b*(*t*) in [5] when the county weights w_i are time-independent (e.g., each county is weighted equally or by area) is

$$b(t) = \frac{2\left[\frac{\left(\sum_{i=1}^{n} w_i r_i \left(D_i(0)\right)^2 e^{2r_i t}\right) - \left(\sum_{i=1}^{n} w_i r_i D_i(0) e^{r_i t}\right)\left(\sum_{i=1}^{n} w_i D_i(0) e^{r_i t}\right)}{\left(\sum_{i=1}^{n} w_i \left(D_i(0)\right)^2 e^{2r_i t}\right) - \left(\sum_{i=1}^{n} w_i D_i(0) e^{r_i t}\right)^2}\right]}{\frac{\sum_{i=1}^{n} w_i r_i D_i(0) e^{r_i t}}{\sum_{i=1}^{n} w_i D_i(0) e^{r_i t}}} \qquad [A1]$$

as $t\to\infty$, $b(t)\to 2$ (Cohen, in press a).

Here we derive *b*(*t*) when $w_i = w_i(t)$, e.g., counties are weighted by their population size or population density at time *t*. From

$$E\big(D(t)\big) = \sum_{i=1}^{n} w_i(t) D_i(t), \qquad [A2]$$

we have

$$\frac{dE\big(D(t)\big)}{dt} = \sum_{i=1}^{n} \left(\frac{dw_i(t)}{dt} \cdot D_i(t) + w_i(t) \cdot \frac{dD_i(t)}{dt}\right). \qquad [A3]$$

From

$$Var\big(D(t)\big) = \sum_{i=1}^{n} w_i(t)\big(D_i(t)\big)^2 - \left(\sum_{i=1}^{n} w_i(t)\, D_i(t)\right)^2, \qquad [A4]$$

we have

$$\frac{dVar(D(t))}{dt} = \sum_{i=1}^{n}\left(\frac{dw_i(t)}{dt}\cdot\left(D_i(t)\right)^2 + 2w_i(t)D_i(t)\frac{dD_i(t)}{dt}\right) - 2\left(\sum_{i=1}^{n} w_i(t)\,D_i(t)\right)\cdot\cdot \sum_{i=1}^{n}\left(\frac{dw_i(t)}{dt}\cdot D_i(t) + w_i(t)\cdot\frac{dD_i(t)}{dt}\right). \quad [A5]$$

When counties are weighted by their population size,

$$w_i(t) = \frac{a_iD_i(t)}{\sum_{i=1}^{n} a_iD_i(t)}. \quad [A6]$$

The exponential model supposes that

$$\frac{dD_i(t)}{dt} = r_iD_i(0)e^{r_it} = r_iD_i(t). \quad [A7]$$

Hence

$$\frac{dw_i(t)}{dt} = \frac{a_i\frac{dD_i(t)}{dt}\cdot\sum_{i=1}^{n} a_iD_i(t) - a_iD_i(t)\cdot\sum_{i=1}^{n} a_i\frac{dD_i(t)}{dt}}{\left(\sum_{i=1}^{n} a_iD_i(t)\right)^2}$$

$$= \frac{a_ir_iD_i(t)\cdot\sum_{i=1}^{n} a_iD_i(t) - a_iD_i(t)\cdot\sum_{i=1}^{n} a_ir_iD_i(t)}{\left(\sum_{i=1}^{n} a_iD_i(t)\right)^2}. \quad [A8]$$

Then

$$\frac{dw_i(t)}{dt}\cdot D_i(t) + w_i(t)\cdot\frac{dD_i(t)}{dt} =$$

$$= \frac{a_ir_iD_i(t)\cdot\sum_{i=1}^{n} a_iD_i(t) - a_iD_i(t)\cdot\sum_{i=1}^{n} a_ir_iD_i(t)}{\left(\sum_{i=1}^{n} a_iD_i(t)\right)^2}\cdot D_i(t) + \frac{a_iD_i(t)}{\sum_{i=1}^{n} a_iD_i(t)}\cdot r_iD_i(t) =$$

$$\frac{a_ir_i\left(D_i(t)\right)^2\cdot\sum_{i=1}^{n} a_iD_i(t) - a_i\left(D_i(t)\right)^2\cdot\sum_{i=1}^{n} a_ir_iD_i(t) + a_ir_i\left(D_i(t)\right)^2\cdot\sum_{i=1}^{n} a_iD_i(t)}{\left(\sum_{i=1}^{n} a_iD_i(t)\right)^2} =$$

$$\frac{2a_ir_i\left(D_i(t)\right)^2\cdot\sum_{i=1}^{n} a_iD_i(t) - a_i\left(D_i(t)\right)^2\cdot\sum_{i=1}^{n} a_ir_iD_i(t)}{\left(\sum_{i=1}^{n} a_iD_i(t)\right)^2}. \quad [A9]$$

Also

$$\frac{dw_i(t)}{dt}\cdot\left(D_i(t)\right)^2 + 2w_i(t)D_i(t)\frac{dD_i(t)}{dt} =$$

$$= \frac{a_ir_i\left(D_i(t)\right)^3\cdot\sum_{i=1}^{n} a_iD_i(t) - a_i\left(D_i(t)\right)^3\cdot\sum_{i=1}^{n} a_ir_iD_i(t)}{\left(\sum_{i=1}^{n} a_iD_i(t)\right)^2} + 2\,\frac{a_iD_i(t)}{\sum_{i=1}^{n} a_iD_i(t)}\cdot D_i(t)r_iD_i(t) =$$

$$\frac{a_ir_i\left(D_i(t)\right)^3\cdot\sum_{i=1}^{n} a_iD_i(t) - a_i\left(D_i(t)\right)^3\cdot\sum_{i=1}^{n} a_ir_iD_i(t) + 2a_ir_i\left(D_i(t)\right)^3\cdot\sum_{i=1}^{n} a_iD_i(t)}{\left(\sum_{i=1}^{n} a_iD_i(t)\right)^2} =$$

$$\frac{3a_ir_i\left(D_i(t)\right)^3\cdot\sum_{i=1}^{n} a_iD_i(t) - a_i\left(D_i(t)\right)^3\cdot\sum_{i=1}^{n} a_ir_iD_i(t)}{\left(\sum_{i=1}^{n} a_iD_i(t)\right)^2}. \quad [A10]$$

Substituting [A9] into [A3] gives

$$\frac{dE\big(D(t)\big)}{dt} = \frac{\sum_{i=1}^{n} 2a_i r_i\big(D_i(t)\big)^2 \cdot \sum_{i=1}^{n} a_i D_i(t) - \sum_{i=1}^{n} a_i\big(D_i(t)\big)^2 \cdot \sum_{i=1}^{n} a_i r_i D_i(t)}{\big(\sum_{i=1}^{n} a_i D_i(t)\big)^2}. \qquad \text{[A11]}$$

Substituting [A9] and [A10] into [A5] gives

$$\frac{dVar\big(D(t)\big)}{dt} =$$

$$= \frac{\sum_{i=1}^{n} 3a_i r_i\big(D_i(t)\big)^3 \cdot \sum_{i=1}^{n} a_i D_i(t) - \sum_{i=1}^{n} a_i\big(D_i(t)\big)^3 \cdot \sum_{i=1}^{n} a_i r_i D_i(t)}{\big(\sum_{i=1}^{n} a_i D_i(t)\big)^2} - 2\frac{\sum_{i=1}^{n} a_i\big(D_i(t)\big)^2}{\sum_{i=1}^{n} a_i D_i(t)} \cdot$$

$$\frac{\sum_{i=1}^{n} 2a_i r_i\big(D_i(t)\big)^2 \cdot \sum_{i=1}^{n} a_i D_i(t) - \sum_{i=1}^{n} a_i\big(D_i(t)\big)^2 \cdot \sum_{i=1}^{n} a_i r_i D_i(t)}{\big(\sum_{i=1}^{n} a_i D_i(t)\big)^2} =$$

$$\frac{\sum_{i=1}^{n} 3a_i r_i\big(D_i(t)\big)^3 \cdot \big(\sum_{i=1}^{n} a_i D_i(t)\big)^2 - \sum_{i=1}^{n} a_i\big(D_i(t)\big)^3 \cdot \sum_{i=1}^{n} a_i r_i D_i(t) \cdot \sum_{i=1}^{n} a_i D_i(t)}{\big(\sum_{i=1}^{n} a_i D_i(t)\big)^3} -$$

$$\frac{2[\sum_{i=1}^{n} a_i\big(D_i(t)\big)^2 \cdot \sum_{i=1}^{n} 2a_i r_i\big(D_i(t)\big)^2 \cdot \sum_{i=1}^{n} a_i D_i(t) - \Big(\sum_{i=1}^{n} a_i\big(D_i(t)\big)^2\Big)^2 \cdot \sum_{i=1}^{n} a_i r_i D_i(t)]}{\big(\sum_{i=1}^{n} a_i D_i(t)\big)^3} \qquad \text{[A12]}$$

Then combining [A2], [A6], and [A11] with [7] gives

$$\frac{\frac{dE(D(t))}{dt}}{E\big(D(t)\big)} =$$

$$= \frac{\frac{\sum_{i=1}^{n} 2a_i r_i\big(D_i(t)\big)^2 \cdot \sum_{i=1}^{n} a_i D_i(t) - \sum_{i=1}^{n} a_i\big(D_i(t)\big)^2 \cdot \sum_{i=1}^{n} a_i r_i D_i(t)}{\big(\sum_{i=1}^{n} a_i D_i(t)\big)^2}}{\sum_{i=1}^{n} \frac{a_i D_i(t}{\sum_{i=1}^{n} a_i D_i(t)} D_i(t)} =$$

$$\frac{\sum_{i=1}^{n} 2a_i r_i\big(D_i(t)\big)^2 \cdot \sum_{i=1}^{n} a_i D_i(t) - \sum_{i=1}^{n} a_i\big(D_i(t)\big)^2 \cdot \sum_{i=1}^{n} a_i r_i D_i(t)}{\big(\sum_{i=1}^{n} a_i D_i(t)\big)^2} \cdot \frac{\sum_{i=1}^{n} a_i D_i(t)}{\sum_{i=1}^{n} a_i\big(D_i(t)\big)^2} =$$

$$\frac{\sum_{i=1}^{n} 2a_i r_i\big(D_i(t)\big)^2 \cdot \sum_{i=1}^{n} a_i D_i(t) - \sum_{i=1}^{n} a_i\big(D_i(t)\big)^2 \cdot \sum_{i=1}^{n} a_i r_i D_i(t)}{\sum_{i=1}^{n} a_i\big(D_i(t)\big)^2 \cdot \sum_{i=1}^{n} a_i D_i(t)} = \frac{2\sum_{i=1}^{n} a_i r_i\big(D_i(0)\big)^2 e^{2r_i t}}{\sum_{i=1}^{n} a_i\big(D_i(0)\big)^2 e^{2r_i t}} -$$

$$\frac{\sum_{i=1}^{n} a_i r_i D_i(0) e^{r_i t}}{\sum_{i=1}^{n} a_i D_i(0) e^{r_i t}}. \qquad \text{[A13]}$$

In the first fraction on the right side, divide the numerator and the denominator by $\exp(2r_1 t)$. In the second fraction on the right side, divide the numerator and the denominator by $\exp(r_1 t)$. Because we assumed $+\infty > r_1 > \cdots > r_n > -\infty$, taking the limit $t \to \infty$ yields

$$\frac{\frac{dE(D(t))}{dt}}{E\big(D(t)\big)} \to \frac{2a_1 r_1\big(D_1(0)\big)^2}{a_1\big(D_1(0)\big)^2} - \frac{a_1 r_1 D_1(0)}{a_1 D_1(0)} = 2r_1 - r_1 = r_1. \qquad \text{[A14]}$$

From 0, [A4], and [A6], we have

$$\frac{1}{Var(D(t))} = \frac{1}{\frac{\sum_{i=1}^{n} a_i(D_i(t))^3}{\sum_{i=1}^{n} a_i D_i(t)} - \frac{\left(\sum_{i=1}^{n} a_i(D_i(t))^2\right)^2}{\left(\sum_{i=1}^{n} a_i D_i(t)\right)^2}} =$$

$$\frac{\left(\sum_{i=1}^{n} a_i D_i(t)\right)^2}{\sum_{i=1}^{n} a_i(D_i(t))^3 \cdot \sum_{i=1}^{n} a_i D_i(t) - \left(\sum_{i=1}^{n} a_i(D_i(t))^2\right)^2}, \qquad \text{[A15]}$$

which combines with [A12] to yield

$$\frac{\frac{dVar(D(t))}{dt}}{Var(D(t))} =$$

$$= \left\{\frac{\sum_{i=1}^{n} 3a_i r_i(D_i(t))^3 \cdot \left(\sum_{i=1}^{n} a_i D_i(t)\right)^2 - \sum_{i=1}^{n} a_i(D_i(t))^3 \cdot \sum_{i=1}^{n} a_i r_i D_i(t) \cdot \sum_{i=1}^{n} a_i D_i(t)}{\left(\sum_{i=1}^{n} a_i D_i(t)\right)^3} - \right.$$

$$\left. \frac{2\left[\sum_{i=1}^{n} a_i(D_i(t))^2 \cdot \sum_{i=1}^{n} 2a_i r_i(D_i(t))^2 \cdot \sum_{i=1}^{n} a_i D_i(t) - \left(\sum_{i=1}^{n} a_i(D_i(t))^2\right)^2 \cdot \sum_{i=1}^{n} a_i r_i D_i(t)\right]}{\left(\sum_{i=1}^{n} a_i D_i(t)\right)^3}\right\} \times$$

$$\frac{\left(\sum_{i=1}^{n} a_i D_i(t)\right)^2}{\sum_{i=1}^{n} a_i(D_i(t))^3 \cdot \sum_{i=1}^{n} a_i D_i(t) - \left(\sum_{i=1}^{n} a_i(D_i(t))^2\right)^2} =$$

$$\frac{1}{\left(\sum_{i=1}^{n} a_i D_i(0) e^{r_i t}\right) \cdot \left[\left(\sum_{i=1}^{n} a_i D_i(0) e^{r_i t}\right) \cdot \left(\sum_{i=1}^{n} a_i(D_i(0))^3 e^{3r_i t}\right) - \left(\sum_{i=1}^{n} a_i(D_i(0))^2 e^{2r_i t}\right)^2\right]} \cdot$$

$$\left[\left(\sum_{i=1}^{n} a_i D_i(0) e^{r_i t}\right)^2 \cdot \left(\sum_{i=1}^{n} 3a_i r_i(D_i(0))^3 e^{3r_i t}\right) - \left(\sum_{i=1}^{n} a_i D_i(0) e^{r_i t}\right) \cdot \right.$$

$$\left(\sum_{i=1}^{n} a_i(D_i(0))^3 e^{3r_i t}\right) \cdot \left(\sum_{i=1}^{n} a_i r_i D_i(0) e^{r_i t}\right) - 4\left(\sum_{i=1}^{n} a_i(D_i(0))^2 e^{2r_i t}\right) \cdot$$

$$\left(\sum_{i=1}^{n} a_i r_i(D_i(0))^2 e^{2r_i t}\right) \cdot \left(\sum_{i=1}^{n} a_i D_i(0) e^{r_i t}\right) + 2\left(\sum_{i=1}^{n} a_i(D_i(0))^2 e^{2r_i t}\right)^2 \cdot$$

$$\left. \left(\sum_{i=1}^{n} a_i r_i D_i(0) e^{r_i t}\right)\right]. \qquad \text{[A16]}$$

To find the limit of [A16], divide the numerator and the denominator by $\exp((4r_1+r_2)t)$ (and not by $\exp(5r_1 t)$, which leads to an indeterminate limit 0/0 as $t\to\infty$). As $t\to\infty$, the new denominator of [A16] converges to

$$a_1 D_1(0) \cdot \left[\left(a_2 D_2(0)\right) \cdot \left(a_1\left(D_1(0)\right)^3\right)\right] = a_1^2 a_2 \left(D_1(0)\right)^4 D_2(0).$$

As $t\to\infty$, the new numerator of [A16] converges to

$$2(a_1D_1(0)a_2D_2(0))\cdot\left(3a_1r_1\left(D_1(0)\right)^3\right)-\left(a_1D_1(0)\right)\cdot\left(a_1\left(D_1(0)\right)^3\right)$$
$$\cdot\left(a_2r_2D_2(0)\right)-\left(a_2D_2(0)\right)\cdot\left(a_1\left(D_1(0)\right)^3\right)\cdot\left(a_1r_1D_1(0)\right)$$
$$-4\left(a_1\left(D_1(0)\right)^2\right)\cdot\left(a_1r_1\left(D_1(0)\right)^2\right)\cdot\left(a_2D_2(0)\right)$$
$$+2\left(a_1^2\left(D_1(0)\right)^4\right)\cdot\left(a_2r_2D_2(0)\right)$$
$$=a_1^2a_2(r_1+r_2)\left(D_1(0)\right)^4D_2(0).$$

Therefore, as $t\to\infty$, [A16] converges to

$$\frac{a_1^2a_2(r_1+r_2)\left(D_1(0)\right)^4D_2(0)}{a_1^2a_2\left(D_1(0)\right)^4D_2(0)}=r_1+r_2. \qquad \text{[A17]}$$

Overall, as $t\to\infty$,

$$b(t)\to\frac{r_1+r_2}{r_1}=1+\frac{r_2}{r_1}. \qquad \text{[A18]}$$

The limiting behavior when $t\to-\infty$ is identical to that when $t\to+\infty$ if the ordering of the exponential parameters is reversed. Hence without further calculation we have immediately, in the limit $t\to-\infty$,

$$\frac{\frac{dE(D(t))}{dt}}{E(D(t))}\to r_n, \qquad \frac{\frac{dVar(D(t))}{dt}}{Var(D(t))}\to r_{n-1}+r_n. \qquad \text{[A19]}$$

From [5], as $t\to-\infty$,

$$b(t)\to\frac{r_{n-1}+r_n}{r_n}=1+\frac{r_{n-1}}{r_n}. \qquad \text{[A20]}$$

Because $b_\pm=\lim_{t\to\pm\infty}b(t)$ is independent of the coefficients a_i, and because population density is the ratio of population size to the area of each county (assumed to be constant in this model), weighting by population density gives exactly the same limiting behavior of $b(t)$ as weighting by population size. That is, the local slope $b(t)$ converges to $1+r_2/r_1$ as $t\to\infty$, and to $1+r_{n-1}/r_n$ as $t\to-\infty$, for both population size and population density.

A.1.2 Correlation between exponential rate and initial county density affects the value of $b(t)$

We observed empirically that counties that had larger initial population density generally had larger growth rate (Figure 2). The slope of the approximately linear relationship between growth rate and the log of the initial population density was statistically significantly positive, and the intercept was statistically indistinguishable from 0. All counties in Norway were used in the analysis ($n = 19$). This relationship may be summarized by

$$r_i = \rho + \sigma \cdot \ln(D_i(0)),\ \sigma > 0,\ \rho \approx 0. \qquad [A21]$$

Here r_i is the exponential rate of county i estimated from the model [7] and $D_i(0)$ is the population density of county i in 1978, $i = 1, 2, \ldots, n$.

In this section we prove that [A21] implies $b(t) > 2$, for equally or area weighted mean and variance of county population density. ln = $\log_e$ is used throughout this section. We assume $\sigma > 0$ and write (Cohen, in press a)

$$\frac{d\ln E\big(D(t)\big)}{dt} = \frac{\sum_{i=1}^{n} p_i D_i(0)(\rho + \sigma \ln D_i(0))\big(D_i(0)\big)^{\sigma t}}{\sum_{i=1}^{n} p_i D_i(0)\big(D_i(0)\big)^{\sigma t}} = \rho + \sigma A,$$

$$A = \frac{\sum_{i=1}^{n} p_i \ln D_i(0)\,\big(D_i(0)\big)^{\sigma t+1}}{\sum_{i=1}^{n} p_i\big(D_i(0)\big)^{\sigma t+1}}.$$

Similarly, we found

$$\frac{d\ln V\big(D(t)\big)}{dt} = 2(\rho + \sigma B),$$

where

$$B = \frac{\sum_{i=1}^{n} p_i \ln D_i(0)\,\big(D_i(0)\big)^{2\sigma t+2} - \Big(\sum_{i=1}^{n} p_i\big(D_i(0)\big)^{\sigma t+1}\Big)\cdot\Big(\sum_{i=1}^{n} p_i \ln D_i(0)\,\big(D_i(0)\big)^{\sigma t+1}\Big)}{\sum_{i=1}^{n} p_i\big(D_i(0)\big)^{2\sigma t+2} - \Big(\sum_{i=1}^{n} p_i\big(D_i(0)\big)^{\sigma t+1}\Big)^2}.$$

Numerical calculation of $\rho + \sigma A$ using parameters estimated from data showed that

$$\rho + \sigma A = \frac{d\ln E(D(t))}{dt} = \frac{\sum_{i=1}^{n} p_i r_i D_i(0) e^{r_i t}}{\sum_{i=1}^{n} p_i D_i(0) e^{r_i t}}$$

is bigger than 0 for all the finite values of t included in our calculation (result not shown).

Then [A21] implies that the local slope of TL, defined in [5], becomes (Cohen, in press a)

$$b(t) = \frac{2(\rho + \sigma B)}{\rho + \sigma A}.$$

Then $b(t) \geq 2$ if and only if one of the following four cases holds (the case for $\sigma = 0$ is excluded, see discussion above):

$$\begin{cases} \rho + \sigma A > 0,\ \sigma > 0 \ and\ B \geq A, \\ \rho + \sigma A > 0,\ \sigma < 0 \ and\ B \leq A, \\ \rho + \sigma A < 0,\ \sigma > 0 \ and\ B \leq A, \\ \rho + \sigma A < 0,\ \sigma < 0 \ and\ B \geq A. \end{cases}$$

Here we analyzed only the case compatible with the data, namely, $\rho = 0$, $\sigma > 0$ and $\rho + \sigma A > 0$. In this case, $b(t) \geq 2$ if and only if $B \geq A$,

$$\frac{\sum_{i=1}^{n} p_i \ln D_i(0) \left(D_i(0)\right)^{2\sigma t+2} - \left(\sum_{i=1}^{n} p_i \left(D_i(0)\right)^{\sigma t+1}\right) \cdot \left(\sum_{i=1}^{n} p_i \ln D_i(0) \left(D_i(0)\right)^{\sigma t+1}\right)}{\sum_{i=1}^{n} p_i \left(D_i(0)\right)^{2\sigma t+2} - \left(\sum_{i=1}^{n} p_i \left(D_i(0)\right)^{\sigma t+1}\right)^2}$$

$$\geq \frac{\sum_{i=1}^{n} p_i \ln D_i(0) \left(D_i(0)\right)^{\sigma t+1}}{\sum_{i=1}^{n} p_i \left(D_i(0)\right)^{\sigma t+1}}.$$

We assume that all $p_i > 0$ and at least two of the initial population density $D_i(0)$, $D_j(0)$, $i \neq j$, are distinct. Define $a_i = \sqrt{p_i}$, $b_i = \sqrt{p_i}\left(D_i(0)\right)^{\sigma t+1}$,

$$\sum_{i=1}^{n} a_i^2 = \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} b_i^2 = \sum_{i=1}^{n} p_i \left(D_i(0)\right)^{2\sigma t+2}$$

and

$$\sum_{i=1}^{n} a_i b_i = \sum_{i=1}^{n} p_i \left(D_i(0)\right)^{\sigma t+1}.$$

Lagrange's identity yields

$$\sum_{i=1}^{n} p_i \left(D_i(0)\right)^{2\sigma t+2} - \left(\sum_{i=1}^{n} p_i \left(D_i(0)\right)^{\sigma t+1}\right)^2 =$$

$$= \sum_{1 \leq i < j \leq n} \left(\sqrt{p_i} \cdot \sqrt{p_j} \left(D_j(0)\right)^{\sigma t+1} - \sqrt{p_j} \cdot \sqrt{p_i} \left(D_i(0)\right)^{\sigma t+1}\right)^2$$

$$= \sum_{1 \leq i < j \leq n} p_i p_j \left(\left(D_j(0)\right)^{\sigma t+1} - \left(D_i(0)\right)^{\sigma t+1}\right)^2.$$

Under the assumption, the denominator of B is always bigger than 0. $B \geq A$ becomes

$$\left[\sum_{i=1}^{n} p_i \ln D_i(0)\left(D_i(0)\right)^{2\sigma t+2} - \left(\sum_{i=1}^{n} p_i\left(D_i(0)\right)^{\sigma t+1}\right)\right.$$
$$\left.\cdot\left(\sum_{i=1}^{n} p_i \ln D_i(0)\left(D_i(0)\right)^{\sigma t+1}\right)\right]\cdot\sum_{i=1}^{n} p_i\left(D_i(0)\right)^{\sigma t+1}$$
$$\geq \sum_{i=1}^{n} p_i \ln D_i(0)\left(D_i(0)\right)^{\sigma t+1}$$
$$\cdot\left[\sum_{i=1}^{n} p_i\left(D_i(0)\right)^{2\sigma t+2} - \left(\sum_{i=1}^{n} p_i\left(D_i(0)\right)^{\sigma t+1}\right)^2\right],$$

$$\left(\sum_{i=1}^{n} p_i \ln D_i(0)\left(D_i(0)\right)^{2\sigma t+2}\right)\cdot\left(\sum_{i=1}^{n} p_i\left(D_i(0)\right)^{\sigma t+1}\right)$$
$$\geq\left(\sum_{i=1}^{n} p_i \ln D_i(0)\left(D_i(0)\right)^{\sigma t+1}\right)\cdot\left(\sum_{i=1}^{n} p_i\left(D_i(0)\right)^{2\sigma t+2}\right).$$

Setting $n_i = (D_i(0))^{\sigma t+1}$,

$$\left(\sum_{i=1}^{n} p_i \ln D_i(0)\, n_i^2\right)\cdot\left(\sum_{i=1}^{n} p_i n_i\right) \geq \left(\sum_{i=1}^{n} p_i \ln D_i(0)\, n_i\right)\cdot\left(\sum_{i=1}^{n} p_i n_i^2\right).$$

The left side is

$$LHS = \sum_{i\neq j}^{n} p_i \ln D_i(0)\, n_i^2 \cdot p_j n_j + \sum_{i=1}^{n} p_i^2 \ln D_i(0)\, n_i^3.$$

The right side is

$$RHS = \sum_{i\neq j}^{n} p_i \ln D_i(0)\, n_i \cdot p_j n_j^2 + \sum_{i=1}^{n} p_i^2 \ln D_i(0)\, n_i^3.$$

$$LHS - RHS = \sum_{i<j}^{n} \left(p_i \ln D_i(0)\, n_i^2 \cdot p_j n_j + p_j \ln D_j(0)\, n_j^2 \cdot p_i n_i \right)$$

$$- \sum_{i<j}^{n} \left(p_i \ln D_i(0)\, n_i \cdot p_j n_j^2 + p_j \ln D_j(0)\, n_j \cdot p_i n_i^2 \right)$$

$$= \sum_{i<j}^{n} \left[\left(p_i \ln D_i(0)\, n_i^2 \cdot p_j n_j - p_j \ln D_j(0)\, n_j \cdot p_i n_i^2 \right) \right.$$

$$\left. + \left(p_j \ln D_j(0)\, n_j^2 \cdot p_i n_i - p_i \ln D_i(0)\, n_i \cdot p_j n_j^2 \right) \right]$$

$$= \sum_{i<j}^{n} \left[p_i p_j n_i^2 n_j \left(\ln D_i(0) - \ln D_j(0) \right) \right.$$

$$\left. - p_i p_j n_i n_j^2 \left(\ln D_i(0) - \ln D_j(0) \right) \right]$$

$$= \sum_{i<j}^{n} p_i p_j n_i n_j \left(n_i - n_j \right) \left(\ln D_i(0) - \ln D_j(0) \right).$$

By definition,

$$\ln n_i = (\sigma t + 1) \cdot \ln D_i(0).$$

Since $\sigma > 0$ and $t \geq 0$, $\sigma t + 1 \geq 1$. So if $\ln D_i(0) \geq \ln D_j(0)$, then $\ln(n_i) \geq \ln(n_j)$. Since $\exp(x)$ is an increasing function, $n_i \geq n_j$. $n_i = n_j$ if and only if $\ln D_i(0) = \ln D_j(0)$. Therefore,

$$p_i p_j n_i n_j \left(n_i - n_j \right) \left(\ln D_i(0) - \ln D_j(0) \right) \geq 0$$

for any pair of i and j, $i < j$. By assumption, there are at least two counties with distinct initial population densities, therefore at least one term in the above summation is strictly positive. This yields that LHS > RHS, $B > A$, and $b(t) > 2$ for any $t \geq 0$. Since the slope of spatial TL b is a weighted average of the local slopes $b(t)$, $b > 2$.

A.2 *Data analysis*

A.2.1 Exponential model of county population density

Table S1 – *Statistics of parameters in linear regressions for exponential model of county populations in Norway from 1978 to 2010 on the $\log_{10}$ scale: $\log_{10}D_i(t) = \log_{10}D_i(0) + (\log_{10}e)r_it$, where i stands for county*

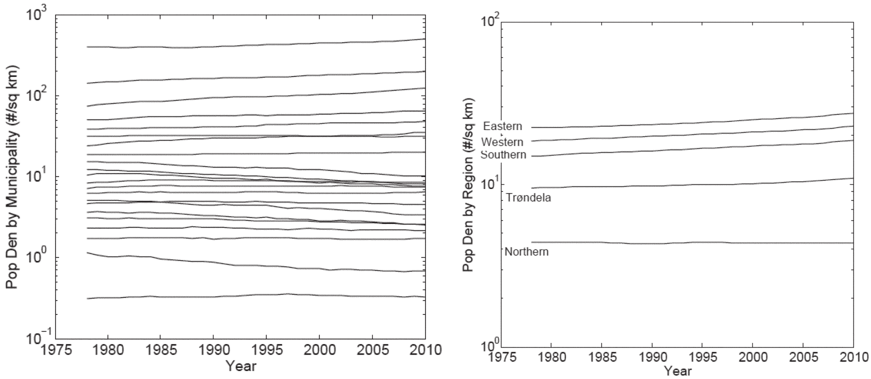
Code	County	Growth rate of county i, r_i			$\log_{10}D_i(0)$			R^2
		Estimate ($\times10^3$)	95% lower bound ($\times10^3$)	95% upper bound ($\times10^3$)	Estimate	95% lower bound	95% upper bound	
01	Østfold	4.6164	4.0251	5.2077	1.7660	1.7612	1.7707	0.8911
02	Akershus	12.2932	12.0233	12.5631	1.8908	1.8886	1.8930	0.9964
03	Oslo	7.7885	6.7430	8.8340	3.0046	2.9962	3.0131	0.8816
04	Hedmark	0.4545	0.2813	0.6278	0.8532	0.8518	0.8546	0.4802
05	Oppland	0.5992	0.4672	0.7313	0.8807	0.8796	0.8818	0.7342
06	Buskerud	5.5754	5.2542	5.8966	1.1830	1.1804	1.1856	0.9759
07	Vestfold	7.1593	6.9058	7.4128	1.9275	1.9255	1.9296	0.9907
08	Telemark	1.1479	1.0241	1.2718	1.0651	1.0641	1.0661	0.9202
09	Aust-Agder	5.7266	5.4072	6.0460	1.0337	1.0311	1.0362	0.9773
10	Vest-Agder	7.1888	6.9504	7.4272	1.2992	1.2972	1.3011	0.9919
11	Rogaland	10.5120	10.2573	10.7667	1.5414	1.5394	1.5435	0.9956
12	Horda land	6.3164	5.9557	6.6770	1.4155	1.4126	1.4184	0.9763
14	Sogn og Fjordane	0.5636	0.3110	0.8163	0.7765	0.7744	0.7785	0.4004
15	Møre og Romsdal	1.7459	1.5939	1.8978	1.2052	1.2040	1.2064	0.9465
16	Sør-Trøndelag	5.1225	4.5977	5.6473	1.1248	1.1206	1.1291	0.9275
17	Nord-Trøndelag	1.0313	0.8001	1.2624	0.7801	0.7783	0.7820	0.7276
18	Nordland	-1.1815	-1.3596	-1.0034	0.8312	0.8297	0.8326	0.8552
19	Troms	2.0962	1.8509	2.3414	0.7646	0.7626	0.7665	0.9074
20	Finnmark	-2.3909	-2.8755	-1.9063	0.2322	0.2283	0.2361	0.7656

Note: For each county i, a linear regression was fitted to independent variable (shifted years $t = 0$ for 1978, 1 for 1979, …, and 32 for 2010) and dependent variable ($\log_{10}$ of county population density $D_i(t)$) across all censuses. The estimated slope and its 95% CI were divided by $\log_{10}e$ to derive r_i and the associated 95% CI. Values in the three columns corresponding to r_i were multiplied by 1,000. For example, $r_{\text{Oslo}} = r_{03}$= 0.0077885 with 95% CI (0.0067430, 0.0088340).

A.2.2 Population density of municipalities and regions

We studied population density at a smaller scale than the county (namely, the municipality) and at a larger scale than the county (namely, the region). On each scale, we analyzed time series of population density, a spatial TL, and discrete local slopes, using data and the exponential model as on the county level.

Figure S1 – *Time series of* log_{10}*(population density) for municipalities (a) and regions (b)*



(a) (b)

Note: 430 municipalities were first sorted according to their initial population densities in a descending order. 22 municipalities as shown in (a) were the 1st, 21st, 41st, …, 401st and 421st municipalities in the sorted list.

Denote population density of region k in year t by $D_k(t)$,

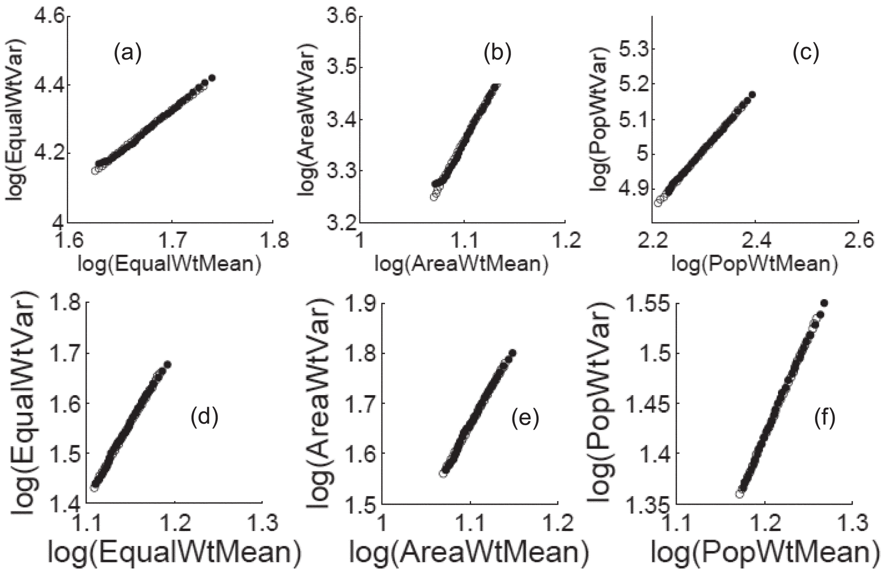
$$D_k(t) = \frac{P_k(t)}{A_k} = \frac{\sum_{i=1}^{n_k} D_i(t) \cdot a_i}{\sum_{i=1}^{n_k} a_i} = \sum_{i=1}^{n_k} \left(\frac{a_i}{\sum_{i=1}^{n_k} a_i} \right) \cdot D_i(t).$$

Here $P_k(t)$ is the population size of region k in year t, A_k is the land area of region k, and n_k is the number of counties in region k. Since our exponential model described the change of county population density well, $D_k(t)$ can be written as a sum of exponential functions of time t,

$$D_k(t) = \sum_{i=1}^{n_k} \left(\frac{a_i}{\sum_{i=1}^{n_k} a_i} \right) \cdot D_i(0) e^{r_i t},$$

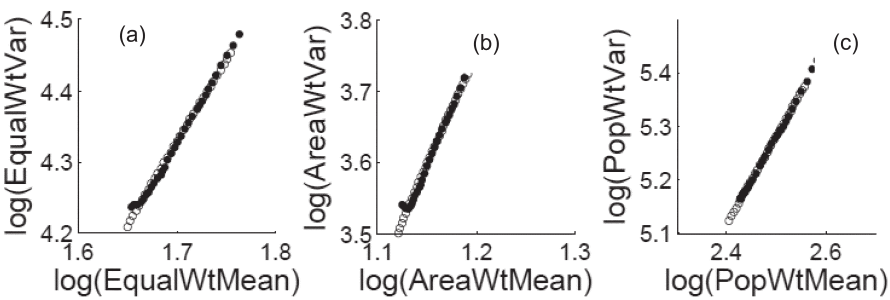
and is therefore a log convex function of t. We found empirically that $\log_{10}(Dk(t))$ behaves as a convex function of time t, as verified by the quadratic least-square regressions (results not shown).

Figure S2 – *Variance and mean (log-log coordinates) of population density for municipalities and regions, excluding Oslo, with 3 weightings*



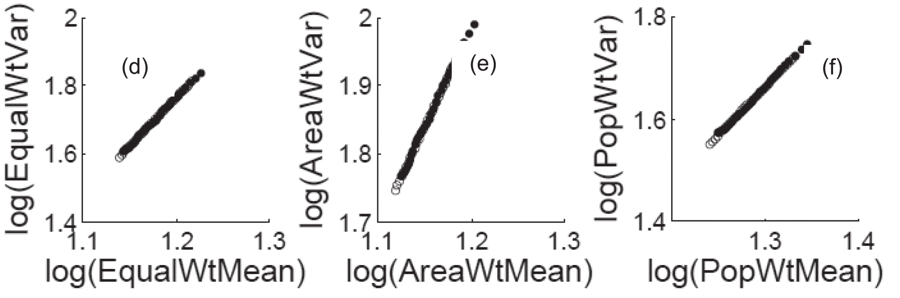
Note: When Oslo was excluded, log(variance of population density) grew as a linear function of log(mean population density) (solid dots), for municipalities (a-c) and regions (d-f), under any weighting method. The exponential model (open circles) described Taylor's law well. Parameters of TL from linear regressions were not statistically different between the data and the model (Table S2).

Figure S3 – *Variance and mean (log-log coordinates) of population density for municipalities and regions, including Oslo, with 3 weightings*



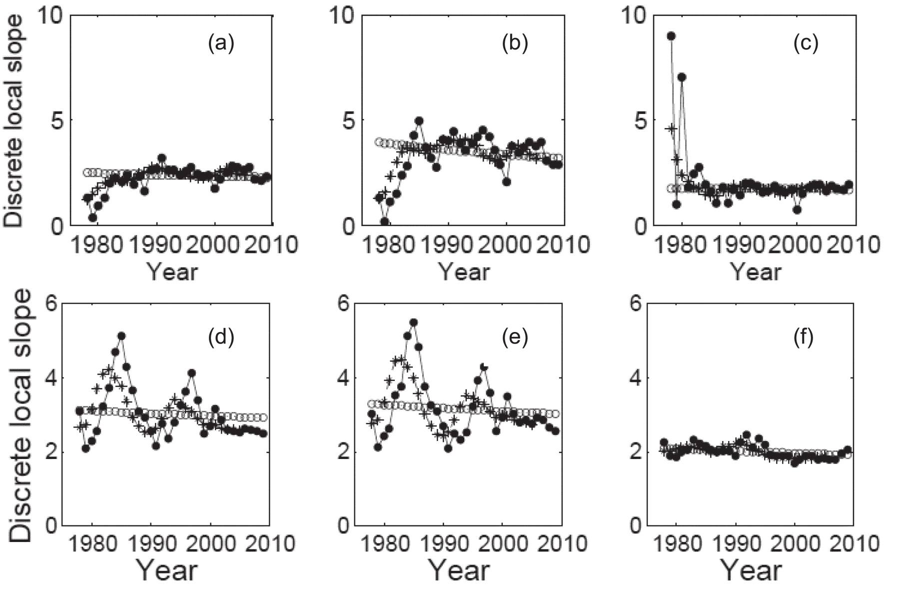
...Cont'd...

Figure S3 – *Cont'd*



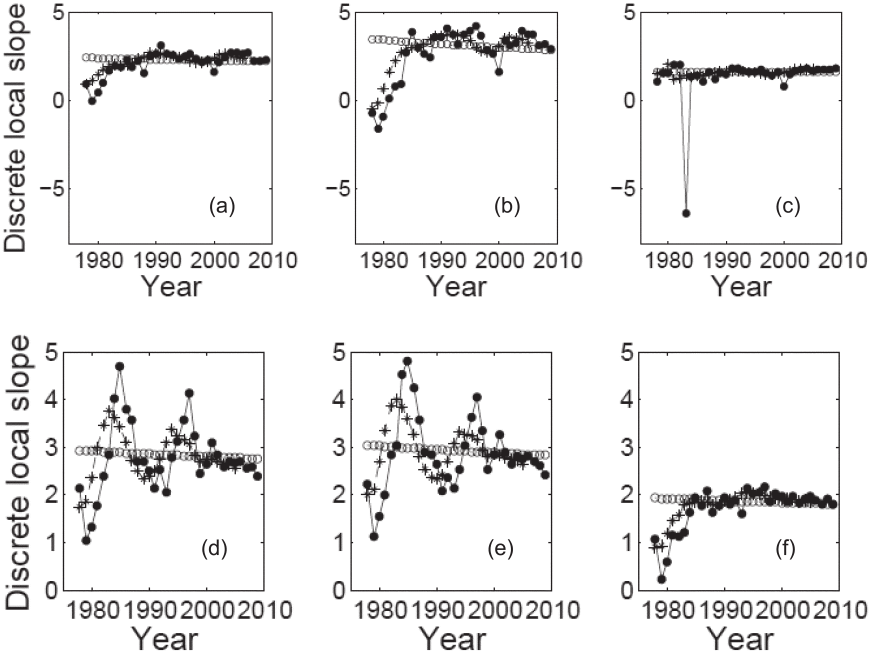
Note: When Oslo was included, log (variance of population density) grew as a linear function of log(mean population density) (solid dots), for municipalities (a-c) and regions (d-f), under any weighting method. The exponential model (open circles) described Taylor's law well. Parameters of TL from linear regressions were not statistically different between the data and the model (Table S3).

Figure S4 – *Discrete local TL slope* [12] *for municipalities (a-c) and regions (d-f) when Oslo was excluded*



Note: Markers were defined in Fig. 5. Local slopes were calculated using equal weights (a, d), area weights (b, e) and population weights (c, f) respectively.

Figure S5 – *Discrete local TL slope* [12] *for municipalities (a-c) and regions (d-f) when Oslo was included*



Note: Markers were defined in Fig. 5. Local slopes were calculated using equal weights (a, d), area weights (b, e) and population weights (c, f) respectively.

Table S2 – *Statistics of linear regressions of dependent variable* log_{10}*(Variance of population density) on independent variable* log_{10}*(Mean population density) excluding Oslo from the data and the model, 1978-2010*

Level	Source	Weighting	Slope			Intercept			R^2
			estimate	95% lower bound	95% upper bound	estimate	95% lower bound	95% upper bound	
Municipality	Data	Equal	2.3473	2.3000	2.3947	0.3337	0.2543	0.4132	0.9970
		Area	3.4617	3.3785	3.5449	-0.4550	-0.5469	-0.3631	0.9957
		Pop.	1.7072	1.6882	1.7262	1.0850	1.0414	1.1286	0.9991
	Model	Equal	2.3487	2.3378	2.3596	0.3303	0.3120	0.3486	0.9998
		Area	3.4775	3.4417	3.5134	-0.4725	-0.5120	-0.4329	0.9992
		Pop.	1.7106	1.7092	1.7120	1.0772	1.0740	1.0804	1.0000
Region	Data	Equal	2.9874	2.9239	3.0508	-1.7761	-1.8487	-1.7034	0.9966
		Area	3.1227	3.0590	3.1864	-1.7791	-1.8495	-1.7087	0.9969
		Pop.	1.9843	1.9595	2.0092	-0.9635	-0.9937	-0.9333	0.9988
	Model	Equal	3.0055	2.9950	3.0160	-1.8938	-1.9058	-1.8818	0.9999
		Area	3.1428	3.1296	3.1561	-1.8013	-1.8160	-1.7867	0.9999
		Pop.	1.9928	1.9845	2.0011	-0.9738	-0.9839	-0.9637	0.9999

Note: $n = 18$; $R^2 > 0.99$ for any weighting method.

Table S3 – *Statistics of linear regressions of dependent variable* log_{10}*(Variance of population density) on independent variable* log_{10}*(Mean population density) includig Oslo in the data and the model, 1978-2010*

Level	Source	Weighting	Slope			Intercept			R^2
			estimate	95% lower bound	95% upper bound	estimate	95% lower bound	95% upper bound	
Municipality	Data	Equal	2.3114	2.2546	2.3682	0.4012	0.3046	0.4978	0.9955
		Area	3.1441	3.0307	3.2574	-0.0178	-0.1487	0.1132	0.9904
		Pop.	1.6433	1.6270	1.6596	1.1740	1.1336	1.2145	0.9993
	Model	Equal	2.3077	2.2971	2.3183	0.4065	0.3884	0.4245	0.9998
		Area	3.1278	3.0972	3.1584	0.0011	-0.0342	0.0365	0.9993
		Pop.	1.6341	1.6340	1.6341	1.1969	1.1967	1.1971	1.0000
Region	Data	Equal	2.8394	2.7955	2.8832	-1.5454	-1.5971	-1.4938	0.9982
		Area	2.9232	2.8796	2.9667	-1.5232	-1.5735	-1.4728	0.9983
		Pop.	1.8829	1.8535	1.9124	-0.7833	-0.8212	-0.7454	0.9982
	Model	Equal	2.8500	2.8415	2.8586	-1.6549	-1.6650	-1.6448	0.9999
		Area	2.9389	2.9282	2.9495	-1.5413	-1.5536	-1.5290	0.9999
		Pop.	1.8764	1.8702	1.8825	-0.7748	-0.7827	-0.7670	0.9999

Note: $n = 19$; $R^2 > 0.99$ for any weighting method.

Table S4 – *Discrete local slopes $b^*(t)$ from the data and from the model using* [12] *and 1-year intervals, for municipalities and regions*

Level	Source	Weighting	$b^*(t)$ excluding Oslo			$b^*(t)$ including Oslo		
			Mean	95% lower bound	95% upper bound	Mean	95% lower bound	95% upper bound
Municipality	Data	Equal	2.2074	1.9948	2.4200	2.1218	1.8721	2.3714
		Area	3.2644	2.8713	3.6576	2.6475	2.0880	3.2069
		Pop.	2.0827	1.5023	2.6631	1.3505	0.8366	1.8645
	Model	Equal	2.3537	2.3298	2.3777	2.3124	2.2891	2.3357
		Area	3.5016	3.4220	3.5811	3.1456	3.0780	3.2132
		Pop.	1.7104	1.7074	1.7135	1.6338	1.6337	1.6341
Region	Data	Equal	3.0191	2.7547	3.2834	2.7697	2.4951	3.0442
		Area	3.1673	2.8671	3.4675	2.8782	2.5945	3.1619
		Pop.	1.9976	1.9301	2.0652	1.7170	1.5570	1.8770
	Model	Equal	3.0068	2.9840	3.0295	2.8510	2.8324	2.8696
		Area	3.1446	3.1158	3.1735	2.9403	2.9171	2.9635
		Pop.	1.9940	1.9759	2.0121	1.8772	1.8639	1.8905

Note: The 95% CIs were calculated using normal theory. For any combination and level and weighting, 95% CI from the model overlapped that from the data, whether Oslo was included or not.