

Can Taylor’s law of fluctuation scaling and its relatives help demographers select more plausible multi-regional population forecasts?

*Joel E. Cohen, Helge Brunborg and Meng Xu**

Abstract

Which of several alternative population forecasts is the ‘best’ or the most plausible? In published work summarized here, we use Taylor’s law (TL) and its quadratic generalization to select the best among six alternative projections (by Statistics Norway) of Norwegian county population density. We consider two time scales: long term (1978–2010 as the historical basis for projections of 2011–2040) and short term (2006–2010 as the historical basis for projections of 2011–2015). We find that the short-term projections selected as ‘best’ by TL are more closely aligned than the four other projections with the recent county density data, and reflect the current high rate of international net immigration to Norway. Our approach needs to be further tested using other data and demographic forecasts.

1 Introduction

Demographers use population projections and population forecasts to draw conclusions relevant to policy and planning, and to provide inputs for other social and natural sciences that depend on demography. But projections and forecasts differ in terms of the justifications of their assumptions.

* Joel E. Cohen (corresponding author), Laboratory of Populations, The Rockefeller University and Columbia University, 1230 York Avenue, New York, NY 10065, USA

Additional affiliations: Earth Institute and Department of Statistics, Columbia University, USA; Department of Statistics, University of Chicago, USA

E-mail: cohen@rockefeller.edu

Helge Brunborg, Previously with Statistics Norway

E-mail: Helge.Brunborg@gmail.com

Meng Xu, Department of Mathematics, Pace University, 41 Park Row, New York, NY 10038, USA

E-mail: mxu@pace.edu

Projections are based on assumptions that are not necessarily intended to be realistic. Alternative projections can illuminate multiple counterfactual worlds by showing the results of ‘what-if’ or *Gedanken* experiments, such as assuming constant fertility or constant mortality in the future, or no migration. For example, Abel et al. (2016) projected the effect on world population growth of implementing the Sustainable Development Goals using a demographic model that stratified national populations by age, sex, and level of education; and that took into account the different levels of fertility and mortality associated with the different levels of education.

Forecasts, by contrast, are based on assumptions about the future that are intended to be realistic (e.g., Lutz et al. 1997; Bongaarts and Bulatao 2000; Gerland et al. 2014; United Nations 2015). Demographers compute alternative forecasts because they are uncertain about the model assumptions and the parameter values in the future. Demographers have to assume when they are making forecasts that some features or trends of the past will continue into the future – because otherwise, they would have no basis for saying anything about the future. The question demographers are attempting to answer is which features or trends, based on how much of the past, will continue into the future.

We have been exploring unconventional answers to this difficult question using the special case of spatially explicit multi-regional forecasts. We summarize here our recent, and still tentative, progress on these issues, which we have published in detail elsewhere (Cohen et al. 2013; Xu et al. 2017). We offer this summary to increase awareness of our explorations, to motivate others to test our approach using other multi-regional data, and to encourage the development of similar methods for other kinds of demographic forecasts. We do not attempt here to provide a review of demographic projections or forecasts.

Our case study of Norwegian counties uses historical population data and a set of population projections prepared by Statistics Norway (2011) based on the Central Population Register (StatBank Norway 2015). Xu et al. (2017) treated the projections as if they were alternative forecasts, i.e., as if they were intended to be realistic; and tried to evaluate their realism *ex ante*, i.e. before observations of the future became available. Unlike stochastic forecasting, our method does not require probabilistic estimates of demographic variables in the future based on observations from the past, and can be implemented using simple statistical procedures.

We use a standard statistical concept called a variance function. In a family of random variables indexed by space and time, such as the population density of a fixed set of counties at different times, a spatial variance function describes the spatial variance (across counties in a given year t) in population density as a function of the spatial mean of population density (across counties in year t) over the course of various years. We call a plot of $\log(\text{variance in year } t)$ on the vertical axis as a function of $\log(\text{mean in year } t)$ on the horizontal axis, with one dot for each year t , a (log-transformed) ‘variance function plot’. Here ‘log’ means ‘ \log_{10} ’. Fig. 1(c, f) are examples of the spatial variance function plots for the historical data and the projections of Norway’s population by county. Our selection criterion for the best

projection(s) assumes that the trend in the historical ‘variance function plot’ will continue through the projection period.

Using methods developed by Cohen et al. (2013), Xu et al. (2017) tested a particular form of variance function known as Taylor’s law (Taylor 1961; review by Eisler et al. 2008; henceforth abbreviated to TL, eqn (1)) and its quadratic generalization (eqn (2)) for six projections of Norwegian county population density from 2011 to 2040. We define as the ‘best’ long-term projections those projections for which the estimated parameters of TL and its generalization for the period from 2011 to 2040 most closely match the corresponding parameters of TL and its generalization fitted to the historical data (1978–2010).

We also define the ‘best’ short-term projections as those projections for which the estimated parameters of TL and its generalization for the period from 2011 to 2015 most closely match the corresponding parameters of TL and its generalization fitted to the historical data (2006–2010). To evaluate the accuracy of the ‘best’ short-term projections, we examine whether the short-term projections our method selected as ‘best’ are most closely aligned with the population density data from 2011 to 2015.

We summarize the procedure for selecting among various multi-regional projections or forecasts using the proposed variance function method so that the method can be tested further and, if it is successful, used. First, historical population multi-regional time series that immediately precede and are comparable in length to the projections should be chosen. Second, for the historical observations and for each of the projections or forecasts separately, the spatial mean (among spatial units) and the spatial variance (among spatial units) should be calculated for each year; and the multiple regression models (or analysis of covariance: eqns (3) and (4), Materials and Methods) should be constructed across years to compare the variance function parameters of the historical observations with the corresponding variance function parameters of each projection or forecast. Third, the projections or forecasts for which the variance function parameters most closely approximate those of the historical observations should be selected. These steps do not depend on heavy computational machinery, and require only a level of statistical analysis that is available in standard statistical software.

2 Materials and methods

Statistics Norway (2011) projected the populations of each county from 2011 to 2040 under L (low), M (medium), H (high), and 0 (no effect) assumptions for four demographic variables: fertility, life expectancy, internal migration, and net immigration. The label of each projection lists the four variables in the order just given. For example, MMM0 assumes medium trajectories for fertility, life expectancy, and internal migration; and 0 net immigration. Statistics Norway (2011) tabulates projected values of the four demographic variables under each assumption. For example, the projected low, medium, and high total fertility rates for 2015 are 1.85, 1.97, and 2.12 children per woman; and for 2060, they are 1.71, 1.93, and

2.08 children per woman. Detailed projected trajectories of all variables are given in Statistics Norway's Statbank. The six projections analyzed here are MMM0, MMML, MMMM, MMMH, LLML, and HHMH.

For each year t in the historical (1978–2010) data and in each of the six long-term (2011–2040) projections, we calculate a population-weighted spatial mean of each county's population density ($D_{t,j}$) in year t weighted by the number of persons $N_{t,j}$ in that county in year t :

$$mean_t = \sum_{j=1}^n \left(\frac{N_{t,j}}{\sum_{j=1}^n N_{t,j}} \right) \times D_{t,j},$$

and a population-weighted spatial variance of county population density

$$variance_t = \sum_{j=1}^n \left(\frac{N_{t,j}}{\sum_{j=1}^n N_{t,j}} \right) \times (D_{t,j} - mean_t)^2.$$

$N_{t,j}$ is the number of persons in year t ($t = 1978, 1979, \dots, 2010$ for historical data; $t = 2011, 2012, \dots, 2040$ for projections) of county j ($j = 1, 2, \dots, n$), and $n = 19$ is the number of counties in Norway. We weight counties by their population size because we are interested in determining the variation in population across Norwegian counties, rather than in examining land use or political issues, for which areal and equal weighting would have been most appropriate (Cohen et al. 2013).

Taylor's law (TL) is a widely applicable empirical variance function that reflects the population density of nonhuman species. For the multi-regional time series analyzed here, the spatial TL reads:

$$\begin{aligned} & \log(\text{variance across counties of population density in year } t) \\ & = \log a_1 + b_1 \log(\text{mean across counties of population density in year } t). \end{aligned} \quad (1)$$

Cohen et al. (2013) tested TL against population data for Norway from 1978 to 2010 at three spatial levels (municipality, county, region) using three weightings (equally, by area, and by population size) of the population density. For each year, the spatial mean and the spatial variance in population density among the studied spatial units were calculated, plotted across years, and fitted on the log-log scale by a least-squares linear regression (eqn (1)). Cohen et al. (2013) found that, regardless of the weighting used, TL accurately described the variation in Norwegian population density at any spatial level (coefficient of determination R^2 of eqn (1) was above 0.96 at the county level; Fig. 4 and Table 2 in Cohen et al. 2013), even though the parameters of TL differed among the three weighting methods. Under each weighting, Cohen et al. (2013) also fitted the log(mean)-log(variance) pairs of Norwegian county population density by a quadratic generalization of TL (due to Taylor et al. 1978, their equation 14),

$$\begin{aligned} & \log(\text{variance of population density in year } t) \\ & = \log a_2 + b_2 \log(\text{mean population density in year } t) \\ & \quad + c_2 [\log(\text{mean population density in year } t)]^2. \end{aligned} \quad (2)$$

We found that the quadratic coefficient c_2 is statistically significantly positive, indicating convexity between $\log(\text{mean})$ and $\log(\text{variance})$ (Table 2 in Cohen et al. 2013).

We fit least-squares regressions (eqns (1) and (2)) to the spatial variance function plots of the historical data (1978–2010) and of each long-term projection (2011–2040) separately, with one point per year. For the linear regression model (TL), we compare the parameters of eqn (1) in the historical data and in the projections using a multiple linear regression model with interaction; which is commonly called analysis of covariance. Specifically, we combine the mean variance pairs from the data and the six projections, and define a categorical variable ‘source’, which specifies the data and the projections, with the level ‘historical’ indicating the historical data, and the projection name (e.g., ‘MMMM’) indicating the corresponding projection. We then incorporate ‘source’ into eqn (1) as:

$$\log(\text{variance}) = a_1 + b_1 \log(\text{mean}) + c_1(\text{source}) + d_1[\log(\text{mean})]:(\text{source}). \quad (3)$$

Here ‘ $\log(\text{mean}):(\text{source})$ ’ represents the interaction between the independent variables ‘ $\log(\text{mean})$ ’ and ‘source’. With ‘historical’ as the reference source level, values of c_1 of eqn (3) show whether the intercept of TL in the historical data differ from the intercept of each projection; and values of d_1 show whether the slope of TL of the historical data differ from the slope of each projection. Similarly, the ‘source’ variable is included in eqn (2) to examine the differences in each parameter (intercept, linear coefficient, quadratic coefficient) of the generalized TL (eqn (2)) in the historical data and in each projection.

$$\begin{aligned} \log(\text{variance}) = & a_2 + b_2 \log(\text{mean}) + c_2[\log(\text{mean})]^2 + d_2(\text{source}) \\ & + e_2[\log(\text{mean})]:(\text{source}) + f_2[\log(\text{mean})]^2:(\text{source}). \end{aligned} \quad (4)$$

We use the values of d_2 , e_2 , or f_2 (eqn (4)) to examine the differences in the parameters in the historical data and in each projection. For each model (eqns (3) and (4)), projection(s) (if any) with parameter estimates that most closely resemble the parameter estimates of the historical data are selected as the ‘best’ projection(s).

We repeat the variance function analysis and the regression diagnoses for the historical data from 2006 to 2010, and for each projection from 2011 to 2015.

To measure the accuracy of a short-term projection selected as ‘best’, we compute for each year t from 2011 to 2015 the mean of absolute percentage error (MAPE) between the observed and the projected population density of each county across Norway,

$$MAPE_{t,k} = \frac{\sum_{j=1}^n \left| \frac{PD_{t,j,k} - OD_{t,j}}{OD_{t,j}} \right|}{n} \times 100\%,$$

and average them over the five years by

$$MAPE_k = \frac{\sum_{t=2011}^{2015} MAPE_{t,k}}{5}.$$

Here $OD_{t,j}$ is the observed population density of county j ($j = 1, 2, \dots, n$) in year t ($t = 2011, 2012, \dots, 2015$); $PD_{t,j,k}$ is the projected population density of county j from projection k in year t ; and n ($n = 19$) is the number of counties in Norway. Each projection k yields five values of $MAPE_{t,k}$, one for each year t . These five values are then compared among the projections by using analysis of variance (ANOVA) to find the most accurate population projection (defined as the projection(s) k with the least value of $MAPE_k$) from 2011 to 2015. To examine the accuracy of our short-term projections, we also use a recently developed modification of MAPE called the rescaled mean absolute percentage error (MAPE-R) (Swanson et al. 2011).

3 Results

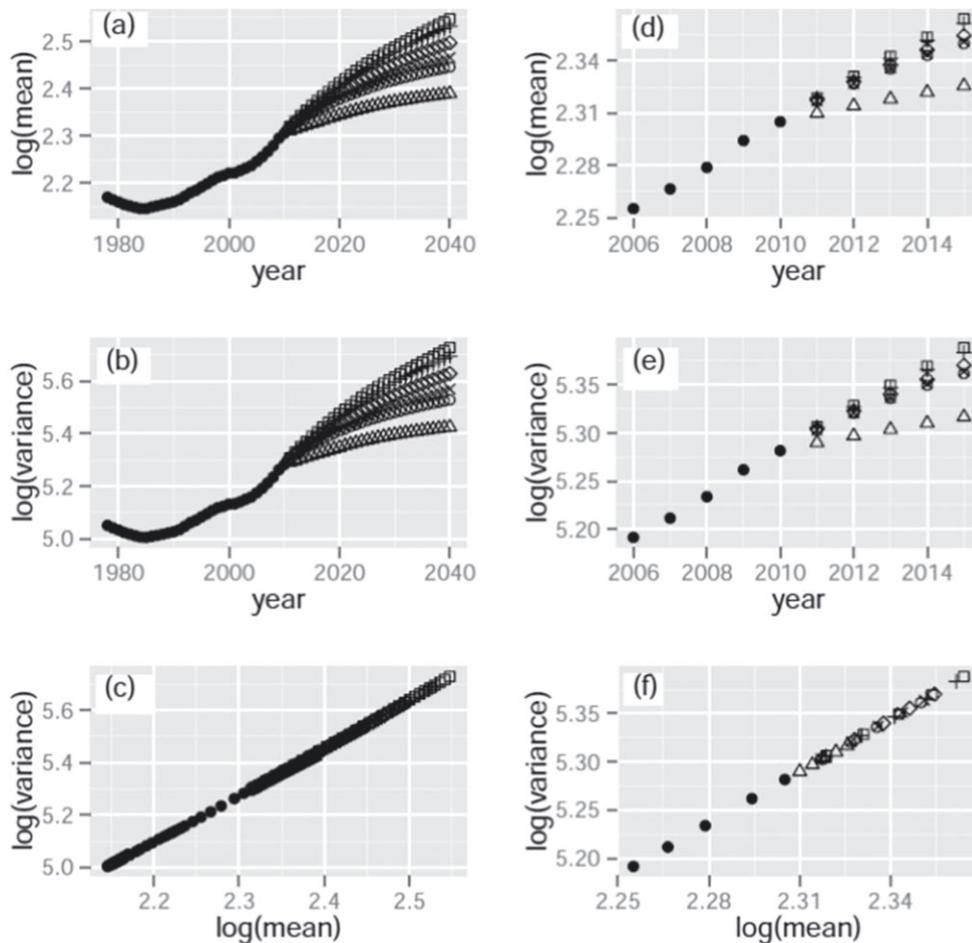
We find that TL describes well the spatial variance function of Norwegian county population density in the historical data from 1978 to 2010 and in each of the six projections from 2011 to 2040 (Fig. 1(c)). In all seven linear regressions (one historical and six projected) that are fitted to the spatial means and the spatial variances, the values of the coefficient of determination R^2 are greater than 0.999. The slopes of TL are between 1.5 and 2.

Further, we find that the slope and the intercept of TL differ between the historical data and five of the six long-term projections. For one projection, labeled MMM0, neither the intercept nor the slope differs significantly from the corresponding value of the parameter of TL fitted to the historical data. MMM0 is selected as the best model based on the similarities of the TL parameters between the data and projections. Because we find that the assumptions of the statistical tests (e.g., homoscedasticity and independence of errors) are not fully satisfied by the data, the alignment of the parameters between MMM0 and the historical data must be considered indicative and approximate.

The generalized TL significantly improves the fit by TL. The results of multiple linear regressions using eqn (4) indicate that in every parameter of the quadratic regression, the parameters of each projection differ significantly from the corresponding parameters of the historical data. No projection is selected as 'best' based on the generalized TL. This conclusion must be tempered by recognizing that the underlying assumptions of the statistical model we use are not always met.

Looking at the short-term historical data (2006–2010) and the six short-term projections (2011–2015) for Norway, we again find that TL describes the spatial variance function of county population density well (Fig. 1(f); $R^2 > 0.9999$). The parameters of TL differ between the historical data and four of the six projections. For the two projections labeled MMMH and HHMH, the outcomes of the multiple linear regression based on eqn (3) do not contradict the hypothesis that TL's intercept is equal to that of the historical data, nor the hypothesis that TL's slope is equal to that of the historical data. The two high-migration variants HHMH and MMMH are selected as the two best short-term projections. The high net immigration rate assumed in the short-term projections HHMH and MMMH reflects

Figure 1:
 Log(mean) against year [a, d], log(variance) against year [b, e], and (log-transformed) variance function plots [c, f] of Norwegian county population density for (a–c) historical data from 1978 to 2010 and six demographic projections from 2011 to 2040, and for (d–f) historical data from 2006 to 2010 and six projections from 2011 to 2015. In a given year, the spatial mean and the spatial variance are calculated among counties weighted by county population sizes. On a log-log scale, the variance function during 1978–1984 overlaps the variance function during the later years (c), and generates an acute hook pattern at the lower left corner of the plot. Markers indicate historical data (●), HHMH (□), LLML (○), MMM0 (Δ), MMMH (+), MMML (×), and MMMM (◇).



Norway's recent high and rapidly increasing levels of immigration due to the country's low unemployment and rapid economic growth rates.

The results of the multiple linear regressions based on eqn (4) show that none of the projections resembles the historical data in any parameter of the least-squares quadratic regression. Thus, no projection is selected using the generalized TL.

Among the six projections, we find that the annual average of the mean absolute percentage errors (MAPEs) and of the rescaled MAPEs (MAPE-R) between each short-term projection and recent (2011–2015) observations follow an identical ascending order: HHMH < MMMH < MMMM < MMML < LLML < MMM0. Under each measure, as HHMH and MMMH are the two projections with the smallest average errors, they are the most accurate. Thus, the two projections identified by TL as being most closely aligned with the historical data are the two projections that most closely approximate the observed county population densities after the projections were published.

4 Conclusion

Variance function analysis of the short-term historical data (2006–2010) and the short-term projections (2011–2015) of Norwegian county populations selected HHMH and MMMH as the two best projections. MAPE and MAPE-R selected the same two projections as being the closest to the recent data (2011–2015). This finding encourages further testing of the use of the variance function in selecting the most accurate short-term demographic projections. The long-term projection selected by the variance function method (i.e., MMM0) seems unrealistic, and its accuracy cannot be evaluated until data of the projection period become available for post hoc comparison. Other high-quality census or historical data and projections should be similarly tested to evaluate how general our method is, and the roles time scales play in the method. The historical data that could be used to further test this method could, for example, come from China's 34 provincial-level administrative units, the 50 states of the USA, or the roughly 200 countries of the world; while the projections that could be used to further test this method could, for example, come from national statistical offices, international statistical and economic agencies, and research institutes.

5 Acknowledgments

This work was partially supported by National Science Foundation grants EF-1038337 and DMS-1225529 to the Rockefeller University. JEC thanks Priscilla K. Rogerson for help, and the organizers of the Wittgenstein Centre Conference 2016 for the privilege of participating.

References

- Abel, G. J., B. Barakat, S. KC and W. Lutz 2016. Meeting the Sustainable Development Goals leads to lower world population growth. *Proceedings of the National Academy of Sciences U.S.A.* 113(50): 14294–14299. <http://www.pnas.org/cgi/doi/10.1073/pnas.1611386113>
- Bongaarts, J. and R. A. Bulatao 2000. *Beyond six billion: Forecasting the world's population*. Washington, D.C.: Panel on Population Projections, Committee on Population, Commission on Behavioral and Social Sciences and Education, National Research Council.
- Cohen, J. E., M. Xu and H. Brunborg 2013. Taylor's law applies to spatial variation in a human population. *Genus* 69(1): 25–60.
- Eisler, Z., I. Bartos and J. Kertész 2008. Fluctuation scaling in complex systems: Taylor's law and beyond. *Advances in Physics* 57(1): 89–142.
- Gerland, P., A. E. Raftery, H. Ševčíková, N. Li, D. Gu, T. Spoorenberg, L. Alkema et al. 2014. World population stabilization unlikely this century. *Science* 346(6206): 234–237.
- Lutz, W., W. Sanderson, and S. Scherbov 1997. Doubling of world population unlikely. *Nature* 387: 803–805.
- Statistics Norway 2011. Population projections, 2011–2100. Report. <http://www.ssb.no/en/befolkning/statistikker/folkfram/aar/2011-06-16>, accessed 7 April 2015.
- StatBank Norway 2015. Table: 05196: Population, by sex, age and citizenship. Data. <https://www.ssb.no/statistikkbanken/selecttable/hovedtabellHjem.asp?KortNavnWeb=folkemengde&CMSSubjectArea=befolkning&PLanguage=1&checked=true>, accessed 7 April 2015.
- Swanson, D. A., J. Tayman and C. F. Barr 2011. MAPE-R: a rescaled measure of accuracy for cross-sectional subnational population forecasts. *Journal of Population Research* 28(2): 225–243.
- Taylor, L. R. 1961. Aggregation, variance and the mean. *Nature* 189: 732–735.
- Taylor, L. R., I. P. Woiwod, and J. N. Perry 1978. The density-dependence of spatial behaviour and the rarity of randomness. *Journal of Animal Ecology* 47(2): 383–406.
- United Nations 2015. *World Population Prospects: The 2015 Revision*. Department of Economic and Social Affairs, Population Division. <http://esa.un.org/wpp/>.
- Xu, M., H. Brunborg and J. E. Cohen 2017. Evaluating multi-regional population projections with Taylor's law of mean–variance scaling and its generalization. *Journal of Population Research* 34(1): 79–99.