CrossMark

Evaluating multi-regional population projections with Taylor's law of mean–variance scaling and its generalisation

Meng Xu[1] · **Helge Brunborg**[2] · **Joel E. Cohen**[3]

Abstract Organisations that develop demographic projections usually propose several variants with different demographic assumptions. Existing criteria for selecting a preferred projection are mostly based on retrospective comparisons with observations, and a prospective approach is needed. In this work, we use the mean–variance scaling (spatial variance function) of human population densities to select among alternative demographic projections. We test against observed and projected Norwegian county population density using two spatial variance functions, Taylor's law (TL) and its quadratic generalisation, and compare each function's parameters between the historical data and six demographic projections, at two different time scales (long term: 1978–2010 vs. 2011–2040; and short term: 2006–2010 vs. 2011–2015). We find that short-term projections selected by TL agree more accurately than the other projections with the recent county density data and reflect the current high rate of international migration to and from Norway. The variance

✉ Joel E. Cohen
cohen@rockefeller.edu

Meng Xu
mxu@pace.edu

Helge Brunborg
Helge.Brunborg@gmail.com

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

2   Statistics Norway, Oslo, Norway

3   Laboratory of Populations, The Rockefeller University and Columbia University, 1230 York Avenue, New York, NY 10065, USA

function method implemented here provides an empirical test of an *ex ante* approach to evaluating short-term human population projections.

Keywords Demographic projection · Net immigration · Population variability · Taylor's law · Power law · Variance function · Fluctuation scaling · Multi-regional projection

Introduction

Demographic projections forecast future population growth under various assumptions or scenarios. Their practical uses include, but are not limited to, fertility policy, immigration policy, pension and retirement planning, social security and planning and siting of facilities for transport, education, water supply, sanitation, and medical care. Global and multi-country projections have been developed by the United Nations (2014), the US Census Bureau (2014), The World Bank (2014), the International Institute for Applied Systems Analysis (Lutz et al. 1997, 2001, 2014; Lutz and Goujon 2004) and some scholars (Dietz et al. 2007), in addition to scores of national projections. These models recognise differing economic, demographic, educational, cultural, and environmental influences on populations. These projections use cohort component methods, age-structured modelling, expert judgment, and stochastic forecasting.

Quantifying uncertainties in population forecasts remains a focal point of research, and multiple approaches have been pursued. Traditionally, e.g. in the projections of the United Nations Population Division up to 2010, variants under different demographic assumptions were constructed for extreme and medium projection scenarios. This approach relied on expert judgment to select the range from high to low projections and gave no probabilistic interpretation of the uncertainty. Stochastic forecasting models based on probabilistic assumptions of demographic parameters were developed to make quantifiable prediction intervals of projected populations (Heyde and Cohen 1985; Cohen 1986; Lee and Carter 1992 [for future mortality only]; Lee and Tuljapurkar 1994; Alho and Spencer 1997; Alkema et al. 2011; Raftery et al. 2013; Gerland et al. 2014). This approach relied on testable modelling of past vital rates and an assumption that the same model form and parameter values would persist in the future. Stochastic multi-regional projections are not easy to produce, and some users may have difficulties using and interpreting the results of probabilistic projections. Another approach compared projections and observations post hoc after the projection period (Stoto 1983; Bongaarts and Bulatao 2000). This last approach does not help projection users, such as city planners, pension policy makers, and immigration policy makers, who cannot wait until after the projection period.

We propose a novel approach to evaluating population projections ex ante and examine its applicability through a case study of population projections for Norwegian counties. Here "ex ante" refers to *evaluating* projections before observations for the projection period become available, not to *generating* projections based on existing conditions (e.g. urban systems models and

microsimulation models, see Smith et al. 2013), although variants generated from such models can be evaluated similarly using our method. Compared to 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 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) of population density as a function of the spatial mean of population density (across counties in the same given year), in the course of various years. We shall call a plot of log(variance in year t) on the vertical axis as a function of log(mean in year t) on the horizontal axis across the years a (log-transformed) "variance function plot". The dependence on year t is often not stated explicitly, and we shall omit that notation henceforth. In this paper, "log" means "$\log_{10}$" and P denotes the p value of a hypothesis test.

A well-tested empirical variance function for population density is Taylor's law (TL) (Taylor 1961). TL can be written as

$$\log(Variance\,of\,Population\,Density) = \log a_1 + b_1 \log(Mean\,Population\,Density).$$
(1)

TL has been confirmed in many very different situations, such as for the number of an individual's sexual partners (Anderson and May 1988), the numbers of measles cases in communities of England and Wales (Keeling and Grenfell 1999), crime reports in England (Hanley et al. 2014), and unemployment in English labour markets (Greig et al. 2014), among thousands of other applications (Eisler et al. 2008).

TL has been little evaluated with human population data. In an early, perhaps the first, empirical test of TL against human population data, Taylor et al. (1978) used decennial censuses of the United States (see their Fig. 6a). Their methods were incompletely described and their results were fragmentary. Cohen et al. (2013) tested TL against data of Norway's population 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. In each year, the spatial mean and the spatial variance of 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 (Eq. 1). Cohen et al. (2013) found that, under any weighting, TL accurately described the variation of Norwegian population density at any spatial level [regression $P < 0.05$ and coefficient of determination (R^2) of Eq. (1) was above 0.96 at the county level, see Fig. 4 and Table 2 in Cohen et al. 2013], although the TL parameters 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 generalisation of TL (due to Taylor et al. 1978, their Eq. 14),

$$\log(Variance\ of\ Population\ Density) = \log a_2$$
$$+ b_2 \log(Mean\ Population\ Density) + c_2 [\log(Mean\ Population\ Density)]^2. \quad (2)$$

They found that the quadratic coefficient c_2 was statistically significantly positive ($P < 0.05$), indicating convexity between log(mean) and log(variance) (see Table 2 in Cohen et al. 2013).

Using the methods developed by Cohen et al. (2013), we tested TL (Eq. 1) and its quadratic generalisation (Eq. 2) for six demographic projections of Norway county population density from 2011 to 2040. We selected as the "best" long-term projections those projections such that their estimated parameters of TL and its generalisation for the future period from 2011 to 2040 most closely matched the corresponding parameters of TL and its generalisation fitted to the historical data (1978–2010). To examine the accuracy of the "best" projections, we repeated the variance function analysis using the historical data on county population densities from 2006 to 2010 and the six projections from 2011 to 2015, and examined whether the short-term projections we selected as "best" by our criterion agreed most closely with recent population density data from 2011 to 2015.

Materials and methods

Data and projections

The population size of each of the 19 counties in Norway was published by Statistics Norway annually from 1978 to 2015 based on the Central Population Register (StatBank Norway 2015). We classified the county population data as historical data (1978–2010) and recent data (2011–2015). County population density was calculated as the number of persons living (from observations) or projected to live (from projections) in a county divided by the county land area (excluding freshwater area, in square kilometres) in a given year. For each county, the land area was measured in 2006 and assumed to be constant from 1978 to 2040.[1]

In 2011, Statistics Norway (2011) projected populations of each county from 2011 to 2040 under L (low), M (medium), H (high) and 0 (no effect) assumptions about four demographic variables: fertility, life expectancy, internal migration and net immigration. The main projection variant MMMM assumed medium fertility (the first M in MMMM), medium life expectancy (the second M in MMMM), medium internal migration (the third M in MMMM) and medium net immigration (the fourth M in MMMM). The assumptions about the variables were (Brunborg and Texmon 2011): the total fertility rate was based on *ad hoc* reasoning based on past trends in Norway and other European countries; the mortality trends were based on a modified Lee-Carter model where low and high series for the death rates were set equal to the upper and lower bounds, respectively, of a 95% confidence interval

---

[1] The geographical boundaries of Hordaland and Rogaland changed on January 1, 2002, with a transfer of one municipality, which led to a 0.5% increase and a 1.2% decrease in the respective population density (Cohen et al. 2013).

(CI); the internal migration rates were constant over time (equal to the rates over the past five years) and with no alternative assumptions; and the future net immigration was derived from an econometric model of migration, relative income and unemployment in Norway and sending countries (Brunborg and Cappelen 2010).

Here we analysed six projections (MMM0, MMML, MMMM, MMMH, LLML, and HHMH) at the county level. These projections all used the medium assumption on internal migration. Any projection(s) selected using our method will not reflect the differences in the internal migration among the counties. (One alternative projection assuming no migration among the counties and no net immigration to Norway [MM00] is highly unrealistic and was made purely for analytical purposes, so we do not analyse it here.) The projections MMM0, MMML, MMMM, MMMH analysed in detail the effect of net immigration by varying its assumption from 0, L, M, to H, while keeping the fertility, life expectancy, and internal migration assumption at M. These projections allowed us to compare different assumptions about net immigration with the continuing international immigration to Norway (Horst et al. 2010). In addition, two variants, LLML and HHMH, were chosen because they yield the lowest and highest population growth, respectively.

Testing spatial Taylor's law

For each year t in the historical (1978–2010) data and in each of the six long-term (2011–2040) projections, a spatial mean and a spatial variance of county population density (D) were calculated across the counties, with each county weighted by the number of persons 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

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

where $N_{t,j}$ is the number of persons in year t ($t = 1978, 1979, \ldots, 2010$ for historical data; $t = 2011, 2012, \ldots, 2040$ for projections) of county j ($j = 1, 2, \ldots, n$), and $n = 19$ is the number of counties in Norway. We weighted counties by their population size because we were interested in the variation of populations among Norway counties, instead of questions about land uses or political issues, for which areal and equal weighting would be the most appropriate (Cohen et al. 2013). Under population weighting, counties with more persons weighed more in the calculated means and variances.

Least-squares regressions (Eqs. 1, 2) were fitted to the spatial variance function plots of the historical data and of each projection separately, one point per year. The Breusch–Pagan test (Breusch and Pagan 1979) and the Durbin–Watson test (Fox 2008) were used to examine the null hypotheses, respectively, of constant regression error variance and no lag-1 regression error autocorrelation. For the linear

regression model (TL), we compared the parameters of Eq. (1) between the historical data and the projections using a multiple linear regression model with interaction. Namely, we combined the mean–variance pairs from data and six projections, and defined a categorical variable "source" that specifies data and projections, with level "historical" indicating historical data and projection name (e.g. "MMMM") indicating the corresponding projection. We then incorporated "source" into Eq. (1) as

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

Here "[log(mean)]:source" represents the interaction between the independent variables "log(mean)" and "source". With "historical" as the reference level of source, values of c_1 of Eq. (3) show whether the intercept of TL of the historical data differed from the intercept of each projection and values of d_1 show whether the slope of TL of the historical data differed from the slope of each projection. Parameters of TL differed statistically significantly between data and a projection if the corresponding P was less than 0.05, and resembled each other if P was greater than or equal to 0.05. Similarly, the "source" variable was included in Eq. (2) to examine the difference in each parameter (intercept, linear coefficient, quadratic coefficient) of the generalised TL (Eq. 2) between the historical data and each projection.

$$\log(variance) = a_2 + b_2 \log(mean) + c_2[\log(mean)]^2 + d_2(source)$$
$$+ e_2[\log(mean)]:(source) + f_2[\log(mean)]^2:(source). \quad (4)$$

Values of d_2, e_2, or f_2 (Eq. 4) and their P were used to compare and examine the significance of the difference in the parameters between the historical data and each projection. For each model, projection(s) (if any) with variance function parameters that resembled the variance function parameters of the historical data were selected as the best projection(s).

We repeated the variance function analysis and regression diagnoses for the data from 2006 to 2010 and for each projection from 2011 to 2015. Separately, parameter estimates of TL and its quadratic generalisation were compared between the data and projections using Eqs. (3) and (4) respectively, and the best projections (if any) were selected based on the resemblance of parameters. Confidence intervals (95%) of the regression parameters were obtained from normal theory.

We did not correct the critical values of any hypothesis tests for simultaneous statistical inference because our approach selects projections with variance function parameters not significantly different from those of the historical data, which are "negative" results of the hypothesis tests. Existing correction methods for simultaneous inference are conservative in controlling false positives (Type-I error), and hence increase the chance of false negatives (Type-II error). This means that more stringent significance level after correction for simultaneous statistical inference would classify more projections falsely in agreement with the historical data. Because we made no correction for multiple tests, approximately 5% of

apparently significant differences in parameters arose as a result of chance fluctuations if the null hypothesis of no difference held true.

Accuracy of 2011–2015 projections

We had five years of observations (2011–2015) since the publication of these projections in 2010. To measure the accuracy of a projection that our methods selected as "best," we computed, for each year t from 2011 to 2015, the mean of absolute percentage error (MAPE) between observed and each projected county population density across all counties of 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 averaged 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 ($j = 1, 2, …, n$) in year t ($t = 2011, 2012, …, 2015$), $PD_{t,j,k}$ is the projected population density of county j from projection k ($k = $ MMM0, MMML, MMMH, MMMM, HHMH, and LLML) in year t, and n ($n = 19$) is the number of counties in Norway. Each projection k yielded five values of $MAPE_{t,k}$, one for each year t. These five values were then compared among projections 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.

We also used a recently developed modification of MAPE, called rescaled mean absolute percentage error (MAPE-R) (Swanson et al. 2011), to examine the accuracy of short-term projections. The advantage of MAPE-R over MAPE is that the former can successfully accommodate substantial outliers in the percentage errors (Swanson et al. 2011). The disadvantage of MAPE-R lies in its non-trivial implementation and calculation. Here, we followed the steps outlined by Swanson et al. (2011) and applied MAPE-R to the six short-term (2011–2015) projections of Norway county population density in comparison to the recent (2011–2015) observations (see "Appendix").

If the projection(s) selected as most accurate by MAPE or MAPE-R matched those selected as "best" using the variance function analysis for the 2006–2010 observations[2] and short-term (2011–2015) projections, the concordance supported the variance function as an *ex ante* method of selecting the most accurate short-term demographic projections and encouraged further testing of the variance function method as a means of selecting among alternative projections of other time series.

---

[2] 2006–2010 data were selected so that historical period and projection period were of equal length. Choosing historical time series as long as the projections is a common practice in forecast accuracy evaluation (Smith et al. 2013).

Least-squares linear regressions, quadratic regressions, multiple linear regressions, and ANOVA were performed in R version 3.2.2 (R Core Team 2015). The Breusch–Pagan test and the Durbin–Watson test were carried out using the 'car' package in R (Fox and Weisberg 2011). Unless otherwise specified, the significance level of all statistical tests was set at 0.05. The confidence level in all interval estimates was set at 0.95.

Results

Analysis of 1978–2010 data and 2011–2040 projections

TL described well the spatial variance function of Norway county population density for the historical data from 1978 to 2010 and for each of the six projections from 2011 to 2040 (Fig. 1c). In all seven linear regressions (1 historical and 6 projected) fitted to the spatial means and spatial variances across years, regression P were less than 0.0001 and R^2 were greater than 0.999. Slopes of TL were statistically significantly less than two and greater than 1.5 (Table 1). Except for MMM0, the Breusch–Pagan test did not reject the null hypothesis that the linear regression error variances were constant. The Durbin–Watson test rejected the hypothesis of no lag-1 autocorrelation in the regression error for the historical data and for each projection. Hence the use of normal theory, which assumes uncorrelated residuals, to estimate confidence intervals for the parameters of TL and its quadratic generalisation must be considered indicative and approximate.

Multiple linear regression using Eq. (3) showed that the slope and intercept of TL were statistically different between the historical data and each of the six long-term projections except MMM0, for which neither the intercept ($P = 0.0740$) nor the slope ($P = 0.1040$) differed from the corresponding value of the historical data. MMM0 was selected as the best model based on the resemblance of TL parameters between the data and projections. However, because the Breusch–Pagan test above rejected the null hypothesis that the linear regression error variances were constant for projection MMM0, the agreement of parameters between MMM0 and the historical data must again be considered indicative and approximate.

The generalised TL showed that the quadratic curvature of the variance function was statistically significant in the historical data and all six projections (Table 1). The generalised TL significantly improved the fit by TL. The Breusch–Pagan test rejected the hypothesis of constant regression error variance for the historical data, HHMH, MMMH, and MMML, but not for LLML, MMM0, or MMMM. The Durbin–Watson test rejected the hypothesis of no lag-1 autocorrelation in the residuals for the historical data and for each projection.

Multiple linear regressions using Eq. (4) indicated that the parameters of each projection differed significantly from the corresponding parameters of the historical data in every parameter of the quadratic regression. No projection can be selected as "best" based on the generalised TL. This conclusion must be tempered by the recognition that the underlying assumptions of the statistical model we used were not always met.
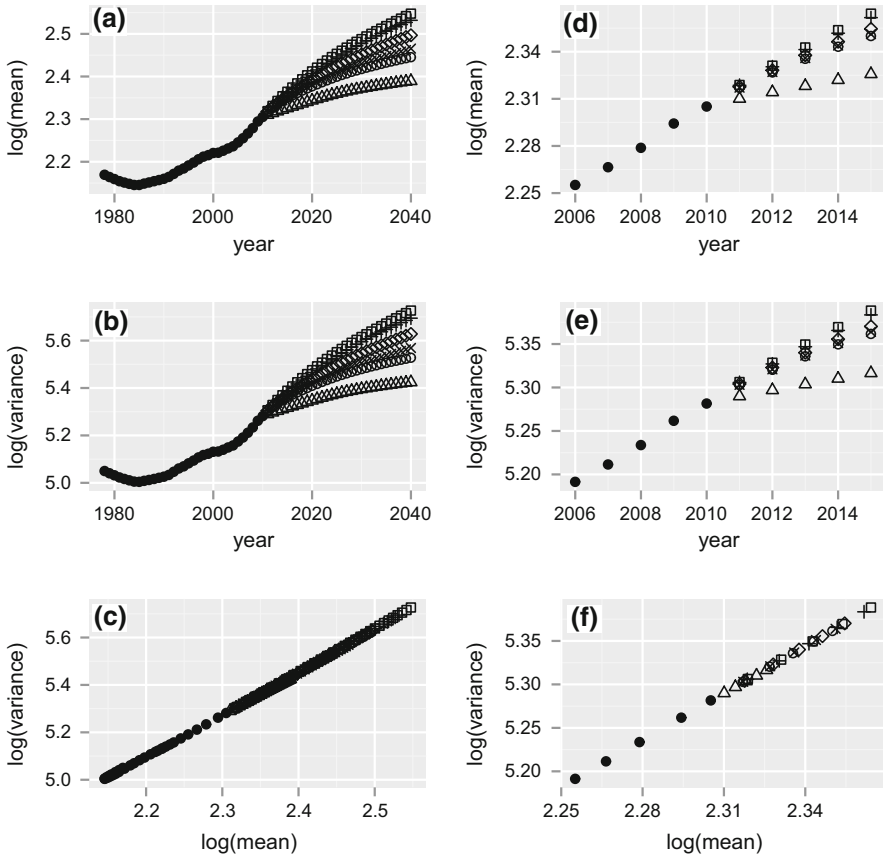
Fig. 1 Log(mean) against year (**a**, **d**), log(variance) against year (**b**, **e**), and (log-transformed) variance function plots (**c**, **f**) of Norwegian county population density for (**a–c**) historical data from 1978 to 2010 and six demographic projections from 2011 to 2040, and for (**d–f**) historical data from 2006 to 2010 and six projections from 2011 to 2015. In a given year, the spatial mean and spatial variance were calculated among counties weighted by county population sizes. On log–log scale, variance function during 1978–1984 overlapped the variance function during the later years (**c**), and generated an acute hook pattern at the *lower left corner* of the *plot*. *Markers* indicate historical data (*filled circle*), HHMH (*open square*), LLML (*open circle*), MMM0 (*triangle*), MMMH (*plus symbol*), MMML (*times symbol*), and MMMM (*diamond symbol*). Figure was produced using "ggplot2" package (Wickham 2009) in R

Analysis of 2006–2010 data and 2011–2015 projections

TL again described well the spatial variance function of Norway's historical (2006–2010) and predicted (2011–2015) county population density using the six projections (Fig. 1f; regression $P < 0.0001$ and $R^2 > 0.9999$, see Table 2). Constancy in regression error variances was not rejected in the historical data or in any projection. Absence of lag-1 error autocorrelation was rejected in the historical data, LLML, MMML, but not in HHMH, MMM0, MMMH, and MMMM. Multiple linear regressions showed that the parameters of TL differed significantly between the historical data and each projection except MMMH and HHMH. For

Table 1 Long-term variance function analysis of data and projections by Taylor's law and generalised Taylor's law

Variance function	Coefficient and statistics	Historical data [1978–2010]	Historical data [1985–2010]	MMM0 [2011–2040]	MMML [2011–2040]	MMMM [2011–2040]	MMMH [2011–2040]	HHMH [2011–2040]	LLML [2011–2040]
TL	Intercept	1.3171	1.2688	**1.3530**	1.1510	1.1084	1.0762	1.0384	1.2464
		(1.2790, 1.3552)	(1.2468, 1.2908)	**(1.3428, 1.3633)**	(1.1476, 1.1544)	(1.1025, 1.1143)	(1.0687, 1.0836)	(1.0286, 1.0483)	(1.2386, 1.2542)
	Slope	1.7183	1.7398	**1.7042**	1.7918	1.8100	1.8238	1.8399	1.7511
		(1.7010, 1.7356)	(1.7298, 1.7497)	**(1.6999, 1.7086)**	(1.7904, 1.7932)	(1.8076, 1.8124)	(1.8207, 1.8268)	(1.8358, 1.8439)	(1.7478, 1.7544)
	Slope P	<0.0001	<0.0001	**<0.0001**	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
	R^2	0.9992	0.9998	**1**	1	1	1	1	1
Generalised TL	Intercept	3.7126	2.7104	−0.6225	1.6554	1.9023	1.9288	2.0910	−0.1513
		(1.9591, 5.4661)	(1.8132, 3.6076)	(−1.5368, 0.2919)	(1.5739, 1.7369)	(1.8601, 1.9444)	(1.8977, 1.9598)	(2.0596, 2.1224)	(−0.2914, −0.0112)
	Linear coefficient	−0.4466	0.4393	3.3844	1.3706	1.1513	1.1216	0.9757	2.9229
		(−2.0311, 1.1379)	(−0.3699, 1.2485)	(2.6068, 4.1620)	(1.3026, 1.4387)	(1.1164, 1.1863)	(1.0961, 1.1472)	(0.9499, 1.0014)	(2.8054, 3.0403)
	Quadratic coefficient	0.4889	0.2932	−0.3572	0.0879	0.1366	0.1445	0.1772	−0.2455
		(0.1311, 0.8467)	(0.1108, 0.4756)	(−0.5225, −0.1919)	(0.0737, 0.1021)	(0.1293, 0.1438)	(0.1392, 0.1497)	(0.1720, 0.1825)	(−0.2701, −0.2209)
	Quadratic coefficient P	0.0119	0.0045	0.0002	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
	R^2	0.9993	0.9999	1	1	1	1	1	1

Parameter estimates of linear (Eq. 1, TL) and quadratic (Eq. 2, generalised TL) regressions fitted to the dependent variable log variance and the independent variable log mean of observed (1978–2010 and 1985–2010) and projected (2011–2040) county population density from six projections. Point estimates were followed by their corresponding 95% CIs in the parentheses. Statistics of selected projection using variance functions for the 1978–2010 data are indicated by bold face

Table 2 Short-term variance function analysis of data and projections by Taylor's law and generalised Taylor's law

Variance function	Coefficient and statistics	Historical data [2006–2010]	MMM0 [2011–2015]	MMML [2011–2015]	MMMM [2011–2015]	MMMH [2011–2015]	HHMH [2011–2015]	LLML [2011–2015]
TL	Intercept	1.1180	1.3718	1.1591	1.1465	**1.1242**	**1.1055**	1.1728
		(1.0906, 1.1453)	(1.3698, 1.3737)	(1.1525, 1.1657)	(1.1423, 1.1507)	(**1.1233, 1.1252**)	(**1.1042, 1.1068**)	(1.1613, 1.1842)
	Slope	1.8062	1.6961	1.7884	1.7939	**1.8035**	**1.8116**	1.7825
		(1.7942, 1.8182)	(1.6953, 1.6970)	(1.7856, 1.7912)	(1.7921, 1.7957)	(**1.8031, 1.8039**)	(**1.8110, 1.8121**)	(1.7776, 1.7874)
	Slope P	< 0.0001	< 0.0001	< 0.0001	< 0.0001	**< 0.0001**	**< 0.0001**	< 0.0001
	R^2	1	1	1	1	**1**	**1**	1
Generalised TL	Intercept	5.0267	1.6677	−0.1890	0.3610	0.9844	1.2798	−1.2633
		(4.2476, 5.8058)	(0.5348, 2.8006)	(−0.2197, −0.1582)	(0.1822, 0.5398)	(0.9076, 1.0612)	(1.1585, 1.4011)	(−1.4830, −1.0436)
	Linear coefficient	−1.6223	1.4408	2.9436	2.4664	1.9230	1.6627	3.8706
		(−2.3056, −0.9390)	(0.4632, 2.4183)	(2.9172, 2.9699)	(2.3133, 2.6194)	(1.8573, 1.9887)	(1.5591, 1.7663)	(3.6823, 4.0589)
	Quadratic coefficient	0.7518	0.0551	−0.2475	−0.1439	−0.0255	0.0318	−0.4474
		(0.6019, 0.9016)	(−0.1558, 0.2660)	(−0.2531, −0.2418)	(−0.1767, −0.1112)	(−0.0396, −0.0115)	(0.0097, 0.0539)	(−0.4878, −0.4071)
	Quadratic coefficient P	0.0102	0.6596	0.0001	0.0132	0.0704	0.1064	0.0021
	R^2	1	1	1	1	1	1	1

Parameter estimates of linear (Eq. 1, TL) and quadratic (Eq. 2, generalised TL) regressions fitted to the dependent variable log variance and the independent variable log mean of observed (2006–2010) and projected (2011–2015) county population density from six projections. Point estimates were followed by their corresponding 95% normal CIs in the parentheses. Statistics of selected projections are indicated by bold face

these two projections, which did not reject an absence of lag-1 error autocorrelation or the constancy in regression error variances, multiple linear regression based on Eq. (3) did not reject the hypothesis that TL's intercept was equal to that of the historical data ($P = 0.1576$ for HHMH and $P = 0.4810$ for MMMH), nor did it reject the hypothesis that TL's slope was equal to that of the historical data ($P = 0.1560$ for HHMH and $P = 0.4857$ for MMMH). The two high-migration variants HHMH and MMMH were selected as the two best projections.

The generalised TL significantly improved TL for the historical data, MMML, MMMM, and LLML (quadratic coefficient $P < 0.05$, see Table 2), but not for MMM0, MMMH, or HHMH (quadratic coefficient $P \geq 0.05$, see Table 2). Constancy in regression error variances was not rejected in the historical data or in any projection. Absence of lag-1 error autocorrelation was rejected in LLML, but not in the historical data or any other projection. Multiple linear regressions based on Eq. (4) showed that none of the projections resembled the historical data in any parameter of the least-squares quadratic regression. Thus no projection can be selected using the generalised TL.

Accuracy of 2011–2015 projections

Among the six projections, 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 followed an identical ascending order: HHMH < MMMH < MMMM < MMML < LLML < MMM0. Under each measure, HHMH and MMMH are the two projections with the least average errors and therefore the most accurate. ANOVA showed that annual average of MAPE (and of MAPE-R) differed significantly among the six projections ($P = 0.0004$ and 0.0002 respectively), with MMM0 having the largest errors under both measures (Fig. 2). Excluding MMM0, annual average of MAPE (and of MAPE-R) among the five remaining projections was not significantly different according to ANOVA ($P = 0.1023$ and $P = 0.0579$ respectively). The similarity between the results using MAPE and those using MAPE-R was in agreement with Rayer (2007).

Discussion and conclusion

We summarize the procedure of selecting among various projections using the proposed variance function method so that the method can be tested further and used if it is successful. First, choose historical population time series that immediately precede and are comparable in length with the projections (Smith et al. 2013). Second, for the historical observations and for each of the projections separately, calculate for each year the spatial means (among spatial units) and spatial variances (among spatial units), and construct the multiple regression models (Eqs. 3 and 4, Materials and methods section) to compare the variance function parameters of the historical observations with the corresponding variance function parameters of each projection. Third, select projections of which the variance function parameters do not differ significantly from those of the historical observations. These steps do not
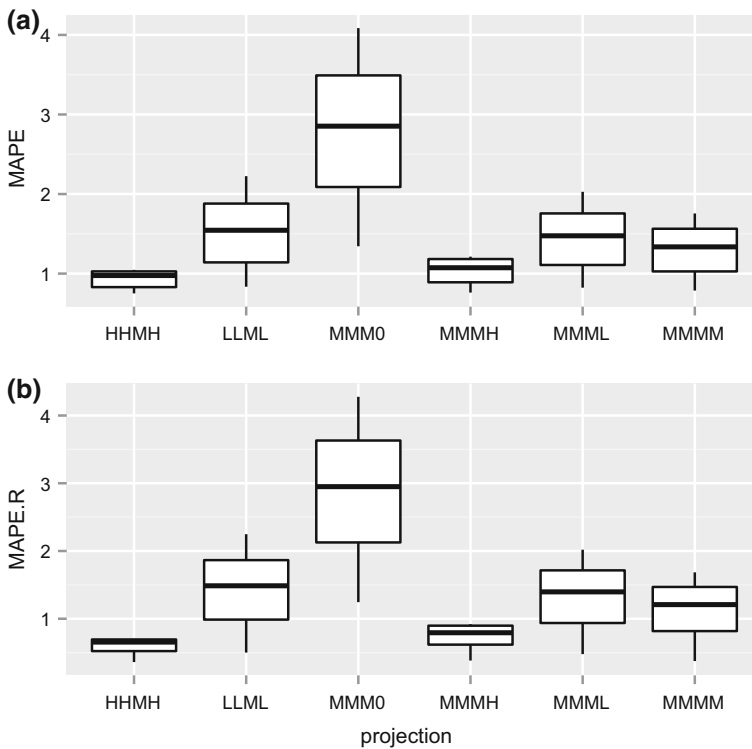
Fig. 2 Boxplots of **a** mean absolute percentage error (MAPE) and **b** rescaled mean absolute percentage error (MAPE-R) between observed and projected county population density from 2011 to 2015 for each of the six demographic projections. The *line* in the *middle* of *each box* represents the median within a projection. The *upper* bound and *lower* bound of *each box* represent the 3rd and 1st quartile, respectively, for each projection. The whiskers of *each box* extend up and down to the maximal and minimal values, respectively

depend on heavy computational machinery and require statistical analysis that is available in standard statistical software.

In testing the assumptions of our regression models, we found sometimes that the Breusch–Pagan test rejected the hypothesis of constant variance of the residuals and the Durbin–Watson test rejected the hypothesis of no lag-1 autocorrelation in the residuals. Because heteroscedasticity and serially correlated residuals do not bias estimates of the coefficients in ordinary least squares regression (Bailey 2016, p. 69), we do not think these violations are likely to invalidate our major conclusions. Nevertheless, if the residuals are positively serially correlated, the standard errors of coefficient estimates are too small and R^2 is biased upward, so the differences between historical and projected variance functions could be overstated. Future analyses should test the variance function approach using methods that make fewer assumptions contradicted by the historical observations and projections. For example, Huber–White standard errors are consistent with heteroscedasticity (Bailey 2016, p. 68, his Eq. 3.12) and the Cochrane and Orcutt (1949) and Prais

and Winsten (1954) estimators adjust for serial correlation of the residuals (Bailey 2016). These refinements are available from standard software (Bailey 2016, pp. 82–85).

Variance function analysis of the short-term historical data (2006–2010) and 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 the closest to the recent data (2011–2015). This finding encourages further testing and possible use of the variance function in selecting the most accurate short-term demographic projections. The same analysis can be applied to Norwegian county population projections in the near future (e.g. 2016–2020) based on recent data (2011–2015) to choose among more recent short-term projections.

The high net immigration rate in the selected short-term projections (HHMH and MMMH) reflected Norway's recent economic and political experiences. In the past decade, Norway experienced high and rapidly increasing immigration due to low unemployment and rapid economic growth. In 2004 and 2007 together, ten east European countries were admitted to the European Union (Cappelen et al. 2011). There has been a continuing influx of immigrants from Poland, Lithuania, Sweden and other countries of the European Union since 2007 (Statistics Norway 2015). Although Norway is not a member of the European Union, Norway has been a member in the EEC (European Economic Cooperation) area since 1994, which allows EU citizens to work and live in Norway with almost no restrictions.

Comparison of variance functions between the long-term historical observations (1978–2010) and the long-term demographic projections (2011–2040) favored MMM0 as the best projection. This choice differed from the projections selected in the short-term analysis (HHMH and MMMH). The absence of net immigration to Norway in the selected projection from the long-term analysis seems unrealistic and contradicts the recent high immigration to Norway. Three reasons may contribute to this finding.

First, it is possible that the effect of net immigration to Norwegian counties was balanced out in the spatial variance used in the analysis, i.e. if we treat county population density as a random variable X, and assume that the net immigration to Norwegian counties is a constant c for all counties, then given all other demographic variables being equal, neither the presence nor the size of net immigration will affect the spatial variance because variance($X + c$) = variance(X). This property of the variance could prevent us from differentiating the assumptions (0, L, M, and H) about net immigration, and consequently we would select poor projections. However, this potential explanation is inadequate because Norway's cities, particularly Oslo, have received many more immigrants, compared to their population size, than the more rural counties compared to theirs. Given the accuracy of selected projections in the short-term analysis and the unrealistic assumption of equal immigration into Norwegian counties, we doubt that the selection of MMM0 can be attributed to this limitation of the variance function technique.

Second, in the variance function plot (Fig. 1c) of historical observations (1978–2010), we observed an acute "hook" pattern during 1978–1984 (in Fig. 1c,

the "hook" during 1978–1984 is superimposed on the main line of the data after 1984). This is caused by the decline of the mean and the variance of population densities across all counties (Fig. 1a, b) due to the drop in Oslo's population density during 1978–1984 (Cohen et al. 2013). Statistically, the overlapping log(mean)-log(variance) pairs near the earlier historical years could bias estimates of the regression parameters, and therefore distort model selection. Repeating our variance function parameter comparison between the historical data from 1985 to 2010 and the six projections from 2011 to 2040, we found that none of the projections agreed with the historical data in the parameters of TL (Eq. 1) or the generalised TL (Eq. 2). In this case, the consistent trend in the variance function plot helped identify a better time interval (1985–2010) for comparison, based on which the unrealistic projection MMM0 was rejected. The sensitivity of our variance function method to the length of historical time series requires further statistical testing using other population data.

A third possible explanation, which we favour over the two prior explanations, of why MMM0 was selected by long-term analysis is that a long times series of historical data (1978–2010) may be more misleading than the recent past. Before 2000, Norway had a stringent immigration policy (Horst et al. 2010) and received significantly less immigration than in the recent decade (Fig. 3). In addition, the projections with low or zero net immigration were based on an assumed declining relative income level in Norway due to lower revenue from the oil and gas sector. These assumptions agreed with the registered net immigration data prior to 2004 (Fig. 3), and reflected Norway's economic and political situations during that period (Cohen et al. 2013). Our variance function was heavily influenced by the low level
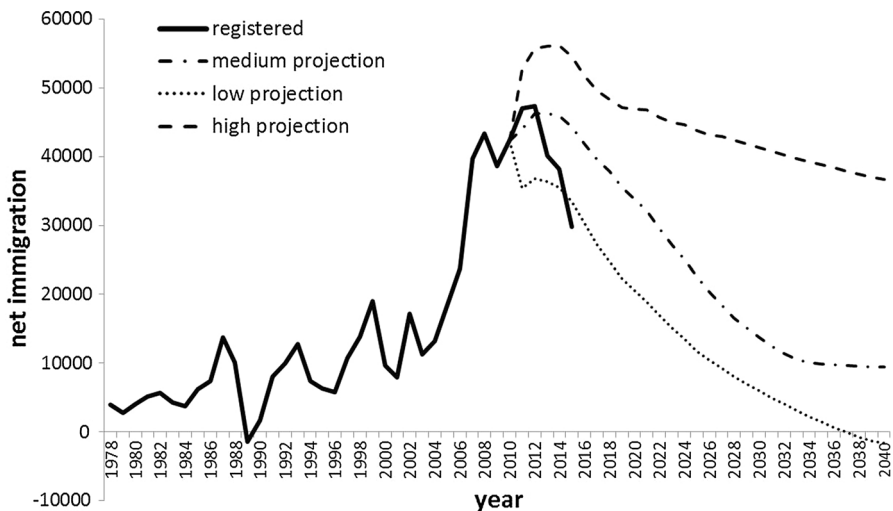

Fig. 3 Net immigration to Norway registered between 1978 and 2015 and projected between 2011 and 2040 under high (*H*), medium (*M*), and low (*L*) net immigration assumptions separately. Net immigration rose (irregularly) from 1990 to about 2010. It rose especially quickly from about 2004 to 2010 and fell sharply since then

of net immigration in the earlier historical years as assumed in projection MMM0. The variance function did not anticipate the recent major change in net immigration.

In a study of how much of the history of international migratory flows is relevant to the future of international migratory flows, Cohen et al. (2008, p. 15272) found that data from the most recent five years (which were then 2000–2004) gave a higher multiple R^2 in modelling migratory flows than data from 1960 to 2004. That is, fitting their model to data from 2000 to 2004 gave a higher correlation with the migrant flows reported from 2000 to 2004 than fitting their model to data from 1960 to 2004 gave to data reported from 1960 to 2004. The dynamics of migration shifted over the years from 1960 to 2004, but the parameters of the model remained constant, leading to a poorer fit of the model to data over the longer period. In influenza epidemiology, Goel et al. (2010, p. 17489) found that a simple autoregressive model to predict a current week's influenza caseload from the prior week's influenza caseload had a higher correlation (0.95) than an autoregressive model based on the influenza caseload two and three weeks before the current week. Though international migration and influenza caseloads change on different timescales (years or decades vs. weeks or months), the present or immediate future is best predicted from the immediate past on the appropriate time scale in both cases. This claim is supported by the accuracy of the selected short-term Norwegian county projections in the current work.

Studies on TL using demographic models are limited. To our knowledge, the first work in this direction was by Anderson et al. (1982), who proposed that the random birth–death–emigration–immigration models yielded TL slopes similar to those from ecological data. They also argued that TL is a natural consequence of random demographic events, questioning the species-specific theory of TL proposed by Taylor and his colleagues (Taylor and Taylor 1977; Taylor et al. 1988). Our results showed that TL and its generalisation can be useful in selecting short-term demographic projections of Norwegian counties based on short-term historical data. Comparisons of parameters of TL or its quadratic generalisation identified some short-term projections of future population growth that were consistent with recent data. However, in our Norwegian example, the variance function method did not select well among long-term projections. To understand how general our method is and the roles of time scales in this method, other high-quality census data and projections should be tested similarly.

Appendix

Overview of MAPE-R

Rescaled mean absolute percentage error (MAPE-R) is a modified post hoc measure of population forecast accuracy based on mean absolute percentage error (MAPE) developed by Swanson et al. (2000). Since absolute percentage errors (APEs)

between observed and projected population sizes are often right-skewed, the central tendency measure, MAPE, is heavily influenced by outliers and tends to overestimate the true error (Rayer 2007). To describe the percentage errors more robustly, taking account of their large outliers, Swanson et al. (2000) first calculated a modified Box-Cox transformation of APEs. The parameter of the modified Box-Cox transformation (Box and Cox 1964) was selected to maximize the likelihood that the transformed APEs were symmetrical, so that the mean became appropriate. Symmetry of APE was evaluated using D'Agostino's test of skewness (D'Agostino 1970), where a rejection of the null hypothesis of zero skewness indicated that APE was not symmetrically distributed and the Box-Cox transformation was necessary. Then Swanson et al. (2000) calculated the inverse of the mean transformed APEs (MAPE-R) to return them to the original data scale.

Details of the theory and empirical test of MAPE-R can be found in Coleman and Swanson (2007) and Swanson et al. (2000, 2011).

Implementation of MAPE-R

We followed the procedure outlined in Swanson et al. (2011) to implement MAPE-R. To measure the accuracy of Norwegian county population projections between 2011 and 2015 for each of the six demographic projections and each year from 2011 to 2015, we first examined the distribution of APE between the projected and observed county population density and tested its skewness using the right-sided D'Agostino's test of skewness with significance level 0.1 (Table S1). Here a 0.1 significance level was used instead of a lower value to avoid ''a greater cost in terms of a downwardly biased measure of accuracy in not transforming a potentially skewed distribution'' (Swanson et al. 2011). If the hypothesis that the skewness of APEs was zero was rejected ($P < 0.1$), then for each projection from each year, we performed the Box-Cox transformation for APE of each county, estimated the transformation parameter (λ from -2 to 2) that maximized the likelihood function for symmetry, averaged the transformed APEs over counties, and calculated MAPE-R using the inverse of Box-Cox transformation (Box and Cox 1964). Finally, the five MAPE-Rs (one from each year from 2011 to 2015) of each projection were compared among the six projections using the one-way ANOVA for the equality of their inter-annual average (see Materials and methods). We used MAPE for comparison if the hypothesis that the APEs had zero skewness was not rejected. Skewness and the corresponding P of APE and transformed APE (APE-T) were listed in Table S1.

D'Agostino's test of skewness was performed using the function ''agostino.test'' from R package ''moments'' (Komsta and Novomestky 2015). Parameter λ at which the likelihood function achieved maxima was found using R function ''optimize'' (R Core Team 2015). R code for the calculation of MAPE-R was given below.

```
===========================================================

library(moments)

load "ape" data set, which contains the absolute percentage error (APE) of county
population density between each projection and the data for each year from 2011 to 2015.

apebyyear<-subset(ape, Year==year)              # select APEs from each year

vects <- list(apebyyear$APE_HHMH, apebyyear$APE_LLML,
apebyyear$APE_MMML, apebyyear$APE_MMMH,
apebyyear$APE_MMM0,apebyyear$APE_MMMM)          # split APEs by projection

for(v in vects){

print(agostino.test(v, alternative="less"))     # perform D'Agostino test of skewness

# create Box-Cox transformation of each county for the given projection and given year
(lambda is the parameter of Box-Cox transformation)

BC1<-function(lambda){ifelse(lambda!=0,(v[1]^lambda-lambda)/lambda,log(v[1]))}

BC2<-function(lambda){ifelse(lambda!=0,(v[2]^lambda-lambda)/lambda,log(v[2]))}

BC3<-function(lambda){ifelse(lambda!=0,(v[3]^lambda-lambda)/lambda,log(v[3]))}

BC4<-function(lambda){ifelse(lambda!=0,(v[4]^lambda-lambda)/lambda,log(v[4]))}

BC5<-function(lambda){ifelse(lambda!=0,(v[5]^lambda-lambda)/lambda,log(v[5]))}

BC6<-function(lambda){ifelse(lambda!=0,(v[6]^lambda-lambda)/lambda,log(v[6]))}

BC7<-function(lambda){ifelse(lambda!=0,(v[7]^lambda-lambda)/lambda,log(v[7]))}

BC8<-function(lambda){ifelse(lambda!=0,(v[8]^lambda-lambda)/lambda,log(v[8]))}

BC9<-function(lambda){ifelse(lambda!=0,(v[9]^lambda-lambda)/lambda,log(v[9]))}

BC10<-function(lambda){ifelse(lambda!=0,(v[10]^lambda-lambda)/lambda,log(v[10]))}

BC11<-function(lambda){ifelse(lambda!=0,(v[11]^lambda-lambda)/lambda,log(v[11]))}

BC12<-function(lambda){ifelse(lambda!=0,(v[12]^lambda-lambda)/lambda,log(v[12]))}

BC13<-function(lambda){ifelse(lambda!=0,(v[13]^lambda-lambda)/lambda,log(v[13]))}

BC14<-function(lambda){ifelse(lambda!=0,(v[14]^lambda-lambda)/lambda,log(v[14]))}

BC15<-function(lambda){ifelse(lambda!=0,(v[15]^lambda-lambda)/lambda,log(v[15]))}

BC16<-function(lambda){ifelse(lambda!=0,(v[16]^lambda-lambda)/lambda,log(v[16]))}

BC17<-function(lambda){ifelse(lambda!=0,(v[17]^lambda-lambda)/lambda,log(v[17]))}
```

```
BC18<-function(lambda){ifelse(lambda!=0,(v[18]^lambda-lambda)/lambda,log(v[18]))}

BC19<-function(lambda){ifelse(lambda!=0,(v[19]^lambda-lambda)/lambda,log(v[19]))}


BC_avg<-
function(lambda){(BC1(lambda)+BC2(lambda)+BC3(lambda)+BC4(lambda)+BC5(lambda)+B
C6(lambda)+BC7(lambda)+BC8(lambda)+BC9(lambda)+BC10(lambda)+BC11(lambda)+BC12
(lambda)+BC13(lambda)+BC14(lambda)+BC15(lambda)+BC16(lambda)+BC17(lambda)+BC1
8(lambda)+BC19(lambda))/19}

BC_var<-function(lambda){((BC1(lambda)-BC_avg(lambda))^2+(BC2(lambda)-
BC_avg(lambda))^2+(BC3(lambda)-BC_avg(lambda))^2+(BC4(lambda)-
BC_avg(lambda))^2+(BC5(lambda)-BC_avg(lambda))^2+(BC6(lambda)-
BC_avg(lambda))^2+(BC7(lambda)-BC_avg(lambda))^2+(BC8(lambda)-
BC_avg(lambda))^2+(BC9(lambda)-BC_avg(lambda))^2+(BC10(lambda)-
BC_avg(lambda))^2+(BC11(lambda)-BC_avg(lambda))^2+(BC12(lambda)-
BC_avg(lambda))^2+(BC13(lambda)-BC_avg(lambda))^2+(BC14(lambda)-
BC_avg(lambda))^2+(BC15(lambda)-BC_avg(lambda))^2+(BC16(lambda)-
BC_avg(lambda))^2+(BC17(lambda)-BC_avg(lambda))^2+(BC18(lambda)-
BC_avg(lambda))^2+(BC19(lambda)-BC_avg(lambda))^2)/19}

ml<-function(lambda){(-19/2)*log(BC_var(lambda))+(lambda-1)*sum(log(v))}    #
```
calculating the maximum likelihood function of transformed APEs

```
curve(ml, from=-2,to=2)        # graphing likelihood function to inspect location of its
```
maxima

```
max<-optimize(ml, interval=c(0,2), maximum=TRUE)        # searching for local
```
maximum of likelihood function

```
mape_t<-
(BC1(max$maximum)+BC2(max$maximum)+BC3(max$maximum)+BC4(max$maximum)+B
C5(max$maximum)+BC6(max$maximum)+BC7(max$maximum)+BC8(max$maximum)+BC9(
max$maximum)+BC10(max$maximum)+BC11(max$maximum)+BC12(max$maximum)+BC1
3(max$maximum)+BC14(max$maximum)+BC15(max$maximum)+BC16(max$maximum)+BC
17(max$maximum)+BC18(max$maximum)+BC19(max$maximum))/19   # calculating average
```
of transformed APEs

```
print(mape_r<-(max$maximum*(mape_t+1))^(1/max$maximum)) # print MAPE-Rs

print(mape<-mean(v))        # print MAPEs

}
```

==

References

Alho, J., & Spencer, B. D. (1997). The practical specification of the expected error of population forecasts. *Journal of Official Statistics, 13*, 203–225.

Alkema, L., Raftery, A. E., Gerland, P., Clark, S. J., Pelletier, F., Buettner, T., et al. (2011). Probabilistic projections of the total fertility rate for all countries. *Demography, 48*, 815–839.

Anderson, R. M., Gordon, D. M., Crawley, M. J., & Hassell, M. P. (1982). Variability in the abundance of animal and plant species. *Nature, 296*, 245–248.

Anderson, R. M., & May, R. M. (1988). Epidemiological parameters of HIV transmission. *Nature, 333*, 514–519.

Bailey, M. A. (2016). *Real econometrics: The right tools to answer important questions*. New York, NY: Oxford University Press.

Bongaarts, J., & Bulatao, R. A. (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.

Box, G. E. P., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society: Series B, 26*(2), 211–252.

Breusch, T. S., & Pagan, A. R. (1979). A simple test for heteroscedasticity and random coefficient variation. *Econometrica, 47*, 1287–1294.

Brunborg, H., & Cappelen, Å. (2010). Forecasting migration flows to and from Norway using an econometric model. In *Work session on demographic projections, Eurostat methodological working papers*, pp. 321–344. http://ec.europa.eu/eurostat/documents/3888793/5848129/KS-RA-10-009-EN.PDF/fbaff784-8e78-417b-8b51-7f8fdf2f3031.

Brunborg, H., & Texmon, I. (2011). Befolkningsframskrivning 2011–2100: Modell og forutsetninger. *Økonomiske Analyser, 4*, 33–45.

Cappelen, Å., Ouren, J., & Skjerpen, T. (2011). Effects of immigration policies on immigration to Norway 1969–2010. Report. Statistics Norway. http://www.udi.no/en/statistics-and-analysis/research-and-development-reports/effects-of-immigration-policies-on-immigration-to-norway-1969-2010-2011/. Accessed 8 August 2015.

Cochrane, D., & Orcutt, G. H. (1949). Application of least squares regression to relationships containing auto-correlated error terms. *Journal of the American Statistical Association, 44*(245), 32–61. doi:10.1080/01621459.1949.10483290.

Cohen, J. E. (1986). Population forecasts and confidence intervals for Sweden: A comparison of model-based and empirical approaches. *Demography, 23*(1), 105–126. **(erratum 25(2), 315, 1988)**.

Cohen, J. E., Roig, M., Reuman, D. C., & GoGwilt, C. (2008). International migration beyond gravity: A statistical model for use in population projections. *Proceedings of the National Academy of Sciences USA, 105*(40), 15269–15274.

Cohen, J. E., Xu, M., & Brunborg, H. (2013). Taylor's law applies to spatial variation in a human population. *Genus, 69*(1), 25–60.

Coleman, C. D., & Swanson, D. A. (2007). On MAPE-R as a measure of cross-sectional estimation and forecast accuracy. *Journal of Economic and Social Measurement, 32*(4), 219–233.

D'Agostino, R. B. (1970). Transformation to normality of the null distribution of G1. *Biometrika, 57*(3), 679–681.

Dietz, T., Rosa, E. A., & York, R. (2007). Driving the human ecological footprint. *Frontiers in Ecology and the Environment, 5*, 13–18.

Eisler, Z., Bartos, I., & Kertész, J. (2008). Fluctuation scaling in complex systems: Taylor's law and beyond. *Advances in Physics, 57*(1), 89–142.

Fox, J. (2008). *Applied regression analysis and generalised linear models* (2nd ed.). New York: Sage.

Fox, J., & Weisberg, S. (2011). *An R companion to applied regression* (2nd ed.). Thousand Oaks, CA: Sage.

Gerland, P., Raftery, A. E., Ševčíková, H., Li, N., Gu, D., Spoorenberg, T., et al. (2014). World population stabilization unlikely this century. *Science, 346*(6206), 234–237.

Goel, S., Hofman, J. M., Lahaie, S., Pennock, D. M., & Watts, D. J. (2010). Predicting consumer behavior with Web search. *Proceedings of the National Academy of Sciences USA, 107*(41), 17486–17490.

Greig, A., Dewhurst, J., & Horner, M. (2014). An application of Taylor's power law to measure overdispersion of the unemployed in English labor markets. *Geographical Analysis, 47*(2), 121–133.

Hanley, Q. S., Khatun, S., Yosef, A., & Dyer, R.-M. (2014). Fluctuation scaling, Taylor's law, and crime. *PLoS ONE, 9*(10), e109004. doi:10.1371/journal.pone.0109004.

Heyde, C. C., & Cohen, J. E. (1985). Confidence intervals for demographic projections based on products of random matrices. *Theoretical Population Biology, 27*(2), 120–153.

Horst, C., Carling, J., & Ezzati, R. (2010). Immigration to Norway from Bangladesh, Brazil, Egypt, India, Morocco and Ukraine. In *PRIO Paper*. Oslo: Peace Research Institute Oslo.

Keeling, M., & Grenfell, B. (1999). Stochastic dynamics and a power law for measles variability. *Philosophical Transactions of the Royal Society B, 354*(1384), 769–776.

Komsta, L., & Novomestky, F. (2015). Moments: Moments, cumulants, skewness, kurtosis and related tests. R package version 0.14. http://CRAN.R-project.org/package=moments

Lee, R. D., & Carter, L. R. (1992). Modelling and forecasting US mortality. *Journal of the American Statistical Association, 87*(419), 659–671.

Lee, R. D., & Tuljapurkar, S. (1994). Stochastic population forecasts for the United States: Beyond high, medium, and low. *Journal of the American Statistical Association, 89*(428), 1175–1189.

Lutz, W., & Goujon, A. (2004). Literate life expectancy: charting the progress in human development. In W. Lutz & W. Sanderson (Eds.), *The end of world population growth in the 21st century: new challenges for human capital formation and sustainable development* (pp. 159–186). London: Earthscan.

Lutz, W., Butz, W. P., & KC, S. (2014). *World population and human capital in the twenty first century*. New York: Oxford University Press.

Lutz, W., Sanderson, W., & Scherbov, S. (1997). Doubling of world population unlikely. *Nature, 387*, 803–805.

Lutz, W., Sanderson, W., & Scherbov, S. (2001). The end of world population growth. *Nature, 412*, 543–545.

Prais, S. J., & Winsten, C. B. (1954). Trend estimators and serial correlation. Cowles Commission Discussion Paper No. 383. Chicago, IL.

Raftery, A. E., Chunn, J. L., Gerland, P., & Ševcíková, H. (2013). Bayesian probabilistic projections of life expectancy for all countries. *Demography, 50*, 777–801.

Rayer, S. (2007). Population forecast accuracy: Does the choice of summary measure of error matter? *Population Research and Policy Review, 26*(2), 163–184.

R Core Team. (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Smith, S. K., Tayman, J., & Swanson, D. A. (2013). *A practitioner's guide to state and local population projections*. New York: Springer.

Statistics Norway. (2011). Population projections, 2011–2100. Report. http://www.ssb.no/en/befolkning/statistikker/folkfram/aar/2011-06-16. Accessed 7 April 2015.

Statistics Norway. (2015). Immigrants and Norwegian-born to immigrant parents. Report. http://www.ssb.no/en/befolkning/statistikker/innvbef. 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.

Stoto, M. A. (1983). The accuracy of population projections. *Journal of the American Statistical Association, 78*(381), 13–20.

Swanson, D. A., Tayman, J., & Barr, C. F. (2000). A note on the measurement of accuracy for subnational demographic estimates. *Demography, 37*(2), 193–201.

Swanson, D. A., Tayman, J., & Bryan, T. M. (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., Perry, J. N., Woiwod, I. P., & Taylor, R. A. J. (1988). Specificity of the spatial power-law exponent in ecology and agriculture. *Nature, 332*, 721–722.

Taylor, L. R., & Taylor, R. A. J. (1977). Aggregation, migration and population mechanics. *Nature, 265*, 415–421.

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*(2), 383–406.

The World Bank. (2014). Population estimates and Projections. Report. http://data.worldbank.org/data-catalog/population-projection-tables. Accessed 7 April 2015.

United Nations. (2014). World population prospects: The 2012 revision. Report. Department of Economic and Social Affairs, Population Division. http://esa.un.org/wpp/. Accessed 7 April 2015.

US Census Bureau. (2014). 2014 National population projections. Report. http://www.census.gov/population/projections/data/national/2014.html. Accessed 7 April 2015.

Wickham, H. (2009). *ggplot2: Elegant graphics for data analysis*. New York: Springer.

Table S1. Mean absolute percentage error (MAPE) and rescaled mean absolute percentage error (MAPE-R) of each projection in each year between 2011 and 2015. Skewness and *P* of the D'Agostino's test of skewness for absolute percentage error (APE) and transformed absolute percentage error (APE-T) are shown in the last four columns. Transformation reduced the skewness value but did not change its significance (*P* < 0.1).

projection	year	MAPE	MAPE-R	Skewness (APE)	*P* (APE)	Skewness (APE-T)	*P* (APE-T)
HHMH	2011	0.7512	0.3606	2.5957	<0.0001	1.4654	0.0027
LLML	2011	0.8351	0.5008	2.4200	<0.0001	1.5024	0.0023
MMML	2011	0.8231	0.478	2.4325	<0.0001	1.4834	0.0025
MMMH	2011	0.7609	0.383	2.5592	<0.0001	1.5332	0.0020
MMM0	2011	1.3435	1.2449	2.0226	0.0002	1.7939	0.0006
MMMM	2011	0.7869	0.3753	2.4616	<0.0001	1.1924	0.0091
HHMH	2012	0.8298	0.5225	2.4871	<0.0001	1.4877	0.0024
LLML	2012	1.1408	0.9872	2.0743	0.0002	1.5531	0.0018
MMML	2012	1.1083	0.9371	2.0981	0.0001	1.4942	0.0023
MMMH	2012	0.8908	0.6177	2.3615	<0.0001	1.4931	0.0024
MMM0	2012	2.0884	2.1265	1.2031	0.0087	1.2838	0.0061
MMMM	2012	1.0275	0.8177	2.1809	0.0001	1.4436	0.0030
HHMH	2013	0.9755	0.7024	2.1317	0.0001	1.3665	0.0042
LLML	2013	1.5447	1.4864	1.4382	0.003	1.2907	0.0059
MMML	2013	1.4749	1.3967	1.4851	0.0024	1.2805	0.0062
MMMH	2013	1.0736	0.7935	1.8091	0.0006	0.9699	0.0237
MMM0	2013	2.8534	2.9499	0.8504	0.0387	1.1631	0.0104
MMMM	2013	1.3359	1.2088	1.5708	0.0017	1.2243	0.0079
HHMH	2014	1.0276	0.6941	2.0347	0.0002	1.1358	0.0117
LLML	2014	1.8801	1.8653	1.1964	0.0089	1.1657	0.0102
MMML	2014	1.7562	1.7143	1.2497	0.0071	1.1627	0.0104

MMMH	2014	1.1829	0.9198	1.6821	0.0010	1.0720	0.0154
MMM0	2014	3.4913	3.6309	0.8412	0.0402	1.1769	0.0097
MMMM	2014	1.5637	1.4694	1.3471	0.0046	1.1452	0.0112
HHMH	2015	1.0469	0.6555	2.0823	0.0002	1.0447	0.0173
LLML	2015	2.2247	2.2479	1.0573	0.0164	1.1087	0.0131
MMML	2015	2.0282	2.0194	1.1018	0.0135	1.0837	0.0146
MMMH	2015	1.2129	0.8977	1.6339	0.0012	1.0106	0.0200
MMM0	2015	4.0857	4.2763	0.7610	0.0551	1.0425	0.0175
MMMM	2015	1.7558	1.6851	1.2475	0.0071	1.1012	0.0136