MEMORANDUM

No 22/2010

The sensitivity of the Scaled Model of Error with respect to the choice of the correlation parameters: A Simulation Study

Rebecca Graziani
Nico Keilman

ISSN: 0809-8786

Department of Economics
University of Oslo

This series is published by the
University of Oslo
Department of Economics

P. O.Box 1095 Blindern
N-0317 OSLO Norway
Telephone: + 47 22855127
Fax: + 47 22855035
Internet: http://www.sv.uio.no/econ
e-mail: econdep@econ.uio.no

In co-operation with
The Frisch Centre for Economic Research

Gaustadalleén 21
N-0371 OSLO Norway
Telephone: +47 22 95 88 20
Fax: +47 22 95 88 25
Internet: http://www.frisch.uio.no
e-mail: frisch@frisch.uio.no

Last 10 Memoranda

No 21/10	Jennifer L. Castle, Jurgen A. Doornik , David F. Hendry, Ragnar Nymoen *Testing the Invariance of Expectations Models of Inflation*
No 20/10	Erik Biørn *Identifying Trend and Age Effects in Sickness Absence from Individual Data: Some Econometric Problems*
No 19/10	Michael Hoel, Svenn Jensen *Cutting Costs of Catching Carbon* *Intertemporal effects under imperfect climate policy*
No 18/10	Hans Jarle Kind, Tore Nilssen, Lars Sørgard *Price Coordination in Two-Sided Markets: Competition in the TV Industry*
No 17/10	Vladimir Krivonozhko, Finn R. Førsund and Andrey V. Lychev *A Note on Imposing Strong Complementary Slackness Conditions in DEA*
No 16/10	Halvor Mehlum and Karl Moene *Aggressive elites and vulnerable entrepreneurs* *- trust and cooperation in the shadow of conflict*
No 15/10	Nils-Henrik M von der Fehr *Leader, Or Just Dominant? The Dominant-Firm Model Revisited*
No 14/10	Simen Gaure *OLS with Multiple High Dimensional Category Dummies*
No 13/10	Michael Hoel *Is there a green paradox?*
No 12/10	Michael Hoel *Environmental R&D*

Previous issues of the memo-series are available in a PDF® format at:
http://www.sv.uio.no/econ/forskning/publikasjoner/memorandum

The sensitivity of the Scaled Model of Error with respect to the choice of the correlation parameters: A Simulation Study

by

Rebecca Graziani[1]
Department of Decision Sciences, Bocconi University, Milano

Nico Keilman[2]
Department of Economics, University of Oslo

23 November 2010

Abstract: The Scaled Model of Error has gained considerable popularity during the past ten years as a device for computing probabilistic population forecasts of the cohort-component type. In this report we investigate how sensitive probabilistic population forecasts produced by means of the Scaled Model of Error are for small changes in the correlation parameters. We consider changes in the correlation of the age-specific fertility forecast error increments across time and age, and changes in the correlation of the age-specific mortality forecast error increments across time, age and sex. Next we analyse the impact of such changes on the forecasts of the Total Fertility Rate and of the Male and Female Life Expectancies respectively.

For age specific fertility we find that the correlation across ages has only limited impact on the uncertainty in the Total Fertility Rate. As a consequence, annual numbers of births will be little affected. The autocorrelation in error increments is an important parameter, in particular in the long run. Also, the autocorrelation in error increments for age specific mortality is important. It has a large effect on long run uncertainty in life expectancy values, and hence on the uncertainty around the elderly population in the future. In empirical applications of the Scaled Model of Error, one should give due attention to a correct estimation of these two parameters.

Key words: Scaled model of error, Stochastic population forecast, Probabilistic cohort component model, Sensitivity, Correlation

JEL classifications: C15, C49, C63, J4

[1] Mail address: Via Roentgen 1, 20136 Milano, Italy. Email rebecca.graziani@unibocconi.it. Telephone +39 02 58 36 55 53. Fax +39 02 58 36 56 30.
[2] Mail address: P.O. Box 1095 Blindern, 0317 Oslo, Norway. Email nico.keilman@econ.uio.no. Telephone +47 22 85 51 28. Fax +47 22 85 50 35.

1 Introduction

Long-term population forecasts, covering a period of fifty years or more, are useful in a number of fields, two of which are analyses of the impact of population trends on contributions and expenditures for public old age pensions, and studies of demographically induced resource use and climate change. These long-term forecasts are necessarily uncertain: for a given country one may imagine many possible demographic futures, but some of these population developments are more probable than others. This calls for probabilistic forecasts, i.e. forecasts in terms of prediction intervals. Such prediction intervals quantify uncertainty – they express the probability that the future population (or age group, or number of births) is expected to fall within a certain range. Since the early 1990s several methods have been developed for computing probabilistic population forecasts. For an overview, see Alho and Spencer (2005), and the special issue of the International Statistical Review edited by Lutz and Goldstein (2004).

One method for probabilistic population forecasting that has gained considerable popularity in the past decade is Alho's Scaled Model of Error. See Alho and Spencer (1997, 2005) for a detailed description. The model, which is summarized below, is based on assumed statistical distributions for age-specific rates for fertility, age- and sex-specific rates for mortality, and age specific numbers of net migration. In addition to these distributions, the model also requires that one specify a number of correlations: correlation of mortality and fertility across ages and over time, and correlation between male and female mortality. Mortality, fertility, and migration are assumed independent under this model, an assumption that finds some empirical support for developed countries (Keilman 1987; Keilman and Pham 2004). The Scaled Model of error is implemented in the simulation program PEP ("Program for Error Propagation"; see Alho and Mustonen 2003, and Alho and Spencer 2005). PEP simulates prediction intervals for future population by repeatedly drawing parameter values for the statistical distributions of the Scaled Model of Error.

In this report we investigate the sensitivity of probabilistic population forecasts produced by means of the Scaled Model of Error with respect to the choice of the correlation parameters. The sensitivity for other parameters, such as the expected value or the variance in the fertility and mortality rates, or the migation numbers, can be assessed on intuitive grounds. But little is known about the sensitivity for correlations, in spite of the fact that the model has been applied extensively, for instance to the projection of many European populations, that of China, and the world population (e.g. Alho and Nikander 2004; Alho et al. 2006; Borgy and Alho 2007, Li et al. 2009). Yet it is important to know which correlation parameters should be given greatest attention in empirical applications of the model, and which will be of less importance for the results.

2 The Scaled Model of Error

We intend to evaluate the impact that a change in the specification of the correlation of the age-specific fertility forecast error increments across time and age and of the correlation of the age-specific mortality forecast error increments across time, age and sex has on the forecasts of the Total Fertility Rate and of the Male and Female Life Expectancies respectively. In our opinion a sensitivity analysis of this kind is extremely useful, since up to now the relevance and the impact of the choice of the Scaled Model of Error input parameters have not be discussed in detail. Such analysis will provide users with a better understanding of the model itself.

Within the Scaled Model of Error forecasts of the population by age and sex are derived by modeling the logarithm of a generic age-specific rate $R(j,t)$ as follows:

$$\log R(j,t) = \log \hat{R}(j,t) + X(j,t), \quad j = 1 \dots J, t = 1 \dots T$$

where j represents the age class, t the time, $\log \hat{R}(j,t)$ is a point forecast of the logarithm of the rate and the error term $X(j,t)$ is, for given age and time, represented as the following sum of forecast error increments

$$X(j,t) = \sum_{q=1}^{t} \varepsilon(j,q).$$

Finally the increments are modeled as the product of a deterministic scale term and the sum of two shocks, one age dependent and the other time and age dependent:

$$\varepsilon(j,t) = S(j,t)(\eta_j + \delta_{jt})$$

where the following assumptions are made on the error terms:

- the variables η_j's are assumed to have a Normal distribution with mean 0, variance κ_j, and the correlation $\rho(\eta_j \eta_i)$ is either set equal to ρ_η (constant correlation assumption) or it is set equal to $\rho_\eta^{|i-j|}$ (AR(1) structure);
- the variables δ_{jt}'s are assumed to be uncorrelated across time, to have, for every time *t*, a Normal distribution with mean 0 and variance $1-\kappa_j$, while the correlation $\rho(\delta_j \delta_i)$ is treated as for the η_j's terms, therefore either a constant correlation is assumed or an autoregressive structure;
- the variables η_j's and δ_{jt}'s are assumed to be uncorrelated.

In the following, we shall assume a constant κ for each age, and an autoregressive structure for the η_j's and δ_{jt}'s, with ρ_η and ρ_δ having the same value, that we shall denote by ρ. Therefore we shall assume $\rho(\eta_j \eta_i) = \rho(\delta_j \delta_i) = \rho^{|j-i|}$.

Under the previous assumptions the variance and the correlations across age and time of the forecast error increments are given by

- $Var(\varepsilon(j,t)) = S^2(j,t)$,
- $\rho(\varepsilon(j,t)\varepsilon(j,t+h)) = \kappa$,
- $\rho(\varepsilon(i,t)\varepsilon(j,t)) = \rho^{|i-j|}$
- $\rho(\varepsilon(i,t)\varepsilon(j,t+h)) = \rho^{|i-j|}\kappa$.

The variance of the error term $X(j,t)$ and the correlation across ages, for given time, have the following expressions

- $Var(X(j,t)) = \sum_{q=1}^{t} S^2(q,t) + \sum_{t_1 \neq t_2 = 1}^{t} \kappa S(j,t_1) S(j,t_2)$,
- $\rho(X(i,t), X(j,t) = \frac{t\rho^{|i-j|} + t(t-1)\kappa\rho^{|i-j|}}{(Var(X(j,t))Var(X(i,t)))^{1/2}}$

From the previous expressions, we expect increasing values of κ and ρ to be associated with increasing variances of the age-specific rates and therefore of the vital rates, built on them.

In the case of the age-specific mortality rates forecasts, an additional assumption is made: for a given age group j, the male and female η_j's and the male and female δ_{jt}'s terms are assumed to be correlated and we shall denote such correlation by ρ_g. We shall investigate the impact that changing values of ρ_g have on the variance of the difference of Male and Female Life Expectancies. The difference of these life expectancies is a parameter of considerable interest in many developed countries in view of the narrowing mortality sex gap. We expect increasing values of ρ_g to be associated with decreasing values of the variability of such difference.

The computer program PEP has been used, among others, to produce the forecast of 18 European countries, within the Uncertain Population of Europe Program (UPE) project (see Alders et al. 2007, for a description of the application of the Scaled Model of Error in the UPE project).

In order to investigate the sensitivity of the Total Fertility and Male and Female Life Expectancies forecasts to different values of κ and ρ and of ρ_g respectively, a simulation study is run and the results are discussed in the following sections. Our empirical application concerns a probabilistic forecast of the Italian Population, from 2009 to 2050, which is an update of the forecast produced within the UPE project. Expected values for fertility and mortality were taken from the latest (deterministic) projections released by the Italian National Statistical Office, ISTAT. This means that we have applied the following values for the TFR and life expectancies:

	TFR (ch/woman)	Life expectancy men (yrs)	Life expectancy women (yrs)
2009	1.40	78.9	84.4
2050	1.58	84.5	89.5

The assumptions made to derive the UPE forecasts on the scales parameters, the parameters κ and ρ, along with the assumptions on migration are the same as those used in the UPE project. Empirical estimates from six countries with long data series (1750-2000) had revealed that the logarithm of the total fertility was reasonably well modeled as a random walk with innovation variance equal to 0.06^2; see Alho and Spencer (2005 p. 254-255) for details. Thus for total fertility and age-specific fertility it was assumed in the UPE project that κ would be zero, and that the scale of total fertility would be 0.06. To find the scale values for age specific fertility, this value was grossed up by 25% (or S(j,t) was set to 1.25*0.06), to account for less than perfect correlation across ages. The age correlation turned out to be 0.95. The same scales were used for all ages and all forecast years for t=20 and beyond (smaller scale values were used for t<20, to account for smaller volatility in recent years; see Alho et al. (2008, p.44)). For mortality, empirical evidence from nine countries with long data series (1841-2000) showed that volatility had been larger among younger age groups than among the elderly, and hence scales were made age-specific with levels between 3 and 7 %; see Alho et al. (2008, p.46). As to the autocorrelation parameter κ , purely empirical estimates based on data from all 18 countries in the UPE project would have resulted in an average value close to zero. This was thought an unreasonable value, as it would ignore the possibility of a change in the trends of mortality the future. For Finland, empirical analyses had resulted in κ equal to 0.15 (Alho 1998). The UPE value of κ was chosen as 0.05, on judgemental grounds. The correlation between men and women was set to 0.85. This value has empirical ground in earlier findings for Finland (Alho 1998).

3 Analysis of the sensitivity of the Total Fertility Rate forecasts

In this Section we investigate the impact on the variability of the Total Fertility Rate caused by changes in κ and in ρ, the correlation across time and age respectively, of the forecast error increments. The UPE forecasts are derived assuming a zero correlation across time and setting ρ=0.95, values chosen on the basis of the investigation of the time series of the past error forecasts, for several european countries; see above.

We consider several values of κ and in ρ and for each of them we draw 3000 values of the age-specific fertility rates from their joint predictive distribution, in accordance to the Scaled Model of Error and work out for each projection year the corresponding Total Fertility Rates. The variability of the Total Fertility Rate is evaluated in terms of the standard deviation of draws and in terms of the difference of upper and lower bound of the derived 80% forecast interval divided by the median of draws. We refer to such measure in terms of relative width of the forecast interval.

Tables 1 and 2 show the values of the standard deviation (std) and of the relative width of the 80% forecast interval (width) for different choices of κ and ρ in 2015, 2030 and 2050. The reference values κ=0 and ρ=0.95 are given in the first column.

Table 1: Total Fertility Rate's standard deviation and relative 80 % interval width for different values of κ ($\varrho = 0.95$)

Year	κ=0		κ=0.02		κ=0.05		κ=0.1	
	Std	width	std	width	std	width	std	width
2015	0.1136	0.196	0.1226	0.2103	0.1281	0.223	0.1418	0.2498
2030	0.3689	0.5779	0.4205	0.6435	0.5195	0.8043	0.6600	1.0063
2050	0.6487	0.976	0.925	1.2791	1.4095	1.7109	2.1424	2.3952

Table 2: Total Fertility Rate's standard deviation and relative 80 % interval width for different values of ρ ($\kappa = 0$)

year	ρ=0.95		ρ=0.9		ρ=0.8		ρ=0.7		ρ=0.6		ρ=0.5	
	std	width	std	width	std	width	std	width	std	width	std	width
2015	0.1136	0.196	0.0996	0.1734	0.0791	0.1339	0.0679	0.1187	0.0582	0.0996	0.051	0.0893
2030	0.3689	0.5779	0.3152	0.5082	0.2415	0.3806	0.2004	0.3177	0.1777	0.2819	0.1496	0.237
2050	0.6487	0.9760	0.551	0.807	0.4238	0.6353	0.3512	0.5218	0.3071	0.4607	0.2617	0.3973

As expected the variability of the Total Fertility Rate increases as the value of κ or ρ increases. A change in autocorrelation κ from zero to 0.1 (which is a fairly small change since in principle, κ may take on any value between zero and one) has strong consequences for the uncertainty in the TFR in the long run: in 2050 both the standard deviation and the width of the prediction interval are larger by a factor 2½. The uncertainty in the TFR is much more sensitive for changes in κ in the long run than in the short run. This is the result of the accumulation of error increments over time. Changes in the age correlation ρ have a modest effect on the uncertainty in the TFR.

4 Analysis of the sensitivity of the Male and Female Life Expectancies forecasts

In this section we investigate the impact on the variability of Male and Female Life Expectancies caused by changes in κ and in ρ, the correlations across time and age, respectively, of the forecast error increments. Moreover we evaluate the impact that a change in ρ_g, the correlation across sexes of the error increments has on the variability of the difference between the Male and Female Life Expectancies. The UPE forecasts are derived setting κ=0.05 and ρ_g=0.85.

Again we draw 3000 values of the male and female age-specific mortality rates for each projection year and work out the corresponding Male and Female Life Expectancies. Tables 3 and 4 display the values of standard deviation and the relative 80 % interval width of the Male and Female

Life Expectancies (denoted respectively EM and EF) for several values of κ and ρ, respectively . Table 5 shows the values of the standard deviation and the interval width of the differencies between Female and Male Life Expectancies for several values of ρ_g. Reference values are given in columns for κ=0.05, ρ=0.95 and ρ_g=0.85.

Table 3: Male and Female Life Expectancies standard deviation and relative 80 % interval width for different values of κ ($\rho = 0.95$)

year	κ=0				κ=0.02				κ=0.05				κ=0.1			
	EM		EF		EM		EF		EM		EF		EM		EF	
	std	width	std	Width	std	width	std	width	std	width	std	width	std	width	std	width
2015	0.6857	0.0211	0.6097	0.0186	0.7264	0.0236	0.6465	0.0186	0.771	0.0241	0.6841	0.0209	0.8428	0.0260	0.7513	0.0232
2030	1.4151	0.0446	1.3201	0.0385	1.6887	0.0531	1.5642	0.0453	2.216	0.069	2.048	0.060	2.6263	0.0797	2.4273	0.0703
2050	1.8368	0.0552	1.8036	0.0498	2.7144	0.08	2.5992	0.0721	3.77	0.112	3.67	0.102	5.1364	0.1465	5.0499	0.1325

Table 4: Male and Female Life Expectancies standard deviation and relative 80 % interval width for different values of ρ (κ=0.05)

year	ρ=0.95				ρ=0.9				ρ=0.8				ρ=0.7			
	EM		EF		EM		EF		EM		EF		EM		EF	
	std	width	std	width	std	width	std	width	std	width	std	width	std	width	std	width
2015	0.771	0.0241	0.6841	0.0209	0.601	0.0186	0.542	0.0162	0.437	0.0136	0.403	0.0116	0.349	0.0112	0.324	0.0092
2030	2.216	0.069	2.048	0.060	1.611	0.0496	1.517	0.0442	1.214	0.0375	1.162	0.0341	0.970	0.0310	0.929	0.0261
2050	3.772	0.112	3.67	0.102	2.913	0.0863	2.898	0.0862	2.135	0.0639	2.151	0.0602	1.724	0.0499	1.782	0.0491

year	ρ=0.6				ρ=0.5			
	EM		EF		EM		EF	
	std	width	std	width	std	width	std	width
2015	0.295	0.0099	0.276	0.0081	0.259	0.0074	0.236	0.0069
2030	0.814	0.0255	0.787	0.0227	0.719	0.0230	0.697	0.0204
2050	1.444	0.0439	1.485	0.0413	1.267	0.0381	1.299	0.0358

Table 5: Life Expectancies Differences standard deviation and relative 80 % interval width for different values of ρ_g ($\kappa = 0.05, \rho = 0.95$)

year	ρ_g=0.95		ρ_g=0.9		ρ_g=0.88		ρ_g=0.85		ρ_g=0.82	
	std	width	std	width	std	width	std	width	std	width
2015	0.2962	0.1250	0.3662	0.1607	0.3919	0.1786	0.4308	0.1964	0.4638	0.2143
2030	0.8716	0.4000	1.0887	0.52	1.1559	0.5556	1.3656	0.6545	1.3859	0.6481
2050	1.6035	0.7358	1.9661	0.9245	2.1123	1.0385	2.3344	1.1154	2.4892	1.1923

year	ρ_g=0.8		ρ_g=0.78		ρ_g=0.75		ρ_g=0.7	
	std	width	std	width	std	width	std	width
2015	0.4785	0.2143	0.4985	0.2321	0.5341	0.2500	0.5807	0.2727
2030	1.4553	0.6852	1.5134	0.7037	1.6014	0.7593	1.7372	0.8148
2050	2.599	1.2075	2.7842	1.3077	2.927	1.3585	3.1142	1.4808

As expected the variability of the Male and Female Life Expectancies increase as the values of κ and ρ increase. As with fertility, the long run effect of even a small change in κ (as compared to its possible range) on the volatility of male and female life expectancies is large. The sex gap in life expectancies is somewhat less sensitive for changes in ρ_g.

5 Concluding remarks

For the Scaled Model of Error, we have analysed the impact of small changes in a number of correlation parameters for the uncertainty in total fertility and life expectancy. For age specific fertility we find that the correlation across ages has only limited impact on the uncertainty in the Total Fertility Rate. As a consequence, annual numbers of births will be little affected. The autocorrelation in error increments is an important parameter, in particular in the long run. Also, the autocorrelation in error increments for age specific mortality is important. It has a large effect on long run uncertainty in life expectancy values, and hence on the uncertainty around the elderly population in the future. In empirical applications of the Scaled Model of Error, one should give due attention to a correct estimation of these two parameters.

6 References

Alho, J. (1998). *A stochastic forecast of the population of Finland.* Reviews 1998/4. Helsinki: Statistics Finland.

Alders M., Keilman, N. and Cruijsen, H. (2007) "Assumptions for long-term stochastic population forecasts in 18 European countries", *European Journal of Population*, **23**, 33--69.

Alho J.M. and Spencer, B. D. (1997) "The practical specification of the expected error of population forecasts", *Journal of Official Statistics*, **13**, 204--225.

Alho, J.M. and Spencer, B. D. (2005) *Statistical demography and forecasting* New York: Springer.

Alho, J. M., Alders, H. C., Keilman, N., Nikander, T., & Pham, D. Q. (2006). New forecast: Population decline postponed in Europe. *Statistical Journal of the United Nations* ECE, 23, 1–10.

Alho, J. M., Cruijsen, H., & Keilman, N. (2008). Empirically based specification of forecast uncertainty. In J. M. Alho, S. E. Hougaard Jensen, & J. Lassila (Eds.). *Uncertain demographics and fiscal sustainability* (pp. 34–54). New York: Cambridge University Press.

Alho, J. M., & Mustonen, M. (2003). Tutorial for PEP—program for stochastic forecasts of population. http://www.joensuu.fi/statistics/pep/peptutorial.pdf.

Alho, J. M., & Nikander, T. (2004). Uncertain population of Europe—summary results from a stochastic forecast. http://www.stat.fi/tup/euupe/rp_reports_and_pub.html.

Borgy, V., & Alho, J. M. (2007). Macro-economic consequences of demographic uncertainty in world regions. In J. M. Alho, S. E. Hougaard Jensen, & J. Lassila (Eds.) *Uncertain demographics and fiscal sustainability* (pp. 195–218). New York: Cambridge University Press.

Keilman, N. (1987). Ex-post errors in official population forecasts in industrialized countries. *Journal of Official Statistics* 13 (3): 245-277.

Keilman, N. & D.Q. Pham (2004). Empirical errors and predicted errors in fertility, mortality and migration forecasts in the European Economic Area. Discussion Paper 386 August 2004, Statistics Norway.

Li, Q., M. Reuser, C. Kraus & J. Alho (2009). Ageing of a giant: A stochastic population forecast for China, 2006-2060. *Journal of Population Research* 26 (1): 21-50.

Lutz, W. & J. R. Goldstein (2004). Introduction: How to deal with uncertainty in population forecasting? *International Statistical Review* 72(1): 1-4.