98/19 November 1998 Documents

Statistics Norway
Statistisk sentralbyrå

Statistics Norway
Department for Social Statistics

Nico Keilman and Pham Dinh Quang

Documents

Predictive Intervals for Age-Specific Fertility

Nico Keilman and Pham Dinh Quang

Predictive Intervals for Age-Specific Fertility

Abstract:
A Vector Autoregression (VAR-) model is combined with a Gamma curve to predict confidence intervals for age-specific birth rates by one-year age groups. The method is applied to observed age-specific births in Norway between 1900 and 1995, and predictive intervals are computed for each year up to 2050. The predicted confidence intervals for Total Fertility (TF) agree well with TF-errors in old population forecasts made by Statistics Norway. The method gives useful predictions for age-specific fertility up to around 2030. For later years, the intervals become too wide. Methods which do not take account of estimation errors in the VAR-model coefficients underestimate the uncertainty for future TF-values. The findings suggest that the margin between high and low fertility variants in official population forecasts for many Western countries are too narrow.

Keywords: Population forecasts, Time series, Fertility, Confidence intervals, Uncertainty

Acknowledgement: Programming assistance by Arve Hetland is gratefully acknowledged. Evert van Imhoff kindly provided us with an improved version of his program Profile. We had useful discussions with Bill Bell, and, in particular, with Juha Alho. Joop de Beer, Lars Østby, and Øystein Kravdal commented upon an early version of this paper. The research was supported by grant nr. 114055/730 from the Norwegian Research Council.

Address: Nico Keilman, Statistics Norway, Department for Social Statistics, P.O.Box 8131 Dep., N-0033 Oslo, Norway. E-mail: nico.keilman@ssb.no.

1. The need for stochastic forecasts

It is easier to predict the main results of the next population forecast for a certain country, than to predict the population itself. This is one of the main conclusions from analyses into the accuracy of national population forecasts published by statistical offices of Western countries after the Second World War. In other words: real developments for fertility, mortality and migration are much more volatile than usually assumed in forecasts. When studying the population forecasts of Western countries, one notes how strikingly similar the predicted values in a new forecast are compared to the previous forecast. The actual variables, however, often show a very different development. While forecasts are surprise-free, reality is not. The rapid fall in fertility in many Western countries in the 1970s came as a surprise for most population forecasters and other demographers. The result was too high birth rates, and an overestimation of the number of young children. Other forecast variables with large errors are the predicted numbers of elderly persons (in particular the oldest old), which were far too low because of too pessimistic mortality assumptions, and the size of immigration, which is determined by largely unforeseeable political, economic and legal factors.

The reason for those forecast errors is our limited understanding of demographic behaviour. Valid behavioural theories that explain birth, death, or migration have not yet been found. Existing theories have limited validity in time or space, or they are strongly conditional or partial, or both (Keyfitz 1982). When it is difficult to explain demographic processes, then it is even more problematic to predict them. The current practice among forecasters is to study regularities and irregularities in the historical developments of major demographic variables such as the TF and the life expectancy, to understand observed trends, and to extrapolate them into the future.

Extrapolation, as next best to prediction on the basis of causal explanation, implies that population forecasts are inherently uncertain. Any serious forecaster will attempt to include that uncertainty in the forecast in such a way, that it will become clear to the user. The standard approach in national population forecasts is to formulate two or more sets of assumptions for those key variables of which the future development is difficult to predict. Examples are the Total Fertility, and the Life Expectancy at Birth. This approach dates back to at least 1933, when Pascal Whelpton computed a population forecast for the United States, in which he presented several fertility variants. But the use of forecast variants is presumably much older, compare the intervals which Spengler (1935) reports for the results of a number of forecasts for the US. Nowadays, statistical agencies in 15 of the 18 member countries of the European Economic Area (EEA, that is the EU- and EFTA-countries, except Switzerland) produce forecasts with between two and four fertility variants (Eurostat 1997a, Table 35). In the forecasts made in the beginning of the 1990s in those countries, the low and high fertility variants defined intervals of between 0.3 and 0.6 children per woman wide for a forecast duration of approximately 10 years. Moreover, the first co-ordinated population forecast for all 18 EEA-countries has a difference between the high and the low fertility variant of 0.5 children per woman in the year 2035 (Eurostat 1997b, Table 3).

In spite of the general use of forecast variants to express uncertainty, this approach is unsatisfactory from a statistical point of view. For instance, Statistics Norway (1997) assumes in the current Norwegian official population forecast that the period Total Fertility in the year 2010 will be between 1.7 (low variant) and 2.1 (high variant) children per woman. However, they do not give the probability that the real TF in 2010 will have a value of 1.7-2.1, or perhaps fall outside that range. Now take the example of an educational planner with an interest in nursery and primary school. For this planner it must be of great importance to know whether the estimated probability of a TF between 1.7 and 2.1 in the year 2010 is 30, or 60, or perhaps even 90 per cent. In the former case, he should incorporate much more flexibility into the school planning process, than in the latter.

Because the traditional approach is so unsatisfactory, some statistical agencies have in recent years attempted to compute stochastic population forecasts with predictive intervals. See for instance Hanika et al. 1997 and Lutz and Scherbov 1998 for Austria, and Alders and De Beer 1998 for the Netherlands. These methods have been inspired by earlier work on stochastic forecasts by for instance Lee and Tuljapurkar 1994 and Alho 1990. The current paper is a contribution to the literature in this area. Its purpose is to generate predictive intervals for age-specific fertility rates. It is part of a larger project of which the aim is to compute predictive intervals for the future population of Norway, broken down by age and sex.

2. The method - generalities

We assume that births are generated by a Poisson process. The intensity of that process depends strongly on the age of the mother. Thus the parameter of the Poisson process, that is, the fertility rate, varies by age. For a certain year, or a given birth cohort of women, we assume that the age pattern of fertility follows a Gamma curve. This is a mathematical function which consists of a Gamma density and a scaling parameter. The Gamma curve has four parameters: the Total Fertility (TF), the Mean Age at Childbearing (MAC), the Variance in that age (VARG), and the minimum age. The four parameters are estimated on the basis of annual Norwegian data from 1900 to 1995. This results in a time series of parameter estimates. The series for three of the four parameters (TF, MAC, VARG) are modelled by means of a multivariate time series model. The minimum age is kept constant at its value as estimated for recent years. Predictions are made for the period 1995-2050 for all four parameters. The Gamma curve is used to transform the parameter predictions back into future age-specific fertility rates.

Since the ultimate purpose is to generate stochastic population forecasts, much attention is given to an appropriate quantification of uncertainty. In the present approach, there are four main sources of uncertainty attached to future birth rates:
1. Sample variation in the historical age-specific birth rates
2. Errors in the parameter estimates of the Gamma curve
3. Residual variance in the time series model

3. The age-specific fertility rate: The Poisson model

Assume that a group of $Y(x)$ women aged x give birth to $B(x)$ children in a certain year. Assume further that the births are generated by a Poisson process. When f_x represents the intensity of the process, the probability of exactly $B(x)$ births among the $Y(x)$ women equals

$$\frac{\exp\{-f_x.Y(x)\}.\{f_x.Y(x)\}^{B(x)}}{B(x)!}$$

This is also the likelihood of observing the data, given the model. The first and second derivatives of the logarithm of the likelihood result in the following estimators for the intensity f_x and the corresponding variance:

$$F_x = B(x)/Y(x), \text{ and}$$

(1) $$Var(F_x) = B(x)/Y^2(x) = F_x/Y(x)$$

The estimator F_x is the traditional age-specific rate. Its variance $Var(F_x)$ is small when the rate itself is small, or when the women at risk $Y(x)$ are numerous, or both.

We have used the Norwegian age-specific birth rates for the years 1900-1993 computed by Brunborg and Mamelund (1994). These were supplemented with rates for the years 1994 and 1995. The age range was from 16 to 44. When computing person years of exposure, we ignored mortality and international migration, and used the population of women at the beginning of each year, broken down in one-year age groups.

4. The age pattern of fertility: The Gamma curve

The array of 29 rates for each year can be summarized by means of a parametric curve, which is a function of age. Thanks to the regular shape of the age pattern of fertility, such a function contains only a few parameters, usually 3-5. Various curves fit the data well: normal, lognormal, double exponential, Coale-Trussell, Hadwiger, polynomial, gamma, and logistic curves, to name the most important ones. Several authors have noted the attractive properties of the Gamma curve. The fit is usually good, and the parameters can be interpreted, after an appropriate transformation, in a straightforward way. See Bell 1997, Hoem et al. 1981, and Duchêne and Gillet-De Stefano 1974. Denote the fertility intensity for age x as f_x, as before. The Gamma curve is usually defined as

(2) $$f_x = \frac{1}{\Gamma(\alpha_3)}\alpha_1\alpha_2^{\alpha_3}(x-\alpha_4)^{\alpha_3-1}\exp[-\alpha_2(x-\alpha_4)] + \varepsilon_x, x \geq \alpha_4$$

The four parameters α_i are to be estimated from the data, ε_x is a residual term, and $\Gamma(.)$ is the Gamma function defined by

$$\Gamma(p) = \int_0^\infty u^{p-1}\exp(-u)du.$$

α_1 represents the TF, whereas α_4 is the minimum age at childbearing. The parameters α_2 and α_3 have no immediate demographic interpretation. However, f_x/α_1 is the Gamma density, with mean $\alpha_4+\alpha_3/\alpha_2$ and variance α_3/α_2^2. Therefore we introduce the following transformation

$$\beta_1 = \alpha_1$$

$$\beta_2 = \alpha_4 + \alpha_3/\alpha_2$$

$$\beta_3 = \alpha_3/\alpha_2^2$$

(3) $$\beta_4 = \alpha_4$$

Hence parameters β_1 and β_4 have the same interpretation as α_1 (TF) and α_4 (minimum age). β_2 represents the mean age at childbearing, while β_3 is the variance in that age. The β-parameters have been estimated by means of non-linear regression, by minimizing the following weighted sum of squares

(4) $$\sum_x w_x(F_x - f_x)^2$$

F_x is the estimated rate for age x as introduced in the previous section, f_x is the underlying theoretical intensity given by expressions (2) and (3), while w_x is the inverse value of the variance of F_x, see expression (1). The latter variance reflects the "measurement error" for the rate F_x: a small variance indicates a precise estimate for the intensity, and vice versa. Hence ages for which the variance is large get less weight in the regression than those with smaller variances. Weighted least squares estimation is approximately equivalent with Maximum Likelihood estimation of the parameters β_i, with births $B(x)$ and exposure time $Y(x)$ as data (Van Imhoff 1991). *Unweighted* least squares (w_x=1) would imply that one regards the birth rates F_x as data. This would give relatively much weight to imprecise rate estimates. In a similar curve fitting exercise for US fertility, Bell (1992, p. 192) used weights equal to four for ages 18-32, and one for all other ages. His aim was to give more weight to ages with high fertility. Note the difference with our approach: we give more weight to the ages where the rate variance is low, usually ages up to 22 and beyond 33 for a country like Norway.

The four β-parameters and the corresponding covariance matrix, together with their variances were estimated on the basis of Norwegian birth rates for each of the years 1900-1995, see Section 3. We used a recent update of Van Imhoff's program Profile (Van Imhoff 1991) for small scale experiments, and the NLIN-procedure in SAS for the complete data set. In the SAS computations we selected the Marquardt algorithm for minimization of expression (4), with appropriate non-negativity constraints for the parameters. The symbolic program MAPLE computed analytical first-order derivatives. For each year, the weights w_x are U-shaped. For instance, in 1995 the minimum is 237,500 at age 28 (F_{28}=0.141; $Y(28)$=33,569). The left and right branches of the w_x-curve increase rapidly to values exceeding 1,000,000 for ages below 20 or above 38. At ages 16 and 44, the weights are 25.8E6 and 14.3E6, respectively. The consequence of these extremely high values is that the fit is very bad around the top of the curve, since the curve is determined strongly by the estimated rates at young and old ages. In order to avoid this effect, we have censored extreme weights. After some experimentation we concluded that a maximum weight value of 3,000,000 gives a good fit around the top of the curve. Therefore, all weights (for all ages and calendar years) exceeding this value were made equal to 3,000,000. For 1995 this was the case for ages 16, 17, and 42-44.

Figures 1-3 give estimates for β_1, β_2, and β_3 with corresponding 95-per cent confidence intervals. The TF, mean age at childbearing, and variance in that age computed in the traditional demographic manner (i.e. moment estimators TF=$\Sigma_x F_x$, $m=\Sigma_x x.F_x$/TF, and $s^2=\Sigma_x (F_x-m)^2$/TF^2) are also given. The latter estimators are only influenced by Poisson variability, not by the fit of the Gamma curve. The minimum age β_4 is not included in these figures, because estimates in recent years were invariably equal to the boundary value of zero (after initial values around 14-15 years of age in the first half of the century).[1]

[1] Estimates for the minimum age β_4 fell below 14 in 1975, and decreased further to reach zero in 1991. During the same period, the estimates for the mean age at childbearing β_2 rose from 26.6 to 28.3 years. Together with the relatively low estimates for the Total Fertility β_1 during these years (< 2), the predicted birth rates at ages below 16 are still negligible, in spite of the unrealistic estimate for the minimum age.

Figure 1. Total Fertility (TF) estimates and 95 per cent confidence interval

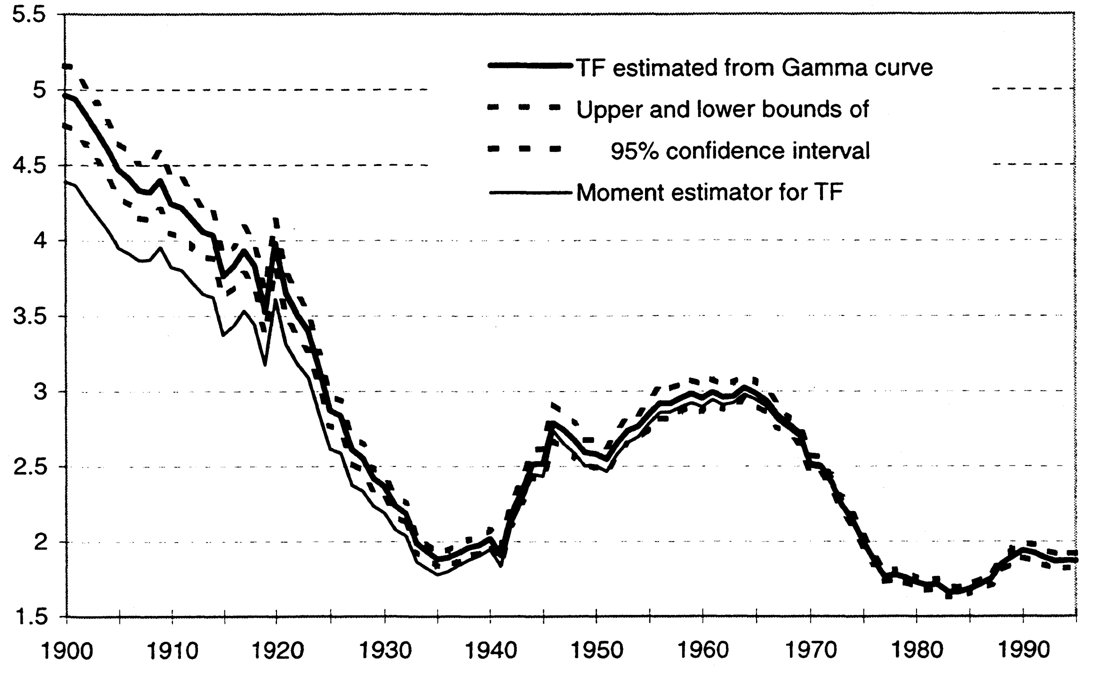


Figure 2. Estimates for mean age at childbearing, and 95 per cent confidence interval

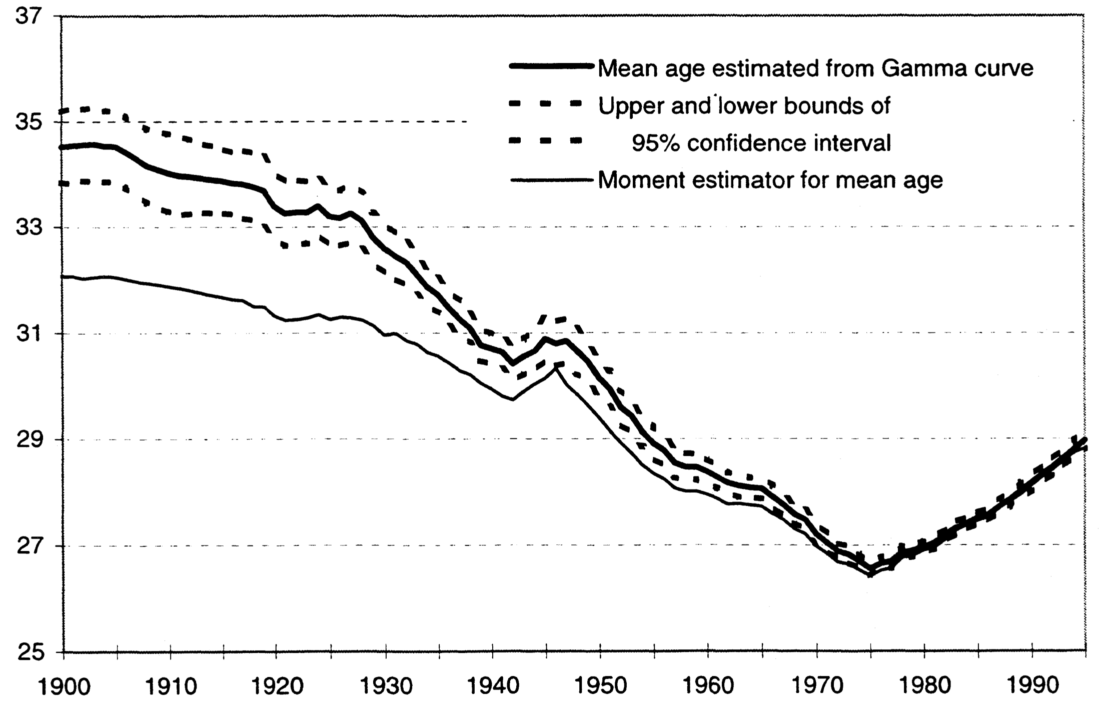
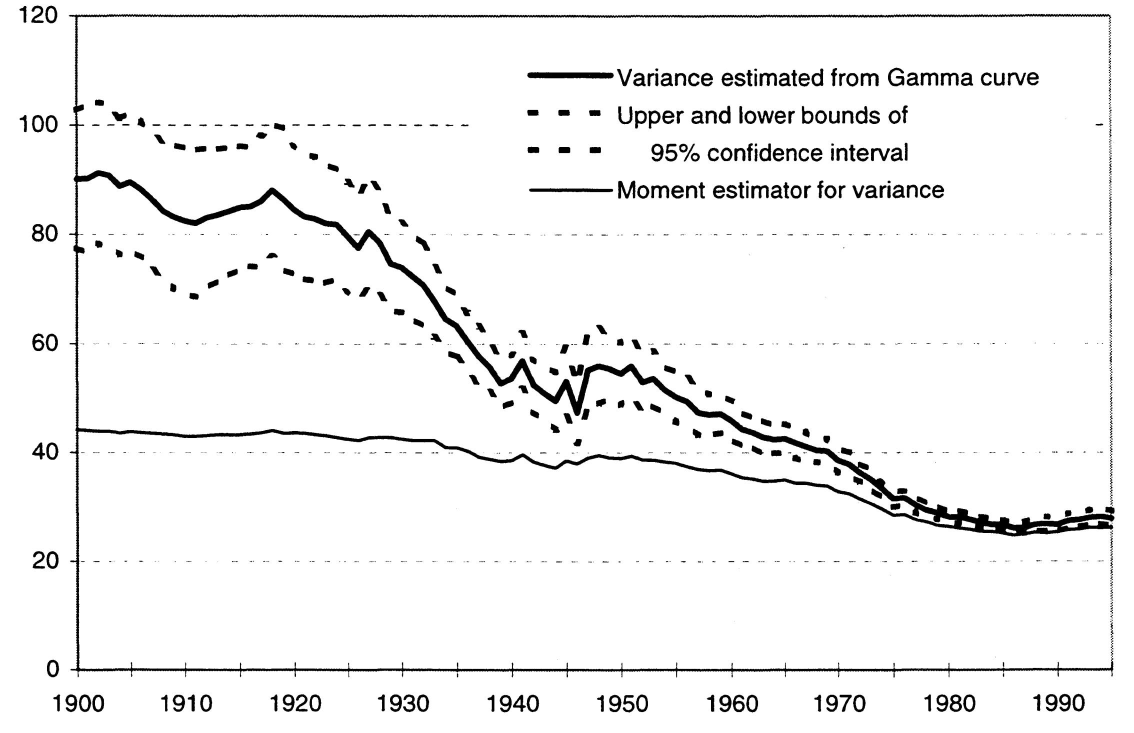


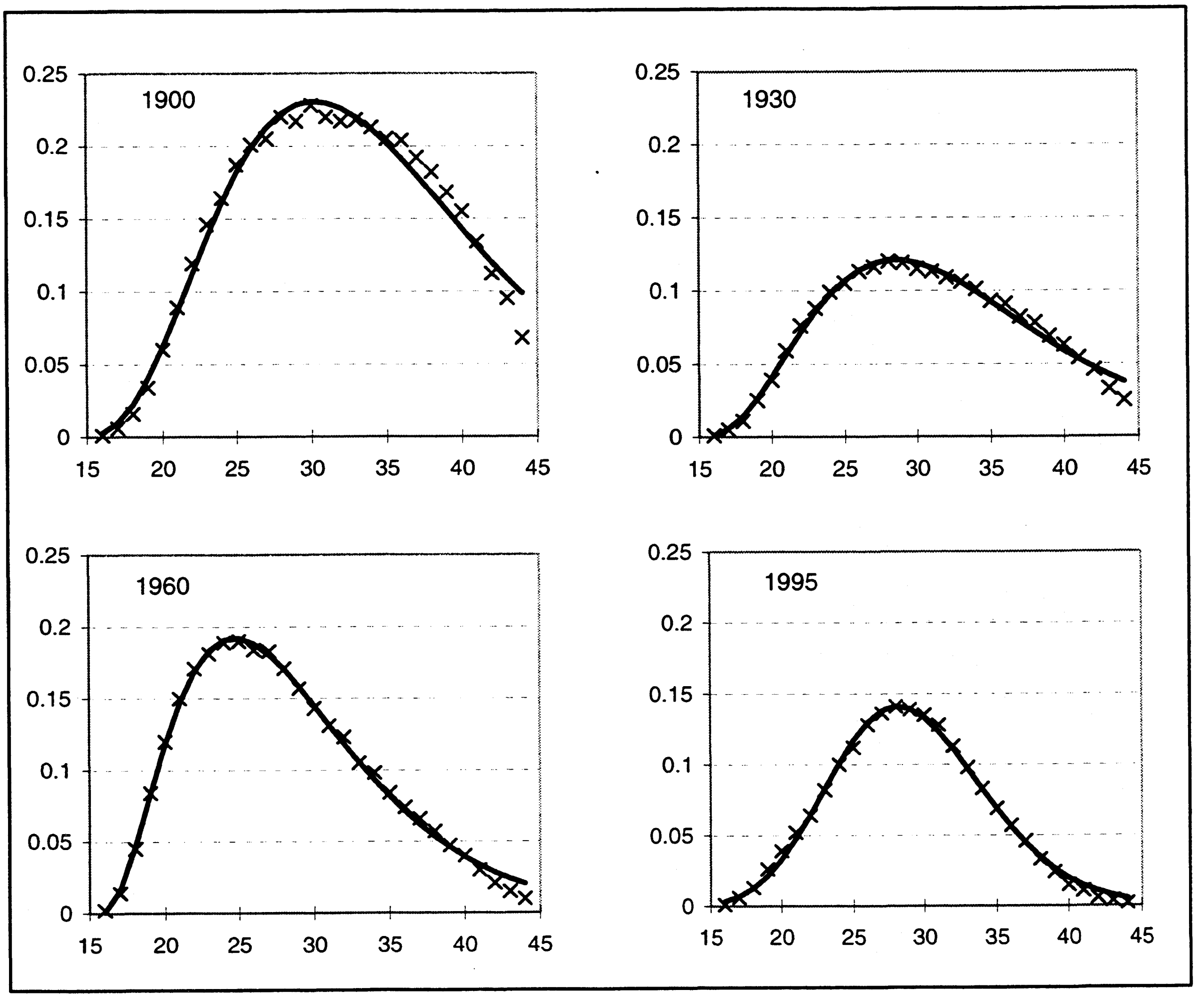
Figure 3. Estimates for variance in childbearing age, and 95 per cent confidence interval



Similar to many other Western countries in this century, Norway had two periods with a strong fertility decrease (Figure 1). The first one, which had begun around 1880, ended in the 1930s, whereas the second one took place at the end of the 1960s and during the 1970s. The baby boom of the 1950s and 1960s was to a large extent the consequence of a decrease in the mean age at childbearing for women born in the years 1920-1945. This led not only to a fall in the period mean age (see Figure 2), but also to high period-TF values (Figure 1). The period-TF attained its minimum in the years 1983 and 1984, when it was as low as 1.66 children per woman. After a rise towards 1.9 children per woman in 1990, the TF has been rather constant. But in recent decades, women get their children at increasingly higher ages, compare the strong rise in the period mean age at childbearing in Figure 2. Facilitated by modern contraceptive methods, growing shares of young Norwegian adults postponed the birth of their first child and took some form of education at the tertiary level during the 1970s and early 1980s. Next they worked some years before they entered parenthood (Kravdal 1994). Much of the fertility decrease during this century was caused by a reduction of higher-parity births, which generally take place at high ages. Together with a reduction in childbearing at young ages, the births were more and more concentrated around the mean age at childbearing - hence the fall in the variance in Figure 3.

The 95%-confidence intervals for the β_i are rather wide in the years 1900-25 and 1945-65, indicating a relatively bad fit. After 1980, the fit is excellent. The traditional TF coincides with the estimates from the Gamma curve from 1970 onwards, and it falls within the 95 per cent confidence bounds from 1940. However, it is much lower in the first half of the century, because of the bad fit of the Gamma curve in those years. Figure 4 illustrates how the fit improves over the years.

Figure 4. Age-specific birth rates, empirical values (xxx) and Gamma curve fit (solid line)



The smooth pattern of the empirical rates causes autocorrelation in the residuals. This systematic bias can be seen very clearly for women older than 25 in the year 1900, but also in the other three years. It means that the assumption of non-correlated residuals is not fulfilled. Thompson et al. (1989) and Bell (1992, 1997) have suggested to account for the systematic deviations by including what they call a "bias adjustment" in predictions of the age-specific fertility rates. In Section 5.2 we will come back to this issue. Meanwhile we note that an assumption on zero correlation between residuals could have been avoided by using non-parametric estimation, although resulting residuals still would have been systematic.

5. A modified VAR-model for the parameters of the Gamma curve

The result of the curve fitting exercise in the previous section is a series of estimates for the four parameters of the Gamma curve for each year between 1900 and 1995, and the corresponding estimated covariance matrix for each year. A modified Vector Autoregression (VAR-) model has been used to predict three of the four parameters: β_1, β_2, and β_3. As noted earlier, the minimum age of childbearing β_4 fell from 14 in 1975 to zero in 1991, and remained at that level since. We predict that β_4 will be zero in the future, too.

5.1. The model

The expressions used here are standard in time series literature, for example Cryer 1986 and Lütkepohl 1993. Let $C_t = (\ln(\beta_{1,t}), \ln(\beta_{2,t}), \ln(\beta_{3,t}))'$ be a column vector with the Gamma curve parameters in year t in logarithmic form. First differences of C led to stationarity, and we found that a multivariate ARIMA (1,1,0) model fitted the data well. The model is of the form

$$Z_t = \phi Z_{t-1} + \varepsilon_t, \tag{5}$$

where $Z_t = C_t - C_{t-1}$, ϕ is a fixed 3x3-matrix of coefficients, and $\varepsilon_t = (\varepsilon_{1,t}, \varepsilon_{2,t}, \varepsilon_{3,t})'$ is a multivariate normal column vector with mean 0 and constant covariance matrix Σ_ε. The model contains no intercept, so that we avoid predicting an indefinitely increasing or decreasing pattern in β_t. An equivalent expression for (5) is the non-stationary process

$$C_t = (I + \phi)C_{t-1} - \phi C_{t-2} + \varepsilon_t. \tag{6}$$

Least Squares estimates of ϕ and Σ_ε in (5) are the same as those in (6). So far, the model can be considered as a VAR-model. In our case, the model was modified: the estimates of ϕ and Σ_ε were obtained from a version of expression (6) estimated by Weighted Least Squares (WLS), giving less weight to years in which the estimated variances for the β_i parameters were high. Therefore, both C_t, C_{t-1}, and C_{t-2} in expression (6) were pre-multiplied with the inverse of the standard deviation of C_t. In principle, we could also have used WLS of expression (5), with the inverse of the standard deviation of Z_t as weights. However, the latter weights are a complicated expression due to the covariance between C_t and C_{t-1}.

Let $\Omega_{\beta,t}$ be the covariance matrix of the vector β_t. Then, by the delta method, we find an approximate value for the covariance matrix of C_t as

$$\Omega_{C,t} \approx \begin{pmatrix} \beta_{1,t}^{-1} & 0 & 0 \\ 0 & \beta_{2,t}^{-1} & 0 \\ 0 & 0 & \beta_{3,t}^{-1} \end{pmatrix} \Omega_{\beta,t} \begin{pmatrix} \beta_{1,t}^{-1} & 0 & 0 \\ 0 & \beta_{2,t}^{-1} & 0 \\ 0 & 0 & \beta_{3,t}^{-1} \end{pmatrix}.$$

Pre-multiplying C_t, C_{t-1}, and C_{t-2} in expression (6) by $(\Omega_{C,t})^{-1/2}$ results in

$$Y_t = (I + \phi)X_t - \phi U_t + \eta_t, \tag{7}$$

with $Y_t = (\Omega_{C,t})^{-1/2} C_t$, $X_t = (\Omega_{C,t})^{-1/2} C_{t-1}$, and $U_t = (\Omega_{C,t})^{-1/2} C_{t-2}$. Ordinary Least Squares (OLS) estimation of (7) is equivalent with WLS estimation of (6), with weights $(\Omega_{C,t})^{-1}$ (Draper and Smith 1981, p. 108). OLS estimation of (7) results in estimates $\hat{\phi}$ for the coefficients, and $\hat{\Sigma}_\eta$ for the covariance matrix of η_t.

From the latter matrix we find an estimate for the covariance matrix of ε_t as $\hat{\Sigma}_\varepsilon = (\Omega_C)^{1/2} \hat{\Sigma}_\eta (\Omega_C)^{1/2}$, where Ω_C is a symmetric weight matrix to be discussed in Section 7.1.

The minimum Mean Squared Error predictor for C at forecast horizon l starting from forecast origin t is

$$\hat{C}_t(l) = E(C_{t+l} | C_t, C_{t-1}, C_{t-2}, \ldots).$$

The forecast error is

$$e_t(l) = C_{t+l} - \hat{C}_t(l) = \sum_{j=0}^{l-1} \psi_j \varepsilon_{t+l-j} ,$$

where the matrices ψ_j are coefficient matrices to be computed from

$$(I + \psi_1 B + \psi_2 B^2 + \psi_3 B^3 + \ .. \ \psi_l B^l)(I - (I + \phi)B + \phi B^2) = I,$$

by equating powers of B, and $\psi_0 = I$. Given our assumption for the distribution of ε_t, we find

(8) $$e_t(l) \sim N(0, \sum_{j=0}^{l-1} \hat{\psi}_j \hat{\Sigma}_\varepsilon \hat{\psi}_j^{'})$$

For each individual component i=1,2,3 we have

(9) $$\frac{C_{i,t+l} - \hat{C}_{i,t}(l)}{\hat{\sigma}_{ii}(l)} \sim N(0,1),$$

where $\hat{\sigma}_{ii}(l)$ is the i-th diagonal element of the covariance matrix of $e_t(l)$ given in expression (8).

5.2. Results

We have limited the time series analysis to the years 1945-1995. On the one hand, a long series is desirable on statistical grounds. On the other hand, there is little reason to believe that the childbearing behaviour of women in the first half of the century was so similar to that in more recent decades, that both can be captured by one model. Moreover, the fit of the Gamma curve was much better in the second half of this century than in the first half. However, we have also investigated the sensitivity of our predictions for choosing the shorter periods 1960-1995 and 1975-1995 (see Section 7.4). Post-war effects in 1946 and 1947 have been removed: new estimates for β and the corresponding covariances in those years were computed by linear interpolation between 1945 and 1948.

Estimates for the elements of ϕ and corresponding standard errors are listed in Table 1.

Table 1. Estimates of ϕ_{ij}

	$\hat{\phi}_{11}$	$\hat{\phi}_{12}$	$\hat{\phi}_{13}$	$\hat{\phi}_{21}$	$\hat{\phi}_{22}$	$\hat{\phi}_{23}$	$\hat{\phi}_{31}$	$\hat{\phi}_{32}$	$\hat{\phi}_{33}$
Estimate	0.6694	0	0	0	0.8852	0	0.0909	0	0.3089
Standard Error	0.1044	-	-	-	0.0735	-	0.0419	-	0.1337

Diagonal elements are high and strongly significant. All but one ($\hat{\phi}_{31}$) of the off-diagonal elements turned out to be non-significant at the five per cent level in a trial calculation. Hence these were set equal to zero, and the model was re-estimated with those restrictions, using Restricted Least Squares. Predictions for β_2 based on the *unrestricted* model showed unrealistically wide 95 per cent confidence intervals, including mean ages well over 50 years already in the years 2020 and later. This is caused by an unusually large value of the elements (3,2) and (2,3) of the symmetric weight matrix Ω_C. By restricting non-significant elements of ϕ to zero, the confidence intervals for β_2, as well as those for β_1 and β_3, became realistic.

The estimated covariances for the non-zero $\hat{\phi}_{ij}$-elements are given in Table 2, whereas Table 3 contains estimates for the residual covariances Σ_ε. Much of the uncertainty, relatively speaking, concerns the TF, as witnessed by the high estimate of $\sigma_{\varepsilon,11}$. This is caused by the large fluctuations in the TF since 1945 (see Figure 1), whereas those in the mean age or in the variance were much smaller (Figures 2 and 3).

Predictions for the TF, the mean age, and the variance have been computed on the basis of expression (6), with $C_{t-1} = C_{1995}$ and $C_{t-2} = C_{1994}$ as starting values. Corresponding confidence intervals for these three indicators can be found using expressions (8) and (9). However, these intervals only reflect uncertainty around the predicted values provided that the matrix ϕ is known. In practice, however, this matrix is estimated, and each $\hat{\phi}_{ij}$ has its own distribution (except for those elements which were fixed to zero). Expressions for confidence intervals around future values of C_t which take the distribution of the ϕ-estimates into account are not known. Therefore we used simulation for the determination of predictive intervals of the elements of C_t, based on random draws of both the distribution for the ϕ-estimates *and* that for the residual vector ε. Predictive intervals based on the assumption that ϕ is known, using expressions (8) and (9), are reported in Section 7.1.

Table 2. Covariance estimates for non-zero elements of $\hat{\phi}$ [1]

	$\hat{\phi}_{11}$	$\hat{\phi}_{22}$	$\hat{\phi}_{31}$	$\hat{\phi}_{33}$
		$\times 10^{-3}$		
$\hat{\phi}_{11}$	10.185	0.011	-0.354	-0.000
$\hat{\phi}_{22}$	0.011	4.547	0.023	-0.130
$\hat{\phi}_{31}$	-0.354	0.023	1.644	-2.484
$\hat{\phi}_{33}$	-0.000	-0.130	-2.484	16.588

[1] Covariances between other elements of $\hat{\phi}$ are zero.

Table 3. Estimates of $\Sigma_\varepsilon=(\sigma_{\varepsilon,ij})$ [1]

$\hat{\sigma}_{\varepsilon 11}$	$\hat{\sigma}_{\varepsilon 12}$	$\hat{\sigma}_{\varepsilon 13}$	$\hat{\sigma}_{\varepsilon 22}$	$\hat{\sigma}_{\varepsilon 23}$	$\hat{\sigma}_{\varepsilon 33}$
		$\times 10^{-3}$			
0.703	0.005	0.105	0.007	0.015	0.309

[1] Only the upper triangular part of the symmetric matrix is given.

We simulated 10,000 sample paths for the vector β, each one from 1996 until 2050. For every sample path we drew one value for the matrix ϕ, and 55 values for the vector ε, one for each year. The estimates of both ϕ and ε follow a multivariate normal distribution. The mean and the covariance of $\hat{\phi}$ are given in Tables 1 and 2. The mean of $\hat{\varepsilon}$ is the null-vector, while its covariance estimates are contained in Table 3. Multivariate normally distributed numbers were drawn from these two distributions using Cholesky decomposition of the covariance matrices (Bratley et al. 1983).

The predicted values $\hat{C}_{i,1995}(l)$ follow a normal distribution. Transforming them back to predictions in terms of the β_l implies that the latter variables are unbounded from above. Therefore, draws that resulted in too large values for the three parameters were rejected. By the middle of the next century, the childbearing behaviour of Norwegian women may be very different from today's. Medical technology may have made it possible to postpone childbearing to ages well beyond 50. But even then, a TF of, say, 11 children per woman, or a mean age at childbearing of 55 years, or a variance in the age at childbearing of 400, are clearly unrealistic. At the same time it is unrealistic to assume that teenage fertility has become so important that the mean age falls below 20. Thus we used the following, quite liberal, restrictions: 0<TF<10, 20<MAC<50, and 0<VAR<250. In addition, a restriction was imposed on the elements of the matrix ϕ: each of those was required to lie between minus one and one, in order to ensure stationarity. Almost 9 per cent of the simulations (967 of 10,967 simulations) had to be rejected on the basis of the restrictions for the elements of ϕ, or for the three parameters. New sample paths were generated until we had obtained 10,000 paths with admissible values. This resulted in 10,000 values for each of the three β_l-parameters for every year between 1996 and 2050. These were ordered by size, and the lowest and highest 250, 1000, and 1666 values were taken as the lower and upper bounds of the 95, 80, and 67 per cent confidence intervals, respectively. Figures 5-7 illustrate these confidence bounds.

Figure 5. Total Fertility

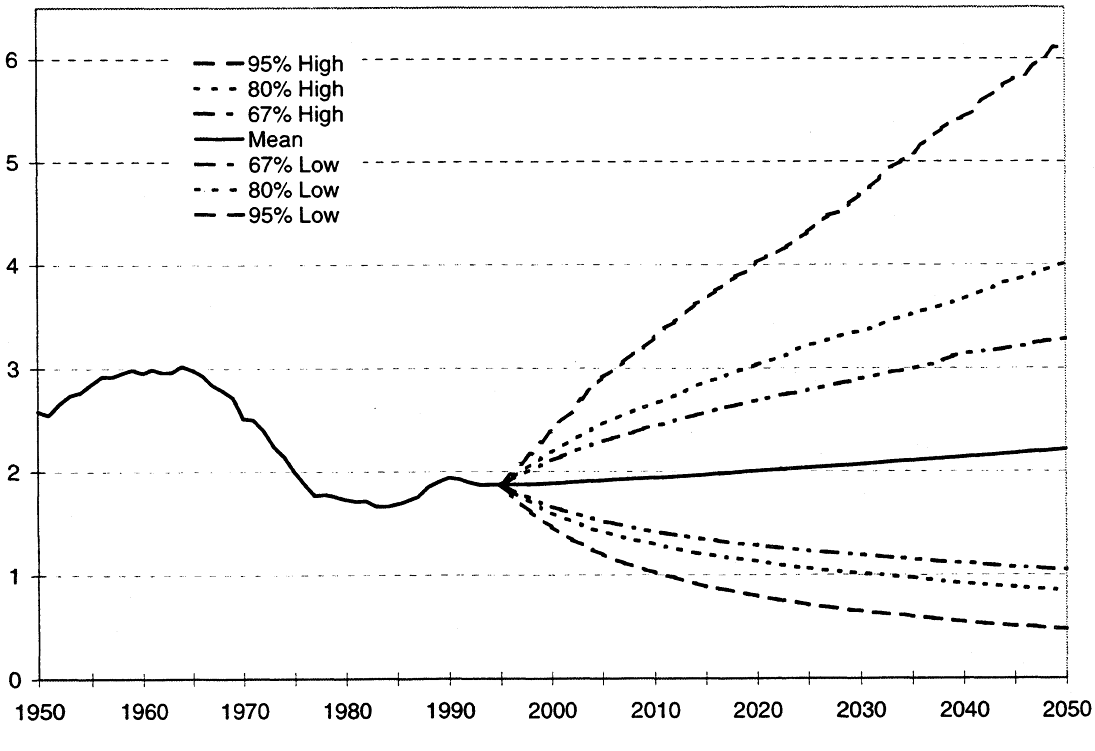


Figure 6. Mean age at childbearing

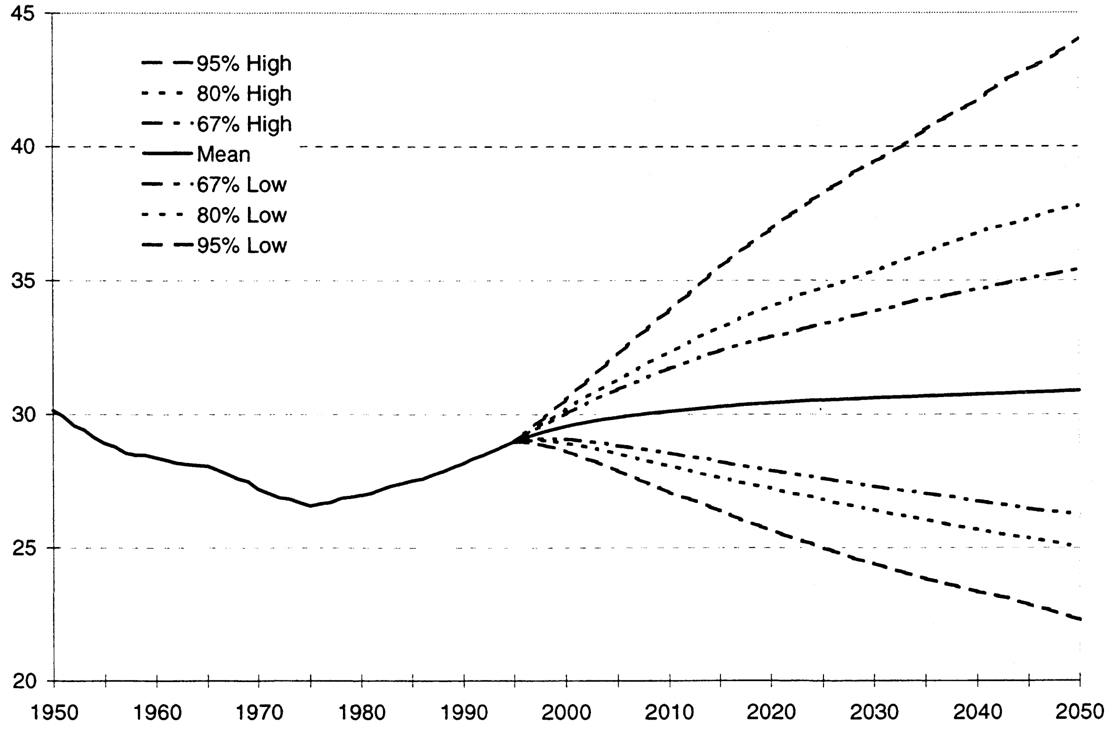
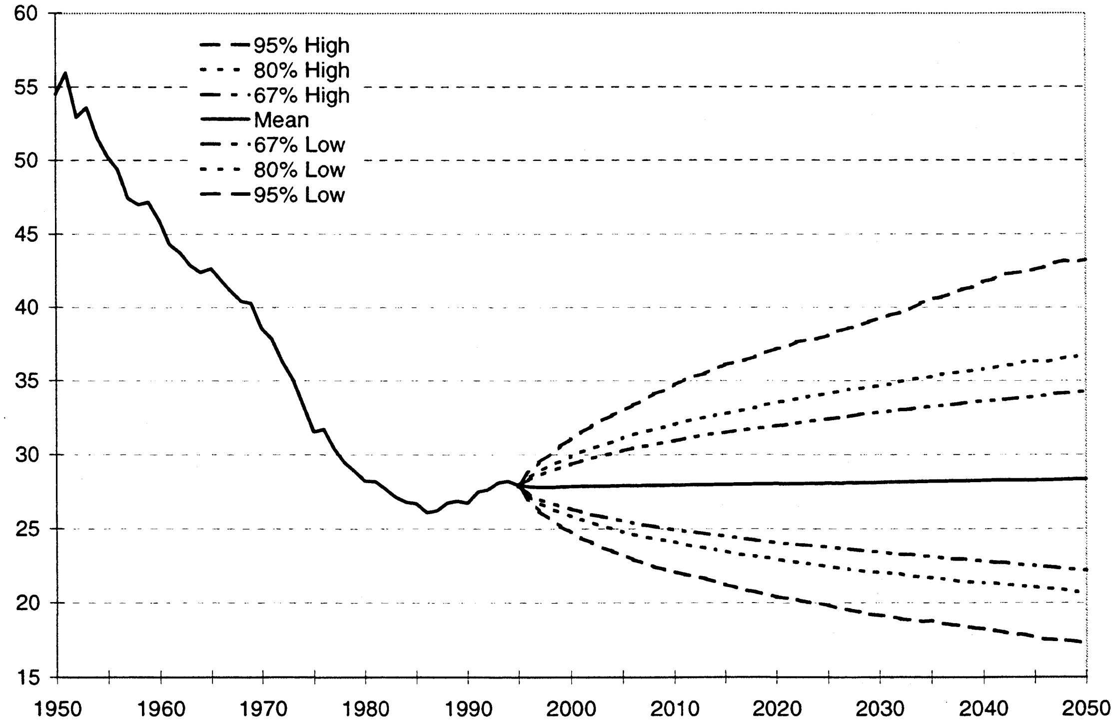


Figure 7. Variance in childbearing age



The average values for the three parameters in 2050 are 2.21 (TF), 30.9 (mean age) and 28.3 (variance), while the medians are 1.86 children per woman, 30.2 years, and 27.6 years^2, respectively. The exponential transformation from C_t back to β_t causes a relatively large difference between mean and median for the TF, but much less so for the mean age or the variance. At the same time, the transformation results in a-symmetric confidence intervals around mean values. This reflects the fact that, say, a doubling of the TF (although a rare event) is much more probable than a drop to a level of zero children per woman - the latter probability is zero. The odds are two against one that the TF in 2050 will lie between 1.1 and 3.3 children per woman, while the 95 per cent confidence interval is (0.5, 6.1) in that year.

The expected probability that the TF will exceed 6.1 children per woman in 2050 is only 2.5 per cent. Yet it is difficult to imagine such extremely high levels of childbearing in a country like Norway. The highest TF ever recorded for Norway is (the moment estimate of) 4.8 in 1879, the highest value since 1845, see Brunborg and Mamelund 1994. With a TF of 6.1 children per woman or more, fertility would exceed the historical maximum by more than one child, and it would exceed the current level in many less developed countries. It would imply a rise by four children over a period of fifty years. One may wonder under what circumstances women would get over six children on average. Activities connected to bringing up those children would have major effects on labour market and educational behaviour of these women and their partners. It is hard to imagine the kind of reasons couples could have when they were to opt for so large families. Clearly, the model predictions for the middle of the next century cannot be considered as realistic. If we somewhat subjectively assume that a fertility level of more than four children per woman on the medium and long term should be rejected, even when the probability of such a level is only a few per cent, we see that the model gives reasonable results up to around 2020 or perhaps 2030. Beyond that, confidence intervals are too wide. One has to take recourse to other methods when predictions so far ahead are required. The easiest one is to assume that in 2030, say, uncertainty is already so large that it will not increase any more. In that case confidence intervals are constant after 2030. A more sophisticated one is to assume that there is an upper bound to fertility levels in Norway. Considerations of this sort led Lee (1993) to include an upper (and lower) bound in his univariate ARIMA-based TF-predictions for the United States, following a suggestion first made by Alho (1990). To that end, he transformed the annual TF into

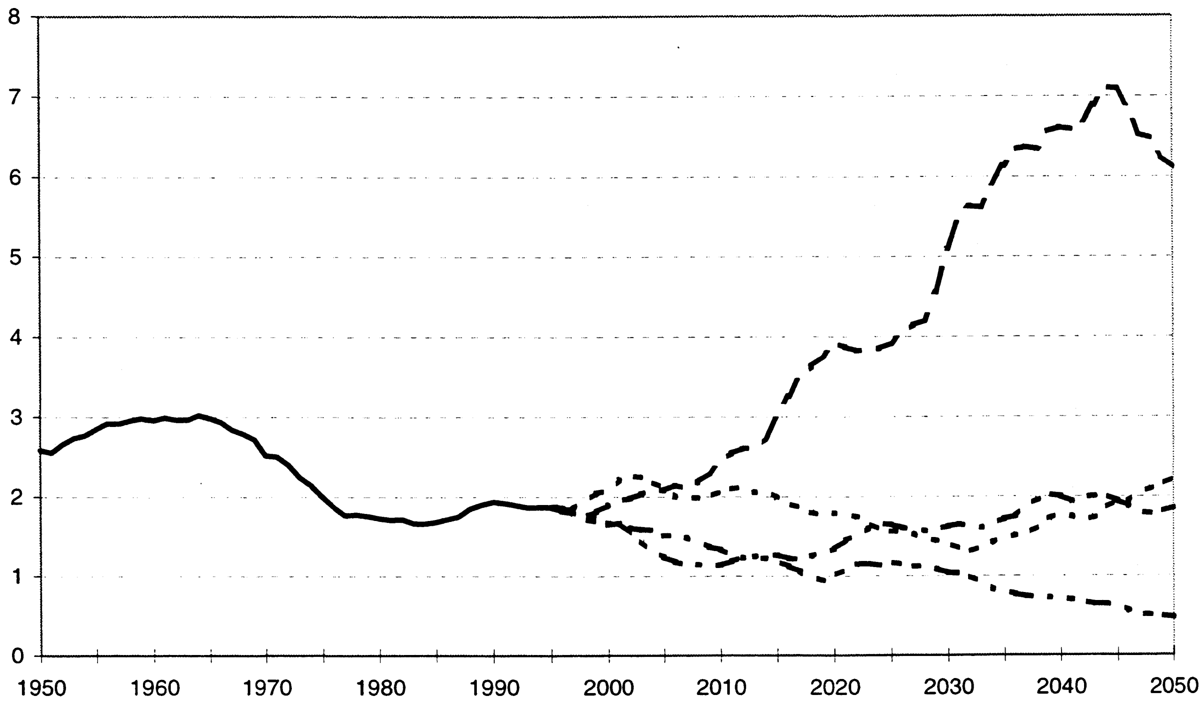
$$g_t = \ln\{(\beta_t - L)/(U - \beta_t)\},$$

where β_t is the TF in year t, g_t is the fertility index to be modeled and forecasted, and U and L are the upper and lower bounds for the TF. This logit transformation will produce a forecast for β_t which will never exceed U or fall below L. However, as noted by Alho and Spencer (1997), such a model may have undesirable consequences. They demonstrated that when g_t follows a random walk process, then β_t will eventually be "absorbed" close to U or L for large enough t. This anomaly also showed up in our case. We selected L=1.0 and U=3.1, and identified a univariate ARIMA (2,1,0)-process for the logit transformed Norwegian TF. (The maximum TF-value in the period 1945-1995 was 3.02 in 1964. Hence an upper bound of 3.0 would cause the transformation to break down in that year. Clearly, the bounds must be strictly larger than the largest observed value, and smaller than the smallest one.) Confidence intervals were computed analytically, assuming that the estimated coefficients of the ARIMA-process are equal to the real ones. In 2050, the bounds of the 67 per cent confidence interval (1.12, 2.86) were very close to those of the 95 per cent confidence interval (1.01, 3.08). In the long run, the boundaries of any interval approach the upper and lower bounds U and L arbitrarily closely. The conclusion is that the logit transformation cannot be used for constraining our confidence bounds.

Another possibility is to assume that the likelihood of a drop in TF is larger the closer the predicted TF approaches a pre-specified upper limit. While such a non-linear model is beyond the scope of the current paper, it would be worthwhile to investigate the statistical properties in future research efforts. For the time being we conclude that we can have confidence in the model up to around the years 2020-2030, but not for the more distant future.

Figure 8 illustrates the volatility in the simulated TF. It shows four sample paths, namely the two paths that came closest to the upper and lower bounds of the 95 per cent interval in 2050, and the paths which hit the average and the median TF-values in 2050.

Figure 8. Four sample paths for the Total Fertility



We experimented also with 5,000 and 1,000 simulations, instead of 10,000, but in those cases the upper bounds of the 95 per cent intervals looked a bit ragged, in particular those for the TF.

The predictions in Figures 5 and 6 are to be compared with the fertility assumptions used by Statistics Norway in their 1996-based population forecast. In that forecast, the year 2010 was chosen as the so-called target year for fertility, that is, the year after which no change was assumed for fertility parameters. Table 4 shows that the TF and the mean age in the Medium Variant agree very well with our median predictions. The constant TF is explained, at least qualitatively, by cohort developments. The cohort TF in Norway decreased rapidly for women born between 1935 (2.57 children per woman) and 1954 (2.05). After some fluctuations in the TF among women born in the second half of the 1950s, this indicator is assumed to fall slightly for women born after 1960, from 2.10 (generation 1960) to 1.97-2.05 children per woman (generation 1971). The decrease is the result of two opposite forces. On the one hand, childlessness increases slightly, from 12 per cent for women born in the 1950s, to an assumed 15-20 per cent for the generations born 30 years later. This presses the cohort TF downwards. At the same time there is an increased propensity among mothers with two children to have a third one. Therefore Statistics Norway decided to opt for a constant period TF in the Medium Variant of a little less than 1.9. But the upward slope in the mean age (26.8 years in 1980, and 28.8 in 1995) was extrapolated, so that a level of 30 years was reached in 2010.

The official forecast's High-Low gap in the TF in 2010 is 0.42 child per woman, and that for the mean age is 1 year. These intervals are rather narrow, compared to the confidence intervals in Figures 5 and 6. For instance, the two-thirds confidence interval for the TF in 2010 is 1.0 children per woman wide, and that for the mean age 3.4 years. Given the normal distribution in our model of log(TF) in 2010, we can conclude that the probability that the real TF in 2010 will lie between the values assumed in the Low and the High Variant of the 1996-based forecast is only 29 per cent. For the mean age the expected probability is no more than 23 per cent. In the concluding section we shall comment these low probabilities.

Table 4. Fertility assumptions in Statistics Norway's 1996-based population forecast

	TF (children per woman)			Mean age at childbearing (yrs)		
	Low variant	Medium variant	High variant	Low variant	Medium variant	High variant
1995[1]	1.86	1.86	1.86	28.79	28.79	28.79
2000	1.79	1.86	1.95	29.34	29.20	29.03
2010 and beyond	1.68	1.86	2.10	30.50	30.00	29.50

[1] Values computed from births statistics.
Source: Statistics Norway 1997.

6. Predictive intervals for age-specific fertility

Figure 9 gives confidence bounds for predicted age-specific fertility rates in the years 2010, 2030 and 2050. They are based on the 10,000 simulations described in the previous section, assuming a minimum age at childbearing equal to zero (compare footnote 2) and an age pattern which follows a Gamma curve. Clearly, the 95 per cent confidence interval is so wide in 2030, that it is not very informative. For the year 2050 it is useless to work with 95 per cent bounds. We have deliberately chosen not to censor rates beyond age 50. Recent medical advances have led to an increased demand for Assisted Reproductive Technology (such as *in vitro* fertilization) after age 30 in Western countries. We cannot exclude the possibility that in the middle of the next century childbearing will be an option for women older than 50. It has been suggested that frozen *ova* may taken from the woman at age 22, say, and that these may be fertilized and implanted later (Beets 1996). The predicted rates, however, are small. At the other end of the age scale, there are some very low rates at ages below 15. These would be ignored in a cohort-component projection.

The Gamma curve fit in Section 4 resulted in deviations for women above age 25. Since these deviations are systematic, they could be included in the predictions. Thompson et al. (1989) and Bell (1992, 1997) describe such a bias adjustment. They took the residuals from the Gamma curve in the last year of data, and extrapolated them forward as constant deviations of future age-specific rates from the forecasted Gamma curves. They report improved accuracy for the first few forecast years, but diminishing effects as the forecast horizon increased. The explanation is that the bias adjustments were dwarfed by the errors in predicting the Total Fertility, the mean age, and the variance. We have not applied any form of bias adjustment, for two reasons. First, the focus in our analysis is on long-term uncertainty, and second, the fit of the Gamma curve was relatively good in recent years.

Figure 9a. Predictive intervals for age-specific birth rates, 2010

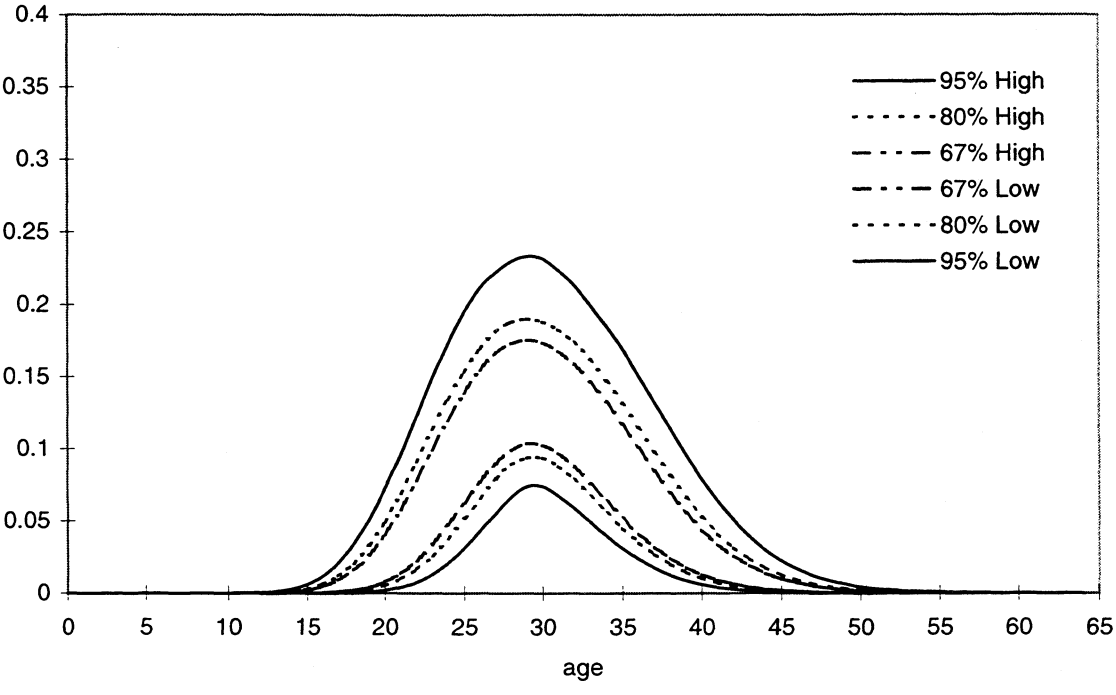


Figure 9b. Predictive intervals for age-specific birth rates, 2030

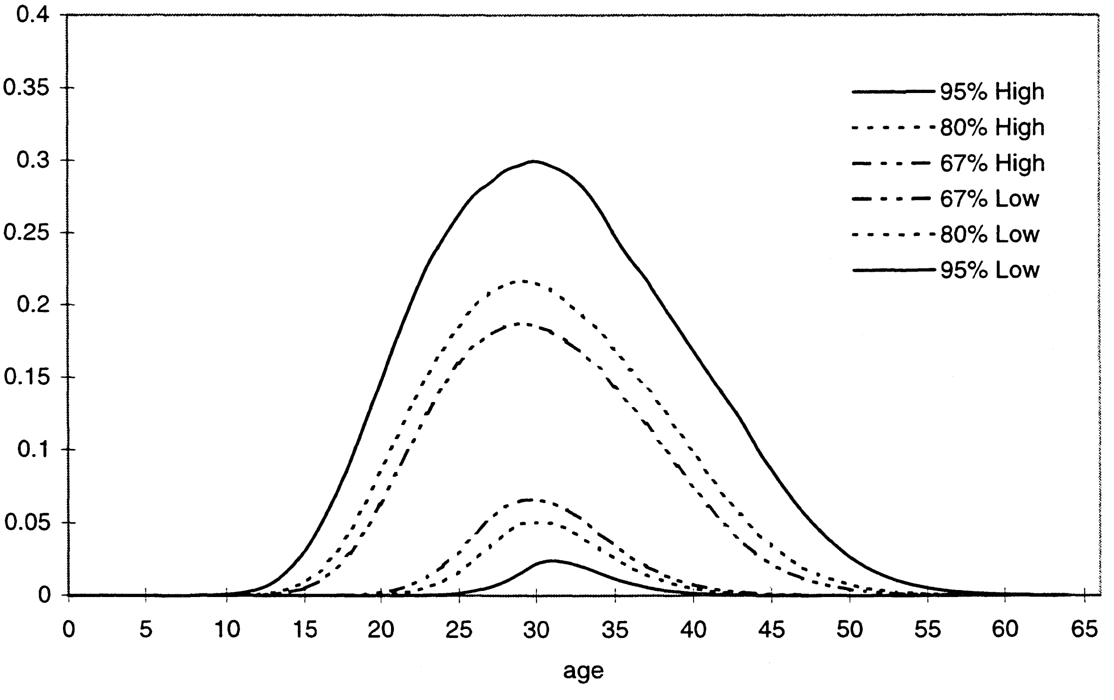
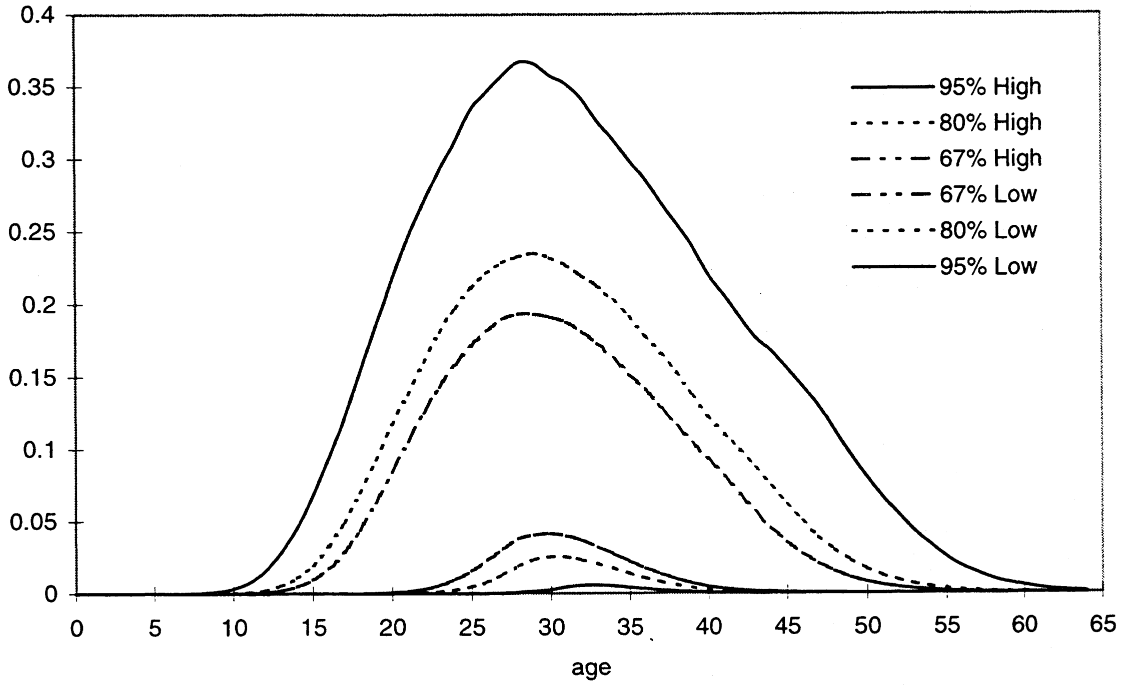


Figure 9c. Predictive intervals for age-specific birth rates, 2050



7. Sensitivity analysis

The analysis in the previous sections differs from earlier studies in three respects. First, as Van Imhoff (1991) has noted, curve fitting exercises almost invariably ignore the fact that a birth rate is not an observed quantity, but an estimate of the parameter of an underlying model. See, for example, Bell 1992, 1997; De Beer 1992; Duchêne and Gillet-De Stefano 1974; Knudsen et al. 1993; Thompson et al 1989; and the references they contain. (Note, however, Hoem (1976) and Hoem et al. (1981) who make a similar point, and establish an interesting link between weighted least squares estimation for the intensities F_x and minimum chi-square estimation for the counts of births B_x.) Second, ARIMA models for the TF or other summary indicators show the same defect, compare, for example, Knudsen et al. 1993; Bell 1997; and Lee 1993. Finally, predictions on the basis of such time-series models assume that the parameters of the model are given, whereas in reality these are only estimates, with corresponding confidence intervals. How serious are these omissions and assumptions? To what extent do they lead to smaller confidence intervals and predictive intervals? We have opted for an empirical analysis of this question, and compared the results reported in the previous sections with corresponding results based on traditional assumptions. In the next three sections we look at predictions with known ϕ-matrix, at unweighted Gamma curve estimates, and at unweighted VAR-model estimates. The final Section 7.4 contains predictions by a model that was estimated for relatively short periods, i.e. 1960-1995 and 1975-1995.

7.1. Predictions with known ϕ-matrix

When the ϕ-matrix is known, simulation is unnecessary, and predictions for the TF, the mean age, and the variance and corresponding confidence intervals can be computed analytically on the basis of expressions (6), (8), and (9). To compute the latter intervals, one needs an assumption for the weight matrix Ω_C which is contained in Σ_ε. (The predictions themselves are not influenced by Ω_C.) For a historical period, $\Omega_{C,t}$ indicates the precision of the estimates of the vector $C_t=ln(\beta_t)$. Future values of $\Omega_{C,t}$ thus express our belief of how well the Gamma curve will fit age-specific birth rates. Since we have annual estimates for the matrix $\Omega_{C,t}$, we could have predicted future values by means of a multivariate time series model. Instead, we have simply kept it constant for future years ($\Omega_{C,t}=\Omega_C$), and experimented with four values: its most recent value $\Omega_{C,1995}$, its average value for the years 1945-1995, as well as the minimum and the maximum values (according to the L_2-norm). The minimum and the maximum gave unreasonably narrow or wide confidence intervals. The average value of Ω_C is very close to the 1995-value, and these two choices for Ω_C resulted in similar future values for the TF, the mean age, and the variance. $\Omega_{C,1995}$ has been used in the predictions. Hence we assume that the future fit of the Gamma curve will be the same as that in 1995.

Table 5 shows that the three parameters level off to values of 1.87 (TF), 30.3 (mean age) and 27.8 (variance) in 2050, close to the medians in Figures 5-7. The 95 per cent predictive intervals in 2050 are somewhat narrower: by 0.5 child for the TF, by 3.9 years for the mean age, and by 1.6 years^2 for the variance. Thus assuming the ϕ-matrix as given leads to 95 per cent bounds for the TF and the mean age in 2050 that are too narrow by 8 per cent and 18 per cent, respectively. The 95 per cent bounds of the variance are very little affected.

Table 5. Predictions and predictive intervals for the TF, the mean age, and the variance, assuming known ϕ-matrix[1]

	TF				Mean age				Variance			
	1995	2010	2030	2050	1995	2010	2030	2050	1995	2010	2030	2050
Prediction	1.87	1.87	1.87	1.87	28.97	30.06	30.25	30.27	27.92	27.79	27.79	27.79
67% Low	-	1.42	1.19	1.05	-	28.50	27.13	26.10	-	24.94	23.42	22.38
67% High	-	2.46	2.94	3.33	-	31.70	33.74	35.10	-	30.97	32.99	34.52
95% Low	-	1.09	0.77	0.60	-	27.08	24.43	22.64	-	22.48	19.87	18.18
95% High	-	3.21	4.53	5.79	-	33.36	37.47	40.47	-	34.36	38.88	42.50

[1] Values for 1995 are Gamma curve estimates.

7.2. Unweighted Gamma curve estimates

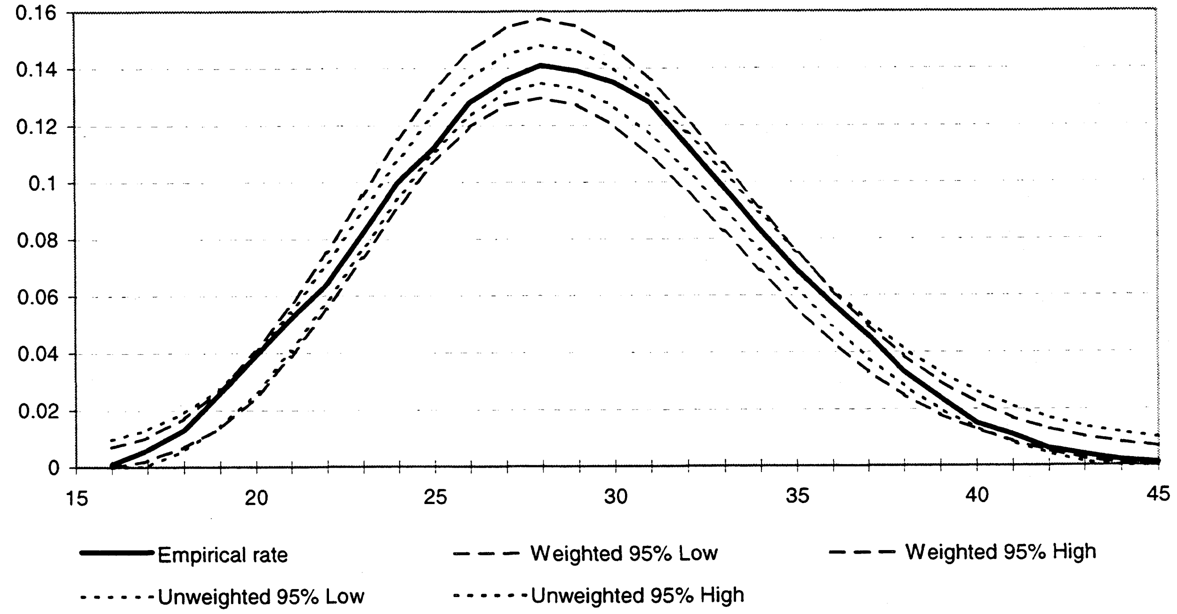
In the weighted case, one takes account of the fact that low birth rates at young and old ages have small variances, and thus these rates get more weight than high rates at intermediate ages. In the unweighted case, rate variances are ignored, and all weights w_x in expression (4) are chosen equal to one. Then it is assumed that birth rates are observed, instead of parameter estimates for the Poisson model of Section 3. The consequence is that the estimated TF becomes higher compared to the weighted case, since the fitted curve follows the high rates more closely. To what extent the mean age and the variance are influenced, is an empirical matter.

We fitted the Gamma curve using unweighted least squares to the data for the years 1945-1995, and found that estimates for all three parameters are lower than in the weighted case in almost every year. The difference is very small after 1970, but larger between 1945 and 1970, when the fit of the Gamma curve was somewhat less (see Section 4). Table 6 compares weighted and unweighted parameter estimates in those years in which differences were smallest and largest. The results indicate that weighting has only had minor impact on the estimates for the TF and the mean age, and a bit more for the variance in the first two decades after the war. However, the estimated parameter variances for the three parameters are relatively small in the unweighted case. This results in narrow confidence intervals for many predicted age-specific rates (see for instance Figure 10 for the year 1995). During the ages of high childbearing, the confidence intervals around the predicted age-specific rates are nearly half as large in the unweighted case compared to the weighted case. For younger and older ages the differences are much smaller - at certain ages the confidence intervals become even *larger* when one does not weight.

Table 6. Comparison of parameter estimates in the unweighted and the weighted case. Years in which differences were largest or smallest

	Year	Unweighted estimate (1)	Weighted estimate (2)	Difference (1)-(2)	Difference (1)-(2) as a % of (2)
TF-largest	1952	2.73	2.66	0.07	2.6
TF-smallest	1985	1.69	1.68	0.01	0.4
MAC-largest	1956	29.24	28.79	0.45	1.6
MAC-smallest	1979	26.90	26.90	0.00	0.0
VAR-largest	1956	57.90	49.37	8.53	17.3
VAR-smallest	1984	26.98	26.80	0.18	0.7

Figure 10. Age-specific birth rates, 1995



7.3. Unweighted VAR-model estimates

When we base the VAR-model on the assumption that the weights w_x are equal to one, and that $\Omega_{\beta,t}$ equals the identity matrix I, we ignore the fact that the birth rates and the three parameters of interest are estimates, each with their own variance. Table 7 gives the predictive intervals for an unrestricted multivariate ARIMA (1,1,0) model that was estimated for the three log-transformed parameters of interest. The intervals are computed analytically on the basis of expressions (8) and (9) assuming that the estimated ϕ-matrix is the real one. Therefore these intervals should be compared with those in Section 7.1.

Table 7. Predictions and predictive intervals for the TF, the mean age, and the variance, assuming w_x=1 and $\Omega_{\beta,t} = I$ [1]

	TF				Mean age				Variance			
	1995	2010	2030	2050	1995	2010	2030	2050	1995	2010	2030	2050
Prediction	1.870	1.877	1.876	1.876	28.97	30.86	31.30	31.36	27.92	30.84	31.27	31.32
67% Low	-	1.501	1.301	1.177	-	29.34	28.14	27.05	-	25.54	23.18	21.45
67% High	-	2.348	2.704	2.991	-	32.45	34.82	36.36	-	37.23	42.18	45.75
95% Low	-	1.211	0.916	0.752	-	27.96	25.40	23.47	-	21.32	17.39	14.91
95% High	-	2.910	3.841	4.680	-	34.05	37.58	41.90	-	44.60	56.22	65.82

[1] Values for 1995 are unweighted Gamma curve estimates.

Whereas the predicted TF in 2050 is almost the same as that in the weighted case, the mean age, and particularly the variance increase to higher equilibrium levels. This is explained by the high estimates for ϕ_{22} (0.986, for the mean age) and ϕ_{32} (0.804, for the variance) that we obtained. The 95 per cent interval for the TF is much narrower than in the weighted case (by 1.7 child in 2050), and that for the mean age a little so (by almost 4 years). The interval for the variance, however, has become much wider, by no less than 25 years^2 in 2050. The reason is the fact that the estimate for the corresponding residual variance $\sigma_{\varepsilon,33}$ is 0.000769, which is more than twice as large as in the weighted case, compare Table 3. The consequence of ignoring weights thus is that we are too optimistic about the future TF in the sense that the predictive intervals are too narrow. At the same time we are too pessimistic regarding the future variance in the age at childbearing. In traditional cohort-component forecasting an error in the TF is more important for the number of births, and hence for the population at young ages, then an error in the variance. If we would have ignored the sample variation in the historical age-specific birth rates in this particular empirical application, and also ignored the errors in the parameter estimates of the Gamma curve, we would have put too much confidence in subsequent births predictions.

7.4. Estimation periods 1960-1995 and 1975-1995

Figures 1-3 show that Norwegian fertility was far from stable during the post-war period, when both the mean age and the TF show major trend shifts. The model estimates in Tables 1-3 were obtained on the basis of data for the years 1945-1995, and hence those estimates as well as the predictions in Figures 5-7 reflect these rather turbulent years. We have re-estimated the model on the basis of two shorter periods, namely the years 1960-1995 and 1975-1995. Would a model estimated for these periods lead to different predictions? The predicted levels for the TF, the mean age, and the variance are probably hardly affected, since the ARIMA-model tends to pick up the trends for the most recent period. Whether the confidence intervals become wider or smaller is an empirical issue. On the one hand, fertility is less volatile during the shorter periods than in the years 1945-1995. This would lead to narrower intervals. On the other hand, when the estimation period is reduced, the estimated residual variance increases (other things remaining the same), and the intervals widen.

Identifying an ARIMA-model on the basis of less than fifty data points is problematic. Therefore we assumed that the (modified) ARIMA-(1,1,0) model for the vector C_t used in Section 5 would also be a good candidate for the periods 1960-1995 and 1975-1995. The results in Table 8 are computed analytically, i.e. assuming known ϕ-matrix. Thus they should be compared with those in Table 5.

The predicted values for the three variables of interest are very similar to those obtained earlier, except for the mean age extrapolated from the years 1975-95. In this case the predictions are up to two years higher, caused by a relatively high estimate for ϕ_{22} (0.96, rather than 0.89 in Table 1). However, since the 67 per cent confidence intervals for the mean age show a great deal of overlap with those in

Table 5, we cannot give much weight to this difference. The confidence bounds for the TF based on the period 1975-1995 are relatively narrow. By 1975, the fall in birth rates had almost come to an end, and fertility was more or less back to normal levels again. The rather small fluctuations in the last two decades increase the predictability of the TF, and this effect is stronger than the fact that the period is only 21 years long. When the TF-bounds based on the shortest period are compared with those starting from 1960-1995, one is struck by the dramatic impact the rapid fall in the TF between 1965 and 1975 has had. Uncertainty is so large (reflected by a large residual variance), that already in 2030 the confidence interval (0.6-5.6 children per woman with 95 per cent probability) is not very informative.

Table 8. Predictions and predictive intervals for the TF, the mean age, and the variance, estimation periods 1960-1995 and 1975-1995

	TF			Mean age			Variance		
	2010	2030	2050	2010	2030	2050	2010	2030	2050
1960-1995									
Prediction	1.87	1.87	1.87	30.00	30.17	30.18	27.92	27.92	27.92
67% Low	1.34	1.07	0.92	28.30	26.82	25.75	25.20	23.70	22.68
67% High	2.61	3.26	3.81	31.81	33.93	35.37	30.94	32.90	34.38
95% Low	0.97	0.63	0.46	26.76	23.96	22.11	22.83	20.24	18.57
95% High	3.59	5.55	7.54	33.64	37.99	41.19	34.15	38.51	41.98
1975-1995									
Prediction	1.87	1.87	1.87	30.81	32.04	32.57	27.92	27.92	27.92
67% Low	1.52	1.32	1.21	29.14	27.63	25.85	25.86	24.83	24.10
67% High	2.30	2.63	2.90	32.57	37.15	41.05	30.16	31.41	32.36
95% Low	1.24	0.95	0.79	27.63	23.97	20.71	24.02	22.18	20.92
95% High	2.81	3.66	4.40	34.36	42.83	51.25	32.47	35.16	37.27

The mean age gives a completely different picture. In this case, intervals based on the period 1960-1995 are very similar to those in Table 5, whereas the intervals are much wider when extrapolations are made from the period 1975-1995. In the latter case, the model just picks up the more or less linear increase since 1975 (see Figure 2). The result is the relatively high estimate for $\hat{\phi}_{22}$ of 0.96, as mentioned earlier. It is much higher than that based on the period 1960-1995 (0.88), when initially the trend was downwards. An increase in $\hat{\phi}_{22}$ leads to an increase in the corresponding covariance of the forecast error $\hat{\sigma}_{22}(l)$ in expression (9), and hence to wider confidence intervals, other things being equal.

We conclude that the predicted levels for the TF, the mean age, and the variance are not significantly affected when we base the extrapolations on the shorter periods 1960-1995 or 1975-1995. As to the confidence intervals, the picture is somewhat mixed. In three of the six cases the intervals are comparable width those based on the period 1945-1995. The TF-intervals are sensitive for the choice of estimation period. It seems as if opting for the period 1945-1995 strikes a good balance between a high residual variance (1960-1995), and an imprecise estimate for the autoregressive coefficient (1975-1995).

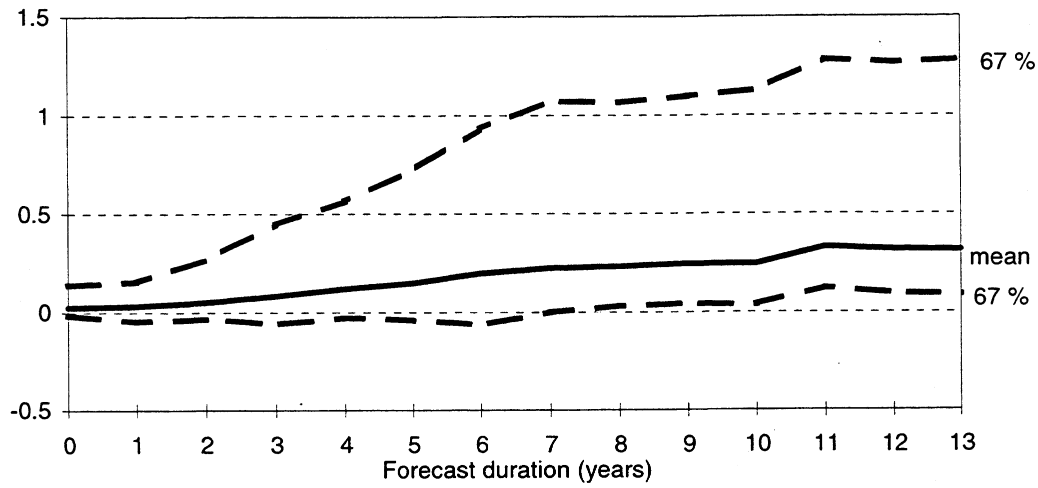
8. Comparison with historical TF-errors

Confidence intervals determine the *expected* errors in the *current* forecast. An independent check of these errors can be provided by investigating *observed* errors in *historical* forecasts. We have looked at the errors in the TF-forecasts that Statistics Norway has published between 1969 and 1993. (Errors in the mean age or in the variance have not been computed, because the assumed values for these indicators in historical forecasts have not been documented.) There have been eleven of such forecasts, with jump-off years 1969, 1970, 1972, 1975, 1977, 1979, 1982, 1985, 1987, 1990, and 1993. We have omitted the 1996-forecast because it is of too recent date. For the other ten we have compared assumed TF-values for each forecast from the jump-off year until 1995 with "observed" values (moment estimators). We have updated the data originally assembled by Texmon (1992), who collected, among others, TFR errors for the forecasts of 1969-1987 during the years 1969-1989. Most forecasts had more than one fertility variant, often two or three. In that case we included all variants in the data, because none of the forecasts, except one, contained a clear advice as to which of the variants was considered by Statistics Norway as the most probable one at the time of publication. Hence it was left to the user to pick one of the variants. The exception was the 1993-based forecast, for which it was clearly indicated that the Medium fertility variant was considered as more probable than the High or the Low variant (Statistics Norway 1994, p. 14). However, users were also advised to investigate the consequences of choosing the other two variants as an input to their own plan or analysis. Only in case the user's conclusions depended little on the choice for the two extreme variants, the user was advised to employ the Medium variant. Thus we may assume that all variants have been used for the forecasts published between 1969 and 1993, although the middle one probably more often than the high or the low one (in case there were three variants).

The error in the TF was simply defined as the assumed minus the "observed" value. Hence a positive or negative error indicates a value that is too high or too low. We have 25 series of TF errors, with a length of between three (the 1993 forecast) and 26 years (the 1970-forecast). The 1969-forecast contained results up to the year 1990 only. Next all errors were ordered by forecast duration, where the jump-off year was defined as duration 0. Hence we had 25 errors for each of the durations 0, 1, and 2 years; 22 for the durations 3-5 years, 19 for durations 6-8, 18 for durations 9-10, and 15 for durations 11-13 years. Errors for longer durations were so few that these were not analysed. For each duration, the errors were ordered from low (including negative values) to high. Finally we selected, by linear interpolation, if necessary, two error values, such that one-sixth of the errors for each duration were lower, and one-sixth were higher than these values. Hence these two values can be interpreted as the bounds of an empirical 67 per cent "confidence" interval. Two out of three errors are within these bounds, and one-sixth of the errors are higher or lower.

Figure 11 shows the bounds of the 67 per cent interval for the TF errors, together with the mean error. The lower bound is close to zero: the error was positive in five out of every six cases. This reflects the fact that the strong fertility decline in the 1970s (see Figure 1) came as a surprise for Norwegian population forecasters, as was the case for demographers in many other Western countries. The distance between the mean and the upper bound is much larger than that between mean and lower bound, indicating that large errors were much more frequent than small ones in the historical forecasts. The width of the 67 per cent interval grows from 0.15 children per woman in the jump-off year to 1.13 at duration 13. The historical errors in Figure 11 increase somewhat faster than the expected ones do. The 67 per cent confidence interval in Figure 5 is 1.13 children wide after a duration of 17 years, instead of 13 years, but the agreement between the two types of errors is striking. Thus the analysis in this section supports the main findings concerning the width of the predictive TF-intervals in Section 5.2, at least for a duration of 10-15 years.

Figure 11. Mean error in historical TF-forecasts, and 67 per cent interval for those errors, by forecast duration



9. Conclusion and discussion

We have shown how statistical techniques can be used to quantify uncertainty connected to age-specific fertility in the future. Unlike traditional analyses, the method takes due account of (1) sampling variability in the birth rates, (2) errors in the parameter estimates of the Gamma curve used for the fertility age pattern, (3) errors in the parameter estimates of the time series model for the prediction of fertility, and (4) residual variance in the time series model. Previous studies have dealt almost exclusively with the fourth source of uncertainty. The method was applied to data for Norway during the period 1945-1995. We found that when error source 3 is ignored, the 95 per cent predictive intervals for the TF and the mean age at childbearing in 2050 become too narrow by 13 per cent and 17 per cent, respectively. In case the first error source is not taken into account when fitting the Gamma curve, confidence bounds around predicted birth rates at ages where fertility is high, may be half as wide as they ought to be. When both the first and the second error sources were ignored, the predictive interval for the TF in 2050 became 1.7 child per woman too narrow.

The predictive intervals for the TF, the mean age at childbearing, and the age-specific fertility rates in this paper appear rather wide. For instance, around the year 2040 the 95 per cent interval for the TF is between 0.6 and 5.4 children per woman on average. The interval for the mean age at childbearing ranges from 23.3 to 41.7 years. These intervals are so wide, that the model gives no useful information any longer. The method produces reasonable results up to the year 2020, or perhaps 2030. In the latter year, the 95 per cent intervals for the TF and the mean age are (0.7, 4.7) children per woman, and (24.4, 39.4) years, respectively. For the more distant future, the method cannot be used.

Meanwhile one should be aware that our confidence bounds, wide as they already are, may be too narrow. After all, they are *predicted* intervals, and these are conditional upon the models that were applied (Poisson model, Gamma curve, and time series model). There is no guarantee that the childbearing behaviour of Norwegian women is in conformity with these models, in particular with the time series model. If one were to take the probability of alternative models into account, the confidence intervals may well become even wider. On the other hand, we cannot disregard the possibility that information *not* included in the statistical models will imply narrower intervals. For instance, if we are 100 per cent certain that Total Fertility in Norway in 2050 will not fall outside a range of, say, between 0.5 and 5 children per woman, this restriction can be included in the simulations, instead of the more liberal restriction of 0-10 children used in Section 5.2[2].

The computed 67 per cent interval for the TF in 2010 is 1.0 children per woman wide, and that for the mean age is 3.2 years wide. In contrast, the gaps between TF and mean age values in 2010 in the High and the Low variant of Statistics Norway's 1996-based population forecast are only 0.4 children per woman and 1.0 years, respectively. On the basis of our model we must conclude that the chances are no more than approximately one in four that the real TF and the real mean age in 2010 will lie between the high and the low values assumed in the official forecast. This means that Norwegian forecasters have been too optimistic when they assumed so narrow bounds between their High and Low fertility variants. Our conclusion on wide bounds is supported, at least for the TF on the medium term, by an independent analysis of historical errors in TF-forecasts since 1969.

The optimism (or self-confidence) among Norwegian population forecasters is not unique, we believe, for two reasons.

First, forecasters in *other Western countries* do not have a much better record concerning the accuracy of their births forecasts during the last thirty years (Keilman 1997). Fertility trends in those countries appear to follow a more or less similar pattern. This pattern is characterized by high birth rates in the 1950s and in the first half of the 1960s, a steep fall in the 1970s, followed by smaller fluctuations during the last fifteen years when postponement and catching up processes played their part. The high-low gap between TF variants in those countries often amounts to 0.3-0.6 children per woman on the medium term (approximately 10 years, see Cruijsen and Keilman 1994). These two facts together make it rather improbable that the real future TF will lie between the levels assumed for the High and the Low variant.

Second, forecast evaluations for *other disciplines* have shown that experts often are too confident (Armstrong 1985, p. 143). Forecasters who state that they have confidence in their forecast, do not predict more accurately than those who say they are uncertain. Furthermore, self-reported confidence increases when a forecasting task is done more often. On the other hand, when forecasters are informed about the accuracy of earlier predictions, they become less confident.

The overoptimism among population forecasters is strikingly illustrated by a recent stochastic forecast for Austria (Hanika et al. 1997; Lutz and Scherbov 1998). The assumption in that forecast is that the TF in the year 2020 will lie with 90 per cent probability between 1.2 and 1.8 children per woman. This assumption is based on expert opinion, not on a statistical analysis. On the basis of the experiences reported in this paper, we expect that the probability for an interval of only 0.6 children per woman wide in 2020 should be in the order of magnitude of 50 per cent, instead of the assumed 90 per cent. The Austrian experts are far too optimistic, in our view.

[2] When this was implemented in a trial simulation, we noted that the 67 per cent confidence interval for Total Fertility in 2010 was still 0.9 children per woman wide (between 1.43 and 2.37), which is to be compared to the 1.0 children gap in Figure 5. In the long run, the difference with Figure 5 was somewhat larger, in particular for the more extreme TF-values. For instance, the 95 per cent confidence bounds became (0.7, 4.3), instead of (0.5, 6.1) in Figure 5.

References

Alders, M. and J. de Beer. 1998. "Kansverdeling van de bevolkingsprognose" ("Probability distribution of population forecasts"), *Maandstatistiek van de Bevolking* **46**(4): 8-11.

Alho, J. M. 1990. "Stochastic methods in population forecasting", *International Journal of Forecasting* **6**: 521-530.

Alho, J. and B. Spencer. 1997. Statistical demography and forecasting. Unpublished manuscript.

Armstrong, J. 1985. *Long-range forecasting: From crystal ball to computer.* New York: Wiley (2nd ed.)

Beets, G. 1996. "Does the increasing age at first birth lead to increases in involuntary childlessness?" Proceedings European Population Conference Milano 4-8 September 1995, Volume 2. Milano: FrancoAngeli, pp. 15-30.

Bell, W. 1992. "ARIMA and Principal Components models in forecasting age-specific fertility", in N. Keilman and H. Cruijsen (eds.) *National Population Forecasting in Industrialized Countries.* Amsterdam etc.: Swets & Zeitlinger, pp. 177-200.

Bell, W. 1997. "Comparing and assessing time series methods for forecasting age-specific fertility and mortality rates", *Journal of Official Statistics* **13**(3): 279-303.

Bratley, P., B. Fox, and L. Schrage. 1983. *A Guide to Simulation.* New York: Springer-Verlag.

Brunborg, H. and S.-E. Mamelund. 1994. "*Kohort- og periodefruktbarhet i Norge 1820-1993*" ("Cohort and Period Fertility for Norway 1820-1993"), Report no. 94/27. Oslo: Statistics Norway.

Cruijsen, H. and N. Keilman. 1994. "Population projections by statistical agencies in countries of the European Economic Area", Paper Annual Meeting of the Population Association of America, Miami, May 1994.

Cryer, J. D. 1986. *Time Series Analysis.* Boston: Duxbury Press.

De Beer, J. 1992. "General time-series models for forecasting fertility", in N. Keilman and H. Cruijsen (eds.) *National Population Forecasting in Industrialized Countries.* Amsterdam etc.: Swets & Zeitlinger, pp. 148-175.

Duchêne, J. and S. Gillet-De Stefano. 1974. "Ajustement analytique des courbes de fécondité générale", *Population et Famille* **32**(2): 53-93.

Eurostat 1997a. "Latest national population forecasts for the countries of the European Economic Area and Switzerland - Reviewed by a set of tables", Background Document Meeting Working Party on Demographic Projections. Luxembourg, September 1997.

Eurostat 1997b. "Beyond the predictable: Demographic changes in the EU up to 2050", *Statistics in Focus, Population and Social Conditions* 1997/7.

Hanika, A., W. Lutz and S. Scherbov. 1997. "Ein probabilistischer Ansatz zur Bevölkerungs-vorausschätzung für Österreich", *Statistische Nachrichten* **12**/1997: 984-988.

Hoem, J. M. 1976. "On the optimality of modified chi-square analytic graduation", *Scandinavian Journal of Statistics* **3**: 89-92.

Hoem, J. M., H. O. Hansen, D. Madsen, J. Løvgreen Nielsen, E. M. Olsen, and B. Rennermalm. 1981. "Experiments in modelling recent Danish fertility curves", *Demography* **18**(2): 231-244.

Keilman, N. 1997. "Uncertainty in national population forecasting", *Journal of Official Statistics* **13**(3): 245-277.

Keyfitz, N. 1982. "Can knowledge improve forecasts?" *Population and Development Review* **8**(4): 729-751.

Knudsen, C., R. McNown, and A. Rogers. 1993. "Forecasting fertility: An application of time series methods of parameterized model schedules", *Social Science Research* **22:** 1-23.

Kravdal, Ø. 1994. "The importance of economic activity, economic potential and economic resources for the timing of first births in Norway", *Population Studies* **48:** 249-267.

Lee, R. 1993. "Modeling and forecasting the time series of US fertility: Age distribution, range, and ultimate level", *International Journal of Forecasting* **9:** 187-202.

Lee, R. and S. Tuljapurkar. 1994. "Stochastic population forecasts for the United States: Beyond High, Medium, and Low", *Journal of the American Statistical Association* **89**(428): 1175-1189.

H. Lütkepohl. 1993. *Introduction to Multiple Time Series Analysis*. Berlin etc.: Springer-Verlag (second edition).

Lutz, W. and S. Scherbov. 1998. "An expert-based framework for probabilistic national population projections: The example of Austria", *European Journal of Population* **14**(1): 1-17.

Spengler, J. 1936. "Population prediction in nineteenth century America", *American Sociological Review* **1**: 905-921.

Statistics Norway. 1994. *Framskriving av folkemengden 1993-2050* ("Population Projections 1993-2050"). Oslo/Kongsvinger: Statistics Norway.

Statistics Norway. 1997. *Framskriving av folkemengden 1996-2050* ("Population Projections 1996-2050"). Oslo/Kongsvinger: Statistics Norway.

Texmon, I. 1992. "Norske befolkningsframskrivinger 1969-1990" ("Norwegian population projections 1969-1990"), in O. Ljones, B. Moen, and L. Østby (eds.) *Mennesker og Modeller*. Oslo/Kongsvinger: Statistics Norway, pp. 285-311.

Thompson, P., W. Bell, J. Long, and R. Miller. 1989. "Multivariate time series projections of parametrized age-specific fertility rates", *Journal of the American Statistical Association* **84**: 689-699.

Van Imhoff, E. 1991 *Profile: A program for estimating the coefficients of demographic age-intensity profiles*. NiDi-Report no. 15. The Hague: NIDI.

Recent publications in the series Documents

97/1 T.C. Mykkelbost and K. Rypdal: Material Flow Analysis of Cadmium and Di-2-etylhexylphthalate (DEHP) in Norway

97/2 S. Grepperud: The Impact of Policy on Farm Conservation Incentives in Developing countries: What can be Learned from Theory

97/3 M. Rolland: Military Expenditure in Norway's Main Partner Countries for Development Assistance. Revised and Expanded Version

97/4 N. Keilman: The Accuracy of the United Nation's World Population Projections

97/5 H.V. Sæbø: Managerial Issues of Information Technology in Statistics Norway

97/6 E.J. Fløttum, F. Foyn, T.J. Klette, P.Ø. Kolbjørnsen, S. Longva and J.E. Lystad: What Do the Statisticians Know about the Information Society and the Emerging User Needs for New Statistics?

97/7 A. Bråten: Technical Assistance on the Jordanian Consumer Price Index

97/8 H. Brunborg and E. Aurbakken: Evaluation of Systems for Registration and Identification of Persons in Mozambique

97/9 H. Berby and Y. Bergstrøm: Development of a Demonstration Data Base for Business Register Management. An Example of a Statistical Business Register According to the Regulation and Recommendations of the European Union

97/10 E. Holmøy: Is there Something Rotten in this State of Benchmark? A Note on the Ability of Numerical Models to Capture Welfare Effects due to Existing Tax Wedges

97/11 S. Blom: Residential Consentration among Immigrants in Oslo

97/12 Ø. Hagen and H.K. Østereng: Inter-Baltic Working Group Meeting in Bodø 3-6 August 1997 Foreign Trade Statistics

97/13 B. Bye and E. Holmøy: Household Behaviour in the MSG-6 Model

97/14 E. Berg, E. Canon and Y. Smeers: Modelling Strategic Investment in the European Natural Gas Market

97/15 A. Bråten: Data Editing with Artificial Neural Networks

98/1 A. Laihonen, I. Thomsen, E. Vassenden and B. Laberg: Final Report from the Development Project in the EEA: Reducing Costs of Censuses through use of Administrative Registers

98/2 F. Brunvoll: A Review of the Report "Environment Statistics in China"

98/3: S. Holtskog: Residential Consumption of Bioenergy in China. A Literature Study

98/4 B.K. Wold: Supply Response in a Gender-Perspective, The Case of Structural Adjustments in Zambia. Tecnical Appendices

98/5 J. Epland: Towards a register-based income statistics. The construction of the Norwegian Income Register

98/6 R. Chodhury: The Selection Model of Saudi Arabia. Revised Version 1998

98/7 A.B. Dahle, J. Thomasen and H.K. Østereng (eds.): The Mirror Statistics Exercise between the Nordic Countries 1995

98/8 H. Berby: A Demonstration Data Base for Business Register Management. A data base covering Statistical Units according to the Regulation of the European Union and Units of Administrative Registers

98/9 R. Kjeldstad: Single Parents in the Norwegian Labour Market. A changing Scene?

98/10 H. Brüngger and S. Longva: International Principles Governing Official Statistics at the National Level: are they Relevant for the Statistical Work of International Organisations as well?

98/11 H.V. Sæbø and S. Longva: Guidelines for Statistical Metadata on the Internet

98/12 M. Rønsen: Fertility and Public Policies - Evidence from Norway and Finland

98/13 A. Bråten and T. L. Andersen: The Consumer Price Index of Mozambique. An analysis of current methodology – proposals for a new one. A short-term mission 16 April - 7 May 1998

98/14 S. Holtskog: Energy Use and Emmissions to Air in China: A Comparative Literature Study

98/15 J.K. Dagsvik: Probabilistic Models for Qualitative Choice Behavior: An introduction

98/16 H.M. Edvardsen: Norwegian Regional Accounts 1993: Results and methods

98/17 S. Glomsrød: Integrated Environmental-Economic Model of China: A paper for initial discussion

98/18 H.V. Sæbø and L. Rogstad: Dissemination of Statistics on Maps

98/19 N. Keilman and P.D. Quang: Predictive Intervals for Age-Specific Fertility

Documents

PORTO BETALT
VED
INNLEVERINGA
P.P.
NORGE/NOREG

Tillatelse nr.
159 000/502

B

Returadresse:
Statistisk sentralbyrå
Postboks 8131 Dep.
N-0033 Oslo

Statistics Norway
P.O.B. 8131 Dep.
N-0033 Oslo

Tel: +47-22 86 45 00
Fax: +47-22 86 49 73

ISSN 0805-9411

Statistisk sentralbyrå
Statistics Norway