POPULATION FORECASTS AND CONFIDENCE INTERVALS FOR SWEDEN: A COMPARISON OF MODEL-BASED AND EMPIRICAL APPROACHES

Joel E. Cohen
    Rockefeller University, 1230 York Avenue, New York, New York 10021

A population projection (or forecast, here synonymous) is an estimate of the numbers, composition, or distribution of a future population. Recent systematic analyses indicate that past forecasts generally failed to recognize realistically the extent of their own uncertainty (Henry and Gutierrez 1977; Ascher 1978; Keyfitz 1982; Stoto and Schrier 1982; Stoto 1983; Smith 1985).

This paper compares several methods of generating confidence intervals for forecasts of population size. These methods are of two main kinds: model-based methods, derived from an explicit demographic model; and empirical methods, based on the distribution of errors of forecasts. The model-based methods and one of the empirical methods are used here with human population data for the first time.

Of the two model-based methods used here, the first (Heyde and Cohen 1985; Heyde 1985) rests on a generalization of a demographic model (Cohen 1976a) for age-structured populations. The model, a stochastic version of familiar component projection methods, is based on a demographic interpretation of products of random matrices (Furstenberg and Kesten 1960). The second model-based method uses the estimators of the long-term growth rate and long-term variability derived by Heyde and Cohen but combines these estimators differently to estimate confidence intervals. These two methods give both projections and confidence intervals for those projections. They do not give confidence intervals for projections derived from other models.

The third method adapts an empirical approach of Williams and Goodman (1971) to constructing confidence limits for economic forecasts. It requires that population projections be derived from another source. A fourth method uses Stoto's (1983) analysis of errors of published projections. These projections were prepared by the traditional component method.

In comparing the confidence intervals generated by the four methods, I forecast according to Heyde and Cohen in all cases. The Heyde-Cohen forecasting procedure boils down to fitting a simple exponential curve exactly through the first and last data points of a time-series of population sizes. This forecasting procedure, while familiar and appealing (e.g., Henry and Gutierrez 1977; Stoto and Schrier 1982; and Stoto 1983), is not an a priori assumption of the demographic model but a maximum likelihood estimator derived from martingale limit theorems (Hall and Heyde 1980) by Heyde and Cohen (1985). This demographic model, which provides a rich representation of age-structured populations with fluctuating vital rates, admits such a simple forecasting procedure because the forecasting procedure is derived not from the short-term but from the asymptotic behavior of the model, which is exponential growth. In this respect and others, the demographic model on which the forecasts of Heyde and Cohen are based is a stochastic generalization of stable population theory.

The model-based methods of estimating confidence intervals are not universally

applicable. They apply where the model's underlying assumptions are consistent with historical time-series of appropriate data, or where, in the absence of appropriate historical data, the underlying assumptions may be presumed relevant, pending data to the contrary.

As Stoto (1983) has pointed out, demographers have analyzed the accuracy of population projections in one of two usually disjoint ways: either by analyzing mathematical models for population growth, or by empirical analysis of past projections. This paper is one of apparently few to compare approaches based on a stochastic population model and on empirical estimates of uncertainty.

I first review the demographic model and the methods of estimating confidence intervals that are based on it. I test the assumptions of the model with the Swedish data and use appropriate portions of the data to estimate confidence intervals. I then adapt the method of Williams and Goodman (1971) and compare the confidence intervals estimated by the empirical methods of Williams and Goodman (1971) and Stoto (1983) with those estimated by the first two methods. Where possible, I also compare the confidence intervals estimated here with those based on previous analyses of Saboia (1974). I conclude with tentative recommendations and with unanswered questions.

MODEL-BASED APPROACHES TO ESTIMATING CONFIDENCE INTERVALS

Theoretical background

The classical theory of populations with age structure assumes that age-specific birth and death rates are constant in time (Keyfitz 1968) or change deterministically (Coale 1972). Various models have been developed for age-structured populations with age-specific vital rates that vary stochastically, i.e., with a random component (see Alho and Spencer 1985, and Cohen 1985, for reviews). Here, for the first time with human population data, I use a model of age-structured populations with stochastic vital rates based on products of random matrices. Various cases of this model have been studied by Cohen (1976a, 1977a,b, 1979a,b,c,d, 1980, 1982), Lange (1979), Lange and Hargrove (1980), Lange and Holmes (1981), Charlesworth (1980), Slade and Levenson (1982), Tuljapurkar and Orzack (1980), and Tuljapurkar (1982). See Tuljapurkar (in press) for a recent review. The model was first used to analyze population data by Cohen, Christensen and Goodyear (1983) and Goodyear, Cohen and Christensen (1985) in a study of time-series of striped bass numbers. Heyde and Cohen (1985) analyze a general form of this model, which I now describe.

Consider a large age-structured population with stochastically fluctuating vital rates. As in the usual matrix formulation of the component method of population projection (see Keyfitz 1968), the number of individuals in each of d age classes at time t is represented by a column d-vector $Y(t)$. The age-specific rates that would project the population forward by one time unit from t to $t + 1$, in the absence of immigration, are represented by a $d \times d$ matrix $X(t + 1)$. Commonly, each element of the first row of $X(t + 1)$ represents the effective fertility of the corresponding age class during the period from t to $t + 1$. The subdiagonal commonly contains the survival proportions. The following analysis applies not only to this standard interpretation of the matrix $X(t + 1)$, but also to all linear generalizations, e.g., to multiregional or (linear) two-sex or parity-structured populations.

The number of net migrants in each age class at time $t + 1$ is represented by a column d-vector $V(t + 1)$. This vector needs to be added to the model only if net migration to or from the population between t and $t + 1$ is not a linear function of $Y(t)$, the population existing at time t. Net migration that is a linear function of $Y(t)$

can be absorbed into the elements of $X(t + 1)$ without adding any $V(t + 1)$.

Given an initial population $Y(1)$ at time 1, the dynamics of the population are modeled by

$$Y(t + 1) = X(t + 1)Y(t) + V(t + 1), \quad t = 1, 2, \ldots \tag{1}$$

I assume that the projection matrices $\{X(t)\}$ form a stochastically stationary ergodic sequence that is uniformly mixing in a sense specified by Heyde and Cohen. Here "stochastically stationary" means that the joint probability distribution of any finite number of the matrices X is invariant with respect to shifts in time. By contrast, "demographically stationary" refers to a population that is constant exactly or on average in total size and, sometimes, in age-specific vital rates. Roughly speaking, the sequence $\{X(t)\}$ is "ergodic" if, for every event defined in terms of a finite number of matrices X, the average frequency of the event converges almost surely to the probability of the event. Qualitatively, uniform mixing means that the vital or other rates in the matrix $X(t_1)$ at time t_1 approach independence of the rates in the matrix $X(t_2)$ at time t_2 as the times t_1 and t_2 become farther apart. The sequence of projection matrices will be uniformly mixing in the sense specified if the matrices are independently and identically distributed or if they are finite in number and determined by an ergodic Markov chain of arbitrary finite order. These special cases probably cover most cases of practical demographic interest.

I also assume that there is a uniform upper bound on the largest vital or other rates that can occur in the matrices X; that for some integer K any product of K matrices has all of its elements positive with probability 1; and that there is another constant C, a positive number greater than 1, such that, with probability 1, the ratio of the largest element of X to the smallest *positive* element of X is less than or equal to C. All three of these assumptions are plausible for the projection matrices used with age-structured human populations and are true for the historical projection matrices that have governed the Swedish population.

The total size of the population at time t may be written $Z(t) = \underline{1}'Y(t)$, where $\underline{1}$ is the vector with every element equal to 1. The methods to be described apply not only to total population size but to any age group within the population, and more generally to any linear function $a'Y(t)$ of the population vector $Y(t)$ where a is a nonnegative nonzero d-vector. For example, if a were the vector of labor-force participation rates, the methods could be applied to the size $a'Y(t)$ of the labor force. For simplicity, I analyze here only the total population size.

Let $W(t) = \log Z(t)$ be the natural logarithm of the population size at time t. (I use natural logarithms throughout.) In the absence of nonlinear net migration, i.e., with $V = 0$ always, if the model (1) satisfies the above assumptions, then asymptotically (i.e., after a long time) the population size $Z(t)$ almost surely becomes proportional to λ^t. This is equivalent to saying that, with probability 1, $W(t)$, the logarithm of population size, asymptotically grows linearly at a rate $\log \lambda$ per unit time. If $\log \lambda > 0$, then the population is increasing; if $\log \lambda < 0$ then the population is decreasing; and if $\log \lambda = 0$ then on average the population is demographically stationary in total size. For the stochastic model (1), $\log \lambda$ is the precise analog of the intrinsic rate of natural increase r for classical stable population theory. Moreover, after a long time, $W(t) - t \log \lambda$ becomes normally distributed with mean 0 and standard deviation equal to $t^{1/2}\sigma$. Stable population theory might be thought of as the special case of this stochastic model in which $\sigma = 0$.

Parameter estimators and point estimators for projections

For present purposes, the two most important consequences of the model described above are that: (a) asymptotically (after a long time), the average growth rate of the logarithm of population size is a constant, log λ per unit of time, and (b) asymptotically, the deviation between the logarithm of population size and its expected value given in (a) is normal, with a zero mean and a standard deviation equal to σ times the square root of time.

Heyde and Cohen (1985) have proposed estimators of the crucial parameters log λ and σ. To describe the estimators, I need some terminology. Three points in time, denoted by t, and three time intervals are involved in most population projections. If $t = 1$ is the epoch (point in time) of the earliest observed population size on which a projection is based, $t = T$ is the epoch of the last observed population size on which a projection is based, and $t = \tau$ is the epoch that the projection is supposed to describe, I call $t = 1$ the base, T the launch, and τ the target. Obviously $1 \leq T < \tau$. I call the difference $T - 1$ the span of the projection (like the span of a bridge, because the projection is supported by data at both ends). I call the difference $\tau - T$ the gap of the projection, and the difference $\tau - 1$ the range of the projection.

For the case without migration, i.e., $V = 0$, Heyde and Cohen have shown that the maximum likelihood estimator of the asymptotic average growth rate log λ of population size is

$$\log L = [W(T) - W(1)]/(T - 1), \tag{2}$$

where I use L for the estimator and λ for the parameter that L estimates. The coefficient σ can be estimated from the data by the estimator

$$s = (1/2) (\pi/2)^{1/2} \times$$
$$\{[\log(T - 1)]^{-1} \sum_{j=1}^{T-1} j^{-3/2} |W(1+j) - W(1) - j \log L| \tag{3}$$
$$+ [\log(T - 2)]^{-1} \sum_{j=1}^{T-2} j^{-3/2} |W(2+j) - W(1) - j \log L|\}.$$

This estimator is consistent, i.e., it converges in probability to the true value of σ as T increases to infinity. The estimator s is sensible only if $T - 2 \geq 3$, so that log $(T - 2) > 0$, i.e., only if $T \geq 5$. The estimator s will not give sensible numbers with fewer than $T = 5$ observations. The finite-sample properties of the estimators (2) and (3) remain to be investigated analytically. Heyde (1985) showed that, if the distribution of the net migration vector V is independent of prior and present projection matrices and population censuses and prior net migration vectors, the identical formulas (2) and (3) can be applied to the model (1) *with* migration. This conclusion probably holds under much weaker assumptions about net migration.

As an estimator of a standard deviation σ, s is unusual in two respects: it is based on the weighted absolute deviations (rather than the more usual square root of the weighted squared deviations), and its weights occur with exponent $-3/2$ (rather than the more usual exponent -1). Heyde and Cohen show that estimators with either of the more usual alternative deviations or weights are not consistent for σ.

The maximum likelihood estimator $w(\tau)$ of the logarithm of population size $W(\tau)$ at future time $\tau > T$ implied by (2) is

$$w(\tau) = W(T) + (\tau - T) \log L$$
$$= W(T) + [(\tau - T)/(T - 1)][W(T) - W(1)]. \tag{4}$$

(Observe: the notation uses small w for estimates or estimators; capital W for

observations.) I use (4) for projections (point estimates of the logarithm of future total population size), in combination with all methods of estimating confidence intervals except Saboia's.

M. Stoto (personal communication, 16 July, 1984) raises an important question: what qualifies as data? For example, given decennial censuses and intercensal interpolations, should the estimators (2) and (3) be applied to the census data only or to the interpolations as well? The answer depends on how the intercensal figures were computed. Linear or exponential interpolations should not be treated as data because they do not reflect the random vital and migratory processes presumed to be at work between censuses. Intercensal estimates derived from a demographic balance of birth, death, and migration could reasonably be used, provided that the vital and migratory data are independently reliable and not simply adjustments derived from some linear or exponential model to match the next census.

Confidence interval estimators

In a projection according to (4), two sources of uncertainty are, first, the uncertainty about the values of the future vital rates contained in the matrices X that will occur from the launch T to the target τ, and second, the uncertainty about the possible difference between the past observed growth rate log L and the true asymptotic growth rate log λ. A third source of uncertainty is model misspecification: the model (1) from which the projection formula (4) is derived may not describe reality. The Williams-Goodman approach below is one way of dealing with this third source of uncertainty. In this section, I take the model (1) as given. The confidence interval estimator presented by Heyde and Cohen (which I sometimes refer to as estimator 1) and another to be developed here (estimator 2) agree on the magnitude of the two sources of uncertainty and differ only in how they are combined. Both approaches give an approximate $100(1 - \alpha)$ percent confidence interval (e.g., a 95 percent confidence interval if $\alpha = 0.05$) for $w(\tau)$, where $\tau > T$, by

$$w(T) + (\tau - T)(T - 1)^{-1}(W(T) - W(1)) \pm s\, F_i(\alpha, T - 1, \tau - T), \qquad (5)$$

where $i = 1$ or 2 according to which estimator is used.

The Heyde-Cohen confidence interval estimator

Define z_β to be that number such that a fraction β of the probability density of a standard normal distribution is to the right of z_β. E.g., $z_{0.05} = 1.64$ approximately. Heyde and Cohen (1985) derive

$$F_1(\alpha, T - 1, \tau - T) = \min_{0 < q < \alpha} \{(\tau - T)(T - 1)^{-1/2} z_{q/2}$$
$$+ (\tau - T)^{1/2} z_{(\alpha-q)/[2(1-q)]}\} \qquad (6)$$

as follows. Consider the two events (possible states of affairs)

$$E_1 = \{|\log L - \log \lambda| < s(T-1)^{-1/2} z_{p/2}$$
$$\text{and } |W(\tau) - W(T) - (\tau-T)\log L| < s(\tau-T)^{1/2} z_{q/2}\},$$
$$E_2 = \{w(\tau) \text{ is in the interval } W(T) + (\tau - T)(T - 1)^{-1}(W(T) - W(1))$$
$$\pm [s(\tau - T)(T - 1)^{-1/2} z_{p/2} + s(\tau - T)^{1/2} z_{q/2}]\}.$$

The lower limit of the interval in the event E_2 occurs if both log L and the increments in population size between T and τ are low, while the upper limit occurs if both log L

and the increments are high. It is easy to see that whenever E_1 occurs, E_2 also occurs, but not necessarily conversely. Therefore, the probability of E_2 is at least as big as the probability of E_1, or symbolically,

$$P(E_2) \geq P(E_1). \tag{7}$$

Because of asymptotic normality and mixing, one has for large T and large $\tau - T$ approximately

$$P(E_1) \simeq (1 - p)(1 - q). \tag{8}$$

The confidence interval estimator suggested by Heyde and Cohen depends on two approximations: approximating the right side of (7) by the left, and approximating the left side of (8) by the right. Heyde and Cohen set $1 - \alpha = (1 - p)(1 - q)$ and solved for the value of p determined by each choice of q, then chose the value of q that minimizes the width of the interval in the event E_2, giving (6). The minimization in (6) can be carried out numerically by scanning a grid of values of q and using a polynomial approximation for z_β.

A new confidence interval estimator: estimator 2

The confidence interval estimator which I now propose,

$$F_2 \, (\alpha, \, T - 1, \, \tau - T) = [(\tau - T)^2 (T-1)^{-1} + \tau - T]^{1/2} z_{\alpha/2}, \tag{9}$$

uses only the approximation (8) but not (7). Since the inequality (7) is bypassed, the confidence intervals based on estimator 2 should be narrower, i.e., include less probability density, than those based on estimator 1, the Heyde-Cohen estimator.

To derive (9), let $Z^{(1)}$ and $Z^{(2)}$ be two independent standard (mean zero, variance one) normal random variables. Then an equivalent formulation of (8) is

$$w(\tau) \simeq W(T) + (\tau - T) \log \lambda + (\tau - T)(T-1)^{-1/2} s Z^{(1)} + (\tau - T)^{1/2} s Z^{(2)}. \tag{10}$$

The term $(\tau - T)(T - 1)^{-1/2} s Z^{(1)}$ represents the variation in $w(\tau)$ contributed by the possible deviation between $\log \lambda$ and $\log L$, while the term $(\tau - T)^{1/2} s Z^{(2)}$ represents the variation in $w(\tau)$ contributed by the increments to population size between the launch and the target. The mixing hypothesis stands behind the assumption that asymptotically these two terms are independent. Now for any real numbers a and b, $a Z^{(1)} + b Z^{(2)}$ has the same distribution as the random variable $(a^2 + b^2)^{1/2} Z$, where Z is also a standard normal random variable. The estimator (9) follows.

EMPIRICAL APPROACHES TO ESTIMATING CONFIDENCE INTERVALS

The Williams-Goodman confidence interval estimator

Williams and Goodman (1971) had long time-series of economic data (the gain in number of telephones at main telephone stations). They fitted a parametric *forecasting* model to the first 24 points of the series, forecasted 18 months ahead, and called the absolute value of the difference between the forecast and the corresponding observed value of the series the absolute forecast error. They then dropped the first point in the series, added the 25th point, re-estimated the parameters of the forecasting model, computed another forecast 18 months ahead, and from comparison with the corresponding observation found another absolute forecast error. Continuing in this way, they obtained a series of absolute forecast errors.

To analyze the distribution of absolute forecast errors, Williams and Goodman (1971) computed the confidence intervals in the conventional way prescribed by the statistical theory of the forecasting model. The computed confidence intervals were

too narrow in comparison with the empirical distribution of absolute forecast errors. (This means, e.g., that the computed 80 percent confidence interval covered only about 70 percent of the forecast errors.) Williams and Goodman (1971) then estimated empirical confidence intervals by fitting a parametric *error* model (a gamma distribution) to the distribution of absolute forecast errors and computing empirical confidence intervals from the fitted error model.

In adapting the idea of Williams and Goodman (1971), I shall omit the fitting of a parametric error model because my time-series has far fewer points than theirs. I shall show instead that, for the Swedish data, it is reasonable to describe the distribution of signed (i.e., positive or negative) forecast errors by a summary statistic, the standard deviation. Also, while Williams and Goodman considered only a single gap (corresponding to the gap required to install additional telephone switching capacity), I shall consider a variety of gaps.

In the absence of guiding theory, I have considered two ways of determining the span to be used in obtaining empirical forecast errors according to the Williams-Goodman approach: *maximal spans* and *fixed spans*. Let θ equal the length (number of observations) of the time-series. To compute empirical forecast errors using maximal spans, I set span = $\theta - 8 -$ gap, so that the span is as large as possible to permit 8 forecasts for the chosen gap. To compute empirical forecast errors using fixed spans, I set span = 8. This seems a reasonable number of censuses to expect many countries to have. With fixed spans, the number of empirical errors that can be computed depends on the gap.

The model (1) of Heyde and Cohen (1985) assumes stationarity in the process that generates the vital rates $\{X(t)\}$. The empirical approaches of Williams and Goodman (1971) and of Stoto (1983) rest on a different assumption of stationarity, namely, that a time-homogeneous process generated past forecast errors and will continue to generate future forecast errors. Whereas the confidence interval estimators 1 and 2 can only be applied when the assumptions underlying the model (1) are plausible, the Williams-Goodman approach to estimating confidence intervals can be used in combination with a wide range of projection techniques, including cohort-component methods.

Stoto's confidence interval estimator

Stoto (1983:17–18) analyzed the errors in the growth rate implied by past population projections. He suggested that for projections $w(\tau)$ made according to (4), approximate 68.3 percent confidence intervals could be computed (in my notation) as

$$W(T) + (\tau - T) \log L \pm (\tau - T)S, \tag{11}$$

where Stoto's optimistic estimate of S for developed countries is $S_{opt} = 5 \times 0.003$ and his pessimistic estimate of S for developed countries is $S_{pes} = 5 \times 0.005$. I inserted a factor of 5 to transform Stoto's time unit of one year to my time unit of five years. For confidence intervals other than 68.3 percent, Stoto suggests multiplying S by a coefficient derived from normal theory.

SWEDISH POPULATION SIZES

The data to be analyzed (Appendix) are the 41 (estimated) total population sizes of Sweden at five-year intervals from 1780 to 1980 inclusive. The estimates from 1780 to 1965 are taken from Keyfitz and Flieger (1968:36). The estimates for 1970 and 1975 are taken from United Nations (1979:168). The estimate for 1980 is taken from United Nations (1983:186).

The data I shall use differ from those available from some other sources. Hofsten (1972:92) reviews the corrections that G. Sundbärg made at the beginning of the twentieth century to the official total population sizes from 1750 to 1810. Neither the official nor the estimated figures of Sundbärg, as quoted by Hofsten, coincide with those of Keyfitz and Flieger (1968) for the corresponding dates but the biggest differences are 1.2 percent of the figures I use. The United Nations' (1979) population estimate for 1970, namely, 8,043,000, differs from that used by Saboia (1974), namely, 8,046,000, which was taken from Keyfitz and Flieger (1971). I treat the data as exact and ignore any possible errors in them. If errors are identified, the methods should be reapplied to corrected figures (as suggested by N. Bennett, personal communication, 5 May 1984).

Testing the model based on products of random matrices

Before using estimators 1 and 2 with the Swedish data, I use the Swedish data to test assumptions of the demographic model (1). First, if the model is approximately correct, then over a long span, $W(t)$, the logarithm of total population size, plotted against time t, should fluctuate around a straight line. A plot (not shown) of $W(t)$ against t, $t = 1780, \ldots, 1980$, appears roughly linear. The slope of the straight line through the first and last logarithms of population size is used in (2) to estimate the asymptotic growth rate of population size. The average growth rate log λ from 1780 to 1980 has been 0.0343 per 5 years, or roughly 0.7 percent per year. Over the same span, the estimate of σ from (3) is 0.0236 (taking five years as the unit of time).

A graph that is more sensitive to possible changes in the growth rate of the Swedish population plots the forward first differences $W(t + 1) - W(t) = \log (Z(t + 1)/Z(t))$ against t (figure 1). (The unit of time here being 5 years, $t + 1$ refers to the epoch of observation 5 years after the epoch t.) Since $Z(t + 1)/Z(t)$ never differs greatly from 1 for the Swedish population, $\log (Z(t + 1)/Z(t))$ is very close to $(Z(t + 1)/Z(t)) - 1$, the fractional change per 5 years. Figure 1 shows no pronounced increasing or decreasing trend in the first differences, especially over the last century. This absence of trends is consistent with the stationarity assumed in the model. However, M. Stoto suggests (personal communication, 16 July 1984) that if one ignores the point for 1805, there may be a downward trend, as is also suggested by figure 2.

In figure 2, I fix 1980 as the launch ($T = 41$) and compute log L and s for spans of 4 (with base 1960) to 40 (with base 1780). The values of log L suggest declining growth. If the peak on the right with base 1940 is viewed as a temporary interruption of the downward trend from around 1815 to around 1925, then figure 2 displays evidence against the stationarity assumed in the model (1). This evidence against stationarity is not conclusive in the absence of a formal statistical test justified by adequate theory. Nevertheless, it hardly seems reasonable to assume that the growth rate of total Swedish population size has been constant since 1780, and the model (1) should therefore not be applied to the Swedish data over this entire span. At the level of intuitive judgment to which I must now resort, it seems more reasonable to apply the model separately, for example, to the data from 1780 to 1875 and to the data since 1880.

Figure 2 suggests the possibility of a downward trend in s with more recent bases, but the scatter in s for bases in this century is so great that this conclusion is far from definite. If the underlying model (1) were correct, there would be no clear trend in s. As M. Stoto (personal communication, 16 July 1984) suggests, s may be generally higher for bases prior to 1860 and lower for bases after 1860. Such a difference would provide an additional rationale for dividing the data into two roughly equal pieces.

I offer no formal statistical tests of stationarity because I know of none that does not rest on a parametric stochastic model such as, e.g., an autoregressive moving
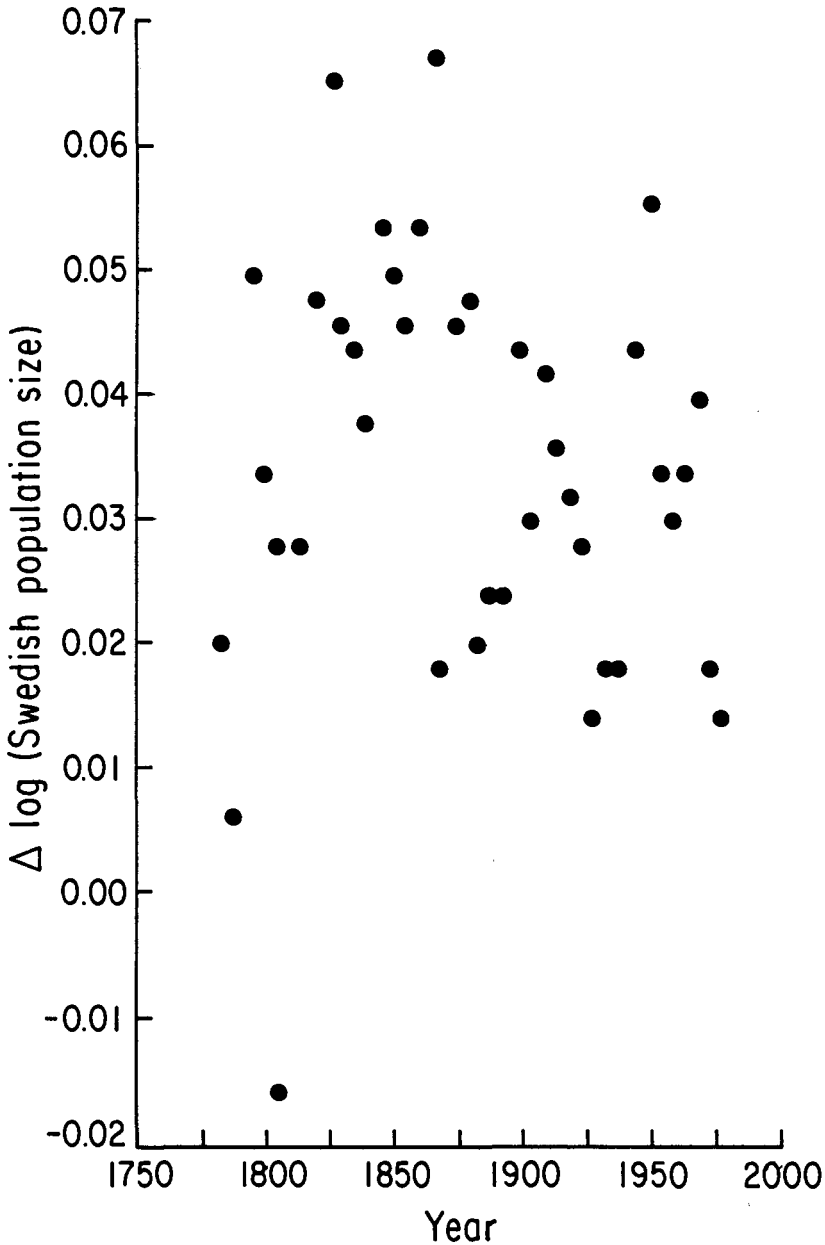
Figure 1. — Forward first differences of the natural logarithm of Swedish total population size (in thousands) from 1780 to 1975.

NOTE: If $Z(t)$ is the total population size (in thousands) at one point of observation and $Z(t + 1)$ is the total population size (in thousands) at the next quinquennium, then the ordinate is $\log Z(t + 1) - \log Z(t)$.

average process. Here I need formal inferential methods that assume only stationarity and mixing.

So far I infer that the model (1), which assumes stationarity, probably does not apply to the whole history of Swedish total population sizes since 1780. The model
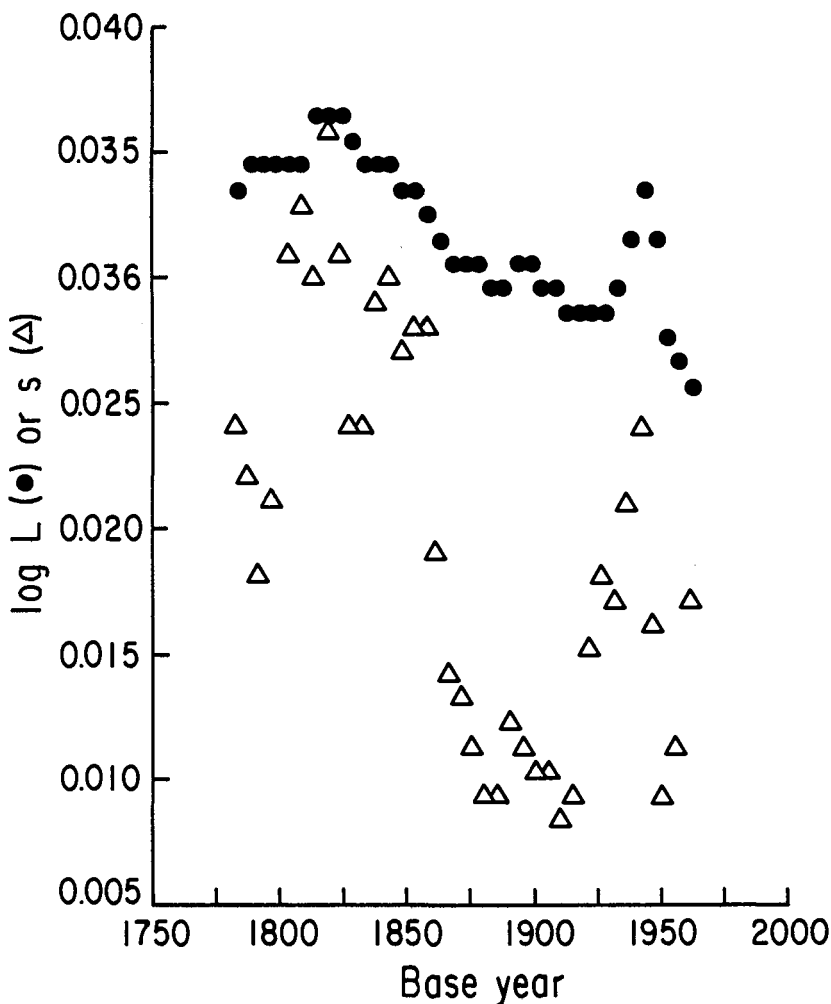
Figure 2. — Estimated growth rates (log L) and coefficients of dispersion(s) for Swedish population projections with a launch of 1980 and various base years.
NOTE: Log L from equation (2); s from equation (3). Population in thousands, observed at quinquennial intervals.

(1) may provide a reasonable approximation to the history of Swedish population sizes from 1780 to 1875 and separately since 1880.

I now examine whether the data display the predicted normality of deviations from strictly exponential growth. For each span and for each gap, the number of forecasts $w(\tau)$ and forecast errors $W(\tau) - w(\tau)$ that can be computed using the 41 Swedish observations is 41 − span − gap. Figure 3 shows that the frequency distribution of the 37 errors of the forecasts with span 3 and gap 1 is bell-shaped, roughly symmetrical with a center near 0. I used the forecast errors over the whole two centuries of Swedish data because the change in the true parameters log λ and σ over any range of four quinquennia is likely to be very small. The frequency distributions (not shown here) of the errors of the forecasts with larger spans and larger gaps are more
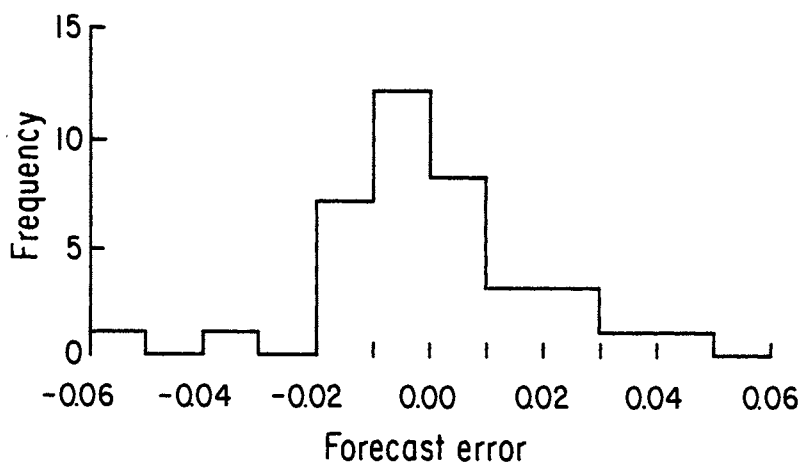
Figure 3. — Empirical frequency distribution of the projection errors in the natural logarithm of Swedish total population based on 37 exponential growth projections with span 3 and gap 1.

irregular, being based on fewer forecasts, but still look roughly normal. Hence the standard deviation seems a reasonable empirical measure of the spread of forecast errors (on a logarithmic scale).

COMPARISON OF CONFIDENCE INTERVALS

I now compare confidence intervals using three subsets of the Swedish population data. First, I use the data from 1780 to 1875 to make projections and confidence intervals from one to ten quinquennia forward, for targets from 1880 to 1925 (table 1).

Table 1. — Comparison of observed population and various confidence intervals[a] for the projected population[b] of Sweden, 1880–1925, based on observed population, 1780–1875.

Entry	Year									
	1880	1885	1890	1895	1900	1905	1910	1915	1920	1925
Exponential point estimate	4533	4710	4894	5086	5285	5491	5706	5929	6161	6402
Heyde-Cohen										
Upper bound	4776	5114	5453	5802	6164	6542	6938	7352	7786	8243
Lower bound	4302	4338	4393	4458	4530	4609	4693	4782	4875	4973
Estimator 2										
Upper bound	4694	4956	5216	5483	5759	6045	6343	6654	6978	7317
Lower bound	4376	4477	4592	4717	4850	4988	5133	5284	5440	5602
Williams-Goodman max. span										
Upper bound	4607	4823	5034	5303	5579	5857	6225	6759	7385	8216
Lower bound	4460	4600	4758	4877	5006	5148	5230	5201	5141	4989
Williams-Goodman fixed span										
Upper bound	4614	4835	5034	5265	5500	5678	5903	6243	.6630	6761
.Lower bound	4453	4588	4758	4912	5078	5311	5516	5631	5726	6063
Stoto optimistic										
Upper bound	4601	4853	5119	5400	5696	6008	6338	6685	7052	7438
Lower bound	4465	4571	4679	4789	4903	5019	5137	5259	5383	5511
Stoto pessimistic										
Upper bound	4647	4951	5275	5621	5988	6380	6797	7242	7716	8221
Lower bound	4421	4480	4541	4602	4664	4726	4790	4855	4920	4986
Observed population sizes	4572	4664	4780	4896	5117	5278	5499	5696	5876	6045

[a] 68.3 percent confidence intervals

[b] In thousands

If 1880 approximates the date of significant changes in Swedish demographic processes, as figure 2 suggests, it is of interest to see how well point and interval estimates based on one set of demographic parameters describe population growth generated under changed demographic parameters. For the population sizes from 1780 to 1875, (2) and (3) yield log L = 0.0384 (corresponding to average annual growth of roughly 0.8 percent per year) and s = 0.0342.

Second, I use the data from 1880 to 1960 to make projections and confidence intervals from one to five quinquennia forward, for targets from 1965 to 1985 (table 2). For the population sizes from 1880 to 1960, (2) and (3) yield log L = 0.0308 (corresponding to average annual growth of roughly 0.6 percent per year) and s = 0.0100. Between 1780–1875 and 1880–1960, while the average growth rate of population size declined by nearly one-quarter, the variability of increments in population size declined much more dramatically, by more than two-thirds, as is suggested by figure 2.

If the underlying parameters of demographic growth (not age-specific vital rates, which varied markedly, but log λ and σ) varied little from 1880 to 1980, it is of interest to compare the success of the forecasts in table 2 with those in table 1. For a given gap, the *absolute* errors of the exponential point estimates in table 1 are comparable to those in table 2: for gaps of 5, 10, 15 and 20 years, the absolute errors in table 1 (vs. those in table 2 in parentheses) are, in thousands, 39 (vs. 20), 46 (vs. 88), 114 (vs. 10), 190 (vs. 150). However, the *relative* errors (table 5) of the exponential point estimates are larger in the earlier period; for gaps of 20 years, more

Table 2. — Comparison of observed population and various confidence intervals[a] for the projected population[b] of Sweden, 1965–1985, based on observed population, 1880–1960.

Entry	Year				
	1965	1970	1975	1980	1985
Exponential point estimate	7714	7955	8203	8460	8724
Heyde-Cohen					
Upper bound	7835	8154	8475	8803	9140
Lower bound	7594	7760	7940	8129	8327
Estimator 2					
Upper bound	7794	8075	8359	8651	8950
Lower bound	7635	7837	8050	8273	8504
Williams-Goodman max. span					
Upper bound	7832	8193	8564	8916	9238
Lower bound	7598	7724	7858	8027	8238
Williams-Goodman fixed span					
Upper bound	7834	8194	8539	8833	8976
Lower bound	7595	7723	7881	8102	8479
Stoto optimistic					
Upper bound	7830	8197	8581	8983	9403
Lower bound	7599	7720	7842	7967	8094
Stoto pessimistic					
Upper bound	7909	8363	8842	9349	9886
Lower bound	7523	7567	7611	7655	7699
Observed population sizes	7734	8043	8193	8310	n.a.

[a] 68.3 percent confidence intervals

[b] In thousands

than twice as large (1895 vs. 1980). This decline in relative errors over time suggests that the absolute errors may be comparable for the earlier and later exponential point estimates because the earlier population size was smaller.

Though the observed population sizes exceed the exponential point estimates for a gap of 5 years in table 1 and for gaps of 5 and 10 years in table 2, for larger gaps the observed population sizes consistently fall below the point estimates. It is hardly surprising that exponential extrapolation fails to capture the long-term trend of declining rates of growth in Sweden's population size.

Of greater interest is the ability of the confidence intervals to contain the realized population sizes. In tables 1 and 2, the observed population sizes strayed outside the estimated 68.3 percent confidence intervals only once: in 1925, the observed population size of 6,045 thousand fell slightly below the lower bound of 6,063 thousand estimated using the Williams-Goodman method with fixed spans. The target 1925 is 50 years after the launch 1875. I conclude that none of the methods generates confidence intervals that are, for these data, unreasonably narrow.

By picking a launch of 1960 for table 2, I can also compare my projections and confidence intervals with those of Saboia (1974). Saboia fitted two time-series models to Swedish data (identical to those used here) from 1780 to 1960. For 1965 and 1970 his ARIMA(1,1,0) model gives point estimates of 7,666 and 7,835 thousands, while his ARIMA(0,2,1) model gives point estimates of 7,716 and 7,953 thousands, respectively. The absolute errors of the ARIMA(1,1,0) model, 68 and 208 thousands, are substantially larger than the corresponding absolute errors of the exponential point estimates, namely 20 and 88 thousands. The absolute errors of the

Table 3. — Comparison of observed population and various confidence intervals[a] for the projected population[b] of Sweden, 1985–2005, based on observed population, 1880–1980.

Entry	Year				
	1985	1990	1995	2000	2005
Exponential point estimate	8562	8822	9089	9365	9649
Heyde-Cohen					
Upper bound	8685	9022	9362	9708	10063
Lower bound	8440	8625	8824	9033	9251
Estimator 2					
Upper bound	8645	8946	9250	9560	9879
Lower bound	8480	8699	8931	9173	9424
Williams-Goodman max. span					
Upper bound	8680	9023	9347	9687	10053
Lower bound	8446	8624	8838	9053	9260
Williams-Goodman fixed span					
Upper bound	8688	9052	9375	9687	9977
Lower bound	8438	8598	8812	9053	9331
Stoto optimistic					
Upper bound	8691	9090	9508	9944	10400
Lower bound	8435	8561	8689	8819	8952
Stoto pessimistic					
Upper bound	8779	9274	9797	10350	10934
Lower bound	8351	8391	8432	8474	8515

[a] 68.3 percent confidence intervals

[b] In thousands

ARIMA(0,2,1) model, 18 and 90 thousand, are practically the same as those of the exponential point estimates. In these cases, the computational complexity of the time-series methods did not yield better predictive accuracy than exponential extrapolation. This confirms a general finding of Stoto (1983).

Using the parameter estimates given by Saboia (1974), I computed 68.3 percent confidence intervals. For the ARIMA(1,1,0) model, the confidence intervals are (in thousands) for 1965 from 7,598 to 7,734 and for 1970 from 7,739 to 7,931. These confidence intervals are narrower than the confidence intervals for corresponding dates given by any method in table 2. The observed population size in 1965 was just at the upper bound of the given interval and in 1970 was outside it. For the ARIMA(0,2,1) model, the confidence intervals are from 7,640 to 7,792 for 1965 and from 7,818 to 8,088 for 1970, both of which contain the observed population sizes. These confidence intervals are about as wide as those based on estimator 2, which are the narrowest of the intervals in table 2. The comparison with the results of

Table 4. — Swedish population projections[a]: Upper half-width of confidence intervals[b] and error[c] of exponential point estimates.

Target year	Heyde-Cohen	Estim-ator 2	WG max. span	WG fixed span	Stoto optim.	Stoto pessim.	Error
Based on observed populations, 1780–1875							
1880	243	162	74	81	69	115	39
1885	404	246	113	125	143	241	−46
1890	559	322	140	140	225	381	−114
1895	717	397	218	179	314	535	−190
1900	880	474	294	215	412	704	−168
1905	1051	554	366	187	517	889	−213
1910	1231	637	519	197	632	1091	−207
1915	1423	724	830	314	756	1313	−233
1920	1625	817	1224	468	891	1555	−285
1925	1840	915	1814	359	1036	1818	−357
Based on observed populations, 1880–1960							
1965	122	80	118	120	117	195	20
1970	199	120	238	239	242	408	88
1975	272	156	360	335	378	639	−10
1980	344	191	456	374	523	890	−150
1985	416	226	514	252	679	1162	n.a.
Based on observed populations, 1880–1980							
1985	123	83	118	126	129	217	n.a.
1990	201	124	202	230	269	452	n.a.
1995	273	161	258	286	418	708	n.a.
2000	344	196	323	323	579	985	n.a.
2005	414	230	405	328	751	1285	n.a.

[a] In thousands

[b] Equals upper bound of 68.3 percent confidence intervals, minus exponential point estimate.

[c] Equals observed population size minus exponential point estimate.

Saboia shows that while the time-series methods are not more accurate than simple exponential extrapolation, and are sometimes less accurate, the confidence intervals derived from time-series methods are not wider than those derived from the other methods considered here and may be too narrow to contain the observed population sizes.

Third and finally, I use the data from 1880 to 1980 to make projections and confidence intervals from one to five quinquennia forward, for targets from 1985 to 2005 (table 3). Planners may find the results of interest now. Future scholars will be able to compare them with subsequent observations. For the population sizes from 1880 to 1980, (2) and (3) yield log $L = 0.0299$ (corresponding to average annual growth of roughly 0.6 percent per year) and $s = 0.0094$.

To facilitate comparison of the width of the confidence intervals generated by different methods, table 4 shows the *difference* between each upper bound of a confidence interval and the corresponding exponential point estimate. I shall refer to

Table 5. — Swedish population projections[a]: Upper half-width of confidence intervals[b] and error[c] of exponential point estimates as a percent of the exponential point estimate.

Target year	Heyde- Cohen	Estim- ator 2	WG max. span	WG fixed span	Stoto optim.	Stoto pessim.	Error
Based on observed populations, 1780–1875							
1880	5	4	2	2	2	3	1
1885	9	5	2	3	3	5	-1
1890	11	7	3	3	5	8	-2
1895	14	8	4	4	6	11	-4
1900	17	9	6	4	8	13	-3
1905	19	10	7	3	9	16	-4
1910	22	11	9	3	11	19	-4
1915	24	12	14	5	13	22	-4
1920	26	13	20	8	14	25	-5
1925	29	14	28	6	16	28	-6
Based on observed populations, 1880–1960							
1965	2	1	2	2	2	3	0
1970	3	2	3	3	3	5	1
1975	3	2	4	4	5	8	0
1980	4	2	5	4	6	11	-2
1985	5	3	6	3	8	13	n.a.
Based on observed populations, 1880–1980							
1985	1	1	1	1	2	3	n.a.
1990	2	1	2	3	3	5	n.a.
1995	3	2	3	3	5	8	n.a.
2000	4	2	3	3	6	11	n.a.
2005	4	2	4	3	8	13	n.a.

[a] In thousands

[b] Equals upper bound of 68.3 percent confidence intervals, minus exponential point estimate.

[c] Equals observed population size minus exponential point estimate.

this difference as a *half-width*. Since these are 68.3 percent confidence intervals, the half-width would correspond to one standard deviation on the logarithmic scale if the logarithm of population size were normally distributed. To illustrate, for 1900 the projected total Swedish population size is 5,285 thousand and the upper bound of the Heyde-Cohen confidence interval is 6,164 thousand, according to table 1. For 1900, table 4 shows 880 thousand (instead of 879 = 6164 − 5285) as the half-width for the Heyde-Cohen confidence interval for 1900 because the half-width in table 4 was computed before rounding. Where the data are available, table 4 also shows each observed population size minus the corresponding point estimate.

The half-widths and errors, expressed in table 4 in thousands of people, are expressed in table 5 as a percent of the corresponding exponential point estimates from tables 1, 2 and 3. For example, the exponential point estimate for 1985 is 8,562,000 (from table 3) and the upper half-width of the Heyde-Cohen 68.3 percent confidence interval is 123,000 (from table 4). Under the Heyde-Cohen column in table 5 the entry for 1985 is 123/8562 = 1 percent to the nearest whole percent.

Before making comparisons between methods, I compare variants of the Williams-Goodman approach. The half-widths of the confidence intervals generated using maximum spans and using fixed spans differ from one another by approximately 10 percent for gaps up to 15 years. I consider this an acceptably low level of difference. Beyond gaps of 25 years (i.e., for the targets 1905 to 1925), the half-widths generated by the two variants differ by a factor of two or more. Even for a gap of 25 years, the half-widths of the two variants may differ markedly (as for the target year 1985 in the middle of table 4 or table 5). I consider this difference too large to be acceptable. I conclude that whichever variant is used, the Williams-Goodman approach should be considered reliable for gaps of up to 15 to 25 years with demographic time-series comparable to these Swedish data.

Comparison of Stoto's optimistic and pessimistic confidence intervals shows, as expected, that the pessimistic intervals are wider than the optimistic by a factor that increases with increasing gaps, starting from 5/3 for a 5-year gap.

To compare the two empirical methods, that of Williams-Goodman and that of Stoto, I use only gaps of 5 to 15 years to avoid unreliable results from the variants of the Williams-Goodman method. For gaps of 5 years, Stoto's optimistic half-widths correspond remarkably well with the Williams-Goodman half-widths, but the pessimistic half-widths are larger by the expected factor of 5/3. For gaps of 10 or 15 years, all of the Stoto half-widths are larger than those of the Williams-Goodman methods, and in some cases dramatically larger. For example, for the target 1995, the Stoto optimistic half-width is 418 thousand (5 percent of the exponential point estimate), while the larger half-width of the two Williams-Goodman variants is 286 thousand (3 percent of the exponential point estimate).

I now compare the model-based confidence intervals based on (6) and (9). In every case, the confidence intervals given by the new estimator 2 are between two-thirds and one-half as wide as those given by the Heyde-Cohen estimator. The half-widths according to estimator 2 become relatively narrower as the gap increases. That the half-widths according to estimator 2 would be narrower than those according to estimator 1 was anticipated theoretically above.

To summarize for gaps of 5 to 15 years, the half-widths of the empirical methods may be ranked as: Williams-Goodman ≤ Stoto optimistic < Stoto pessimistic. (For larger gaps, the estimated half-widths of different variants of the Williams-Goodman approach are too divergent to be reliable.) The half-widths of the model-based methods may be ranked as: estimator 2 < Heyde-Cohen.

When the half-widths of the empirical and model-based methods are compared,

the importance of using data from different historical periods emerges, because the results of comparisons depend on which data are used. In the confidence intervals for 1880–1900, the Williams-Goodman half-widths are substantially smaller than those of estimator 2, and Stoto's pessimistic half-widths are substantially smaller than those of Heyde-Cohen. In the confidence intervals for the twentieth century, the reverse holds for both comparisons. The present analysis does not permit me to rank the half-widths of the empirical and model-based methods independently of the data to which the methods are applied.

Since Stoto (1983) analyzed only data and projections from the twentieth century, perhaps I should not apply his method to the Swedish data from 1780 to 1875. Then for targets from 1880 to 1900, the half-widths of the remaining methods are ranked as: Williams-Goodman < estimator 2 < Heyde-Cohen, roughly in the ratios 1:2:3.

In the twentieth-century confidence intervals, the half-widths given by the Heyde-Cohen, Williams-Goodman, and Stoto optimistic methods are all remarkably close for a gap of five years. For larger gaps, these three methods rapidly diverge, but none yields confidence intervals nearly as wide as the Stoto pessimistic estimates. If I had to pick one method to discard in analyzing modern data for a developed country, such as Sweden, I would discard Stoto's pessimistic method as giving intervals that are too wide.

A perspicacious referee (undoubtedly M. Stoto) argues, on the contrary, that Stoto's pessimistic method gives confidence intervals of reasonable width, and that the other methods give intervals that are too narrow. Just as the regime of demographic growth appears to have changed in Sweden in the second half of the nineteenth century, so the regime of demographic growth could change again between the launch and the target, the referee argues. By fitting my model to a period of more or less homogeneous growth (1880–1980) chosen after the fact, the referee says, I have left its projections vulnerable to surprise—an unexpected change in demographic regime. While the model-based and Williams-Goodman methods assume no change in regime between the base and the target, says the referee, Stoto's pessimistic confidence intervals allow for the historically experienced amount of surprise.

In response, I think that Stoto's optimistic and pessimistic intervals are just as vulnerable to surprise as are the intervals derived from other methods because Stoto's method assumes that whatever generated past forecast errors will also generate future forecast errors. If a new kind of demographic forecasting or a new kind of demographic change or both arise, whatever generated past forecast errors will no longer generate future forecast errors. The relative merits of Stoto's vs. other methods can only be resolved quantitatively, I believe: do the 68.3 percent confidence intervals, no matter how derived, contain the actual population size neither more or less than 68.3 percent of the time? My case against Stoto's pessimistic intervals, based in part on the predictions for 1880–1925 that use data from the different regime of 1780–1875, is that Stoto's pessimistic intervals appear to contain the actual population sizes (at least for Sweden—perhaps not in general) more than 68.3 percent of the time. However, I also think the referee's argument is worth stating here to show that reasonable and informed people have yet to reach complete agreement.

The optimistic intervals of Stoto grow wider much more rapidly than those of Heyde-Cohen, estimator 2, or Williams-Goodman. In the very limited experience provided by table 4, the narrower confidence intervals of the latter three methods are wide enough to contain the realized population sizes. Estimator 2, which gives the narrowest confidence intervals, is valuable in suggesting a minimum level of

uncertainty associated with exponential point estimates: for 1985, for example, given the data up to 1980, there is at least one chance in three that the Swedish population will number more than 8,950,000 or less than 8,504,000, according to table 2.

To evaluate the confidence intervals for future population sizes, I assess their demographic plausibility. The upper bounds for the year 2005 of the Heyde-Cohen, estimator 2, Williams-Goodman fixed span, and Stoto optimistic confidence intervals imply average annual growth rates of 0.8 percent, 0.7 percent, 0.7 percent and 0.9 percent, respectively, while the lower bounds of the corresponding intervals imply average annual growth rates of 0.4 percent, 0.5 percent, 0.5 percent and 0.3 percent. As noted above, from 1880 to 1980, the average annual growth rate was roughly 0.6 percent. Over all possible 25-year spans between 1880 and 1980 (i.e., 1880–1905, 1885–1910, and so on up to 1955–1980), the maximum and minimum average annual growth rates were 0.78 percent and 0.44 percent. Roughly one in three *independent* observations should fall outside a 68.3 percent confidence interval, but all possible 25-year spans are hardly independent. This analysis of implied growth rates therefore suggests, but does not prove, that the Heyde-Cohen and Stoto optimistic confidence intervals seem unnecessarily wide, and that the estimator 2 and Williams-Goodman fixed span confidence intervals seem about right.

DISCUSSION AND RECOMMENDATIONS

Nearly 30 years ago, Hajnal (1957) wrote an eloquent and profoundly skeptical review of mathematical models in demography that still merits reading. He indicated that the practical achievements of models at that time were quite modest and gave reasons why. In consonance with Hajnal and many others, Siegel (1972:51) observed: "Elaboration of projection methodology has not resulted in any great increase in the precision of the projections, largely because birth rates have fluctuated widely, and the fluctuations have proven difficult, if not impossible, to predict."

More recently, the forecasting errors of projections and extrapolations have been analyzed quantitatively in several studies (Henry and Gutierrez 1977; Keyfitz 1982; Stoto 1983; Stoto and Schrier 1982; Ascher 1978; Smith 1985). While limited space does not permit me to review the details of these studies here, I draw three simple lessons from them.

1. The longer the gap of a population forecast, the lower its accuracy. When they are accurate at all, population forecasts are usefully accurate for less than a generation.

2. For short-term forecasts, simple projection methods, such as assuming constant geometric growth, are at least as good as complicated ones.

3. Forecasters generally underestimate both the uncertainty of the forecasts they produce and the instability of the core assumptions from which those forecasts are derived.

This paper compares two kinds of techniques for estimating confidence intervals of population forecasts. The first is derived from the application of martingale limit theory to a stochastically explicit demographic model for age-structured populations, and is appropriate for populations with an exponential trend of increase over the range from the earliest datum to the furthest target of projection. The second kind of technique rests on the empirical distribution of forecast errors.

Numerous other techniques for estimating confidence intervals of population forecasts have been proposed (Alho and Spencer 1985). Lee (1974) emphasized that different forecasting models imply dramatically different confidence intervals. My results confirm Lee's observation.

To prepare a population forecast and confidence intervals based on a time-series of population sizes (or time-series of sizes of any linear combination of age-classes, e.g., a time-series of numbers of children or of labor force sizes), I tentatively recommend several steps. Because the theoretical and empirical underpinnings of these recommendations need much further development, I expect these recommendations to evolve.

1. Examine the data graphically, and especially the first differences of the logarithms of the population sizes, to see whether a trend (with possible fluctuations) of exponential growth is plausible for the entire span of the data or for a recent portion of the span. If exponential growth is not plausible over at least a recent portion of the span, the two methods based on model (1) should not be used. The empirical methods of this paper may still be used. Other, e.g., time-series methods (Lee 1974; Alho and Spencer 1985; and those reviewed by Cohen 1985) may be appropriate.

2. Using that portion of the recent data for which exponential growth is plausible (at a minimum, the last two observations), prepare exponential point estimates for future population sizes, but not more than 25 years ahead.

3. Estimate log λ and σ using the Heyde-Cohen estimators (2) and (3) above, and compute confidence intervals according to estimator 2, using (5) and (9).

4. Also compute confidence intervals using fixed spans and maximal spans in the Williams-Goodman procedure and discard any confidence intervals for which the two variants diverge markedly.

5. If the population of interest is a developed country, compute Stoto's optimistic confidence intervals for developed countries.

6. Evaluate the demographic plausibility of the average population growth rates implied by the upper and lower bounds of the confidence intervals that result from steps 3, 4, and 5 and retain those confidence intervals with plausible implied growth rates. Realistic confidence intervals probably fall within the range of confidence intervals that survive.

Given the substantial practical stakes in reliable projections and confidence intervals, each of these steps needs to be justified and made more specific by future research. 1. What graphical methods can best distinguish exponential from nonexponential trends over varying spans of past data? 2. How far ahead should projections be computed for any desired level of reliability, and how much of the past is an optimal span? 3. Are there better estimators of log λ and of σ and are there better ways of combining them to get confidence intervals for the model (1)? 4. What are the theoretical underpinnings of the Williams-Goodman procedure and which variant (fixed spans vs. maximal spans) is preferable under what conditions? 5. Can Stoto's optimistic confidence intervals be refined for a particular developed country that lacks many prior population projections? 6. What other tests of demographic plausibility should be applied to confidence intervals? How reliable is the test of demographic plausibility described here?

If it is disappointing that the above recommendations do not lead to unique confidence intervals, in that disappointment lies a major lesson of this analysis. I sought confidence intervals for population projections in the first place because finitely many observations of the past and incomplete theoretical understanding of the present and future can justify at best interval, not point, estimates of future population sizes. It is impossible to know precisely what the future population will be. Having compared four or more methods of making interval estimates of future population sizes, I now recognize that the same two limitations (finite knowledge of the past, and less than perfect theory of the present and future) make it impossible to

specify a unique population model with unique parameter values from which unique interval estimates of future population sizes (or compositions or distributions) could be derived. Consequently I must be satisfied with a range of confidence intervals for population projections. Uncertainty attaches not only to the point forecasts of future population but also to the estimates of those forecasts' uncertainty. Progress in research, through improved population theories, improved data and improved statistical estimators, can aim to reduce the uncertainty of interval estimates but cannot eliminate it.

This uncertainty of interval estimates is one of several limits on the accuracy of population forecasts (Land 1985). Collectively these limits are one of several kinds of limits on the possible precision and certainty of demographic knowledge (Cohen 1976b, 1983, in press).

Most papers on confidence intervals for population projections (unlike Lee 1974) advocate a single method without testing that method against the merits of others. Since confidence intervals do vary, as Lee first and now I have observed, real progress in measuring the uncertainty of population forecasts requires the direct comparison of competing approaches, empirically and theoretically. In the conviction that real progress is possible, I may exemplify a remark of Hajnal (1957:103): ". . . perhaps, in this as in other fields, it is a good idea to have a few people engaged in striving for the unattainable."

Appendix: Swedish Population Size Estimates 1780-1980

The estimated total size (in thousands) of the Swedish population at 5-year intervals from 1780 to and including 1980 is listed below, reading from left to right within each row, and by rows from top to bottom. E.g., the estimated total population of Sweden in 1790 is 2,161,000.

2104	2147	2161	2274
2352	2418	2380	2450
2573	2749	2876	3004
3123	3296	3462	3625
3824	4092	4164	4362
4572	4664	4780	4896
5117	5278	5499	5696
5876	6045	6131	6242
6356	6636	7017	7262
7480	7734	8043	8193
8310			

REFERENCES

Alho, J. M., and B. D. Spencer. 1985. Uncertain population forecasting. Journal of the American Statistical Association 80:306–314.

Ascher, W. 1978. Forecasting: An Appraisal for Policy-Makers and Planners. Baltimore: Johns Hopkins University Press.

Charlesworth, B. 1980. Evolution in Age-Structured Populations. Cambridge: Cambridge University Press.

Coale, A. J. 1972. The Growth and Structure of Human Populations: A Mathematical Investigation. Princeton: Princeton University Press.

Cohen, J. E. 1976a. Ergodicity of age structure in populations with Markovian vital rates, I: Countable states. Journal of the American Statistical Association 71:335–339.

———. 1976b. Irreproducible results and the breeding of pigs; or nondegenerate limit random variables in biology. BioScience 26:391–394.

———. 1977a. Ergodicity of age structure in populations with Markovian vital rates, II: General states. Advances in Applied Probability 9:18–37.

———. 1977b. Ergodicity of age structure in populations with Markovian vital rates, III: Finite-state moments and growth rates; illustration. Advances in Applied Probability 9:462–475.

———. 1979a. Long-run growth rates of discrete multiplicative processes in Markovian environments. Journal of Mathematical Analysis and Applications 69:243–251.

———. 1979b. Ergodic theorems of demography. Bulletin of the American Mathematical Society N.S. 1:275–295.

———. 1979c. Contractive inhomogeneous products of non-negative matrices. Mathematical Proceedings of the Cambridge Philosophical Society 86:351–364.

———. 1979d. Comparative statics and stochastic dynamics of age-structured populations. Theoretical Population Biology 16:159–171.

———. 1980. Convexity properties of products of random non-negative matrices. Proceedings of the National Academy of Sciences 77:3749–3752.

———. 1982. Multiregional age-structured populations with changing rates: weak and stochastic ergodic theorems. Pp. 477–503 in K. C. Land and A. Rogers (eds.), Multidimensional Mathematical Demography. New York: Academic Press.

———. 1983. How is the past related to the future? Pp. 59–71 in Center for Advanced Study of the Behavioral Sciences Annual Report 1982. Stanford, California.

———. 1985. Stochastic demography. In S. Kotz and N. L. Johnson (eds.), Encyclopedia of Statistical Sciences. New York: John Wiley and Sons.

———. In press. An uncertainty principle in demography and the unisex issue. American Statistician.

Cohen, J. E., S. W. Christensen, and C. P. Goodyear. 1983. A stochastic age-structured population model of striped bass (*Morone saxatilis*) in the Potomac River. Canadian Journal of Fisheries and Aquatic Sciences 40:2170–2183.

Furstenberg, H., and H. Kesten. 1960. Products of random matrices. Annals of Mathematical Statistics 31:457–469.

Goodyear, C. P., J. E. Cohen, and S. W. Christensen. 1985. Maryland striped bass: recruitment declining below replacement. Transactions of the American Fisheries Society 114:146–151.

Hajnal, J. 1957. Mathematical models in demography. Cold Spring Harbor Symposia on Quantitative Biology 22:97–103.

Hall, P., and C. C. Heyde. 1980. Martingale Limit Theory and Its Applications. New York: Academic Press.

Henry, L., and H. Gutierrez. 1977. Qualité des prévisions démographiques à court terme. Étude de l'extrapolation de la population totale des départements et villes de France, 1821–1975. Population (Paris) 32:625–647.

Heyde, C. C. 1985. On inference for demographic projection of small populations. Pp. 215–223 in L. M. LeCam and R. Olshen (eds.), Proceedings of the Berkeley Conference in honor of Jerzy Neyman and J. Kiefer, vol. 1. Monterey, Calif.: Wadsworth and Hayward: Institute of Mathematical Statistics.

Heyde C. C., and J. E. Cohen. 1985. Confidence intervals for demographic projections based on products of random matrices. Theoretical Population Biology 27:120–153.

Hofsten, E. 1972. The Swedish Population 1750–1970. Stockholm: National Bureau of Statistics.

Keyfitz, N. 1968. An Introduction to the Mathematics of Population. Reading, Mass.: Addison Wesley.

———. 1982. The limits of population forecasting. Chapter 13 in Keyfitz, N., 1982. Population Change and Social Policy. Cambridge, Mass.: Abt Books.

Keyfitz, N., and W. Flieger. 1968. World Population: An Analysis of Vital Data. Chicago: University of Chicago Press.

———. 1971. Population: Facts and Methods of Demography. San Francisco: W. H. Freeman.

Land, K. C. 1985. Methods for National Population Forecasts: A Critical Review. Population Research
    Center Paper 7.001, University of Texas at Austin.
Lange, K., 1979. On Cohen's stochastic generalization of the strong ergodic theorem of demography.
    Journal of Applied Probability 16:496–504.
Lange, K., and J. Hargrove. 1980. Mean and variance of population size assuming Markovian vital rates.
    Mathematical Biosciences 52:289–301.
Lange, K., and W. Holmes. 1981. Stochastic stable population growth. Journal of Applied Probability
    18:325–334.
Lee, R. D. 1974. Forecasting births in post-transition populations: stochastic renewal with serially
    correlated fertility. Journal of the American Statistical Association 69:607–617.
Saboia, J. 1974. Modeling and forecasting populations by time series: the Swedish case. Demography
    11:483–492.
Siegel, J. S. 1972. Development and accuracy of projections of population and households in the United
    States. Demography 9:51–68.
Slade, N. A., and H. Levenson. 1982. Estimating population growth rates from stochastic Leslie matrices.
    Theoretical Population Biology 22:299–308.
Smith, S. K. 1985. Population projections in the real world: tests of forecast accuracy. University of
    Florida, manuscript.
Stoto, M. A. 1983. The accuracy of population projections. Journal of the American Statistical
    Association 78:13–20.
Stoto, M. A., and A. P. Schrier. 1982. The Accuracy of State Population Projections. John F. Kennedy
    School of Government, Discussion Paper 117D. Cambridge, Mass.: Harvard University.
Tuljapurkar, S. D. 1982. Population dynamics in variable environments II. Correlated environments,
    sensitivity analysis and dynamics. Theoretical Population Biology 21:114–140.
————. Demographic applications of random matrix products. In J. E. Cohen, H. Kesten, and C. M.
    Newman (eds.), Random Matrices and Their Applications. Contemporary Mathematics no. 50.
    Providence, RI: American Mathematical Society. In press.
Tuljapurkar, S. D., and Orzack, S. H. 1980. Population dynamics in variable environments. I. Long run
    growth rates and extinction. Theoretical Population Biology 18:314–342.
United Nations. 1979. Demographic Yearbook Historical Supplement. ST/ESA/STAT/SER.R/8. New
    York: United Nations.
United Nations. 1983. Demographic Yearbook 1981, 33d ed. New York: United Nations.
Williams, W. H., and M. L. Goodman. 1971. A simple method for the construction of empirical
    confidence limits for economic forecasts. Journal of the American Statistical Association 66:752–754.