TFR predictions based on Brownian motion theory

Nico Keilman

Department of Economics

University of Oslo

P.O. Box 11650 Blindern

N-0317 Oslo

Norway

Email: nico.keilman@econ.uio.no

June 2001

Abstract

- **Context**

In stochastic cohort component forecasts, the expected level of future TFR and its predictive distribution are of central concern. Time series models can be used to predict the TFR and its moments on the short run (up to 10 or 20 years). Experience shows that on the long run (40-50 years), such models result in excessively wide prediction intervals. Appropriate logit transformations may keep the TFR bounded within pre-specified limits. However, when using this method predictive densities become bi-modal on the long run, with probability mass concentrated near the upper and the lower bound.

- **Method**

I model the time series of period TFR-values, or transformations thereof, as a Brownian motion with absorbing barriers. I give and analyse
- expressions for the predictive distribution of the TFR assuming it follows a random walk with absorbing ceiling a;
- approximate expressions for the first and second moments of the predictive distribution of the TFR, when first differences of the log of the TFR follow a random walk with absorbing ceiling a.

- **Results**

When first differences of the log of the TFR follow a random walk with absorbing ceiling a, I find that the second moment of the predictive distribution for the long-run TFR in Norway is insensitive for levels of a beyond a threshold of approximately 3 children per woman. The second moment grows rapidly when a increases from TFR(0) to approximately 2.5 children per woman, where TFR(0) is the starting point of the TFR. This conclusion holds for a fairly broad range of innovation variances. The threshold level of a is rather sensitive for the starting values of the time series, but it remains below 4 children per woman over a reasonable range of those starting values.

- **Conclusions**

If the log of the TFR follows an ARIMA(1,1,0) process with autoregressive coefficient close to one and realistic innovation variance, sample paths that exceed approximately 3-4 children per woman may be rejected when simulating future fertility in Western countries. This will not have any major effect on the width of the long-term predictive distribution.

TFR predictions based on Brownian motion theory[1]

1. Introduction

Keilman and Pham (2000) present TFR-predictions for Norway for the years 1996-2050. These are based on an ARIMA (1,1,0) model for the logarithm of the annual TFR, estimated on the basis of data for the years 1945-1995. The long-run predictions for the mean and median values of the TFR appear reasonable: 2.2 and 1.9 children per woman, respectively. However, and not surprisingly, the prediction intervals are excessively wide on the long run. For example, the 95 per cent prediction interval in 2050 ranges from 0.6 to 6.1 children per woman. Around the year 2020, the upper 95 per cent bound exceeds the level of four children per woman, clearly an unrealistic value for Norway, even if the probability is only 2.5 per cent. Thus the model produces informative prediction intervals up to 20-30 years ahead, but in the long run the predictions are unrealistic and the prediction intervals are too wide.

The reason why the ARIMA model on the long run produces wide prediction intervals for the period TFR is that it contains no extraneous information beyond the past data. Several factors are associated with the development of fertility after World War II in Norway, similar to other countries. Important elements in this respect are the introduction of modern contraception, the adoption of new norms and values with respect to childbearing and partnership, increased interest in tertiary education, and growing levels of labour force participation among women (Kravdal 1994; Lesthaeghe and Surkyn 1988). None of these factors is, or could be, explicitly modelled. It is reasonable to expect that they will be in operation in the future, too. Yet we have no idea to what extent these or other factors will constrain Norwegian fertility in the next century, be it period or cohort developments.

One has to take recourse to other methods when predictions so far ahead are required. The easiest one is to assume that in 2020, say, uncertainty is already so large that it will not increase any more (e.g. Alho and Spencer 1997). In that case prediction intervals are constant after 2030. A more sophisticated one is to assume that there is an upper bound to fertility levels in Norway. This can be done in several ways: by means of a logit transformation, by means of simulation in which extreme values are rejected, and analytically, based on the theory of Brownian motion.

---

2. Logit transformation

Lee (1993) has used a logit transformation in order to restrict predicted TFR-values. He first transformed the annual TFR for the United States into

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

where L and U are pre-specified lower and upper limits for the TFR. Next he identified and estimated an ARIMA-model for the transformed variable g_t. This logit transformation will produce a forecast for the TFR that 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 TFR_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 TFR.[2] In 2050, the bounds of the 67 per cent prediction interval (1.12, 2.86) were very close to those of the 95 per cent 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 Lee's logit transformation cannot be used for constraining our prediction intervals.

3. Rejecting extreme values

In Keilman and Hetland (1999) we simulated predictive TFR intervals by means of the ARIMA (1,1,0) model mentioned in Section 1, thereby rejecting extreme values. We experimented with upper bounds ranging from 2 to 5, and inspected the resulting prediction intervals. We found that setting reasonable limits to the predicted TFR has little impact on the bounds of the 67 per cent prediction intervals in the short run. In the year 2020, this interval is 0.80 children per woman wide when the TFR is limited to 3 children, and 0.92 and 0.95 children wide for maximum TFR values of 4 and 5, respectively. In the long run however, uncertainty (as expressed by the width of the 67 per cent prediction interval) grows faster. In 2050, the interval widens from 1.31 children per woman for a maximum TFR of 3, to 1.64 and 1.88 children per woman for maximum TFRs of 4 and 5, respectively. The 95 per cent

interval bounds are rather strongly influenced by the choice for a certain maximum TFR-value, also in the short run. In 2010 the width amounts to 1.60, 1.92, and 2.06 children per woman for maximum TFR-values of 3, 4, and 5. A maximum TFR equal to 3 would imply a median and a mean TFR that decrease over time, which was judged unreasonable. Therefore the general conclusion was that a maximum TFR equal to 4 children per woman appeared to be a reasonable choice.

4. Brownian motion

Choosing upper and lower bounds for the TFR as discussed in the previous sections is largely an arbitrary matter. At the same time, the choice may have important consequences for the shape of the predictive distribution, as was argued in Section 3. In the current section we will discuss an analytical method. The method is based on the theory of Brownian motion. It is exact in case the time-series for the parameter is a random walk, i.e. an AR(1)-process with coefficient equal to 1. In Keilman and Pham (2000), first differences in logs of the TFR and the mean age at childbearing (MAC) are AR(1), with coefficients equal to 0.67 and 0.88. In other words, the process for the MAC is close to a random walk, and that for the TFR is reasonably close. This makes it worthwhile to investigate to what extent Brownian motion theory can be used to select the bounds for the fertility parameters.

Assume that a random variable $Z=Z_t$ follows a Brownian motion, but that it is absorbed as soon as it reaches a ceiling a. The transition density of the process at time t is (Borodin and Salminen 1996, 107)[3]

$$f(z_t; z_0, a) = \frac{1}{\sigma\sqrt{2\pi t}} \left\{ \exp\left(-\frac{(z_t - z_0)^2}{2\sigma^2 t} \right) - \exp\left(-\frac{(z_t - (2a - z_0))^2}{2\sigma^2 t} \right) \right\}, \qquad z_t, z_0 < a, \qquad (1)$$

where z_0 is the starting position at time $t=0$ of the random variable, and σ^2 is the variance of the innovation process. When a approaches infinity, the second exponential vanishes, and the density approaches that of a normal variable with expectation z_0 and variance $\sigma^2 t$, as expected.

---

[2] The maximum TFR-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.

[3] Note that, in contrast to Borodin and Salminen, we have introduced a *ceiling* instead of a *floor* (reversing the inequalities), and a general process variance equal to σ^2 instead of a standard variance equal to one (substituing $\sigma^2 t$ for t).

We want to compute the expected value and the variance of Z_t as a function of the ceiling a, given that the process has not been absorbed.

At time t there are two possibilities: the process moves freely on the line $-\infty < z_t < a$, or it is absorbed at the ceiling a. The probability that we encounter the first situation is $\int_{-\infty}^{a} f(z_t; z_0, a)dz$. Using expression (1), it can be shown that this integral equals

$$\int_{-\infty}^{a} f(z_t; z_0, a)dz = 2\Phi\left(\frac{a - z_0}{\sigma\sqrt{t}}\right) - 1, \tag{2}$$

where $\Phi(u) = \Pr[U \leq u]$ denotes the cumulative distribution of a standard normally distributed variable U. The probability that the process is absorbed at a is the complement of this expression, viz. $2\{1-\Phi(.)\}$, and the transition density for the event that the process is in z_t, given that it has *not* been absorbed, is $g(z_t;z_0,a)=f(z_t;z_0,a)/[2\Phi(.)-1]$. Thus, expressions (1) and (2) give us the probability density of g. Its expected value variance may be computed on the basis of the moment generating function $M(s) = \int_{-\infty}^{a} e^{sz} g(z_t; z_0, a)dz$. Writing u for $(a-z_0)/\sigma\sqrt{t}$, one may readily verify that the expected value of $g(z_t;z_0,a)$ is

$$E_g(Z_t; z_0, a) = M'(0) = \frac{z_0 - 2a\{1 - \Phi(u)\}}{2\Phi(u) - 1} = a - \frac{a - z_0}{2\Phi(u) - 1}. \quad a < \infty \tag{3}$$

For large a, the probability of non-absorption $(2\Phi(u)-1)$ approaches 1, so that the expected value approaches z_0. For finite a, $2\Phi(u)-1<1$, and the expected value is lower than z_0. $\Phi(u)$ falls when t increases, so that the expected value becomes progressively lower than z_0. All these findings are as one would expect.

After some tedious but straightforward algebra one finds for the second moment of $g(z_t;z_0)$

$$M''(0) = (z_0^2 + \sigma^2 t) - 4a(a - z_0)\frac{1 - \Phi(u)}{2\Phi(u) - 1} + \frac{2(a - z_0)}{2\Phi(u) - 1} \cdot \sqrt{\frac{\sigma^2 t}{2\pi}} \exp\left(-\frac{u^2}{2}\right), \qquad a < \infty$$

so that its variance equals

$$\text{Var}_g(Z_t;z_0,a) =$$

$$(z_0^2 + \sigma^2 t) + \left\{ -4a(a-z_0)\frac{1-\Phi(u)}{2\Phi(u)-1} - \left(a - \frac{a-z_0}{2\Phi(u)-1}\right)^2 \right\} + \frac{2(a-z_0)}{2\Phi(u)-1}\cdot\sqrt{\frac{\sigma^2 t}{2\pi}}\exp\left(-\frac{u^2}{2}\right). \quad (4)$$

For large a, $\Phi(u)$ approaches one, while at the same time the exponential vanishes. In that case the second moment is close to $z_0^2+\sigma^2 t$, and the variance approaches $z_0^2 + \sigma^2 t - z_0^2 = \sigma^2 t$, which is the variance of a random walk.

How do the expected value in expression (3) and the variance in expression (4) change when the ceiling a changes? For fixed z_0, t and σ^2, the expected value increases monotonically with growing a. It is lower than the starting value z_0, but approaches that value asymptotically, when a is increased. The variance in expression (3) is the sum of three terms, T_1, T_2, and T_3, say. $T_1 = z_0^2 + \sigma^2 t$ represents the second moment of a random walk - it is independent of the ceiling a. T_2 is written in curled brackets, and T_3 contains the exponential function. Figure 1 illustrates how T_2, T_3, and the variance depend on a, for the case $z_0=0$, $t=1$, and $\sigma^2=1$.

[Figure 1. Variance of a Brownian motion with ceiling a, and its components]

The figure shows that the variance is S-shaped. For small a, it starts a little under .5, and it approaches $\sigma^2 t=1$ asymptotically when a grows.[4] The reason is that, for large enough a, the term T_2 approaches $-z_0^2=0$, and T_3 dies out.

Figure 1 holds for $t=1$ only. When time increases, the S-curve is shifted upwards, and at the same time it is stretched to the right. For very small yet positive a, the variance (by L'Hopital) is close to $t(2-\pi/2)$, and for large a it approaches t (still assuming $z_0=0$ and $\sigma^2=1$), which confirms the earlier finding. Empirically I found that for a given ceiling a, at time $t \approx a$, the variance is close to its asymptotic level t.

5. Integrated random walk

The next step is to derive the expected value and variance for the process $Y_t = \exp(C_t)$. From now on I shall work in discrete time, and speak of "random walk" instead of "Brownian motion". C_t is the integrated random walk, so that $Z_t = C_t - C_{t-1}$. The process Z_t has expected value $E_g(Z_t; z_0, a) = E(Z_t)$, and variance $\text{Var}_g(Z_t; z_0, a) = \text{Var}(Z_t)$, given by expressions (3) and (4) respectively. C_t has expectation $E(C_t) = c_0 + \Sigma_t E(Z_t)$, so that the expectation of Y_t equals $E(Y_t) = \exp[E(C_t) + \frac{1}{2} \text{Var}(C_t)]$. Since the Z-process has independent innovations, the covariance $\text{Cov}(Z_t, Z_{t+h})$, $h \neq 0$, is zero. Thus C_t has variance $\text{Var}(C_t) = \Sigma_t \text{Var}(Z_t)$, and, by the Delta method, Y_t has variance approximately equal to $\text{Var}(C_t).\exp(2E(C_t))$.[5]

Choosing an appropriate value for the ceiling a poses a problem. Assume that the process Y_t starts in a known point y_0, and that it has a ceiling $b > y_0$. Then $c_0 = \ln(y_0)$, and the ceiling for C_t is $\ln(b)$. This means that at time $t=1$, the increment Z_1 should not exceed $a_1 = (\ln(b) - c_0)$, at time $t=2$, Z_2 should not exceed $a_2 = (\ln(b) - C_1)$, and so on. In general, Z_t should not exceed $a_t = (\ln(b) - C_{t-1})$. In this set-up, the ceiling becomes a random variable, so that expressions (3) and (4) for the expected value and the variance of Z_t no longer can be used; instead one should write

$$E_g(Z_t; z_0, A_t) = E_A\left(\frac{z_0 - 2A_t\{1 - \Phi(U_t)\}}{2\Phi(U_t) - 1}\right) = E_A\left(A_t - \frac{A_t - z_0}{2\Phi(U_t) - 1}\right) \qquad A_t < \infty \qquad (3')$$

and

$$\text{Var}_g(Z_t; z_0, A_t) =$$

$$(z_0^2 + \sigma^2 t) + E_A\left(\left\{-4A_t(A_t - z_0)\frac{1 - \Phi(U_t)}{2\Phi(U_t) - 1} - \left(A_t - \frac{A_t - z_0}{2\Phi(U_t) - 1}\right)^2\right\} + \frac{2(A_t - z_0)}{2\Phi(U_t) - 1}\cdot\sqrt{\frac{\sigma^2 t}{2\pi}}\exp\left(-\frac{(U_t)^2}{2}\right)\right),$$

with $U_t = \dfrac{A_t - z_0}{\sigma\sqrt{t}}$.

---

[4] When $a=0$, the ceiling coincides with the starting point z_0, and we have a degenerated process.

[5] When a random variable Y is a function f of some other random variable X, then approximate expressions for the expected value and the variance of $Y = f(X)$ are $E(Y) \approx f(\xi) + \frac{1}{2}f''(\xi).\text{Var}(X)$, and $\text{Var}(Y) \approx [f'(\xi)]^2\text{Var}(X)$, where $\xi = E(X)$.

Next, the Delta method may be invoked to compute the expected value of these two highly non-linear functions of A_t. This requires the second derivative of the expected value and of the variance, see footnote 5. However, preliminary experimentation showed that the second derivative of the expected value is of the order of magnitude of $u_t.exp(-u_t^2/2)$, which is small over a large range of u_t. Thus this second derivative may be ignored in practice, and one can simply replace, in expression (3'), A_t by its expected value $\ln(b)-E(C_{t-1})$. The same is true for the variance of Z_t, which is dominated by the non-random part $z_0^2+\sigma^2 t$. A second problem is that subsequent values of Z_t are negatively correlated, due to the time-dependent ceiling a_t. When this autocorrelation is ignored, for instance because $E(C_{t-1})$ is small compared to $\ln(b)$, the variance of C_t is overestimated.

6. Application

I have analysed the mean value and the variance of the predicted TFR for Norway in the period 1996-2050, using the expressions given in (3') for Y_t. First differences in log-values of the TFR were assumed to follow a random walk, with innovation variance equal to $\sigma^2=0.000703$ (Keilman and Pham (2000), Table 3.3). The starting value z_0 was $\ln(\text{TFR}_{1995}/\text{TFR}_{1994}) = \ln(1.8700/1.8704) = -0.00021$.

For large values of b, the process Z_t is an unconstrained random walk, and Y_t is the exponential of an integrated random walk. In that case, predictions for both Z_t and its integrated process C_t are normally distributed; hence prediction intervals for Y_t are straightforward, once the mean and the variance of Z_t and C_t have been computed. Prediction intervals for Y_t computed this way may be compared with those based on the ARIMA (1,1,0) process of Section 1. Since the AR(1)- coefficient of the latter process was estimated to be 0.67, the variance of this process Y_t grows much slower with time than the variance of the exponential cumulated random walk Y_t of this section. Thus the prediction intervals of the random walk-based process grow faster than those of the ARIMA-based process. Yet I found that the prediction intervals for the two processes agreed quite well in the short run. For instance, the 67% interval and the 95% interval of the random walk-based process in 2010 were equal to [1.39, 2.49] and [1.05, 3.29], respectively; those of the ARIMA-based process were slightly narrower: [1.42, 2.46] and [1.09, 3.21], respectively. In the medium and long run however, the random walk-based process resulted in prediction intervals that were much wider than those of the ARIMA-based process, for the reason just mentioned. Stated

differently, the innovation variance σ^2 of the random walk-based process had to be reduced from 0.000703 to 0.00032 and 0.00021, in order to produce comparable prediction intervals in 2030 and 2050, respectively.

I have analysed the expected value and the standard deviation of Y_t as a function of the ceiling b, for the short-term (2010), the medium-term (2030), and the long-term (2050) situation. Thus I computed the two statistics for the Y_t-process in these three years based on the assumptions σ^2=0.000703, σ^2=0.00032, and σ^2=0.00021, respectively. Figure 2 shows the results. Very remarkably, values for the ceiling b higher than approximately 2.7 children per woman have little or no impact on the expected value and the standard deviation of Y_t. This is explained by the fact that, even for the largest value of the innovation variance σ^2 (i.e. σ^2 =0.000703) the probability of non-absorption ($2\Phi(u)$-1) 55 years ahead is .97 or more, as soon as b exceeds 2.3. Even when b is as low as 2.0, the probability of non-absorption is still .88 at time t=4, and higher for shorter and longer durations. Hence the bound b is hardly effective.

[Figure 2. Expected value and standard deviation of Y_t as a function of the ceiling b]

It should be noted that the findings reported here are quite sensitive for the choice of the initial increment z_0. If z_0 is increased from $\ln(1.87/1.8704)$ = -0.00021 to $\ln(1.87/1.86)$ = 0.00536, say, the threshold b, i.e. that value of b beyond which the standard deviation of Y_t hardly changes, moves up from 2.7 to 3.1 children per woman. More formally, the impact of z_0 can be analysed as follows. The standard deviation of Y_t depends on the ceiling b: $\mathrm{sd}Y_t = \mathrm{sd}Y_t(b)$, but for large enough b it is independent of the ceiling (in practice for b>6). The latter standard deviation may be called the equilibrium value. For a certain level of b, the standard deviation is at 95 per cent of its equilibrium value. Denote this threshold level of b as b^*. For instance, in Figure 2, the equilibrium level of the standard deviation in 2010 (σ^2 =0.000703) is 0.54. At $b=b^*$=2.25, the standard deviation is 0.95 per cent of that equilibrium value, or 0.51. Table 1 illustrates that the threshold level b^* increases by 0.3 to 0.9 children per woman when z_0 increases by 0.005. The acceleration is stronger for small innovation variance. Given the fact that the Norwegian TFR in 1995 was 1.87, a z_0 equal to 0.01 would imply a TFR of 1.85 in 1994. Thus for small increases in the TFR, the threshold b^* moves fast to higher levels, although it remains below 4 children per woman..

Table 1. Threshold levels b^* and corresponding standard deviation values of Y_t for various choices of the initial increment z_0

	$\sigma^2=0.000703$ (2010)				$\sigma^2=0.00021$ (2050)			
z_0	-0.005	0	0.005	0.01	-0.005	0	0.005	0.01
b^*	2.3	2.6	3.1	3.6	2.0	2.4	3.0	3.9
$\mathrm{sd}Y_t(b^*)$	0.99	1.18	1.41	1.68	0.77	1.01	1.33	1.75

7. Conclusion

The simulations for the ARIMA model mentioned in Section 2 suggested that a ceiling equal to 4 children per woman is appropriate. In this paper I have demonstrated that when the underlying process is a random walk, i.e. an AR(1)-proces with autoregressive coefficient equal to one, choosing a ceiling equal to between 3 and 4 children per woman will not affect the expected value and the standard deviation of the predicted TFR. Whether the level of the ceiling b influences the bounds of the predictive interval is a question that could not be answered here: when the ceiling b becomes effective, the predictive distribution of C_t is no longer normal, and the form of the distribution is unknown.

References

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

Borodin, A.N. and P. Salminen. 1996. *Handbook of Brownian Motion: Facts and Formulae.* Basel: Birkhauser Verlag.

Keilman, N. and A. Hetland (1999) "Simulated confidence intervals for future period and cohort fertility". Working Paper No. 6, Joint ECE-Eurostat Work Session on Demographic Projections, Perugia, Italy, May 1999.

Keilman, N. and D.Q. Pham (2000) "Predictive intervals for age-specific fertility". *European Journal of Population.***16**(1): 41-66.

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.

Lesthaeghe, R. and J. Surkyn. 1988. "Cultural dynamics and economic theories of fertility change", *Population and Development Review* 14(2): 3-45.

Figure 1. Variance for Brownian motion with ceiling a, and its components
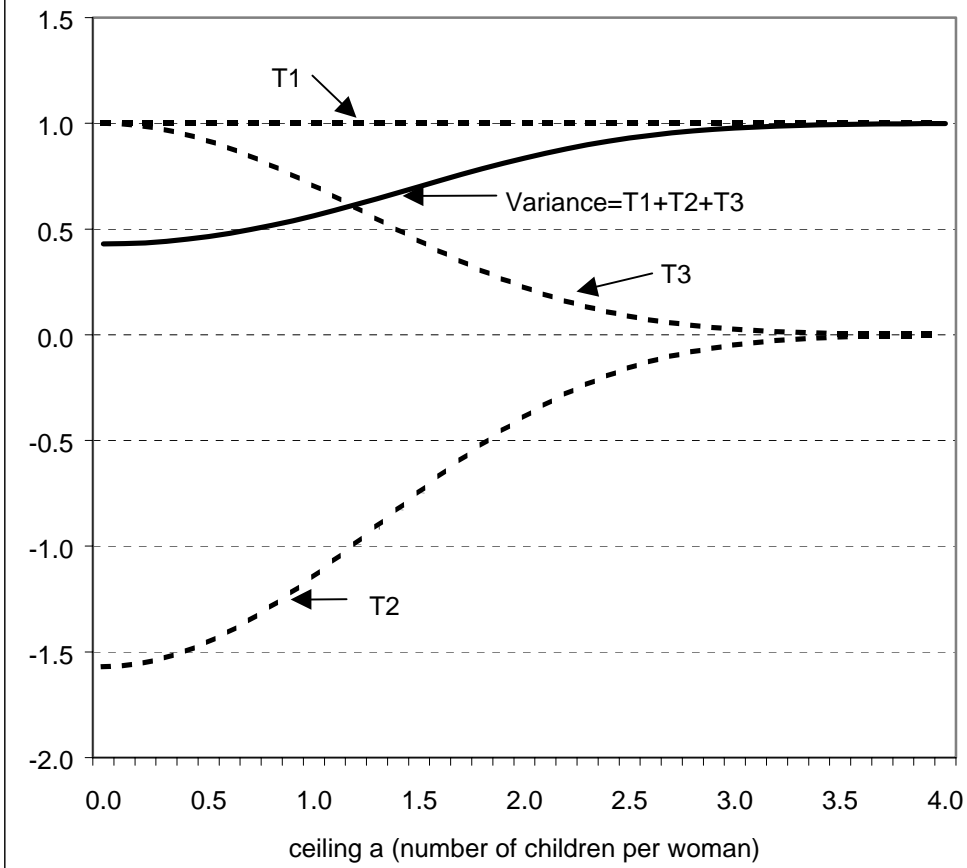
Figure 2. Expected value and standard deviation of Y(t), as a function of ceiling b