Marriageability and divorceability : a simulation of the unobservables through the conditional gaussian diffusion process

FERNANDO RAJULTON

Vrije Universiteit Brussels

IPD-Working Paper 1986-1

Acknowledgements : My sincere thanks to Dr. Anatoli Yashin for the fruitful discussions on the model and for practical guidance in simulation exercises. Thanks are also due to the IPD-staff, VUB, Brussels for the cordial support and computer facilities not easily available elsewhere.

ORDERING INFORMATION

Extra copies of this working paper are available from :

Mr. E. Vanden Balck
Secretary/Librarian
Interuniversity Programme in Demography
c/o Centrum voor Sociologie
Vrije Universiteit Brussel
Pleinlaan 2
B-1050 Brussels
BELGIUM

at the cost of US \$ 5.00 or BF 250. Please attach check or international money order.

ORDER FORM

Name ....
Address ....
....
....
....
....

wishes to receive ...... copies of Working Paper Nr..........
and includes payment of US \$ .... or BF........

INTRODUCTION

Characteristics that heterogenize a population are many. A few are observables, many are not. Making use of observed characteristics, mostly expressed as covariates, analytical tools have long since been developed to study the impact of these covariates on human behaviour. The well-known hazard model, in its various forms (Cox, 1972, 1975; Cox and Oakes, 1984; Miller et al., 1981; Tuma and Hannan, 1984; Vanderhoeft, 1985), is one such analytical tool in survival analysis. Recent researches, however, have revealed that an analysis of observed characteristics is to a certain extent subject to the assumption of the kind of distribution of the unobservables (Heckman and Singer, 1982; Ridder and Verbakel, 1983).[1]

The relevance of unobservable characteristics can be seen through an example. Consider the notion of "frailty" of individuals - a notion introduced by Vaupel, Manton and Stallard (1979) in a study of mortality selection. Frailty at individual level is an unobservable, and yet it is of common knowledge that frailer individuals "die" first, leaving the population to consist of more and more "robust" individuals. Such a population will have patterns of "mortality" different from the ones derived without taking into account the individual heterogeneity in frailty.

The awareness that a consideration of unobserved heterogeneity at individual level would portray reality more adequately led to

1. One could think of, as the most familiar example, the assumption of the normally distributed error terms in a (multiple) regression analysis or of the binomially distributed error terms in a logit analysis.

including the distribution of the unobservables in an analysis of observed characteristics. However, it has been brought to light in recent times that a mere assumption of the kind of distribution of the unobservables without an explicit consideration of the relationship that could exist between the "observed" mortality rate and the parameters of the underlying unobserved characteristics would only lead to wrong inferences.[2] The random walk model of human mortality proposed by Woodbury and Manton (1977) which considered the influence of the randomized physiological characteristics (or variables) such as serum cholesterol, blood pressure,etc. on human mortality served as an example of a correct step towards building a model which purports to examine the influence of the unobservables. Their pioneering work was developed further to a more general model based on the conditional Gaussian diffusion processes, which explicitly formulated the mathematical relationship between the "observed" mortality and the parameters of the underlying physiological variables in terms of the changes over time in the means and covariances of the physiological variables (Yashin, Manton and Vaupel, 1985). This model, known as the conditional Gaussian diffusion model, covers the possibility of analysing the effects of both the observed and the unobserved or partially observed characteristics in their multivariate form (Yashin, 1984; Yashin and Manton, 1984; Yashin, Manton and Stallard, 1985b; Yashin, 1985).

The present study does not intend to give a detailed description of this dynamic model nor its historical development. Readers

---

2. Heckman and Singer (1982), for example, showed how entirely different inferences can be made with the estimated parameters of observed characteristics, depending on the kind of distribution of the unobservables. But they failed to consider the relationship in question and were led to have recourse to non-parametric estimation procedure.

interested in the subject will have a plethora of mathematical studies given in the Bibliography. On the contrary, this study intends to show the flexibility of the model in its application to various analytic contexts. One such context of practical interest, namely marriageability, is discussed in detail in this paper; and another, namely divorceability, is shown to be a straightforward extension - both in the light of changes taking place in marital behaviour especially in the developed countries. Relevant notions and a viable measure (though incomplete) of marriageability and divorceability are introduced in SECTION 1. The model specification based on the ideas of the conditional Gaussian diffusion process, which allows an analysis of heterogeneity dynamics of these unobservables, is explained in brief in SECTION 2. And SECTION 3 gives an illustrative example of a simulation carried out for the case of marriageability with the ideas developed in SECTION 2. The last section points out the extension to divorceability and other demographic variables of interest. Substantive interpretations are given as and when circumstances lend themselves to.

SECTION 1 : MARRIAGEABILITY

Consider a number of individuals (of the same sex) forming a cohort of the single at age 15. Individuals transit to the married state at different ages. Their transit to the married state is comparable to "death" (note that the first marriage is a non-renewable event) and the time of their transition is comparable to "death-time" - the usage in Survival Analysis.

Several researches in the past have attempted to explain the process of entry into first marriage. It is a process that depends on the number of marriageables in the cohorts of the same sex as well as of the opposite sex (the well-known two-sex problem in demography), the marriageability of individuals being determined by various cultural norms as well as by certain demographic determinants such as age and cohort size. Louis Henry (English version 1972) initiated the study of this process through the concepts of "circles" and "interactions" between the marriageables of either sex. Coale (1971) discovered two different aspects of this process: (a) in a cohort of the unmarried, some are destined never to marry, and (b) the process of entry into first marriage can be described by a double exponential which is a convolution of a normally distributed entry into marriage market and an exponentially distributed delay. In an independent study at the same time, Hernes (1972) interpreted the process in terms of two sociological forces that operate on the unmarried: (a) an increasing percent of the cohort already married and (b) the capacity of individuals to marry (which is referred to, in this study, as marriageability).

These studies had the basic notion of marriageability of the marriageables, either implicitly or explicitly, which corresponds to the notion of "frailty" mentioned above. Most eligible candidates marry early, leaving a pool of the less and less eligible not yet married. As the most eligible leave the cohort of the single, the less eligible may become more eligible or even lesser eligible, depending on the availability of partners, for example. Some individuals, on the other hand, whatever be their initial marriageability or however dynamically their marriageability may change over time, are destined never to marry. Heterogeneity in marriageability is thus a subject worth analysing, provided some information is available.

Of the three studies mentioned above, that of Hernes incorporates an explicit treatment of the unobservable marriageability and estimates its average at the initial time when the cohort is formed. A brief review of the Hernes model will be helpful for further analysis presented in this study. Hernes based his arguments on two sociological forces that impinge on the single. The first is the <u>social pressure</u> to marry: this pressure increases with the increasing percent of the cohort already married, as the "psychological experience of the undesirability of celibacy increases as more of the same age group enter into wedlock". The second force is the <u>individual capacity</u> which, in accordance with a two-fold assumption made by Hernes for the sake of a simple model building, <u>declines with age</u> - "He that marries late marries ill", says an English proverb -, this decline being <u>the same</u> for all individuals. Thus, a consideration of the two social forces coupled with the two-fold assumption led Hernes to formulate a simple model[3]

$$P_t = \frac{1}{1 + \frac{1}{k a^{b^t}}} \qquad (1)$$

where P_t = the proportion of the cohort already married at time t;

$\log a = A / \log b$, A being the average initial marriageability;

$b(<1)$ = the constant of deterioration in marriageability

and $k = P_0 / [a(1-P_0)]$

---

3. Certain points are to be noted in an application of this model: (1) P_0 is mathematically well-defined and is not equal to zero.

contd.

This model is, in fact, a type of diffusion models and depends on the three parameters a, b and k. As the expression ka^{b^t} is of a Gompertz form, these three parameters can be estimated with the well-known 3-point procedure used in fitting a logistic curve.

Curiously enough, the type of diffusion model such as this can be applied to various processes which can be explained by similar arguments based on diffusion.[4] Rajulton (1985a) extended the model to cover the cases of proportions divorced and proportions remarried on similar lines of arguments on social forces - hence, the notions of "divorceability" or "remarriageability", as the case may be, of relevant cohorts formed at different ages.

Table 1 presents the observed and estimated proportions[5] of the married and divorced females in Belgium in the year 1981 and the estimated parameters of the model through the 3-point procedure

---

Practically speaking, this means that while fitting the data, the first year of the process would be taken as $t_0 = 0$.

(2) The asymptote of the curve P_t is given by

$$\lim_{t \to \infty} P_t = \frac{1}{1 + \frac{1}{k}}$$

as b is less than unity.

(3) In the expression of P_t, if one lets $g(t) = ka^{b^t}$, then $g(t)$ is a Gompertz function and the parameters a, b and k can be estimated through the 3-point fitting procedure (Croxton et al., 1967). A disadvantage of this fitting procedure is that the estimates will be dependent on the (arbitrary) number of observations considered. In the case of first marriage, for example, 36 observations or even less, would be sufficient as the number of marriages after age 50 (15+36) would be negligible.

4. Hernes himself came out with the same model to explain the diffusion of television sets in Norway (1978).

5. These are in fact the first passage probabilities of transition from the never married to married state and from the presently married state to divorced state obtained through a semi-Markovian multistate analysis (Mode, 1982; Rajulton, 1985a) which considers the effect of duration in the present state on the probabilities of transition to other states.

Table 1: The observed and estimated proportions + married and divorced, females, Belgium, 1981 - the estimates obtained through the Hernes model with 30 observations

Age x	parameters of Hernes' model	FIRST MARRIAGE duration t	obs. prop.	est. prop.	DIVORCE duration t	obs. prop.	est. prop.
	a			.000390			.004766
	b			.813509			.867135
	k			7.101772			.288779
	ability A			1.619981			.762157
	asymptote			.876570			.224072
15		1	.0020	.0028		.0000	.0000
16		2	.0107	.0118		.0000	.0000
17		3	.0332	.0379	1	.0014	.0014
18		4	.1009	.0940	2	.0027	.0028
19		5	.1973	.1858	3	.0051	.0052
20		6	.3186	.3023	4	.0088	.0088
21		7	.4450	.4220	5	.0138	.0139
22		8	.5508	.5274	6	.0200	.0206
23		9	.6333	.6118	7	.0265	.0289
24		10	.6933	.6760	8	.0394	.0387
25		11	.7363	.7239	9	.0518	.0497
26		12	.7661	.7595	10	.0640	.0616
27		13	.7893	.7860	11	.0764	.0740
28		14	.8061	.8060	12	.0902	.0866
29		15	.8202	.8211	13	.1017	.0990
30		16	.8307	.8328	14	.1121	.1111
31		17	.8389	.8418	15	.1218	.1225
32		18	.8455	.8488	16	.1312	.1333
33		19	.8513	.8544	17	.1411	.1433
34		20	.8555	.8587	18	.1488	.1524
35		21	.8598	.8622	19	.1574	.1607
36		22	.8628	.8650	20	.1650	.1682
37		23	.8655	.8672	21	.1719	.1749
38		24	.8676	.8690	22	.1784	.1809
39		25	.8695	.8705	23	.1847	.1863
40		26	.8717	.8716	24	.1904	.1910
41		27	.8734	.8725	25	.1947	.1952
42		28	.8750	.8733	26	.1990	.1988
43		29	.8763	.8739	27	.2032	.2021
44		30	.8776	.8744	28	.2067	.2049
45		31	.8788	.8748	29	.2101	.2074
46		32	.8799	.8751	30	.2131	.2095
47		33	.8808	.8754	31	.2161	.2114
48		34	.8819	.8756	32	.2185	.2131
49		35	.8827	.8758	33	.2205	.2145
50		36	.8833	.8759	34	.2225	.2158
Mean ($\bar{T}$)			8.5696	8.5083		17.6325	17.1254
Variance (σ_T^2)			21.9567	17.2781		56.5138	49.9727

+ Refer Footnote 5.

with 30 observations (from age 15 to age 44 in the case of first marriage and from age 17 to age 46 in the case of divorce). Note the close correspondence between the observed and estimated proportions, particularly up to t =30. In the case of proportions married, 24 or 27 observations would have yielded estimates much closer to the observed ones but with slight variations in the estimates of the parameters. Use of 30 observations seems to give the best estimates in the case of proportions divorced. Note also that the capacity to marry or to divorce and the constant of deterioration should be non-negative; the former can be greater than unity, but the latter always less than unity. Table 1 presents also the mean and variance of transition-time (which are also mean and variance of duration spent in the state of origin).

The main idea of this section was to show that, given the proportions married of a cohort for each age x, there is a model through which marriageability - which cannot be observed at individual level - can be estimated at cohort level at intial time t_0. That this average measure can serve as a point of departure, along with other measures given in Table 1 (such as mean time of marriage, its variance and the observed proportions) for a dynamic analysis will be shown in SECTION 3. It should be noted, however, that unlike the Hernes model, a dynamic analysis does not need to make the two-fold assumption that the initial marriageability declines with age and that this decline is the same for all individuals. But the two social forces that explain the process (of entry into marriage or divorce) can be retained in the dynamic model.

SECTION 2 : A DYNAMIC MODEL - THE CONDITIONAL GAUSSIAN DIFFUSION PROCESS OF FIRST MARRIAGE

The recently developed conditional Gaussian diffusion model is a helpful tool for analysing a cohort's heterogeneity in marriageability. The generality of the model admits the necessary modifications concerning the two social forces which were shown to account well for the process of first marriage: the socio-psychological pressure exerted on an unmarried individual by the increasing percent of his/her peers already married and his/her own capacity to marry. The model with the necessary modifications is given in brief in this section. The next section will outline a procedure for simulating the unobservable marriageability of individuals with a set of parameters such that the end-results (marriage-times of individuals) will correspond very closely to the known data on marriage-times.

The dynamic model under consideration - Gaussian diffusion process - has two key components:[6]

(a) <u>linear dynamics</u> - changes in the unobserved variable (marriageability here) are assumed to follow a linear autoregression; and

(b) <u>quadratic dependency</u> - the mathematical relationship between the observed hazard rate (without considering individual heterogeneity in marriageability) and the values of the unobserved variable can be expressed as a quadratic function.

---

6. These two components have been found to describe adequately the human mortality related to physiological changes (which are either unobserved or at the most partially observed) in a number of epidemiological studies of chronic disease. As the process of first marriage is comparable to a mortality analysis, the influence of the unobserved variable (marriageability) can be studied in a similar fashion.

Suppose that the marriage rate at time t for the i-th individual in a cohort of n unmarried individuals depends on a process y_{it} which evolves over time. Assume that the process y_{it} evolves independently of all other individuals but depends on the cohort's cumulative proportion of the already married (P_t) at time t. Note that P_t is not an individual variable, but a cohort variable (or an environmental variable) which can be observed. For the sake of simplicity in notations, the suffix i denoting the i-th individual will be dropped hereafter, unless necessary.

In accordance with the first component of the diffusion process, y_t is assumed to satisfy the linear diffusion-type stochastic differential equations:

$$dy_t = (a_{0t} + a_{1t}\,y_t)\,P_t\,dt + b_t\,dW_t \qquad (2)$$

where a_{0t}, a_{1t} and b_t are well-bounded functions and W_t is a Wiener process which does not depend on the initial condition y_0 which is Gaussian distributed with known mean m_0 and variance v_0. And, in accordance with the second component, the marriage rate for an individual, denoted by $m(t, y_t, P_t)$ is assumed to be a quadratic function of the set of values of y_t :

$$m(t, y_t, P_t) = (m_{0t} + m_{1t}\,y_t + m_{2t}\,y_t^2)\,P_t \qquad (3)$$

Let the observed (unconditional) rate be denoted by $\bar{m}(t, P_t)$. The relationship between the observed rate and the individual (conditional) rate $m(t, y_t, P_t)$ described by the conditional Gaussian diffusion process can be expressed in the form:

$$\bar{m}(t, P_t) = \mathcal{E}[\, m(t, P_t) \mid T_i \geq t \,] \tag{4}$$

where $\mathcal{E}$ denotes the mathematical expectation,[7] and T_i is the marriage-time (time of transition from the unmarried to the married state) of the i-th individual associated with the rate $m(t, y_t, P_t)$. When y_0 follows Gaussian with mean m_0 and variance v_0 and when $m(t, y_t, P_t)$ is quadratic as expressed in (3), then $\bar{m}(t, P_t)$ is the age-specific marriage rate among the <u>not-yet-married</u> (that is, "survivors") at time t. $\bar{m}(t, P_t)$ can be shown to satisfy the following relationship to the parameters of the distribution of y_t:

$$\bar{m}(t, P_t) = \left[m_{0t} + m_{1t}\, m_t + m_{2t}(m_t^2 + v_t) \right] P_t \tag{5}$$

where m_t and v_t satisfy the non-linear differential equations:

$$\frac{dm_t}{dt} = \left[a_{0t} + a_{1t} m_t - v_t m_{1t} - m_{2t} v_t m_t \right] P_t \tag{6}$$

$$\frac{dv_t}{dt} = \left[2a_{1t} v_t + b^2 - 2v_t^2 m_{2t} \right] P_t \tag{7}$$

with the initial conditions m_0 and v_0.

Three observations are in order:

(a) Note that y_t follows Gaussian with mean m_t and variance v_t. y_t can take negative values and cannot, therefore, be considered describing the marriageability process. However, one can let y_t evolve as specified above and consider marriageability as equal to y_t^2.

---

7. In other words, the observed rate is the conditional expectation of individual rates. In Survival Analysis, this expectation needs the concept of averaging an exponent, as any exponent can be considered as a conditional survival function. The mathematical expectation of an exponent which is a functional of a Wiener process has a long history of development. See Yashin (1984) for details. More recently, Yashin has shown that if the functional is of a quadratic form, the expectation of the exponent using the conditional Gaussian property has many advantages over the earlier techniques.

(b) The above specifications of the conditional Gaussian model, which has taken into account the first force off the process of entry into marriage, can be simplified for the sake of practical application in the next section. The equations (2) through (7) can be simplified by letting

(i) m_{0t} and m_{1t} in Eq.(3) be zero

(ii) m_{2t}, a_{0t}, a_{1t} and b_t be constant over time

whereupon,[8]

$$dy_t = (a_0 + a_1 y_t) p_t \, dt + b \, dW_t \tag{2a}$$

$$m(t, y_t, p_t) = m \, p_t \, y_t^2 \tag{3a}$$

$$\bar{m}_t = m \, p_t \, (m_t^2 + v_t) \tag{5a}$$

$$\frac{dm_t}{dt} = (a_0 + a_1 m_t - m \, v_t m_t) \, p_t \tag{6a}$$

$$\frac{dv_t}{dt} = (2 a_1 v_t + b^2 - 2 m \, v_t^2) \, p_t \tag{7a}$$

(c) The relationship between the observed rate and the underlying marriageability process expressed in Eq.(5) or (5a) can be used to develop the likelihood function which in turn can be used to estimate the parameters of the process if the distribution of marriage-times is known. The likelihood function would be as follows:

---

8. Eq.(3a) is analogous to the individual level analysis which Hernes considered initially in developing the model- see his Eq.1. The rate of change in the probability of the i-th individual's getting married is equal to the proportion married multiplied by a parameter of conversion; that is, $\frac{dp_i}{dt} = qP_t$

Or again, Eq.(5a) can be contrasted to the observed rate $\bar{M}_t$ which can be derived from the Hernes model; that is, $\bar{M}_t = Ab^t P_t$

$$\mathcal{L} = \prod_{i=1}^{n} \bar{m}_{T_i} \exp\left[-\int_0^{T_i} \bar{m}_s \, ds \right]$$

where T_i is the marriage-time of the i-th individual.

And hence,

$$\log \mathcal{L} = \sum_{i=1}^{n} \left[\log m\, P_{T_i} (m^2_{T_i} + v_{T_i}) - \sum_{s=0}^{T_i} m . P_s (m_s^2 + v_s) \right]$$

Some guidelines for estimating the parameters through the maximum likelihood or minimization computer programs can be referred to in Yashin and Rajulton (1985).

If efficient maximum likelihood or minimization programs are not available, or even if available cannot be easily handled, the simulation procedure described in the next section can be of some help. With the available information on the initial average marriageability, mean and variance of marriage-times, a simulation of marriage-times can be carried out in such a manner that these correspond very closely to the observed ones - whereby the optimal estimates of the parameters a_0, a_1, b and m will be indicated. This simulation exercise has an advantage over that of optimization through computer programs in the sense that it reveals how, for changing parameters, "convergence" takes place towards the observed behaviour. However, it depends on the availability of random number generators (of uniform and Gaussian distributions).

Before entering into this simulation exercise, an interpretation of the parameters in question deserves one's attention. As the dynamics of marriageability is assumed to be linear, the interpretations of these parameters are comparable to the usual ones adduced to the parameters of any linear regression model. a_0 denotes the "drift" or the systematic change in mean values; a_1 denotes the "regression effect" or the convergence to mean values (due perhaps to some homeostatic tendencies in the unobserved variable); and b denotes the "diffusion effect" or divergence due to random influences (cf. Woodbury and Manton, 1977). Moreover, the observed marriage rate at time t is the average of individual rates conditional on not being married up to time t. The individual rates at time t are expressed as products of three quantities:

p_t - the proportion married at time t (a cohort variable)

y_t^2 - marriageability at time t (an individual variable)

and m - the marriage parameter (which is the same for all individuals and for all time t).

It is worth emphasizing here that, of the four parameters, m is the most important as it links the marriage rate not only at individual level but also at cohort level to the underlying unobservable process of marriageability and to the environmental (social) pressure - directly in the former case (see Eq.(3)) and indirectly through the parameters of the unobservable process in the latter (see Eq.(5)).

SECTION 3: A SIMULATION OF THE UNOBSERVABLE PROCESS OF MARRIAGEABILITY

A set of normally available data on proportions married of a cohort of the single provides valuable clues for the kind of simulation described in this section. The mean time of marriage ($\bar{T}$), its variance (σ_T^2) and the average initial marriageability (A) estimated through the Hernes model - all the three found in Table 1 - along with the mathematics of the conditional Gaussian diffusion processes help in simulating the unobservable process of marriageability over time.

To begin with, Eqq.(2a), (6a) and (7a) can be expressed in discrete form as follows:

$$dy_t = (a_0 + a_1 y_t) P_t \; dt + b \; dW_t \qquad (2a)$$

that is,

$$y_{t+\Delta} - y_t = (a_0 + a_1 y_t) P_t \Delta \quad + b \; dW_t$$

which can be written through the property of a Wiener process as

$$= (a_0 + a_1 y_t) P_t \Delta \quad + b\sqrt{\Delta} \; Z_{t+\Delta}$$

where $Z_t \sim N(0,1)$. Letting Δ=1, this reduces to

$$y_{t+1} = a_0 P_t + (1 + a_1 P_t) y_t + b \; Z_{t+1} \qquad (2b)$$

Similarly, expanding the terms of the differential equations (6a) and (7a),

$$m_{t+1} = a_0 P_t + (1 + a_1 P_t) m_t - M P_t m_t v_t \qquad (6b)$$

$$v_{t+1} = (1 + 2a_1 P_t) v_t + b^2 P_t - 2M P_t v_t^2 \qquad (7b)$$

Secondly, denoting by y_0^t the process up to time t, let $F(t)$ be the conditional distribution of marriage-time; that is,

$$F(t) = Prob[T \leq t \;|\; y_0^t] \qquad (8)$$

$$= 1 - exp[- \int_0^t M(s, y_s, P_s) \; ds] \qquad (8a)$$

In order to simulate the marriage-time of an individual, therefore, a uniform random number r is generated such that

$$F(t) = Prob\left[r \leq F(t)\right] = Prob\left[F^{-1}(r) \leq t\right] \qquad (8b)$$

For details on simulation, cf.Rubinstein (1981). This uniform random variate leads to the following expression through equations (8a), (8b) and (3a):

$$\begin{aligned} r &\leq 1 - exp\left[-\int_0^t m(s, y_s, P_s)\, ds\right] \\ &= 1 - exp\left[-\int_0^t m\, y_s^2\, P_s\, ds\right] \end{aligned} \qquad (9)$$

which implies

$$- Log\ (1-r) \leq \int_0^t m y_s^2\, P_s\, ds \qquad (9a)$$

If T is the marriage-time of an individual, then

$$T = inf\left[t : \int_0^t m y_s^2\, P_s\ ds \geq - Log(1-r)\right] \qquad (10)$$

In discrete formulation, this means that T is that time t for which the sum $\Sigma\, m y_s^2\, P_s$ becomes greater than or equal to $-Log\,(1-r)$. As this sum, however, requires the auto-regressive values of y_t^2 through eq.(2b) and of the marriage parameter m, the parameters a_0, a_1, b and m have to be fixed for carrying out the simulation. The coefficient b - denoting the "diffusion effect" and the random influences - can be arbitrarily fixed to be very small, say 0.001. We know that $y_0 \sim N(m_0, v_0)$, whose parameters are known. As marriageability is the square of this variate, from the theory of statistical distributions, y_0^2 is known to follow a non-central χ^2 with mean A and variance S^2 which are related to m_0 and v_0 as follows (cf. Lindgren, 1976; Kendall & Stuart, 1977; see also Appendix A):

$$\begin{aligned} A &= m_0^2 + v_0 \\ S^2 &= 2 v_0 (v_0 + 2 m_0^2) \end{aligned} \qquad (11)$$

A is known but S^2 is unknown; Eq.(11) cannot be solved directly. However, under stability conditions when the random disturbances are negligible, Eq.(2a) can be written as

$$\frac{dy_t}{dt} = (a_0 + a_1 y_t) P_t = 0 \tag{2c}$$

which yields the relation

$$y_t = - \frac{a_0}{a_1} = y \tag{12}$$

and hence, the mean-time of marriage would be (cf.Appendix B):

$$\bar{T} = \frac{1}{m y^2 P} = \frac{a_1^2}{a_0^2 m P} \tag{12a}$$

where y and P denote the values of the y_t and P_t in stability. P is known to be about 0.883 and $\bar{T}$ about 8.5696 from Table 1.

As $y = - \frac{a_0}{a_1}$, it would be preferable to give negative values to a_1, thus letting a_0 (denoting the systematic change in mean values) to be positive. As the mean values m_t will normally be around $\frac{a_0}{a_1}$ for initial and higher values of t,

$$m_0 = \frac{a_0}{|a_1|} \tag{13}$$

Therefore, if one fixes the other two unknown parameters m and a_1 arbitrarily for the sake of starting the simulation, a_0, m_0 and v_0 can be successively found through the Eqq.(12a),(13) and (11). A simulation carried out with the rough estimates of the parameters will result on a distribution of marriage-times whose mean and variance can be compared to the actual ones ($\bar{T}$ and σ_T^2). If they correspond closely, these estimates can be taken as the "best" ones; if they do not, new values of m and a_1 can be tried. Trials with various values of m and a_1 will indicate the direction and magnitude of these parameters.

The following guidelines may be of practical use:

(i) The rough estimates obtained through (arbitrary) specifications of $\mathcal{M}$ and a_1, can be used in (6b) and (7b) to examine whether m_t and v_t values converge or not; a divergence of means and (co)variances of the process is clearly out of question. It should be noted, however, that v_t might become negative for higher values of t - in which case, it would be practical to make it equal to zero. The values of m_t will normally be assured to be around $\frac{a_0}{|a_1|}$ especially at the initial and higher values of t.

(ii) For each set of rough estimates, the resulting mean and variance of marriage-times can be compared to the known values. Changes in $\mathcal{M}$ and a_1 will affect the values of mean and variance, thus indicating the direction towards the best estimates. See below for an illustration.

(iii) The final proportion married or remaining single has also to be taken into account; as far as possible, it has to correspond to the observed one. Note, however, that the estimated P_t values for each t are to be calculated within the simulation with the number of individuals who transit to the married state under the influence of the unobserved variable marriageability and the social pressure measured by the proportion married at time t.

To summarize, the algorithm for the simulation exercise is as follows (this algorithm and the above guidelines have been incorporated in the computer program SIMMAR given in Appendix C): The final proportion married (P) being known and fixing b to be very small, say, 0.001,

(1) fix m and a_1 (e.g. m around 1.00 and a_1 around -0.1)

(2) calculate a_0 through (12a)

(3) calculate m_0 through (13)

(4) calculate v_0 and S^2 through (11)

With these values of a_0, a_1, b, m, m_0, v_0, carry out the simulation for each time t and for each individual i as follows:

(5) generate a uniform random variate r_i

(6) calculate $-\log(1-r_i)$

(7) generate $y_0 \sim n(m_0, v_0)$

(8) generate $z_t \sim n(0, 1)$

(9) with the generated y_0 and an initial p_0 (0.002 from Table 1) calculate y_t and $\Sigma\, m y_s^2 p_s$. When this sum exceeds the value of $-\log(1-r_i)$ at a particular value of t, that value of t is the required T_i - the marriage-time of the i-th individual

(10) if the i-th individual transits to the married state at a particular time t, add that individual to the proportion married such that for each t, the simulated proportion p_t' is found out (note that it is this p_t' which is used in the sum used in step number (9).

(11) optional - a control in the algorithm can be inserted such that when the final proportion married under simulation arrives at the actual known value, the simulation can be terminated.

(12) go back to step number (5) till a specified number of t times are over (e.g. t= 1,36 as most individuals would be married by the time they are aged 50).

This algorithm will be repeated if the mean and variance of the simulated distribution of marriage-times (along with the final proportion married obtained through step number (10) if step number (11) has not been carried out) do not correspond closely to the known values (8.57 and 21.96 respectively from Table 1).

The sequence of results shown in Table 2 - obtained for a cohort of 100 individuals - serves as an example of identifying the best estimates of the relevant parameters through repeated application of the above algorithm with different values of m and a_1.[9]

Table 2: Means and variances of distribution of marriage-times obtained for various values of m and a_1 for a cohort of 100 individuals

trial	1	2	3	4	5	6	7
m	1.0	1.0	1.0	1.0	1.0	0.95	0.92
a_1	-0.02	-0.05	-0.10	-0.20	-0.40	-0.40	-0.38
$\bar{T}$	6.9	6.9	7.21	7.7	8.4	8.46	8.71
σ_T^2	7.96	7.82	10.10	16.0	21.47	22.25	21.77

With a fixed value of m (=1.0), trial values of a_1 indicate that at about $a_1 = -0.4$ (the 5th trial), the mean and variance are closer to the actual ones. Therafter, the value of m can be adjusted in order to get a very close approximation to the desired values. At $m = 0.92$ and $a_1 = -0.38$, the simulation results in mean and variance of the simulated marriage-times that are very

9. These results have been obtained using the NAG library subroutines for generating the uniform and Gaussian random variates. However, keeping in mind the non-availability of the NAG library in many centres, the computer program SIMMAR given in the Appendix calls the uniformly distributed random variates from the intrinsic function installation of the Fortran library (mostly available in many computer centres) and provides a subroutine for generating the Gaussian variate.

close to the known values. Estimates of other parameters are inferred from these two parameters (using the steps numbered 1 through 4 in the algorithm). Table 3 presents the estimates of various parameters, as well as the simulated transition times of 100 individuals (the blank '-' indicating a transition time which exceeds 36 years). And Table 4 rearranges the transition times, both observed and simulated for 88 out of 100 individuals (see also Table 1).

Table 3: The estimated parameters of the diffusion process and the simulated transition times of 100 individuals

$a_0 = 0.144$	$a_1 = -0.380$	$b = 0.001$
$m = 0.920$	$m_0 = 0.379$	$v_0 = 1.476$
$A = 1.620$	$s^2 = 5.207$	$P = 0.883$
$\bar{T}' = 8.716$	$\sigma_T'^2 = 21.772$	survivors = 12

12	4	5	10	7	11	13	10	5	7
-	6	4	9	15	-	5	19	7	5
-	-	4	4	10	2	8	10	19	9
20	8	7	-	17	5	-	5	11	-
9	9	12	6	9	5	19	1	10	6
14	-	11	6	3	6	4	17	4	11
-	8	13	7	17	7	8	14	20	5
-	8	7	6	20	16	3	5	4	4
6	5	-	11	7	7	5	17	5	10
9	-	7	15	8	8	6	5	6	3

Appendix D gives each transition time for the i-th individual and her associated values of $-log(1-r)$, y_0 and the current p'_t. It would be worth examining the trajectory of marriageability of

Table 4: Observed and simulated frequency of marriage-times

t	1	2	3	4	5	6	7	8	9	10
obs.	0	1	2	7	10	12	13	10	8	6
sim.	1	1	3	8	13	9	10	7	6	6
t	11	12	13	14	15	16	17	18	19	20+
obs.	5	3	2	2	1	1	1	1	-	6
sim.	5	2	2	2	2	1	4	-	3	3

certain individuals for the sake of an illustration. Table 5 presents the trajectories of individual women numbered 3, 4, 15, 31 and 11 whose marriage-times are 5, 10, 15, 20 and greater than 36 respectively (refer Table 3, individuals row-wise). The 3rd individual who starts with a very high marriageability (5.28) gets married at the 5th year, her potential decreasing over the years. Compare hers with a rather low one of the 48th woman (0.52, Appendix D) who gets married, however, within a year. This is because the latter's initial marriageability, low as it is, is associated with a random number *r* which is also

Table 5: Trajectory of marriageability for chosen individuals

	Individuals numbered				
t	3	4	15	31	11
0	5.284647	0.120229	0.113649	0.643523	0.204212
1	5.277955	0.120246	0.113283	0.643008	0.204162
2	5.237895	0.120339	0.111104	0.639928	0.203856
3	5.164960	0.120507	0.107161	0.634344	0.203291
4	4.996053	0.120924	0.098257	0.521286	0.202063
5	4.589125	0.122055	0.077680	0.589113	0.198857
6		0.123978	0.045341	0.531555	0.192750
7		0.126598	0.017733	0.465707	0.185798
8		0.129561	0.002036	0.396882	0.179082
9		0.132443	0.001598	0.336936	0.172032
10		0.135297	0.013211	0.287817	0.164883
11			0.032304	0.248200	0.159751
12			0.053411	0.219034	0.154794
13			0.073249	0.198260	0.152159
14			0.089365	0.182259	0.149584
15			0.103569	0.171059	0.148212
16				0.162150	0.146479
17				0.157228	0.145580
18				0.153686	0.144412
19				0.151031	0.143986
20				0.150048	0.143229
					....

very low such that the summand $my_0^2 p_0 = 0.92 * 0.52 * 0.002 = 0.0009$ exceeds $-log(1-r) = 0.0004$ at the very first year. The 4th woman who starts with an initial potential of 0.12 gets married at the 10th year, hers however increasing as years go by, though only slightly. The 15th individual getting married at the 15th year starts with an initial marriageability of 0.114 and experiences a fluctuation in her marriageability as years go by.

The estimated parameters help formulating the marriageability process as follows:

$$y_{t+1} = 0.144\, p_t + (1 - 0.38\, p_t) + 0.001\, z_{t+1}$$

whose mean and variance are given by

$$m_{t+1} = 0.144\, p_t + (1 - 0.38\, p_t)\, m_t - 0.92\, p_t m_t v_t$$

$$v_{t+1} = (1 - 2(0.38)\, p_t) + (0.001)^2\, p_t - 2\,(0.92) p_t\, v_t^2$$

Further, the i-th individual's marriage rate at time t is $m(t, y_t, p_t) = 0.92\, p_t y_t^2$ and the corresponding cohort rate is $\bar{m}(t, p_t) = 0.92\, p_t\, (m_t^2 + v_t)$. Note that any individual's rate is 0.92 times the combined influence of the sociological factors considered in the model (one increasing the pressure to marry and the other counter-balancing this pressure through that individual's own potential evolving over time). Moreover, the cohort rate at time t is expressed in terms of the mean and variance of the process of marriageability at time t. Table 6 compares the differences in the cohort rate and individual rates for individuals numbered 3, 4 and 15. The cohort rate at any time t (given in col.5 in Table 6) differs from those of individuals (given in cols.7, 9 and 11) who have different potentials and hence different marriage times. Compare this analysis with that of Vaupel and Yashin(1979, 1982). Fig.1 in Appendix E plots these rates for the sake of visual comparison of the differences in cohort and individual rates.

Table 6: Cohort and Individual rates over time and the simulated proportions married, mean and variance of the process of marriageability

time t	simulated p'_t	process y_t mean m_t	var. v_t	cohort rate $m p'_t(m_t^2 + v_t)$	individual rates ind. no. 3 y_t^2	$m p'_t y_t^2$	ind. no. 4 y_t^2	$m p'_t y_t^2$	ind. no. 15 y_t^2	$m p'_t y_t^2$
0	0.002	.379	1.476	0.0148	5.285	0.0486	0.1202	0.0011	0.1136	0.0010
1	0.01	.377	1.466	0.0148	5.278	0.0486	0.1202	0.0011	0.1132	0.0010
2	0.02	.372	1.415	0.0286	5.238	0.0964	0.1203	0.0022	0.1111	0.0020
3	0.05	.362	1.320	0.0667	5.165	0.2376	0.1205	0.0055	0.1072	0.0049
4	0.13	.340	1.109	0.1465	4.996	0.5975	0.1209	0.0145	0.0983	0.0118
5	0.26	.297	.705	0.1897	4.589	1.0977	0.1221	0.0294	0.0777	0.0186
6	0.35	.255	.328	0.1265			0.1240	0.0399	0.0453	0.0146
7	0.45	.244	.171	0.0954			0.1266	0.0524	0.0177	0.0073
8	0.52	.250	.088	0.0720			0.1296	0.0620	0.0020	0.0010
9	0.58	.215	.046	0.0620			0.1324	0.0707	0.0016	0.0009
10	0.64	.284	.023	0.0610			0.1353	0.0797	0.0132	0.0078
11	0.69	.303	.011	0.0653					0.0323	0.0205
12	0.71	.321	.005	0.0706					0.0534	0.0349
13	0.73	.336	.002	0.0772					0.0732	0.0492
14	0.75	.347	.001	0.0838					0.0894	0.0617
15	0.77	.356	.0004	0.0901					0.1036	0.0334

SECTION 4 : DIVORCEABILITY AND OTHER RELEVANT QUESTIONS

The analysis carried out for the case of the unobservable process of marriageability can be extended to the case of divorceability - a topic becoming more and more relevant for populations in developed countries. The decline in first marriages and the increase in divorces along with the rising trend in cohabitation have changed the life-styles of many western populations. (For details on these behavioural changes in Belgium, see Rajulton (1985b)). And demographers' attention has been drawn to the varied influential factors of changing marital behaviour. The heterogeneous evolution of these factors and their influences on divorce, for example, can be fruitfully analysed by the application of the Gaussian diffusion model, whose flexibility has been illustrated in the last two sections.

The measure of divorceability following on the same lines of arguments used for the case of marriageability can serve as a starting point. However, one obvious caution is that divorce is a renewable event (unlike the first marriage) and a comparison to survival analysis can be made only with each order of divorce. Any type of data other than the life-history data could be contaminated with different orders of divorce. With this caution in mind, the simulation exercise can be carried out in the same manner as described in the last section. As divorceability is a topic more interesting and more useful for policy considerations, the present author hopes to complement this paper with a subsequent one on divorceability, taking advantage of the extensive adaptability of the Gaussian diffusion model to include the influence of socio-economic variables that evolve over time.

The analysis presented in this paper has been restricted to to the simple case when the parameters of the unobservable process are constant over time. This restriction can be relaxed, if parameters changing over time are found tp be necessary. The consequent increase in the number of parameters would certainly call for sophisticated minimization or maximum likelihood programs in the case of a direct estimation of parameters (see p.13) or for more computer memory space in the case of trial and error procedure through simulation. Perhaps, a via media would be to consider parameters changing over time for certain duration groups. What has been presented here, however, is basic to any more sophisticated application of the model; this simple experiment would be sufficient for many practical cases.

More relevant than the question of relaxing the assumption of constancy of parameters over time is that of relaxing the Markovian assumption inherent in the model. The Gaussian model is actually based on the Kolmogorov-Fokker-Planck equation, the application of which requires the assumption that the unobservable process is Markovian:- an individual's future profile (of marriageability or divorceability) is a result of both a deterministic function of his/her current profile and a stochastic term. This can be generalized (Yashin, 1985) to depend on the entire trajectory of the unobserved variable by which past profile of an individual can be made to influence the time of transition.

The relevance of the non-Markovian condition becomes more centralized when various exogenous characteristics and their heterogeneous effects either directly on the behaviour under study or indirectly on the influencing variables (whether observed or unobserved) are taken into consideration. Very often, an evolution of these exogenous factors such as employment, economic condition, education, migration status, urban or rural environment etc. is at the least partially observable. And the endogenous component could be modelled as dependent on the evolution of these factors; for example, changes in employment, migration status and economic conditions are known to have their effects on changes in marriageability or divorceability.

Further, certain exogenous factors could have greater impact on certain subgroups of populations. Thus, divorceability may be higher in urban populations, religion and number of children may affect decisions on the timing of divorces. Certain groups will therefore be rapidly "selected" under favourable changes in these factors and the rapidity of selection may be reflected in the parameters in question of the underlying processes. However, the (linear) dynamic equations describing the changes in these factors would not be affected by selection so that no further modification of the auto-regressive equations for these factors would be necessary.

Finally, the necessity of life-history data (and associated event-history analysis) is more and more accentuated in social analysis. Further research would determine the usefulness of the Gaussian model in analysing such life-history data in which, once again, certain variables are observed and many others are not. The heterogeneity dynamics is thus a field which opens new vistas for further research, and surely the Gaussian model specifically constructed for a dynamic analysis of heterogeneity in behaviour can be of great help. Imagination and application are the only two requisites.

REFERENCES

1. Coale A.J.(1971), "Age Patterns of Marriage", Population Studies, 25(193-214).
2. Coale A.J. and McNeil D.R.(1972), "The Distribution by Age of the Frequency of First Marriage in a Female Cohort", Journal of the American Statistical Association, 67(743-49).
3. Cox D.R.(1972), "Regression Models and Life Tables (with Discussion), J.R.S.S., Series B, vol.34, pp.187-220.
4. Cox D.R.(1975), "Partial Likelihood", Biometrika, 62(269-276).
5. Cox D.R. and D.Oakes(1984), Analysis of Survival Data, Chapman and Hall, N.Y.
6. Croxton F.E., D.J.Cowden and S.Klein(1967), Applied General Statistics, Englewood Cliffs: Prentice Hall.
7. Heckman J.J. and B.Singer(1982), "Population Heterogeneity in Demographic Models", Multidimensional Mathematical Demography, ed.by Land K. and A.Rogers, Academic Press, N.Y., pp.567-599.
8. Henry L.(1972), "Nuptiality", Theoretical Population Biology, 3(135-152).
9. Henry L.(1976), Population Analysis and Models, Edward Arnold (Pub.)Ltd, London.
10. Hernes G.(1972), "The process of Entry into First Marriage", American Sociological Review, 37(173-182).
11. Hernes G.(19

12. Kendall M. and A.Stuart(1977), The Advanced Theory of Statistics, vol.I, Distribution Theory, Charles Griffin &Co.,Lts,London.
13. Lindgren B.W.(1976), Statistical Theory, Macmillan Pub.,N.Y.
14. Miller R.G. et al.(1981), Survival Analysis, John Wiley & Sons, N.Y.
15. Mode C.J.(1982), "Increment-Decrement Life Tables and Semi-Markovian Processes from a Sample Path Perspective", in Land K.C. and A. Rogers (eds.), Multidimensional Mathematical Demography, Academic Press, N.Y.
16. Rajulton F.(1985a), "Heterogeneous Marital Behaviour in Belgium, 1970 and 1977: An Application of the Semi-Markov Model to Period Data", Mathematical Biosciences, 73(197-225).
17. Rajulton F.(1985b), Semi-Markovian Multistate Analysis of Marital Behaviour in Belgium: Theory and Application to Period Data, Ph.D. Dissertation, Vrije Univeristeit Brussel, Brussels.

18. Ridder G. and W.Verbakel(1983), "On the Estimation of the Proportional Hazard Model in the Presence of Unobserved Heterogeneity", unpublished paper, Univ. of Amsterdam, faculty of actuarial sciences and econometrics.
19. Rubinstein R.Y.(1981), Simulation and the Monte-Carlo Method, John Wiley & Sons, N.Y.
20. Tuma N.B. and M.T.Hannan(1984), Social Dynamics, Models and Methods, Academic Press, N.Y.
21. Vanderheoft C.(1985), Stratified Proportional Hazard Models: A GLIM-oriented Approach, with Special Reference to the Problem of Competing Risks, IPD-Working Paper, 1985-5, VUB, Brussels.
22. Vaupel J.W., K.G.Manton and E.Stallard(1979), "The Impact of Heterogeneity in Individual Frailty on the Dynamics of Mortality", Demography, 16(439-454).
23. Vaupel J.W. and A.I.Yashin(1982), The Deviant Dynamics of Death in Heterogeneous Populations, IIASA Working Paper, WP-47, Laxenburg, Austria.
24. Woodbury M.A. and K.G.Manton(1977), "A Random Walk Model of Human Mortality and Aging", Theoretical Population Biology, 11(37-48).
25. Yashin A.I.(1984), Dynamics in Survival Analysis: Conditional Gaussian Property versus Cameron-Martin Formula, WP-84, IIASA, Laxenburg, Austria.
26. Yashin A.I. and K.G.Manton(1984), Evaluating the Effects of Observed and Unobserved Diffusion Processes in Survival Analysis of Longitudinal Data, CP-84-58, IIASA, Laxenburg, Austria.
27. Yashin A.I.(1985), Survival Analysis with Partially Observed Covariates: Discrete Time, unpublished paper.
28. Yashin A.I., K.G.Manton and J.W.Vaupel(1985), "Mortality and Aging in a Heterogeneous Population: A Stochastic Process Model with Observed and Unobserved Variables", Theoretical Population Biology, 27(154-175).
29. Yashin A.I.,K.G.Manton and E.Stallard(1985), The Propagation of Uncertainty in Human Mortality Processes, CP-85-36, IIASA, Laxenburg, Austria.
30. Yashin A.I. and Rajulton F.(1985), A Simulation Study of the Conditional Gaussian Diffusion Process Model of Survival Analysis, forthcoming Working Paper, IIASA, Laxenburg, Austria.

Appendix A: Mean and Variance of Non-Central Chi-Square

When Z follows a Normal distribution with mean M and variance S^2, the square of the variate(Z^2) follows a non-central chi-square distribution, whose moment generating function is given by:

$$\frac{1}{S\sqrt{2\pi}}\int_{-\infty}^{\infty} \exp\left[tZ^2 - \frac{(Z-M)^2}{2S^2}\right] dZ$$

Expanding the exponential term,

$$tZ^2 - \frac{(Z-M)^2}{2S^2} = \frac{2tS^2Z^2 - Z^2 - M^2 + 2MZ}{2S^2}$$

(i.e) $$-\frac{1}{2S^2}\left[Z^2(1-2S^2t) + M^2 - 2MZ\right]$$

(i.e) $$-\frac{1}{2S^2}\left[\left(Z\sqrt{1-2S^2t} - \frac{M}{1-2S^2t}\right)^2 + M^2\left(1-\frac{1}{1-2S^2t}\right)\right]$$

(i.e) $$-\frac{1}{2S^2}\left[\left(Z\sqrt{1-2S^2t} - \frac{M}{1-2S^2t}\right)^2 - \frac{2M^2S^2t}{1-2S^2t}\right]$$

Therefore, the M.G.F. of Z^2 is

$$\frac{1}{S\sqrt{2\pi}}\int_{-\infty}^{\infty} \exp\left(-\frac{M^2t}{1-2S^2t}\right) \cdot \exp\left[\frac{-1}{2S^2}\left(Z\sqrt{1-2S^2t} - \frac{M}{1-2S^2t}\right)^2\right]$$

(i.e) $$\exp\left(\frac{M^2t}{1-2S^2t}\right) \cdot (1-2S^2t)^{-\frac{1}{2}}$$

Expanding the product up to the second degree and rearranging the terms in t,

$$\left[1 + \frac{M^2t}{1-2S^2t} + \frac{M^4t^2}{2(1-2S^2t)^2} + \ldots\right] \times \left[1 + S^2t + \frac{3}{2} S^4t + \ldots\right]$$

(i.e) $$S^2t + \frac{3}{2}S^4t^2 + \frac{M^2t}{1-2S^2t} + \frac{M^2S^2t^2}{1-2S^2t} + \frac{M^4t^2}{2(1-2S^2t)^2}$$

Differentiating this twice with respect to t and at t=0, and denoting the mathematical expectation by Ex, we get

$$\left.\frac{d(M.G.F)}{dt}\right|_{t=0} = S^2 + M^2 = Ex(Z^2)$$

$$\left.\frac{d^2(M.G.F)}{dt^2}\right|_{t=0} = 3S^4 + 6M^2S^2 + M^4 = Ex(Z^4)$$

Hence, the variance = $Ex(Z^4) - (Ex(Z^2))^2$

$$= 2S^2(S^2 + 2M^2)$$

Appendix B : To prove that mean death time $\bar{T} = \frac{1}{MY^2P}$

We have Prob $\left[T > t \mid Y_0^t\right] = \exp\left[-\int_0^t MY_s^2 \, P_s \, ds\right]$

Under stability, $\frac{dY_t}{dt} = 0 = a_0 P_t + a_1 P_t Y_t$

i.e; $Y_t = \frac{a_0}{a_1} = Y$

Therefore, Prob $\left[T=t \mid Y_t=Y\right] = \exp\left[-MY^2Pt\right]$,where P is the asymptote of P_t

Hence, $Ex(T) = \int_0^\infty t \exp(-MY^2Pt)dt = \frac{1}{MY^2P}$

Note that $Ex(T^2)$ will be equal to $\frac{2}{M^2Y^4P^2}$ and hence the variance is given by $\frac{1}{M^2Y^4P^2}$ which is the same as the square of $Ex(T)$. Therefore, variance will not be of much use to determine the value of a_1.

Appendix C

```
      Program SIMMAR ( input, tape1, tape2, output)
c ------------------------------------------------------------------
c   This is a program for simulating the times to death of
c   100 individuals and the unobserved process Y(t).  The
c   input (on the screen) specifies the parameter values
c   initiating the simulation.  The parameters are :
c          n    =  number of individuals to be considered
c                  (the program admits a maximum of 100,
c                   can be increased if memory space is
c                   available)
c          n1   =  number of time units ( a maximum of 36 is
c                  incorporated in the program, can be increased
c                  with the memory space)
c
c          mu   =  the "mortality" parameter (real value)
c
c          a1   =  the second coefficient (of autoregressive
c                  equation for the unobservable process)
c
c          pf   =  final proportion "dead"  up to the time n1
c
c          xbar =  the mean of the distribution of"death times"
c
c          af   =  average initial "frailty"-the unobservable
c
c   The uniform random variates are generated through the
c   intrinsic function RANF(N) of the FORTRAN IV, which returns
c   values uniformy distributed over the range (0,1), the values
c   0 and 1 being excluded.  N  is a dummy argument which is
c   ignored.  Some FORTRAN installations have CALL RANSET(N) and
c   CALL RANGET(N) which initializes the seed of RANF and obtains
c   the current seed of RANF respectively.  Check with the
c   computer installations and make modifications if necessary.
c
c   The normal random variates are produced by the subroutine
c   Gauss which makes use of RANF also.  The subroutine GAUSS has
c   been incorporated in this program. and produces a specified
c   mean and variance( AM  denotes the mean and  S the standard
c   deviation).
c
c   The output (on the screen) gives the parameter values as
c   well as the mean and variance of the simulated "death times"
c   so that a direct comparison can be made with the known values.
c   Tape1 on output contains the sequence of individuals who
c   transit to "death" along with their associated values of
c   -log(1-r), Yo^2  and "death time".  Tape2 contains the trajectory
c   of "frailty" over time for individuals numbered 10,20,30,40
c   and 50.  Change the individuals specified if desired; these
c   trajectories can be used for plotting if comparison of indivi-
c   dual frailty values over time is desired.
c ------------------------------------------------------------------

      real y0(100), y(100,36), p(36), eps(100,36), yy(100),mu ,mØ
      real  m(36), v(36), yØ2(100),  y2(100,36), r(100)
      integer t(100)
      data b/0.001/, pØ/0.002/
      print 11
   11 format (* give parameter values - n, n1,mu, a1,pf,xbar,af*)
      read * , n,n1, mu,a1,pf,xbar,af
      c1 = sqrt(xbar* mu * pf)
      aØ = a1/c1
      mØ = aØ/abs(a1)
```

```
      gØ  =  af - (mØ ** 2)
      s2  =  (2 * gØ) * (gØ + 2*(mØ**2))
      write 12, aØ, a1, b, mu, mØ, gØ, af, s2, xbar, pf
   12 format( 5x,  * the parameters are : *,/,
                 5X,  *  aØ   =* , f8.4,/,
                 5x,  *  a1   =* , f8.4,/,
                 5x,  *  b    =*, f8.4,/,
                 5x,  *  mu   =*, f8.4,/,
                 5x,  *  mØ   =*, f8.4,/,
                 5x,  *  gØ   =*, f8.4,/,
                 5x,  * abil  =*, f8.4,/,
                 5x,  *  s2   =*, f8.4,/,
                 5x,  * xbar  =*, f8.4,/,
                 5x,  * finp  =*, f8.4,/)
      write 66
   66 format(1x)
c ---------------------------------------------------------------
c  The initial parameters having been specified, generate the
c  random variates following uniform distribution and gaussian
c  distribution.  This is done for each time t so that indivi-
c  duals who transit to "death" at a particular time may be
c  eliminated to form the proportion "dead" for every time t.
c ---------------------------------------------------------------
      do 1 k=1,n1
      do 2 i=1,n
c ---------------------------------------------------------------
c  Generate the uniform random variate r(i) for each individual
c  and find the value of -log(1-r).
c ---------------------------------------------------------------
      r(i) = ranf(n)
      r(i) = -alog(1-r(i))
c ---------------------------------------------------------------
c  Generate the initial random unobserved variate YØ(i) follow-
c  ing gaussian with mean mØ and variance gØ
c ---------------------------------------------------------------
      am = mØ
      s  = sqrt(gØ)
      call gauss(ix,s,am,v)
      yØ(i) = v
      yy(i) =  mu * (yØ(i) ** 2) * pØ
c ---------------------------------------------------------------
c  Generate the error random variate eps(i,k) following gaussian
c  with mean Ø and variance 1.
c ---------------------------------------------------------------
      am = Ø.
      s  = 1.
      call gauss(ix,s,am,v)
      eps(i,k) = v
c ---------------------------------------------------------------
c  Find the autoregressive sequence of unobserved Y variate and
c  the summand used for fixing the time to death. cf.text
c ---------------------------------------------------------------
      if(k.eq.1)  y(i,k) = aØ * pØ  + (1+a1*pØ)*yØ(i) + b*eps(i,k),
      if(k.ne.1)  y(i,k) = aØ * pØ  + (1+a1*p(k-1))*y(i,k-1)
     1   b*eps(i,k)
      yy(i)  =  yy(i) + (mu *(y(i,k)**2)*p(k))
      if(yy(i).gt.r(i)) t(i)=k
      if(yy(i).gt.r(i)) p(k)=p(k)+(1./n)
      if(yy(i).gt.r(i)) write(1,64) i,r(i),yØ(i),t(i)
   64 format(3x,i3,2(2x, f10.6),2x,i3)
```

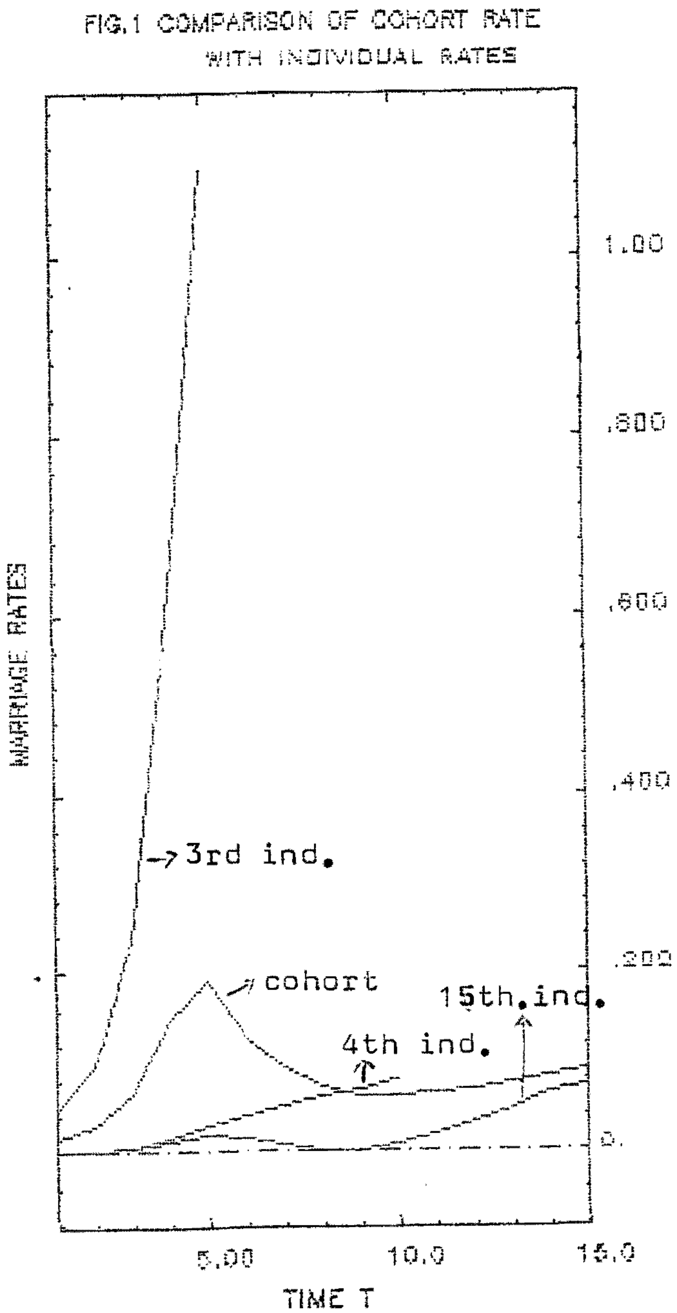
```
      if(p(k).gt.pf) go to 3
    2 continue
    1 continue
    3 continue
c -------------------------------------------------------------
c  Get the various results in Tape1 and Tape2
c -------------------------------------------------------------
      write 14, (t(i),i=1,n)
   14 format(10(2X,i3,2x)

      do 6 i=1,n
      yØ2(i) = yØ(i) ** 2
      do 6 k=1,n1
       y2(i,k) =  y(i,k) ** 2
    6 continue
      k=Ø
      write(2,11) k,yØ2(1Ø),yØ2(2Ø),yØ2(3Ø),yØ2(4Ø),yØ2(5Ø)
      do 7 k=1,n1
      write(2,11) k, y2(1Ø,k), y2(2Ø,k), y2(3Ø,k), y2(4Ø,k), y2(5Ø,k)
    7 continue
   11 format(5x,I3,5(3x,f9.6))
c ------------------------------------------------------------------
c  Find the mean and variance of the distribution of "death-times"
c ------------------------------------------------------------------
      t1=Ø.
      t2=Ø.
      do 5 i=1,n
      if(t(i).ne.Ø) t1 = t1 + t(i)
      if(t(i).ne.Ø) t2 = t2 + t(i)**2
    5 continue
      n2 = pf * n
      tbar = t1/n2
      tsig =sqrt((t2/n2)-(tbar ** 2))
      write 66
      write 15, tbar, tsig
   15 format(5x,6Hmean =,1x,f7.3,5X,9hst.dev. =,1x,f7.3)
c -------------------------------------------------------------
c  Find the mean and variance of the unobserved variate for
c  all t and check whether they converge or diverge. They are
c  stored in Tape1.
c -------------------------------------------------------------
      do 8 k=1,n1
      if(k.eq.1) m(k) = aØ*pØ + (1+a1*pØ)*mØ  - mu*pØ*mØ*gØ
      if(k.eq.1) v(k) = (1+2*a1*pØ)*gØ  +(b**2)*pØ  - 2*mu*pØ*(gØ**2)
      if(k.ne.1) m(k) = aØ*p(k-1) + (1+a1*p(k-1))*m(k-1)
     1  - mu*p(k-1)*m(k-1)*v(k-1)
      if(k.ne.1) v(k) = (1+2*a1*p(k-1))*v(k-1) + (b**2)*p(k-1)
     1   - 2*mu*p(k-1)*(v(k-1)**2)
      write(1,66)
      write(1,17) m(k),v(k)
    8 continue
   17 format(5x,2(f10.6,3x))
      stop
      end
```

```
      subroutine gauss
c --------------------------------------------------------------
c   Purpose : to compute a normally distributed random number
c             with a given mean am and standard deviation s.
c   usage   : call gauss(ix,s,am,v)
c   parameters:  ix - must contain an odd integer number with 9
c             or less digits on the first entry to gauss. Thereafter
c             it'll contain a uniformly distributed integer random
c             number generated by the subroutine for use on the
c             next entry to the subroutine.
c   method  : uses 12 uniform random numbers to compute normal
c             random numbers by central limit theorem. The result
c             is then adjusted to match the given mean and st.dev.
c             The uniform random numbers are generated through the
c             fortran ranf(n)
c --------------------------------------------------------------
      a =Ø.
      do 5Ø i=1,12
      y = ranf(ix)
   5Ø a = a+y
      v = (a - 6.Ø)*s + am
      return
      end
```

Appendix D : Simulated time of marriage (T_i) of the i-th individual and associated values of $-\log(1-r_i)$, Y_{0i} and the proportion married (P_{T_i}) at the time of the i-th individual's transition.

T_i	i	$-\log(1-r_i)$	Y_{0i}	P_{T_i}
1	48	.000404	-.720446	.012
2	26	.023103	3.007357	.022
3	55	.670494	4.709619	.032
3	77	.046247	1.533069	.042
3	100	.069176	2.132158	.052
4	2	.255818	1.949870	.062
4	13	1.019067	3.721305	.072
4	23	.093150	1.674000	.082
4	24	.055316	1.140912	.092
4	57	.035836	-.778841	.102
4	59	.832089	3.610284	.112
4	79	.221535	1.900350	.122
4	80	.223194	1.732732	.132
5	3	.464070	2.298836	.142
5	9	.040166	.627301	.152
5	17	.427122	2.075022	.162
5	20	.109737	.779552	.172
5	36	.589112	2.322386	.182
5	38	.138221	-1.104562	.192
5	46	.122459	-1.246518	.202
5	70	.185044	1.333819	.212
5	78	.042535	.531853	.222
5	82	.036732	.440995	.232
5	87	.203424	1.571260	.242
5	89	.317947	1.717952	.252
5	98	.499497	1.772390	.262
6	12	.978440	2.258425	.272
6	44	.558410	-1.514168	.282
6	50	1.241890	2.070874	.292
6	54	.576213	-1.729562	.302
6	56	.773323	1.697110	.312
6	74	.335750	-1.321418	.322
6	81	2.020342	-2.546284	.332
6	97	.787105	1.601624	.342
6	99	.228378	1.099208	.352
7	5	2.109858	2.221880	.362
7	10	.856068	-1.383450	.372
7	19	.360926	.874441	.382
7	33	.268303	.798016	.392
7	64	.005498	.029093	.402
7	66	.598291	-1.367750	.412
7	73	.715955	1.299349	.422
7	85	.365902	.950845	.432
7	86	.993963	1.566890	.442
7	93	1.391212	1.622854	.452
8	27	.514146	.841932	.462
8	32	3.124072	2.237326	.472
8	62	.538528	-1.052679	.482
8	67	.528340	-1.073198	.492
8	72	1.060165	1.171882	.502
8	95	1.489395	1.477925	.512
8	96	.033511	.144863	.522
9	14	1.165141	1.115685	.532
9	30	.555840	.715359	.542
9	41	.927674	.994666	.552
9	42	.621550	.812812	.562
9	45	.142037	.263897	.572
9	91	.043771	.054192	.582
10	4	.254938	.346740	.592
10	8	.566983	.663891	.602
10	25	.039282	-.021377	.612
10	28	.970199	.885687	.622
10	49	.024429	-.065680	.632
10	90	.076271	.098144	.642
11	6	.048637	-.144737	.652
11	39	.636316	.602214	.662
11	53	2.654615	1.520146	.672
11	60	.493877	.462356	.682
11	84	.045526	-.169602	.692
12	1	1.585350	1.030642	.702
12	43	.670887	.546337	.712
13	7	.199148	-.556891	.722
13	63	3.381429	1.564901	.732
14	51	1.727693	.951776	.742
14	68	.447320	-.856746	.752
15	15	.224694	-.337119	.762
15	94	.717095	.389043	.772
16	76	1.870567	.947472	.782
17	35	1.020738	.416044	.792
17	58	.450365	-.288518	.802
17	65	1.256127	.566083	.812
17	88	2.148201	.981202	.822
19	18	.680864	-.667369	.832
19	29	2.291576	.963892	.842
19	47	.934687	.206113	.852
20	31	2.069976	.802199	.862
20	69	.726452	-.493784	.872
20	75	.885140	-.818428	.882

Appendix E



Comments: The cohort rate, attaining its peak at the 5th year, is entirely different from the individual rates. Though the 3rd woman's marriageability decreases (albeit slightly) over the years, her rate increases exponentially. The 4th woman's marriageability as well as her rate increases steadily, although her rate is much lower than the cohort rate till the 9th year, a year before her transit to the married state. The 15th woman experiences fluctuations both in her marriageability and in her rate of transition and remains single till the 15th year.

IPD-WORKING PAPERS 1983

1983-1. R. Lesthaeghe : "A Century of Demographic and Cultural Change in Western Europe : An Exploration of Underlying Dimensions".

X 1983-2. R. Lesthaeghe, C. Vanderhoeft, S. Becker, M. Kibet : "Individual and Contextual Effects of Education on Proximate Determinants and on Life Time Fertility in Kenya".

1983-3. F. Eelens : "Impact of Breast-feeding on Infant and Child Mortality with varying Incidence of Malaria - Evidence from the Kenya Fertility Survey 1977-78".

1983-4. S. Becker, A. Chowdhury, S. Huffman : "Determinants of Natural Fertility in Matlab, Bangladesh".

1983-5. C. Vanderhoeft : "A Unified Approach to Models for Analysis of Zero-One Data with Applications to Intermediate Fertility Variables".

1983-6. S. Wijewickrema, R. Bulté : "Migration Impact on Population Growth in Belgium : A Multiregional Investigation with Detailed Projection Results for the Period 1976-2001".

1983-7. S. Wijewickrema et Alii : "Marital Status Trends in Belgium, 1961-77 : Application of Multi-State Analysis".

1983-8. H.J. Page : "Fertility and Family : New Currents and emerging Emphases in Research and Policy" (Overview of the UN Expert Group Meeting on Fertility and Family (New Delhi, 1983) preparatory to the International Conference on Population).

1983-9. R. Lesthaeghe, C. Vanderhoeft, S. Becker, M. Kibet : "Individual and Contextual Effects of Female Education on the Kenyan Marital Fertility Transition" (abbreviated non-technical version of Working Paper 1983-2).

X = OUT OF PRINT

IPD-WORKING PAPERS 1984

1984-1. F. Rajulton : "An Age Dependent Semi-Markov Model of Marital Status in Belgium : An Application of Littman's Algorithm to Period Data, 1970".

1984-2. R. Lesthaeghe : "Fertility and its Proximate Determinants in sub-Saharan Africa : The Record of the 1960's & 70's".

1984-3. C. Vanderhoeft : "Life Tables and Relational Models : A Unified Approach, and Applications in Demography".

1984-4. R. Lesthaeghe : "Demographic Renewal and Problems of Social Organization".

1984-5. S. Becker : "A Response Bias in the reporting of Month of Birth in Pregnancy History Surveys".

1984-6. R. Lesthaeghe : "On the Adaptation of sub-Saharan Systems of Reproduction".

IPD-WORKING PAPERS 1985

1985-1. R. Lesthaeghe and F. Eelens : "Social Organization and Reproductive Regimes : Lessons from sub-Saharan Africa and historical Western Europe".

1985-2. P. Willems and S. Wijewickrema : "The Evolution of Nuptiality in Belgium from 1954 to 1981".

1985-3. F. Eelens and L. Donné : "The Proximate Determinants of Fertility in Sub-Saharan Africa: A Factbook based on the Results of the World Fertility Survey".

1985-4. Ronald C. Schoenmaeckers : "Current Fertility Behaviour in Africa. Results from a Birth Interval Analysis of WFS Data".

1985-5. Camille Vanderhoeft : Stratified Proportional Hazards Models. A GLIM oriented approach, with special reference to the problem of competing risks.

1985-6. Stanley Wijewickrema : "Childlessness in Belgium and Flanders".

1985-7. R. Lesthaeghe : "Value Orientations, Economic Growth and Demographic Trends - Toward a Confrontation?

1985-8. Fernando Rajulton and Stan Wijewickrema : "An Enquiry into the Two Basic Assumptions of Multistate Increment-Decrement Life Tables".