Time Dependency in Diffusion Models: Gamma-Diffusion Models as an Alternative to the Hernes Model

Jean-Marie Le Goff

University of Lausanne

Contact: Jean-Marie.LeGoff@unil.ch

Abstract

Marriage process in a cohort is sometimes analyzed with a Hernes model. This model presents a convenient property that clearly allows one to distinguish the quantum effect in the cohort from the tempo effect. In its original paper, Hernes (1972) formulates his model in terms of diffusion or contagion of the idea of marriage from people already married to those not yet married. The spread of the proportion of married people in the cohort has an increasing effect on the risk of marriage. However, this increasing effect is slowed down by the fact that unmarried people progressively cease to be attractive on the marriage market when aging. Hernes mentions that this decreasing force could be related to another mechanism: Each individual in the cohort is heterogeneous in his or her susceptibility to get married. Persons who have a better susceptibility to marry will marry early, and the weight of individuals with worse susceptibility will be higher and higher as time goes on. The consequence of the heterogeneity in susceptibility is a negative effect on the risk to marry, which slows down the increasing effect due to the mechanism of diffusion. We believe that this second mechanism no longer corresponds to the model formally specified by Hernes. This paper shows models corresponding to this mechanism, the gamma-logistic and the gamma-mixed influence diffusion models. These models are estimated on data of the Wisconsin Longitudinal Study.

Keywords: Diffusion models, Hernes model, Gamma-diffusion models, Event history analysis

Time Dependency in Diffusion Models: Gamma-Diffusion Models as an Alternative to the Hernes Model

The Hernes model is classically used to analyze marriage across cohorts (Hernes, 1972; Diekmann, 1989; Goldstein & Kenney, 2001). This model presents a capable property to distinguish clearly the quantum effect in the cohort (i.e., the proportion of married people) from the tempo effect; that is, the timing of marriage (Billari & Toulemon, 2006). At the opposite of other popular models used in the analysis of marriages—for example, in the Coale-McNeil model—the quantum effect is considered as a fraction of people predetermined from the beginning of the marriage process to remain unmarried (Coale & McNeil, 1972). If the Hernes model was originally estimated on aggregated data, it has been incorporated in the corpus of Event History Analysis models and can be estimated on individual history data (Wu, 1990; Rohwer & Pötter, 2002).

In his seminal paper (Hernes, 1972), Hernes formulates his model in terms of diffusion or contagion of the idea of marriage from people already married to those not yet married. Contagion is determined by a mechanism of imitation by non-married persons or by a mechanism of persuasion of married persons on non-married individuals. Whatever the mechanism, the spread in the proportion of married people in the cohort has an increasing effect on the risk of marriage. However, Hernes postulates that an opposing force slows down this diffusion effect because unmarried people progressively cease to be attractive on the marriage market as they age. As a consequence, there is a decreasing effect on the hazard of marriage. The overall hazard of marriage then results in two components: increasing in relation to the diffusion of marriage in the cohort and decreasing as a consequence of the depreciation of the marriageability. Later, Diekmann proposed the log-logistic model as an alternative to Hernes model, with the same opposing mechanisms of diffusion and depreciation of the marriageability.

In its 1972 paper, Hernes mentions that the decreasing force can result from another mechanism; Aging people who remained unmarried are those, for example, who have never held a prestigious job and, for this reason, are unattractive on the marriage market. The point, here is on individual heterogeneity. In another and more precise paper, Hernes (1976) uses the term of *structural heterogeneity*: "when a capacity is differentially distributed in the population" (p. 428). The mechanism underlying the distribution of the risk of marriage for the cohort is different from the precedent described. This mechanism is similar to those described with notions of unobserved heterogeneity or of frailty in unemployment studies and mortality studies, respectively (Heckman & Singer, 1982; Vaupel, Manton, & Stallard, 1979). In the present case, those in the cohort who have a better ability to marry will marry early, and the weight of people with unfavorable capacities will become progressively higher and higher as time goes on. As a consequence, the differential ability results in a negative effect on the risk to marry, which slows the increasing effect due to the mechanism of diffusion. A similar effect could be described if each individual in the cohort differs by his or her own "susceptibility" to adopt the behavior when in contact with someone already married who transmits the idea of marriage (Strang & Tuma, 1993).

As in unemployment or mortality studies, the difficulty is that the differential ability or susceptibility can be due to unobserved characteristics of persons. This unobserved heterogeneity then has to be incorporated in the models. In this paper, we propose two diffusion models that introduce an unobserved ability or susceptibility of persons to adopt behavior or an innovation that can be estimated on individual retrospective data. In the first model, called the *gamma-logistic model*, the diffusion mechanism is described by the classical logistic curve while the unobserved ability or susceptibility of individuals is patterned by a gamma distribution. In the second model, called the *gamma-mixed influence diffusion model,* the unobserved heterogeneity is previously patterned by a gamma distribution while there are two processes of diffusion: The first process is due to internal influence—that is, influence of persons who already adopted the behavior on those who did not yet adopt. The second is due to external influence such as media, advertising, and institutions that diffuse norms about marriage, and so on.

In the first section of the paper, after a recall of the Hernes model specification, we present the gamma-logistic and the gamma-mixed influence models. In the second section, we apply these models in the case of marriage of men and women interviewed in the Wisconsin Longitudinal Study. We are especially interested to compare the fit of these models with the fit of a Hernes model and a log-logistic model.

Time Dependency in Diffusion Models

This section discusses a general family of diffusion model proposed by Hernes in its second influential paper on diffusion models (Hernes, 1976). This general formulation is interesting because all models evoked in the introduction of the present paper—the Hernes model, the log-logistic model, logistic and mixed influence model—are particular forms of this general formulation. This model can be formalized for each kind of diffusion, such as the propagation of an innovation, behavior, a rumor, and eventually a contagious illness. However, we have especially in mind the diffusion of marriage in a cohort. The model can be read as a general mixed-influence diffusion model (Mahajan & Peterson, 1985), in which transmission coefficients associated to external and internal influences vary with time. As originally formulated by Hernes (1976, p 434), this model does not include unobserved susceptibility or ability to adopt the behavior, but we will include it later. Let $F(t)$, the cumulative proportion of persons already married between t_0 and t; $f(t)$, the probability density to get married which is also the derivative of $F(t)$; $S(t)$ the complementary of $F(t)$, i.e., the proportion of people who are not married at time t:

$$f(t) = p(t)S(t) + q(t)F(t)S(t) \quad (1)$$

The product *F(t)S(t)* represents the probability for two persons, one unmarried the other married, to interact in the absence of social barriers between groups of unmarried and married people. *q(t)* is the rate of diffusion or contagion, given that an unmarried person is in contact with a married person in a unit of time. In this general formulation, this rate is a function of time. In the case of marriage, the coefficient of diffusion *q(t)* can be understood as the rate for a single person to get married, given that he or she receives information about marriage from an already married person in the cohort. The level of this rate at time *t* can depend on several elements related to the predisposition of the person to marriage and to his or her position on the marriage market. As formulated in this model, each person already married, even for a long time, is considered to have a potential influence on an unmarried person; this influence is equal among all already married people. *p(t),* also a function of time, is the rate of adoption of the behavior due to external influence such as media, norms on marriage, and so on. Strang and Tuma (1993) suggest another interpretation for *p(t)* in which it is no longer related to external influence but to the effect of individual endogenous characteristics on marriage rates.

This model can be rewritten as a hazard rate function instead of a probability density function. If *h(t)* symbolizes this hazard rate, as *h(t)=f(t)/S(t)*:

$$h(t) = p(t) + q(t)F(t) \tag{2}$$

In the case of *q(t)=0*, adoption of diffusion depends only of intrinsic characteristics of each individual or of external influence. In this case, *p(t)* can be shaped by one of the usual functions applied to a parametric event history model. For example, if *p(t)* is considered to be constant *(p(t)=p*), then an exponential model is estimated. But if *p(t)* is considered to be always increasing or decreasing, it can be estimated by a Weibull or a Gompertz function.

In the case of $p(t)=0$, the process of adoption of the behavior depends only on internal influences. In this case, if $q(t)$ is a constant ($q(t)=q$), the model corresponds to the well-known logistic growth with q as coefficient of diffusion (Coleman, 1964; Griliches, 1957; Mahajan & Peterson, 1985). In his first paper on the diffusion of marriage in American cohorts, Hernes (1972) rather supposes that $q(t)$ is a decreasing function of time due to the fact that unmarried people progressively lose their ability to attract potential partners as they grow older. This depreciation of the "marriageability" of a person on the marriage market draws a force opposed the force due to the process of diffusion. An alternative meaning of this decreasing force on the hazard rate of marriage is proposed by Diekman (1989), who argues that it corresponds to a process of isolation of single persons as they get older and, consequently, a decrease of potential partners on the marriage market due to their isolation. The formulation proposed by Hernes for $q(t)$ is:

$$q(t) = Ab^t \tag{3}$$

A is the initial average "marriageability" ($A>0$), while b is the constant of deterioration of this ability ($0<b<1$). After integration (Hernes, 1972) and reparametrization (Wu, 1990), the model is specified with three parameters (see Table 1): The first is related to the quantum effect of marriage, the second one to the tempo effect, while the third indicates the initial proportion of persons adopting the innovation (Billari & Toulemon, 2006). The Hernes model owns the peculiarity to be defective, which means that the cumulated proportion $F(t)$ of married persons does not necessary reach 1 at the end of the marriage process. A fraction of person in a cohort is excluded from marriage due to the negative force on the marriage hazard rate becoming higher in absolute value to the positive force from the increase of already married persons.

Another decreasing function with time has been proposed by Diekmann (1989) for whom the most used log-logistic model can be interpreted as a diffusion model with a decreasing coefficient of diffusion with time. In the case of the log-logistic model:

$$q(t) = \frac{b}{t} \tag{4}$$

Where *b>0*. The log-logistic model is more parsimonious than the Hernes model with only two parameters to estimate: one related to the initial conditions and the second to the decrease in time. At the opposite of the Hernes model, the log-logistic model is not defective, which means that everyone is considered to have adopted the innovation by the end of the process. Immunity can however be considered with the hypothesis that a fraction of persons in a cohort will never adopt the innovation (Brüederl & Diekmann, 1995). This hypothesis—which corresponds to estimate a split population or a cure model (Schmidt & Witte, 1988)—means that according to some unobserved characteristics, some people are determined from the beginning of the process to remain unmarried. Generalizations of the Hernes and/or the log-logistic models have been proposed by Yamaguchi (1994), Diekmann (1992), and Braun and Engelhard (2004). Whatever these generalizations, the principle of a diffusion process countered by a loss of abilities remains.

The classic mixed influence diffusion model corresponds to the hypothesis in which in equation 1 and 2, $p(t)$ and $q(t)$, are constant ($p(t)=p; q(t)=q$) (Mahajan & Peterson, 1985). If this model has been introduced in mathematical sociology by Coleman (1964), it has been rather diffused in marketing research, consequently to the work of Bass (1969). In this discipline, the peculiarity of this model is generally estimated on aggregated data on the diffusion of innovation products (Mahajan & Peterson). If this model has been less estimated on individual data of adoption, its adaptation to the corpus of parametric methods of event history analysis with a procedure of estimation based on the likelihood maximization does not present many difficulties (Bass, Jain, & Krishnan, 2000; Roberts & Lattin, 2000). This model can be written:

$$f(t) = [p + qF(t)]S(t) \tag{5}$$

It has been suggested that this model means that a fraction of persons adopts the behavior under external influences while others adopt it due to internal influences. According to this point of view, former individuals are considered innovators while the later are imitators (Bass, 1969). However, this point of view has been criticized for the fact that the mixed influence diffusion model, as specified, does not allow a clear distinction between innovators and imitators. As it is formalized, this model means that one person can adopt the innovation either under external influences or internal influences (Tanny & Derzko, 1988). Such a model could take into account the interesting case of demographic behavior such as marriage or union formation, since external influences can be considered as social pressure coming from persons other than pairs who already adopted the behavior or from institutions that diffuse social norms on marriage. External can be for example parents or relatives, but also network channels in which social norms about marriage are diffused. The mixed influence model, as initially formulated by Coleman (1964), is not defective; it does not allow that a fraction of individuals in a cohort remain unmarried. Such a fraction can be introduced by estimating a split-population model in which, for unknown reasons or characteristics, a portion of individuals is determined to remain unmarried. However, the introduction of general unobserved heterogeneity can also be envisioned.

Unobserved Individual Susceptibility

The Hernes and log-logistic models are based on the observation that for most diffusion processes, the shape of the growth curve is an asymmetric S shape, "with the upper shank of the S being more extended" (Lekvall & Wahlbin, 1973, p. 364). Such an asymmetric shape is usually observed in the case of several demographic behaviors, especially marriage. Hypotheses made by Hernes (1972) and by Diekmann (1989) are that this asymmetry corresponds to a decrease in the transmission rate *q(t)* when time increases. However, an alternative hypothesis can be proposed: This hypothesis leans on notions of unobserved heterogeneity or frailty as it is developed in the analysis of unemployment and mortality (Heckman & Singer, 1982; Vaupel et al., 1979; Aalen, Borgan, & Gjessing, 2008). In the domain of mortality, frailty models assume that the general shape of the hazard rate is the same for each individual of a population but that each individual is characterized by his or her own frailty, which remains invariant as time passes. The hazard rate of death for an individual corresponds to the product of the general shape of mortality hazard and the individual frailty. Such a model means that at the beginning of the process, most frail people die while less frail take more and more weight in the population of survivals. As a consequence, the hazard rate of decease is decreasing.

By analogy with frailty models, diffusion processes of an innovation or behavior can be decomposed into two elements: The first is the general shape of the diffusion, which could be called—following Strang and Tuma (1993)—the *infectiousness* from those who have adopted the behavior or, more preferably, the transmissibility of the innovation to persons who did not have yet adopted the behavior. The second element corresponds to the susceptibility or the ability of an individual to adopt the innovation or the behavior. The susceptibility is here the equivalent of the frailty or the unobserved heterogeneity with the common property to be unobserved. This unobserved susceptibility can be related to the ability of a person to be in contact with persons who have already adopted the innovation or to a person's ability to accept an innovation. By hypothesis, it remains invariant as time goes on. As in the case of frailty models, we suppose a proportional effect of the individual susceptibility on the risk of adoption of the behavior. If the starting model is the simplest model of logistic growth:

$$h_i(t \mid u_i) = u_i\left[qF(t)\right] \quad (6)$$

Where u_i represents the individual unobserved susceptibility to adopt the innovation and $h_i(t \mid u_i)$ represents the hazard rate for an individual given that he has susceptibility u_i to adopt. q expresses the coefficient of transmissibility from person who already adopted. If we suppose that u_i is distributed such that its mean is equal to 1, then q represents the coefficient of transmission of the behavior to a person with an average susceptibility of adoption.

However, the nature of individual susceptibility has to be better specified. It is important to underline that this susceptibility is related to a person and not to possible transmitters and their "infectiousness" (Strang & Tuma, 1993). As in the Hernes or the log-logistic models, the model expressed in Eq. (6) supposes that everyone who adopted the behavior has the same infectiousness, whatever the moment of the adoption or the social proximity with potential adopters. Susceptibility can be related to two series of factors: The first type of factor can be related to the level of contact with others and, more generally, to the openness to receive information. The second type of factor is related to the probability that someone adopts the innovation after he has acquired information about it. For example, in the marketing research tradition, the susceptibility is related to the ability of a person to purchase the product (Jeuland, 1981, qtd. in Mahajan & Peterson, 1985; Roberts & Lattin, 2000). This suggests that the susceptibility depends from the properties of the innovation, more generally of the object of diffusion. Individual susceptibility is specific to the innovation and can be different according to what is diffused. It then can be related to the openness toward the innovation but also to the context in which the person is living. For example, if the behavior is the marriage, susceptibility can depend from the aversion degree to the marriage of the person but also of its attractiveness on the marriage market. The more a person is isolated from others or the more aversion he has toward the behavior—or the less the context is favorable for him—the lower his susceptibility to adopting the behavior.

In Eq. (6), the transmissibility process follows a logistic growth. Two opposite "forces" play a role on the process of adoption of the behavior. As before, the first force is related to the increase of people who have already adopted the innovation with the effect of increasing the hazard rate. The second force is related to the differential of susceptibility among individuals. By analogy with frailty models, the more susceptible persons will first experiment the event. Consequently, less susceptible individuals will progressively take more and more weight in the population of people who did not yet experiment the event along the time. Such a model with a constant rate of transmission from people who already adopted the behavior to those who did not, and with an individual susceptibility to adopt the behavior then explains an S growth curve with a more extended upper shank as well as the Hernes or the log-logistic models. This kind of model with individual susceptibility can be extended to the classic mixed influence model (Jeuland, 1981, qtd. in Mahajan & Peterson, 1985). The hypothesis here is that the individual susceptibility to adopt is similar whether under external or internal influence.

$$h_i(t \mid u_i) = u_i\left[p + qF(t)\right] \qquad (7)$$

In this case, p and q represent the average susceptibility of adoption under each of the influences. As before, the model displays two opposing forces—one related to the increase of the force of adoption due to increase of persons that already have adopted the behavior while the second is related to the increase of the weight of the less susceptible to adopt in the population. Finally, by similarity with the general model expressed in Eq. (1), we can write a general model in which external and internal diffusion coefficients are expressions of time and into which is introduced unobserved susceptibility:

$$h_i(t \mid u_i) = u_i \left[p(t) + q(t)F(t) \right] \tag{8}$$

From Individual Hazard to Population Hazard

Expressions (6) to (8) are individual hazard rate that depend of individual susceptibility. Following our analogy between the susceptibility to adopt an innovation or behavior and the frailty in mortality research, we now assume that u_i is gamma distributed with a mean equal to 1 and a variance κ. With the assumption of a gamma distributed unobserved heterogeneity model, it has been shown that whatever the shape of the underlying or basic hazard rate (Aalen et al., 2008):

$$h(t) = \frac{\alpha(t)}{1 + \kappa C(t)} \tag{9}$$

and:

$$S(t) = \left(1 + \kappa C(t)\right)^{-\frac{1}{\kappa}} \tag{10}$$

h(t) and *S(t)* represents respectively the hazard rate at the level of the population and the probability of not having experienced the event or the behavior at time *t* while *α(t)* represents the basic hazard rate and *C(t)* represents the cumulated basic hazard rate from *0* to *t*:

$$C(t) = \int_0^t \alpha(u)du \tag{11}$$

Expressions (9) and (10) mean that:

$$h(t) = \alpha(t)\left(S(t)\right)^{\kappa} \tag{12}$$

As *S(t) = 1-F(t)* and if we consider that *α(t)* is shaped by the Hernes mixed-influence diffusion model as expressed in expression (8), i.e., *α(t) = [p(t)+q(t)]F(t)*, then:

$$h(t) = [(p(t) + q(t))F(t)](1 - F(t))^{\kappa} \quad (13)$$

And then the density of adoption becomes:

$$f(t) = [(p(t) + q(t))F(t)](1 - F(t))^{1+\kappa} \quad (14)$$

As the Hernes (1976) mixed-influence diffusion model is the more general expression, this property of diffusion models with a gamma-distributed susceptibility remains when q is a constant and when p is equal to 0 (logistic diffusion model) or is a constant (classic mixed-influence diffusion model). These two last models will be estimated on a dataset of the Wisconsin Longitudinal Study on marriage.

Application to Wisconsin Longitudinal Study Data

In this section, we estimate different models of diffusion on the Wisconsin Longitudinal Study Sample (WLS) (2006). The Wisconsin Longitudinal Study is a panel in which the sample is composed of one-third of men and women who graduated from a Wisconsin high school in 1957 (N = 10 317). Several interviews were conducted between 1957 and 2005 on this sample but sometimes also with their sibling or marital partner. An interview in 1975 on a subsample of 4,330 men and 4,808 women (N = 9138) reconstitutes the marital history of each individual, especially the date of the first marriage. The estimation of the diffusion model on this sample necessitates the hypothesis that behaviors of the sample reflect ones in the entire cohort. Such a graduate school cohort can be considered a little ancient and, as a consequence, the pattern of marriage cannot reflect contemporaneous marriage behaviors. However, the sample of the WLS presents several advantages: First, as mentioned, it is composed of graduate students of secondary schools in Wisconsin, and we can assume that this population was socialized in a similar manner.

A first examination of data shows that marriages are rare before June 1957 but start to increase at this date, especially in the case of women. We suppose that this is because most of youth left school after their second degree. We decided to consider May 1957 as the starting time (t_0) of the marriage process and discounted all persons who married before this date. The final sample is composed of 4,319 men and 4,786 women. Non-married persons at the time of the 1975 interview were censored (respectively 7.1% of men and 6.0% of women).

In a first step, we estimate four models for each gender: the Hernes (1972) model, the split population log-logistic model (Brüederl & Diekmann, 1995), the logistic-gamma model and the gamma-mixed diffusion model. For each of these models, if three parameters have to be estimated, these parameters have different meanings according to the model. Hernes models is estimated using a specification of the probability density, survival, and likelihood functions proposed by Wu (1990) (Rohwer & Pötter, 2002; Billari & Toulemon, 2006). One parameter is related to the initial marriageability of persons, the second to the speed of decline of this ability, and the third to the "quantum" of the marriage process. Log-logistic models are estimated by specifying a likelihood function allowing for the fact that a fraction, to be estimated, of the cohort will never get married (Schmidt & Witte, 1988; Box-Stephensmeier & Jones, 2004). In this case, the first parameter is related to the initial fraction of persons already married at the beginning of the process, from which starts the diffusion process. The second parameter is related to the decreasing diffusion coefficient. The third parameter indicates the fraction of persons who are determined from the beginning of the process to remain unmarried. Logistic models are unusual in the corpus of event history analysis but imply two parameters estimating the initial fraction of married women and the diffusion coefficient (Banks, 1994; Braun & Engelhart, 2004). Here, a third parameter estimates the variance of the unobserved gamma function of susceptibility. In the case of the gamma-mixed influence diffusion model, parameters allow estimating coefficients of external and internal diffusion (Bass, Jain, & Krishnan, 2000) and the variance of the gamma function.

Models were estimated with TDA version 6.4, especially with the use of the *frml* command, which allows programming its own likelihood function for event history models

(Rohwer & Pötter, 2002). An example for the estimation of the Hernes model is developed in the user's manual. We also use sometimes the function *mle* (maximum likelihood estimation) of the library stats4 in R (Venable & Ripley, 2002). All parameters, except when they are proportions, must be positive, which means that are estimated their logarithm. Proportions must be definite between 0 and 1, and then are estimated their logit transform. For example, in the case of the gamma-mixed influence model, all parameters are positive definite, which means that are estimated their logarithm:

$$\begin{aligned} p &= \exp(a) \\ q &= \exp(b) \\ \kappa &= \exp(\mathrm{c}) \end{aligned} \qquad (15)$$

Where a, b and c are parameters to be estimated. Table 1 gives formulas of models after integration for *F(t)* and *h(t)* and results of their estimation, while Figure 1 draws for men and women cumulated function of marriages as estimated in each model. There are no possibilities to compare models with the deviance or the BIC, for example, since models are very different. However, if we compare the maximum of the logarithm of likelihood obtained for each model, it is the highest in the case of the gamma-mixed influence model for both men and women. This result seems to indicate that this model fits better the data than other estimated models. Estimation of the gamma-logistic model presents the lowest maximum of logarithm of likelihood, which seems to mean that this model is the worst. Hernes model and the log-logistic model present intermediary maximum of log-likelihood, the latter higher than the former. Differences in maximum likelihood are more pronounced for women than for men. We suppose that is because the process of marriage starts from the end of school for women. Fits show that gamma-mixed influence models is very near of the non-parametric cumulative function of occurrence of the event—computed from Kaplan-Meier estimators of the survival function—while there is an overestimation of it at the end of the process in the case of log-logistic model. In the case of the Hernes and the gamma-logistic model, survival curves are first overestimated and after underestimated.

Insert Figure 1 here

Insert Table 1 here

Covariates can be included in the mixed influence diffusion model. In this case, if covariates $x_1, x_2, x_3, .., x_n$ represent characteristics that are supposed to have an influence on the external influence parameter, covariates $y_1, y_2, y_3, .., y_n$ on the internal influence parameter and covariates $z_1, z_2, z_3, .., z_n$ on the variance of the gamma distribution, then different parameters can be estimated assuming that :

$$\begin{aligned} p &= \exp(a_0 + a_1 x_1 + a_2 x_2 + \ldots + a_n x_n) \\ q &= \exp(b_0 + b_1 y_1 + b_2 y_2 + \ldots + b_n y_n) \\ \kappa &= \exp(c_0 + c_1 z_1 + c_2 z_2 + \ldots + c_n z_n) \end{aligned} \quad (16)$$

Where the different parameters *a*, *b* and *c* have to be estimated. The same characteristics can be introduced in each part of the model. This result means, however, that processes of diffusion are independent according to the characteristics introduced in the model. In the present case, we dispose of the year of birth of each respondent. The majority of them were born in 1939 (74.7% and 80.4% of men and women, respectively). However, some of them were born before (21.8% and 14.4%, respectively), while a few were born in 1940 (3.5% and 5.2%). One can suppose that the external pressure to marriage—from family, social network, and so on—will be higher for the oldest because of their age. In this case, we can expect that the p parameter will be higher than in the case of those born in 1939. In contrast, we can suppose that youngest will have less pressure from their family, so the p parameter will be lower than in the case of those born in 1939. However, there is no reason to think that internal influence will be different as well as variation in susceptibility to adopt the behavior changes according to birth year. To verify these hypotheses, for each sex, we estimate a model in which are introduced covariates related to the birth year of respondents in each parameter of the gamma-mixed influence diffusion model. Birth year is dichotomized into three covariates: those born before 1939, those born during this year (reference category), and those born after.

Models estimation confirms the hypothesis for women as well as for men (see Table 2). In the case of women born before 1939, the coefficient associated to the external influence parameter is significant and positive, which means a higher effect of external influence in comparison with women born in 1939. Women born after 1939 present a negative coefficient, which means less external influence on the hazard of marriage. Parameters related to these two covariates are not significant in the case of the internal influence parameter and variation in susceptibility to be influenced. Results are similar for men except that there is a significant effect only in the case of persons born before 1939 and not in those born after.

Conclusion

We propose here—as an alternative to the Hernes model—two models: the gamma-logistic model and the gamma-mixed influence diffusion model. The Hernes model postulates that a diffusion effect is slowed down by the decrease of the marriageability of persons as time goes on. The models we proposed are based on another suggestion of Hernes (1972, 1976) in which the diffusion effect is progressively counterbalanced by heterogeneity in the susceptibility of persons to adopt the behavior. Persons with the higher susceptibility have the higher risk to adopt, and those who have the lower risk take more and more weight in the population when time goes on. The models we developed are similar to models with a gamma frailty in mortality studies. The estimation of one of this model—the gamma-logistic model—on the Wisconsin longitudinal data on marriage fits less well than the Hernes and the log-logistic models; the other—the gamma-mixed diffusion model—fits better on data.

Such results have also been obtained in the case of the National Child Development Study marriage data of regularly interviewed persons born between the third and sixth of March 1958 in Great Britain (not shown here). These results encourage us to estimate this model on the case of diffusion of other behaviors.

References

Aalen O., Borgan, O., & Gjessing, H. K. (2008). *Survival analysis and event history analysis. A process point of view*. New York: Springer.

Banks, R. B. (1994). *Growth and diffusion phenomena. Mathematical frameworks and applications*. Berlin: Springer Verlag.

Bass, F. M. (1969). A new-product growth model for consumer durables. *Management Sciences*, *15*, 215−227.

Bass, F. M., Jain, D., & Krishnan, T. (2000). Modeling the marketing-mix influence in new-product diffusion. In Mahajan Vihay, Muller Eitan, & Wind Yoram (Eds), *New-product diffusion model* (99−122). New York: Springer.

Billari, F., & Toulemon, L. (2006, June). *Cohort childlessness forecasts and analysis using the Hernes model*. Paper presented at the European Population Conference 2006. Liverpool, UK.

Brüederl, J., & Diekmann, A. (1995). The log-logistic model. Two generalizations with an application to demographic data. *Sociological Methods and Research, 24*(2), 158−186.

Box-Stephensmeier, J. M., & Jones, B. S. (2004). *Event history modelling. A guide for social scientists*. Cambridge: Cambridge University Press.

Braun, N., & Engelhard, H. (2004). Diffusion processes and event history analysis. *Vienna Yearbook of Population Research*, 111−131.

Coale, A. J., & McNeil, D. R. (1972). The distribution by age of the frequency of first marriage in female cohort. *Journal of American Statistical Association, 67*, 743−749.

Coleman, J. S. (1964). *Introduction to mathematical sociology*. London: MacMillan, The Free Press of Glencoe.

Diekmann, A. (1989). Diffusion and survival models for the process of entry into marriage. *Journal of Mathematical Sociology*, *14*, 31−44.

Goldstein, J. R., & Kenney, C. T. (2001). Marriage delayed or marriage forgone? New cohort forecasts of first marriage for US women. *American Sociological Review, 66*, 506–519.

Griliches, Z. (1957). Hybrid corn: An exploration in the economics of technological change. *Econometrica, 25*(4), 501–522.

Heckman, J., & Singer, B. (1982). *Population heterogeneity in demographic models*. Orlando, FL: Academic Press.

Hernes, G. (1972). The process of entry into first marriage. *American Sociological Review*, *37*(2), 173−182.

Hernes, G. (1976). Diffusion and growth. The non-homogeneous case. *Scandinavian Journal of Economics*, *78*, 427−436.

Jeuland, A. (1981). *Parsimonious models of diffusion of innovation: Derivation and comparisons*. Working Paper, Marketing Department. Chicago: Graduate School of Business, University of Chicago.

Lekvall, P., & Wahlbin, C. (1973). A study of some assumptions underlying innovation diffusion functions. *Swedish Journal of Economics*, *75*, 326−377.

Mahajan, V., & Peterson, R. A. (1985). *Models for innovation diffusion*. Newbury Park, CA: Sage.

Roberts, J., & Lattin, J. (2000). Disaggregate-level diffusion models. In Mahajan Vihay, Muller Eitan, & Wind Yoram (Eds), *New-product diffusion model* (207−236). New York: Springer.

Rohwer, G., & Pötter, U. (2002). TDA user's manual. *Ruhr Universität Bochum*. Retrieved from http://www.stat.ruhr-uni-bochum.de/tda.html

Strang, D., & Tuma, N. B. (1993). Spatial and temporal heterogeneity in diffusion. *American Journal of Sociology*, *99*(3), 614−639.

Schmidt, P., & Witte, A. D. (1988). Predicting criminal recidivism using split population survival time models. *Journal of Econometrics*, *40*, 141−159.

Tanny, S. M., & Derzko, N. A. (1988). Innovators and imitators in innovation diffusion modelling. *Journal of Forecasting, 7*, 225−234.

Vaupel, J. W., Manton, K., & Stallard, E. (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography*, *16*, 439−454.

WLS. (2006). *Wisconsin Longitudinal Study handbook. Tracking the life course*. Madison: University of Wisconsin. Retrieved from http://www.ssc.wisc.edu/wlsresearch/documentation/handbook/WLS_Handbook.pdf

Venable, W. N, & Ripley, B. D. (2002). *Modern applied statistics with S* (4th ed.). New York: Springer.

Wu, L. L. (1990). Simple graphical goodness-of-fit tests for hazard rate. In Mayer Karl Ulrich & Tuma Nancy (Eds), *Event history analysis in life course research* (184−199). Madison: The University of Wisconsin Press.

Yamaguchi, K. (1994). Some accelerated failure-time regression models derived from diffusion process models. *Sociological Methodology,* 267−300.

Table 1

Diffusion Models and Estimation

Model	**Cumulative function *F(t)* and hazard function *h(t)***	**Estimates Men**	**Estimates Women**
Hernes[a]	$F(t) = \dfrac{\sigma^{-1}\exp(-\beta\lambda^{t})}{1+\sigma^{-1}\exp(-\beta\lambda^{t})}$ $h(t) = \dfrac{-\sigma\beta\lambda^{t}\log\lambda\exp(\beta\lambda)}{1+\sigma^{-1}\exp(-\beta\lambda^{t})}$	log(β) = 2.00*** log(λ) = 4.58*** log(σ) = -2.74*** LMV = -21206.01	log(β) = 1.80 log(λ) = 3.81 log(σ) = -2.68 LMV = -22539.94
Split-Population Log-Logistic[b]	$F(t) = \dfrac{k(\lambda t)^{a}}{1+(\lambda t)^{a}}$ $h(t) = k\dfrac{\lambda\rho(\lambda t)^{a-1}}{\left[1+(\lambda t)^{a}\right]\left[1+(1-k)(\lambda t)^{a}\right]}$	log(λ) = -4.13*** log(a) = 0.88*** logit(k) = -3.67*** LMV = -21199.32	log(λ) = -3.56*** log(a) = 0.52*** logit(k) = -4.09*** LMV = -22415.65
Gamma-Logistic	$F(t) = 1-\left[1-\kappa\log\left[\dfrac{\exp(-qt)-\alpha}{\exp(-qt)}\right]\right]^{-\frac{1}{\kappa}}$ $h(t) = \dfrac{\dfrac{q}{1-\alpha\exp(-qt)}}{1-\kappa\log\left[\dfrac{\exp(-qt)-\alpha}{\exp(-qt)}\right]}$	log(α) = -4.00*** log(q) = -2.63*** log(κ) = -0.03 LMV = -21275.45	log(α) = -3.16*** log(q) = -2.26*** log(κ) = -0.10*** LMV = -22635.22

(Table 1, cont'd)

Model	Cumulative function *F(t)* and hazard function *h(t)*	Estimates Men	Estimates Women
	$F(t)=1-\left[1-\kappa\log\left[\frac{\left(1+\frac{q}{p}\right)\exp[-(p+q)t]}{1+\frac{q}{p}\exp[-(p+q)t]}\right]\right]^{-\frac{1}{\kappa}}$	log(*κ*)=-0.20*** log(p)=-6.19*** log(*q*)=-2.84*** LMV=-21170.53	log(*κ*)=-0.19*** log(p)=-4.737*** log(*q*)=-2.80*** LMV=-22312.46
Gamma-Mixed Influence	$h(t)=\dfrac{\dfrac{p+q}{1+\frac{q}{p}\exp[-(p+q)t]}}{1-\kappa\log\left[\frac{\left(1+\frac{q}{p}\right)\exp[-(p+q)t]}{1+\frac{q}{p}\exp[-(p+q)t]}\right]}$		

Note: Reparametrization of the Hernes model by Wu (1990): λ=b, β=-A/log(b).

[a]Hernes, 1972; Wu, 1990.

[b]Brüederl & Diekmann, 1995.

* Significant at the level of 5 per 100.

** Significant at the level of 1 per 100.

*** Significant at the level of 1 per 1000.

Table 2

Estimated Coefficients of Models Taking into Account the Birth Year of Respondents

	Men			**Women**		
	External Influence	**Internal Influence**	**Susceptibility**	**External Influence**	**Internal Influence**	**Susceptibility**
Intercept (Born in 1939)	-6.42***	-2.79***	-0.14***	-4.81***	-2.78***	-0.21***
Born before 1939	0.90***	-0.17	-0.085	0.58***	-0.08	0.17
Born after 1939	-0.40	0.0427	-0.52	-0.51***	0.10	0.16
LMV		-21127.63			-22284.5	

*** Significant at the level of 1 per 1000.

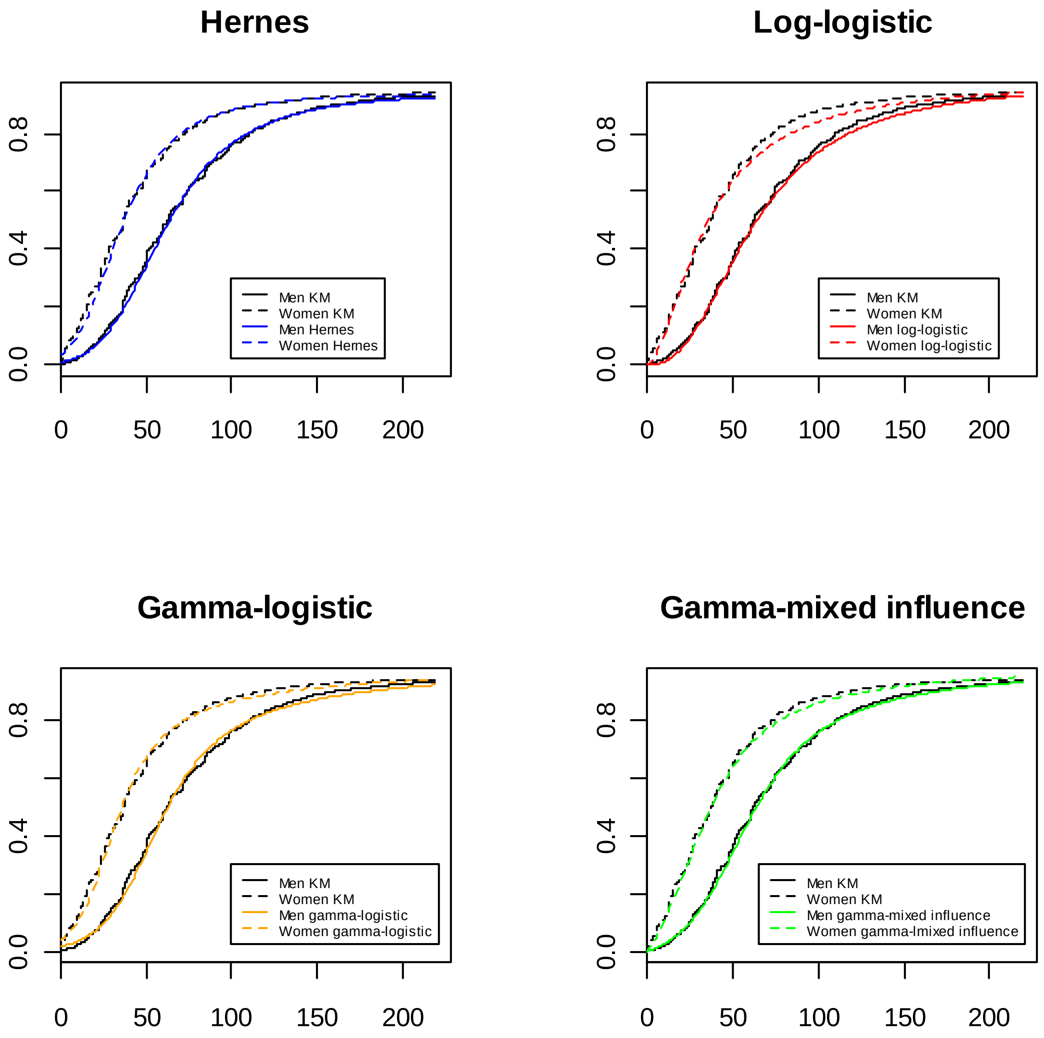

Figure 1. Fit of different diffusion models.