UNIVERSITÉ DE GENÈVE
FACULTÉ DES SCIENCES DE LA SOCIÉTÉ
Institut d'études démographiques et du parcours de vie

LIVES
Pôle de recherche national

Young adults' excess mortality: individual reality or yet another heterogeneity's ruse?

Adrien Remund
University of Geneva
Institute for Demographic and Life Course Studies
NCCR LIVES - Overcoming vulnerability: life course perspectives

June 2014

Abstract

Young adults excess mortality (YAEM) is a long known demographic fact, but its causes remain unclear. The common perception of YAEM presents it as an inevitable phenomenon stemming from the endogenous development of adolescents and the natural turmoil associated with puberty. Although this point of view has been challenged a long time ago by psychologists and anthropologists, it has retained credit, not only in the eyes of the public, but also amongst scholars from other fields. This endogenous hypothesis corresponds in demographic terms, to Vaupels level 1 explanation, i.e. that "the observed change is produced by a corresponding change at the individual level".

In a previous study, I showed that YAEM is not universal, since historically about one quarter of female and 10% of male populations ignore it. Moreover, the periods of absence of excess mortality are concentrated on specific historical contexts (1950s-1970s for women, the 1930 generation for men). Additionally, the causes of death show that the so-called accident hump has not always been one, and that before WW2 it was mainly caused by tuberculosis.

These observations cast doubt on the endogenous hypothesis, and suggest that YAEM could be produced by a level 2 explanation, namely that the force of mortality observed at the population level would be an artifact of a change in the structure of the population. In order to test the validity of this heterogeneity hypothesis, I designed a procedure that allows estimating the force of mortality of two sub-populations exempt of excess mortality resulting in an overall mortality that corresponds to YAEM as it is observed in real populations. The success of this first analysis allows me to pursue my research by using mixture models on micro data in order to distinguish covariates that may play the role of this so-far unobserved heterogeneity.

1 Introduction

The traditional perception of young adults' excess mortality is the one of an inevitable phenomenon stemming from the endogenous development of adolescents and the natural turmoil associated with puberty. In his seminal work (Hall, 1904), Hall highlighted specificities in the behaviours of adolescents, notably a peak of depression, criminality, a tendency to sensations seeking, and a sensitivity to media and peer pressure (Arnett, 2007)[187]. Later, psychoanalysis reinforced this vision of adolescence as a peculiar stage of development marked by "Storm and Stress", due to the resurgence of oedipian conflicts, and by nature source of turmoil (Freud, 1968). For some analysts, the absence of conflict would even be a sign of pathology (Arnett, 1999).

More recently, this perception of adolescence has been renewed by the neuro-psychological literature known in the field as the "adolescent brain" theory. According to its promoters, there are neurological signs of a peculiar

stage of development in adolescents' brain, including a bell-shape evolution of grey matter in frontal, parietal and temporal lobes (Giedd et al., 1999), which corresponds to a pruning process of less used synapses. Some authors have indicated that these changes are localised within regions known to be related to impulsion control, judgement and decision making (Lenroot and Giedd, 2006)[723]. Others have suggested that adolescents would be less able to assess risk because of their unsynchronised development of cognitive and emotional abilities (Steinberg, 2005). At this point, some scholars have bridged these neuronal observations with the earlier psychological theories of adolescence. This link is often suggested in simple qualifications of young adults, such as "adolescents are risk takers" (Spear, 2000)[421], or that "it comes as no surprise to parents of teens that the brain of an 8 year old is different than the brain of a 13 year old" (Giedd, 2004). Some have even extended the concept of adolescence to other species like apes or rats (Spear, 2000).

Taken as a whole, these theories have imposed three postulates according to which adolescence is (1) a separate stage in Life, (2) a universal phenomenon, and (3) systematically problematic (Sercombe, 2010). We will refer them together as the *endogenous* hypothesis. Although this point of view has been challenged a long time ago by other psychologists (Arnett, 1999; Offer and Schonert-Reichl, 1992; Petersen, 1993; Males, 2009a; Bessant and Watts, 2012) and anthropologists (Dasen, 2000; Choudhury, 2010; Mead, 1928), it has retained credit, not only in the eyes of the public, but also among scholars from other fields (Offer et al., 1981).

In demography, where interest for young adults is often limited to the concept of transition to adulthood, few mortality studies have explicitly focused on this particular question. Translated in demographic terms, the implications of the endogenous theory are that all individuals go through a phase of heightened risk taking, emotional turmoil, and social conflict due to the maturation process of their brain, which increases their individual risk of death. This hypothesis has recently served as an implicit justification for the study of secular trends in male sexual maturity, in which the author claims that "risk-taking and surplus mortality (the "accident hump") are signatures of the male human's early adult years" (Goldstein, 2011)[1]. This position suggests that the hump observed in young adults' mortality is due to the fact that all individuals, possibly at different levels, experience a phase of excess risk of death. This corresponds incidentally to Vaupel's level 1 explanation, i.e. that "the observed change is produced by a corresponding change at the individual level" (Vaupel and Yashin, 2006)[272].

This hypothesis can hardly be directly tested, given the latent nature of the individual risk of death. Indeed, there is no such measure as the individual probability of death; since Death is a dichotomous condition one can only say from an individual that he or she is dead or alive. Only aggregated measures can be constructed, for instance age-specific death rates, but this implies to assume a homogeneous population (i.e. that the mean rate applies to all individuals in the same way). Although the shape of the individual risk of death cannot be analysed, it is possible to test several corrolaries from the endogenous hypothesis, namely that

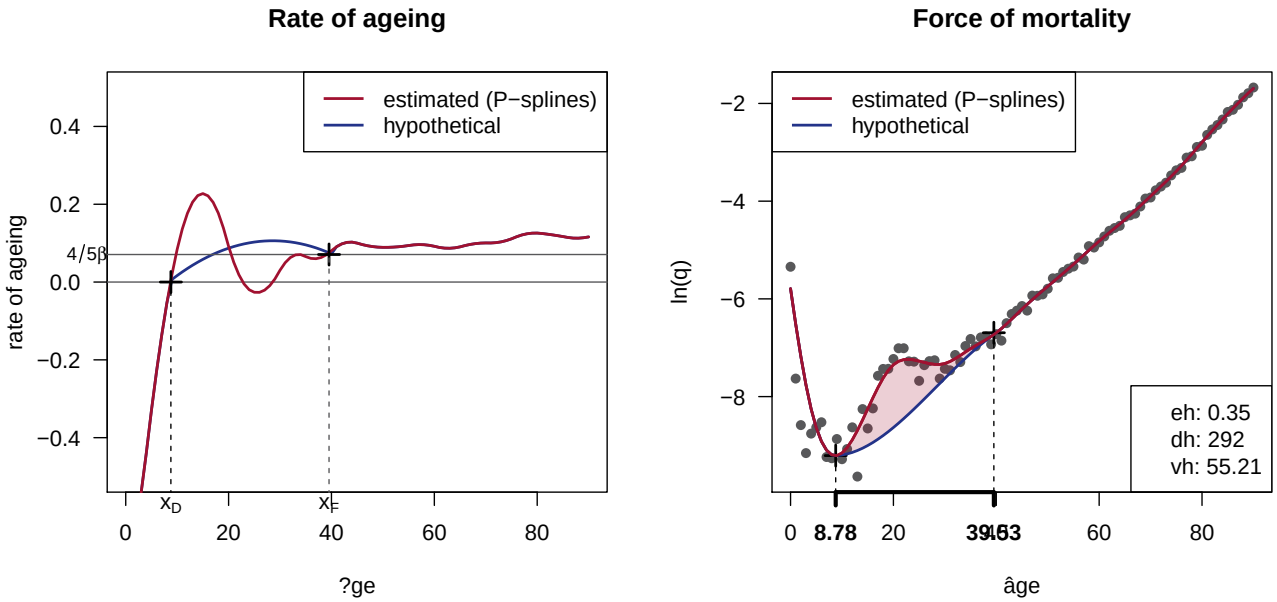
YAEM is a universal phenomenon that can be found in all human populations, since all individuals experience it.

YAEM peaks between 12 and 16 years of age, i.e. the age range during which grey matter undergoes a transformation on the MRI studies.

YAEM is due to 'external' causes of death (mainly accidents, suicides, and homicides), supposedly linked to the ability to assess and deal with risk taking.

Pursuiing this goal, I defined in a previous study young adults' excess mortality (henceforth *YAEM*) as "the deviation of the force of mortality from the evolution due to senescence" (Remund, 2012). I then suggested a measure of *YAEM* that consists in estimating the hypothetical trajectory of the force of mortality in the absence of *YAEM*, by interpolating the rate of ageing between the lowest point of the force of mortality (x_D) and the moment when it reaches a stable level (x_F). By comparing this hypothetical scenario to what is observed in reality, one obtains an estimation of how much life expectancy is lost to *YAEM*. For instance, in the case of Swiss males in 2005, *YAEM* decreased life expectancy by 0.35 years (eh), costed 292 lives (dh), of people who would have lived on average another 55 years (vh) (figure 1).

Figure 1: *YAEM* in a homogeneous population
Swiss males 2005



In order to test the three above statements, I applied this methodology on 10'000 life tables covering four centuries and 32 countries. The results of this first analysis reject all three statements, and hence cast doubt on the endogenous theory. First, *YAEM* is not a universal phenomenon, since about 25% of female and 10% of male populations did not display a significant mortality hump. Moreover, the periods of absence of *YAEM* are concentrated on specific historical contexts (1950s to 1970s for women, the 1930 cohort for men). This suggests that it is a highly socioeconomic and/or epidemiological phenomenon and hence entails at least an important exogenous component. Secondly, the location of the mortality peak is almost constantly 20 years, and the deviation of the force of mortality often goes beyond 40 years of age, discarding the neurological development as the main driving force of the mortality hump. Thirdly, an analysis on Swiss causes of death (Remund, forthcoming) showed that the so-called 'accident hump' has not always been linked to risk taking and/or emotional instability, but was in fact a 'tuberculosis hump' before WW2, and has only become associated with (traffic) accidents and suicides in the last half century.

This apparent failure of the endogenous theory leads us to consider another alternative explanation for *YAEM*, which does not assume that all individuals go through a phase of 'Storm and Stress'. Since at least the 1980s, demographers know how unobserved heterogeneity can trick us by creating at population level patterns that do not exist at the individual level (Vaupel and Yashin, 1985). This corresponds to Vaupel's level 2 explanation, i.e. that "the observed change is an artefact of a change in the structure of the population (i.e., a change in the composition of a heterogeneous cohort)" (Vaupel and Yashin, 2006)[272]. The main difference with the endogenous hypothesis is to relax the homogeneity assumption, and thus consider the possibility that the overall population is heterogeneous.

The aim of this paper is to evaluate the plausibility of the heterogeneity hypothesis, first by proposing an appropriate model design and an efficient estimation procedure, and then by applying it to a real population. If the premise of a heterogeneous population reproduced the hump observed in real populations, then the heterogeneous hypothesis could be called consistent. This would lead to additional analyses, this time necessary on micro data, to test whether this hypothesis is also sound.

2 Model design

In macro-level demographic analysis, populations are traditionnaly considered as homogenous, i.e. we assume individuals equal in their risk of death within a cohort. This is at least still the dominant approach used in traditional life table computation. However, since the end of the 1970s, this theoretical paradigm has been progressively challenged. The seminal work in this development is certainly the 1979 paper of Vaupel, Manton and Stallard

Vaupel1979, whose main value was to highlight the impact of individual heterogeneity on the dynamics of mortality.

Since then, numerous studies have taken over and expanded the concept of unobsetrved heterogeneity and individual frailty [1] Most of the use the same operationalization of frailty, namely a continous distribution of a proportional effect. The individual risk of death is hence related to the population risk of death in the following way.

$$\mu(x)_i = z_i \cdot \mu(x)$$

where z_i is an unobserved multiplicative coefficient indicating the relative probability of death of a the individual i, compared to the *standard* invidual for which $z = 1$ (Vaupel et al., 1979)[440]. In this model, individual frailty is considered as a constant over the life course, which is obviously a simplification. The usual justification of this simplification is that without this hypothesis the model becomes quickly very complex and its data requirements are irealistic.

When frailty is modelled using proportional hazards as here, the distribution of z is usually continous and taken from a gamma law of shape parameter k and rate parameter λ. In practice, one holds $\lambda = k$ in order to force the mean frailty to be equal to one ($\bar{z} = 1$). Other distribution classes have been studied elsewhere (**??**), but the gamma distribution seems to have become an undisputed standard.

Accelerated time failure (or *accelerated-ageing*) models represent the main alternative to proportional hazard models. Studies using ATF settings remain although rare, and when they exist, focus on old-age (**?**) or cause-specific mortality (**?**). The formulation of the model consists in applying the frailty term as a time coefficient, hence its name. Individuals do not differ by their baseline risk of death but by their rate of ageing. Algebaically, the individual probability of death becomes (Vaupel and Yashin, 2006)[273].

$$\mu(x)_i = \mu(z_i \cdot x)$$

where z_i is a multiplicative coefficient of time that captures the speed of deviation of the force of mortality compared to a standard individual for which $z_i = 1$. Other formulation sometimes combine both sources of frailty (**?**).

This double focus, on one hand on continous frailty, and on the other hand on proportional hazard, has negleted to a certain extend the variety of possible ways frailty can affect the relation between individual a population death rate. This is probably linked to the theoretical focus on old age mortality, specifically on plateau effects observed amongst supercentenarians (**??**). The variety of possible models has however been highlighted in one of the most cited articles in Demography (Vaupel and Yashin, 1985), in which the authors show how the existence of two sub-populations with strictly increasing forces of mortality can generate a mortality hump at the population level 2. Strangely enough, this idea has never been applied to the study of young adults' excess mortality.

Figure 2: Hump created by two diverging subpopulations
From Vaupel and Yashin 1985

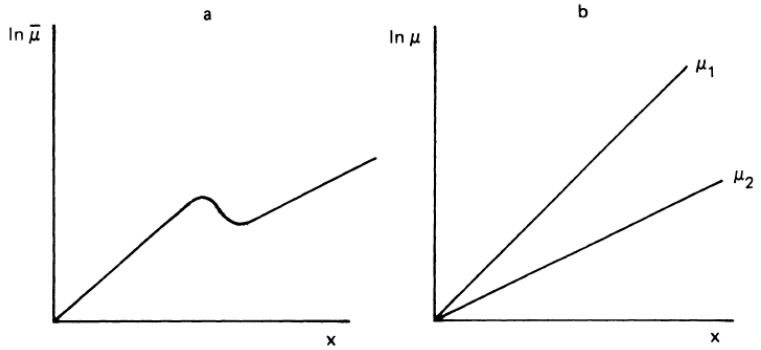


Figure 3. The observed hazard rate may rise steadily, then decline, and then rise again even though the hazard rates for the two subcohorts are steadily rising. The curve for $\bar{\mu}$ was calculated from (2), (3), and (4) using $\mu_1(x) = .0001 \cdot \exp(.2x)$, $\mu_2(x) = .0001 \cdot \exp(.1x)$, and $\pi(0) = .5$. The curves are shown for values of x from 0 to 75. Note that $\bar{\mu}$ and μ_i are plotted on logarithmic scales.

[1] For a general overview, see for example (Vaupel and Yashin, 2006).

Building on this idea, I will develop a discrete frailty model composed of at least two distinct but unobserved subpopulations, whose respective force of mortality diverge. My hypothesis is that this is a sufficient assumption to recreate the real hump observed at the population level. This rests on the following process. From the bottom of the age-specific death rate, at around ten years of age, the mean force of mortality is pushed up by the existence of a small vulnerable subpopulation. As time goes by, the individuals composing this subpopulation die out faster that the ones of the main subpopulation, and the overall force of mortality progressively flattens, and possibly decreases, until it converges to the one of the main subpopulation. Doing so, the overall force of mortality reflects a hump even though none of the subpopulation, and hence its individuals, experiences one.

Treating frailty as a discrete variable that indicates membership to certain subpopulations comes down to treating agregated mortality as a mixture distribution. In statistics, this phenomenon is analysed by means of mixture models, which postulate the existence of subpopulation within the general population. Applied to survival analysis, mixture models have given birth to variants of the Cox model (**?**), such as cure models (**?**).

Let us assume that each of these subpopulations follow a Gompertz law, i.e. its force of mortality progresses exponentially with age. This assumption is common in Demography and allows to rull out any hump at the individual level. This might be too restrictive but is the best way to prove that the endogenous hypothesis is not necessary. Each force of mortality is hence defined by two parameters, α and β, which control respectively the baseline mortality rate and the rate of ageing. Algebraically,

$$\mu_1(x) = \alpha_1 \cdot e^{\beta_1 \cdot x} \text{ et } \mu_2(x) = \alpha_2 \cdot e^{\beta_2 \cdot x}$$

Another advantage to the Gompertz model is that the survival probability of each subpopulation has a closed form as follows.

$$l_1(x) = e^{\int_0^x -\mu_1(t)dt} \text{ et } l_2(x) = e^{\int_0^x -\mu_2(t)dt}$$

$$l_1(x) = e^{-\frac{\alpha_1}{\beta_1} \cdot (e^{\beta_1 \cdot x} - 1)} \text{ et } l_2(x) = e^{-\frac{\alpha_2}{\beta_2} \cdot (e^{\beta_2 \cdot x} - 1)}$$

One obtains the agregated force of mortality by computing the weighted mean of the two subpopulations, using as weights their respective proportion in the surviving population. This proportion not only depends on the survival probility within each subpopulation, but also from their initial share in the overall population, henceforth π_1 et π_2. The overall force of mortality is hence expressed by the following function:

$$\bar{\mu}(x) = \frac{\mu_1(x) \cdot l_1(x) \cdot \pi_1 + \mu_2(x) \cdot l_2(x) \cdot \pi_2}{l_1(x) \cdot \pi_1 + l_2(x) \cdot \pi_2}$$

Once the Gompertz law taken into account, one obtains the following fomula linking the observed mortality rates to the overall force of mortality.

2 subpopulations[2]

$$\begin{aligned} M_x &= \bar{\mu}(x) + \epsilon_x \\ &= \frac{\mu_1(x) \cdot l_1(x) \cdot \pi_1 + \mu_2(x) \cdot l_2(x) \cdot \pi_2}{l_1(x) \cdot \pi_1 + l_2(x) \cdot \pi_2} + \epsilon_x \\ &= \frac{\alpha_1 \cdot e^{\beta_1 \cdot x} \cdot e^{-\frac{\alpha_1}{\beta_1} \cdot (e^{\beta_1 \cdot x} - 1)} \cdot \pi_1 + \alpha_2 \cdot e^{\beta_2 \cdot x} \cdot e^{-\frac{\alpha_2}{\beta_2} \cdot (e^{\beta_2 \cdot x} - 1)} \cdot \pi_2}{e^{-\frac{\alpha_1}{\beta_1} \cdot (e^{\beta_1 \cdot x} - 1)} \cdot \pi_1 + e^{-\frac{\alpha_2}{\beta_2} \cdot (e^{\beta_2 \cdot x} - 1)} \cdot \pi_2} + \epsilon_x \end{aligned}$$

This model can be generalised to higher orders if the data suggest the need of more subpopulations. Adding subpopulations can be useful when the agregated force of mortality show signs of deviation to the Gompertz trend,

[2] Although this formula contains six parameters, only five of them need to be estimated, since π_2 can be expressed as the reciprocal of π_1, i.e. $\pi_2 = 1 - \pi_1$.

others than young adults' excess mortality. One can imagine for instance a bimodal hump or displaying a wider dispersion after its peak (Kostaki, 1992), or a bend in the force of mortality later in life. In order to test the rlevance of increasing the number of subpopulations, I also constructed a 3-subpopulation model, whose algebraical expression is as follows.

3 subpopulations[3]

$$\begin{aligned} M_x &= \bar{\mu}(x) + \epsilon_x \\ &= \frac{\mu_1(x) \cdot l_1(x) \cdot \pi_1 + \mu_2(x) \cdot l_2(x) \cdot \pi_2 + \mu_3(x) \cdot l_3(x) \cdot \pi_3}{l_1(x) \cdot \pi_1 + l_2(x) \cdot \pi_2 + l_3(x) \cdot \pi_3} + \epsilon_x \\ &= \frac{\alpha_1 \cdot e^{\beta_1 \cdot x} \cdot^{-\frac{\alpha_1}{\beta_1} \cdot (e^{\beta_1 \cdot x} - 1)} \cdot \pi_1 + \alpha_2 \cdot e^{\beta_2 \cdot x} \cdot e^{-\frac{\alpha_2}{\beta_2} \cdot (e^{\beta_2 \cdot x} - 1)} \cdot \pi_2 + \alpha_3 \cdot e^{\beta_3 \cdot x} \cdot^{-\frac{\alpha_3}{\beta_3} \cdot (e^{\beta_3 \cdot x} - 1)} \cdot \pi_3}{e^{-\frac{\alpha_1}{\beta_1} \cdot (e^{\beta_1 \cdot x} - 1)} \cdot \pi_1 + e^{-\frac{\alpha_2}{\beta_2} \cdot (e^{\beta_2 \cdot x} - 1)} \cdot \pi_2) + ^{-\frac{\alpha_3}{\beta_3} \cdot (e^{\beta_3 \cdot x} - 1)} \cdot \pi_3} + \epsilon_x \end{aligned}$$

The originality of this model is that discrete frailty has not been used yet to model mortality humps, as far as we are aware of. Some limitations ought to be mentioned though. First, this model should only be applied to cohort data, to avoid mixing age and period effects (see e.g. Frost, 1939). Secondly, this model does not consider migration either between subpopulations or outside the overall population. In other words, it is not possible to jump from the vulnerable to the main population. Other models have been suggested to overcome this limitation and allow people to change their frailty level over time (Vaupel et al., 1988; **?**; **?**; Vaupel and Yashin, 2006), but they introduce a second dimension to the problem which is the transition risks between frailty levels, which make them quasi impossible to estimate as they have no closed form. There is indeed an infinity of possible combinations of transition and death rates capable of creating the oberved rates.

A by-product of this model is that is allows estimating the life expectancy lost to young adults' excess mortality in an alternative way to a homogenous population. It is indeed possible to estimate the life expectancy of each subpopulations, as well as the overall population. The difference between latter, $\bar{e_0}$, and the fittest of the two subpopulations represents what would have been gained if all individuals belonged to the more robust subpopulation. In other words, this estimates the potential gains in life expectancy of avoiding individuals to follow the frailer path, assuming that this is a reasonable public health achievement.

3 Shape of the objective function

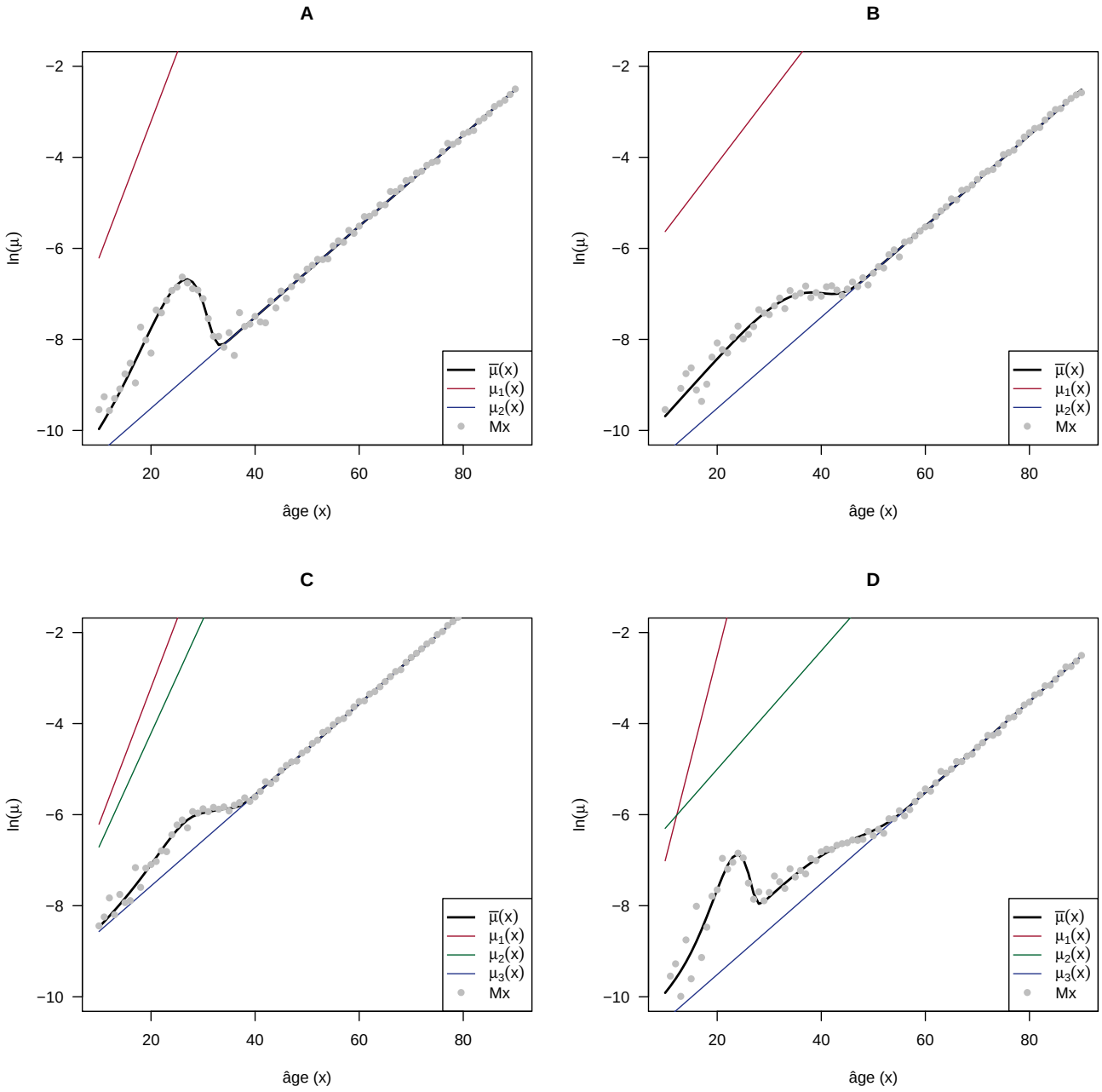
The estimation of the parameters can be made by the weighted least square estimator. In order ot correct the heteroskedasticity crated by the wide range of levels of mortality rates, it has been suggested to use the inverse of the number of deaths as weights. According to (Brillinger, 1986), these are approximately equal to the variance of the residuals. The objective function is thus

$$RSSw = \sum_x (M_x - \mu(x))^2 \cdot \frac{1}{D_x}$$

In order to get a feeling of the model, we will test it on simulated data before applying it to real populations. The main advantage is that in this case the real parameter values are known. Four artificial dataset were created to reflect the variety of possible shapes of the force of mortality (figure 3). The first one (A) uses two subpopulations to depict the simplest mortality hump. The second one (B) also uses two subpopulations, but whose rate of ageing are much closer. The effect is that the hump resembles more a plateau after its peak. The third one (C) postulates the existence of three subpopulations, of which two of them form together the vulnerable share of the population. This can generate bimodal humps or plateau effects similar to (B). The last simulation (D) also uses three subpopulations, but this time only one of them is strongly diverging from the main population. The third population essentially acts here as a bending effect on the force of mortality at later ages.

[3]Although this formula contains nine parameters, only eight of them need to be estimated, since π_3 can be expressed as the reciprocal of $\pi_1 + \pi_2$, i.e. $\pi_3 = 1 - \pi_1 - \pi_2$.

Figure 3: Four simluated datasets based on the heterogeneity hypothesis



The real parameters are shown in table 1. Once the overall forces of mortality computed, stochasticity is introduced by generating random deviates of a binomial distribution, with probability equal to $\bar{mu}(x)$ and trials equal to the age-specific male population size of Switzerland in 2010. The "observed" mortality rates are then computed by dividing those artificial observed death counts by the same population under exposure. This insures a realistic varability in the data, as well as a similar variability in each datasets.

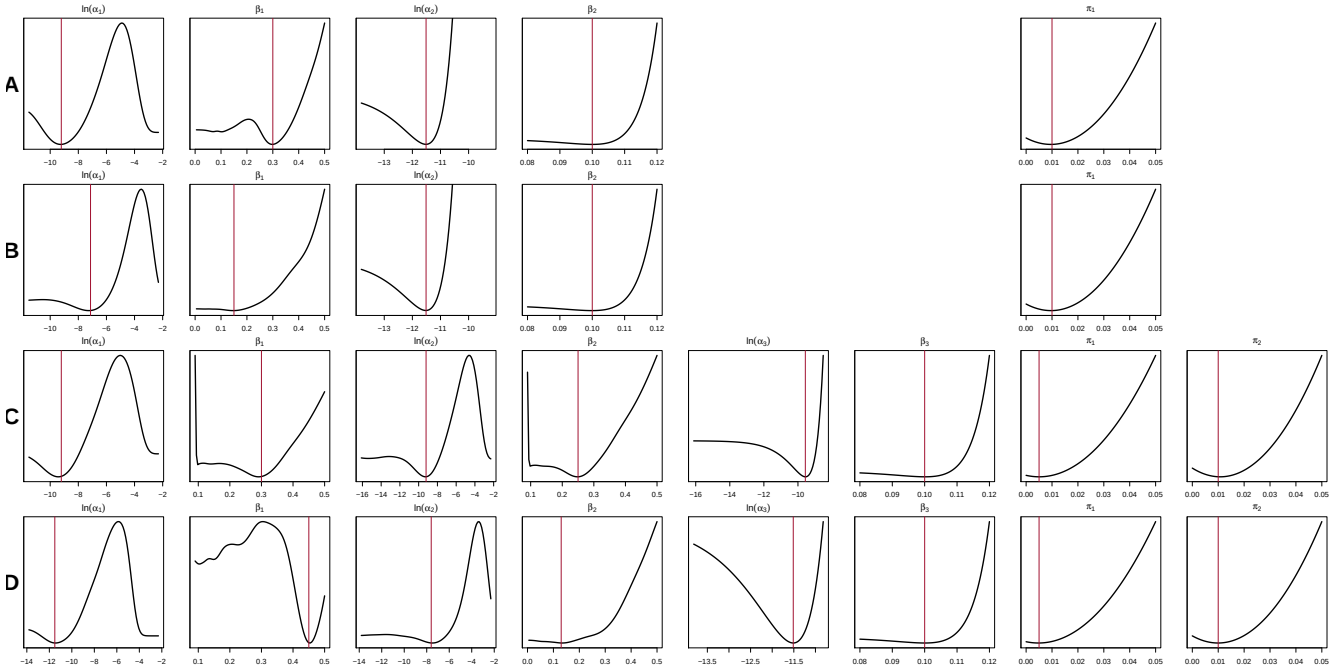
	a1	b1	a2	b2	a3	b3	p1	p2
A	1.0E-04	0.30	1.0E-05	0.10			0.010	
B	8.0E-04	0.15	1.0E-05	0.10			0.010	
C	1.0E-04	0.30	1.0E-04	0.25	7.0E-05	0.10	0.005	0.010
D	1.0E-05	0.45	5.0E-04	0.13	1.0E-05	0.10	0.005	0.010

Table 1: True parameters for each simulation

A first advantage of these simulations is to be able to study the shape of the objective function in the neigh-

bouring of the solution. Since the parmeter space is too big to be represented graphically, one can represent the one-dimensional evolution of the objective function by holding the other parameters to their true values. This way, one gets a fair idea of the shape of the function around the true values, which is particularely important to see whether it is convex. Indeed, most of the classical estimation algorithms use this assumption to find the values of the parameters that minimize the objective function. Figure 4 reprensents this function for each dataset (rows) and each parameter (columns).

Figure 4: Evolution of RSSw around the true values for each dataset



This first analysis shows that the weighted least squares function is not stricly convex for several parameters. Consequently, there are multiple local minimas in addition to the global minimum. The most problematic cases are α_1 for all datasets, as well as α_2, β_1 and β_2 for the 3-subpopulation models. Although the other relations are convex, let us not forget that there representations are only one-dimensional simplifications of a five-, respectively, eight-dimension problem. In other words, there is no garantee that additional local minimas do not exist outside those already identified, further from the true values.

These results are not so surprising given the analytical expression of the objective function. Let us recall that in order to obtain a convex relationship, the gradient of each parameter (i.e. the partial derivatives of the objective function with respect to each parameter) should be strictly increasing. Since each parameter appears both in the numerator and in the denominator of $\bar{\mu}(x)$, this hypothesis seems highly unlikely. We shoud hence conclude that the weighted least square function is probably not convex in all dimensions of the parametes space. The problem of local minimas is thus real and therefore the quetion of the optimisation algorithm is important.

4 Comparison of different optimisation algorithms

We have seen that the convexity hypothesis is highly doubtful in our model. It is therefore important to select an optimisation algorithm that is capable of dealing with such a problem. For this reason, let us compare three algorithms, namely the Levenberg-Marquardt (LM) algorithm, the Simulated Annealing (SANN) technique, and the Differential Evolution (DE) metaheuristic.

At the heart of this problem is the choice of starting values for the parameters, which can have a dramatic influence on the capacity of the algorithm to find the global minimum. The more those values are far from the true ones, the more probable it is that the algorithm will get stuck in one of the local minimas. As we do not want to impose strong limitations to the parameter values, the choice of starting values becomes more difficult. We will

only use logical boundaries for the parameters, defined to limit the parameter space to meaningful values (table 2).

Table 2: Limits to each parameters

		a1	b1	a2	b2	a3	b3	p1	p2
2 sous-pop.	min	1e-06	0.05	1e-06	0.05			0.001	
	max	0.01	0.5	0.01	0.5			0.1	
3 sous-pop.	min	1e-06	0.05	1e-06	0.05	1e-06	0.05	0.001	0.001
	max	0.01	0.5	0.01	0.5	0.01	0.5	0.1	0.1

Amongst all possible tests, we will focus on the capacity of the three algorithms to overcome local minimas whatever the starting values chosen inside the above limits. This is why we generated 100 random initial values from a uniform distribution bounded by the same limits. Given those initial values, each algorithm was run 100 times and the values of the objective function were compared. We also tested other aspects of the algorithm, but the conclusions remained the same.

4.1 Levenberg-Marquardt algorithm

Levenberg Levenberg1944 and Marquardt Marquardt1963 suggested an algotithm based on a mix between Taylor series and the steepest-descent method. In short, the proposed method "shares with the gradient methods their ability to converge from an initial guess which may be outside the region of convergence of other methods. The algorithm shares with the Taylor series method the ability to close in on the converged values rapidly after the vicinity of the converged values has been reached. Thus, the method combines the best features of its predecessors while avoiding their most serious limitations" (Marquardt, 1963)[441].

This method is available in R under the *minpack.lm* package that provides an alternative to the default Gauss-Newton algorithm of the nonlinear least squares *nls* package. Its promoters praise its ability to converge from worse initial values and thuse avoid better local minimas (Spiess, 2012).

Its results are however poor on our four datasets. Out of the 100 attempts, less that a third converge to any value (table 3). When the estimation did not converge, the initial parameters were on average further away from the true values, which is a sign of the sensitivity to starting values[4].

Table 3: Nombre de convergences russies et Somme moyenne des erreurs relatives (SER) des paramtres initiaux

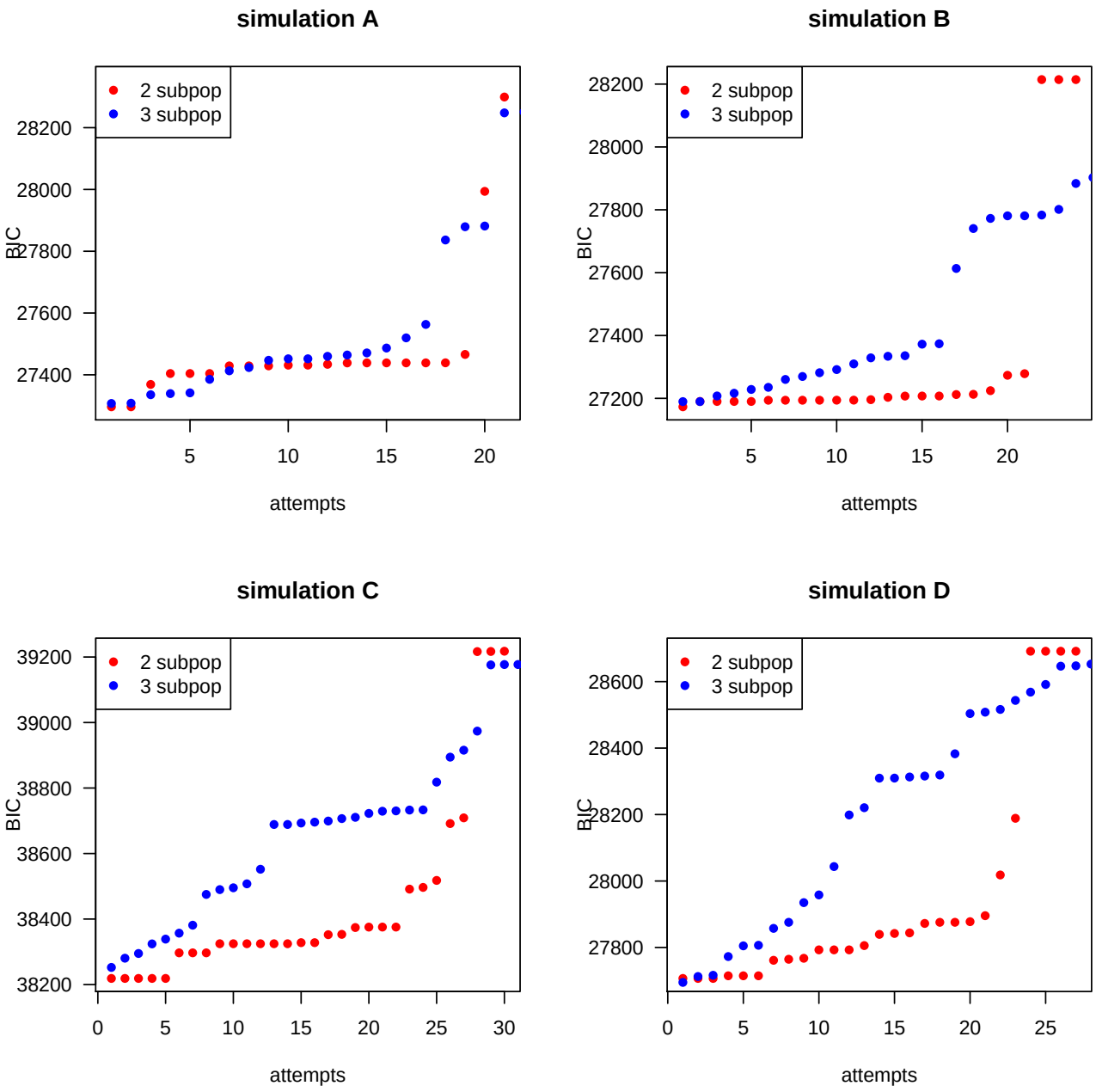
	conv	conv	SER	SER
	2 ss-pop	3 ss-pop	succs	chec
A	21	33	551	536
B	24	32	558	501
C	30	37	185	160
D	27	39	987	944

When the algorithm converges, the results are not even that good. Only a fraction of the attempts manages to reach the value of the objective function computed with the true values[5]. Figure 5 represents the level reached by each attempt on the scale of the BIC, so that the 2-subpopulation and the 3-subpopulation solutions can be compared. On all the datasets, the algorithm fails to avoid local minimas, which can be seen fairly obviously through the flat steps when the attempts are sorted.

[4]
$$SER = \sum_j |\frac{\theta_j}{r_{\theta_j}} - 1|$$
where θ_j is the true value and r_{θ_j} is the initial value of the parameter j

[5]That some solutions are better than the true values is due to the introduction of random variability.

Figure 5: Value of the BIC reached by the Levenberg-Marquardt algorithm on the four artificial datasets



This has serious consequences in the ability to chose between the two models with 2 or 3 subpopulations, due to the incertainty attached to the minimum value of the BIC for each model. Indeed, even in the first dataset, there are plenty of values for the 3-subpopulation model that are better than the worst one reached by the 2 subpopulation model. The probability to chose the wrong model is thus very high. In conclusion, the LM algorithm, as well as probably any other gradient-based deterministic algorithm, is not a good solution to estimate our model.

4.2 Simulated Annealing algorithm

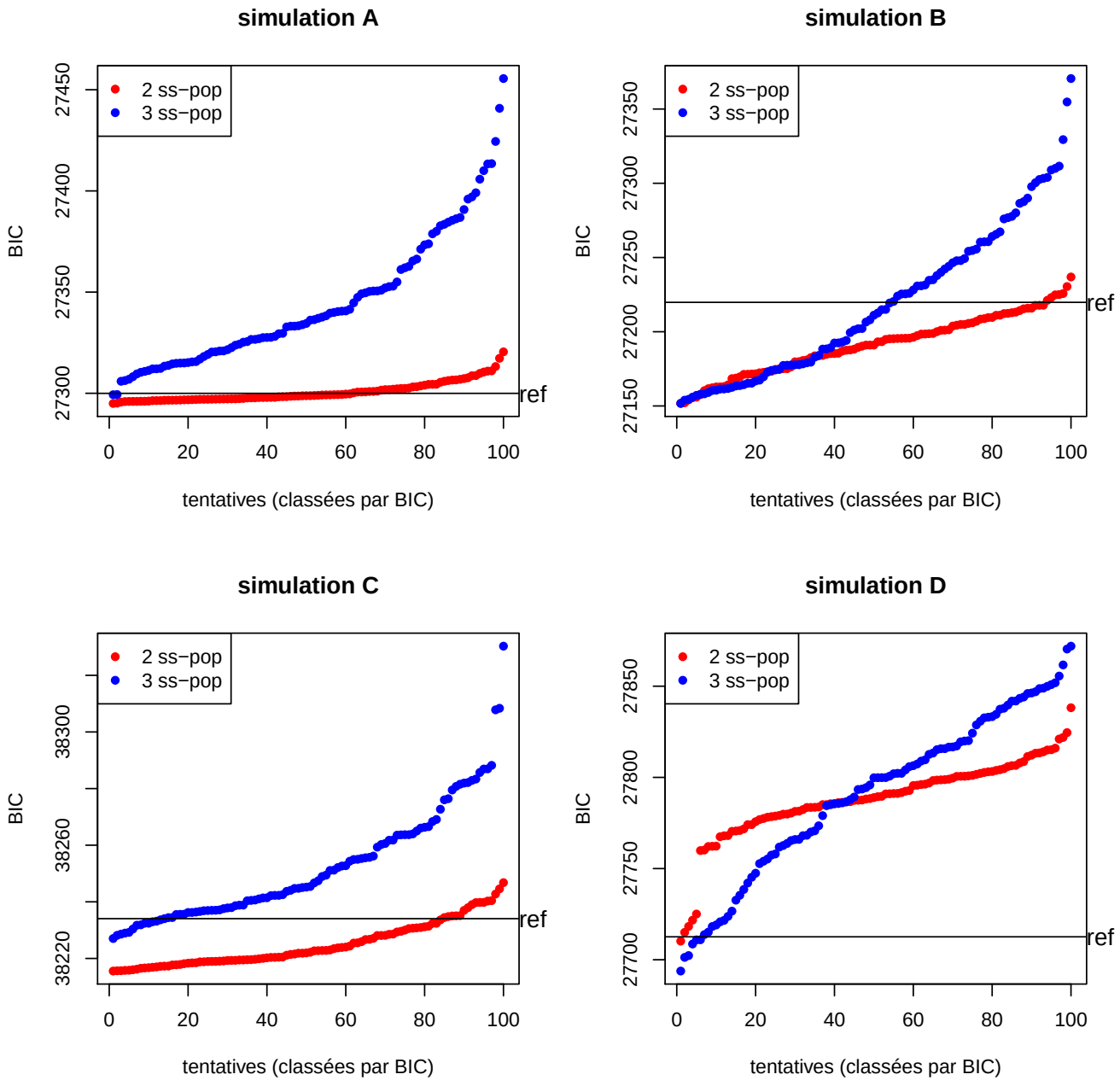
Since the begining of the 1980s, several stochastic algorithm have been developped. One of the most famous one is the Simulated Annealing (SANN) technique, which is inpired by the metalurgic process that consists in succesively heating and cooling the metal in order to change its cristaline structure. In 1983, Kirpatrick et al. suggested to use this princip to build a stochastic optimisation algorithm. In short, it allows worse solutions to be preserved in a first phase, but as the number of iteration increases (symbolised here by the cooling process), the probability to

do so decreases. Eventually, the process should converge to the global minimum.

In R, the SANN method is available in the *GenSA* (Generalized Simmulated Annealing) package. It is konwn to fare well against other methods (**??**), particularely when the dimension of the problem is big (**?**). For this exercise, we kept the defaults values of the tuning parameters, $q_v = 2.62$ et $q_a = -5$, following the recomendations of the authors (**?**)[15].

All the 100 attempts converged for all four datasets, which is in itself an improvement. However, all these attempts did not reach the general optimum (figure 6). What is striking here is the absence of steps as for the LM algorithm, which can be interpreted as a sign of less sensitivity to local minimas. However, since all the attempts do not reach the same (global) optimum, there seems to be an issue with the termination criterion.

Figure 6: Value of the BIC reached by the SANN algorithm on the four artificial datasets



The spread of optimal values reached by the 100 estimations does not allow to make clear decisions on the optimal number of subpopulations. For instance, if we were to pick randomly two values of the BIC reached respectively

by the 2-subpopulation and the 3-subpopulation models, the probability to chose the correct model would only be 63%. This probability is even worse for the datasets B (6%) and C (45%).

An additional weak point of the SANN algorithm is its computing time (**??**). Even though the GenSA pacakge uses a modified algorith supposedly quicker to converge, each estimation takes about 50 seconds for the 2-subpopulation model and 80 seconds for the 3-subpopulation model. The SANN algorithm fare thus better than the LM algorithm, but not well enough to sucessfully estimate our model independently from the chose starting values and in a reasonible time.

4.3 Differential Evolution algorithm

Alternatives to computational optimisation algorithms, known commonly as metaheuristic methods, have been gradually introduced over the last twenty years. Some of them (often referred as 'swarm intelligence') are inspired from biological processes, and usually mimic decentralized and self-organized systems observed in the animal world such as ant colonies (Dorigo and Birattari, 2010), bee hives (Pham et al., 2005), bat colonies (Yang, 2010), or wolf packs (Mirjalili et al., 2014). Alongside these techniques, another branch of optimization algorithms found its inspiration in evolutionary processes. Evolutionary algorithms were suggested several decades ago ago (Mitchell, 1998) and have developed in different branches, which share the same basic rationale.

1. generate a random initial population within a certain solution space
2. evaluate the fitness of each individual of this population
3. repeat the following steps until a fixed criterion (e.g.. number of iterations)
   (a) select the best individuals (parents)
   (b) breed new individuals through crossover and mutation (children)
   (c) evaluate the fitness of the offspring

Since the mid-1990s, differential evolution (DE) optimisation has emerged as a particularly promising subfield and has been applied to numerous contexts (Price et al., 2005). Its principal improvement over other evolutionary algorithms consists in estimating potential offspring for each member of the current population, by combining other members of the current population, and only keep those candidates that have a better fit than the member in question (Storn and Price, 1997)[346][6].

This algorithm has three main parameters, CR, F and NP. CR stands for the *crossover probability* and regulates the probability to perform the crossover between different members of the population, as opposed to keep the current member as it is in the next population. F stands for *differential weight* and represents the weights of the parents in the formation of the child. Finally, NP represents the (constant) size of the population at each generation. Although much attention has been given to the selection of the best parameters (Pedersen, 2010; Das and Suganthan, 2011), the choice largely depends on the function to be minimised. A rule of thumb is that NP should be about ten times the number of parameters, CR should be between 0.8 and 1, and F between 0.5 and 1 (Storn).

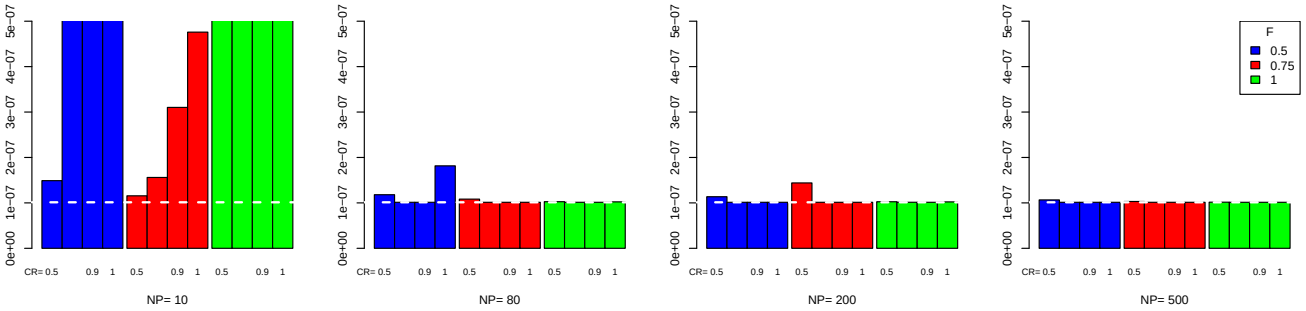
The DE algorithm has several advantages over classical methods of the Gauss-Newton family and even other metaheuristic methods (**?**Mullen et al., 2011). It converges more often, avoids more efficiently local optimums, does not require a derivative of the objective function, nor even a continuous objective function. It also permits a custom definition of the objective function, which is in our case the weighted residual sum of squares, and allows constraints that combine several parameters. This is particularly useful in the case of the 3-subpopulation model, for which $\pi_1 + \pi_2 \leq 1$. Recently, the *DEoptim* package has been developed to perform differential evolution optimisation in R.

The first task is to determine the most efficient values of the three parameters NP, CR and F. With this goal in mind, an optimisation was run for each combination of NP, CR and F, using the following range of values:

[6]This part of the crossover is called *differential mutation* and is what makes it stand out from classical evolutionary algorithms. Different strategies are described for the combination of members (or *crossover*). The classical one is referred in the literature as `DE / rand / 1 / bin`, whereas the favoured one seems to be `DE / local-to-best / 1 /bin`. The first one selects the parents at random, whereas the second one uses a combination of random and currently-best members. (Mullen et al., 2011)

$NP \in \{10, 80, 200, 500\}$, $CR \in \{0.5, 0.8, 0.9, 1\}$ and $F \in \{0.5, 0.75, 1\}$. All estimations were stopped after 200 iterations (generations), which is the default value. As predicted by the literature, a population size (NP) inferior to ten times the number of parameters seems to be too low to reach the optimum (figure **??**). On the contrary, greater population sizes do not seem to improve the quality of the estimation, even though they considerably slow down the process. Indeed, the average computation time increases proportionally to the population size, from 0.4 (NP=10) up to 19 seconds (NP=500), which does not advocate the use of $NP > 80$. Focusing then on the values for $NP = 80$, values for F and CR of respectively 0.75 and 0.9 seem to be good candidates.

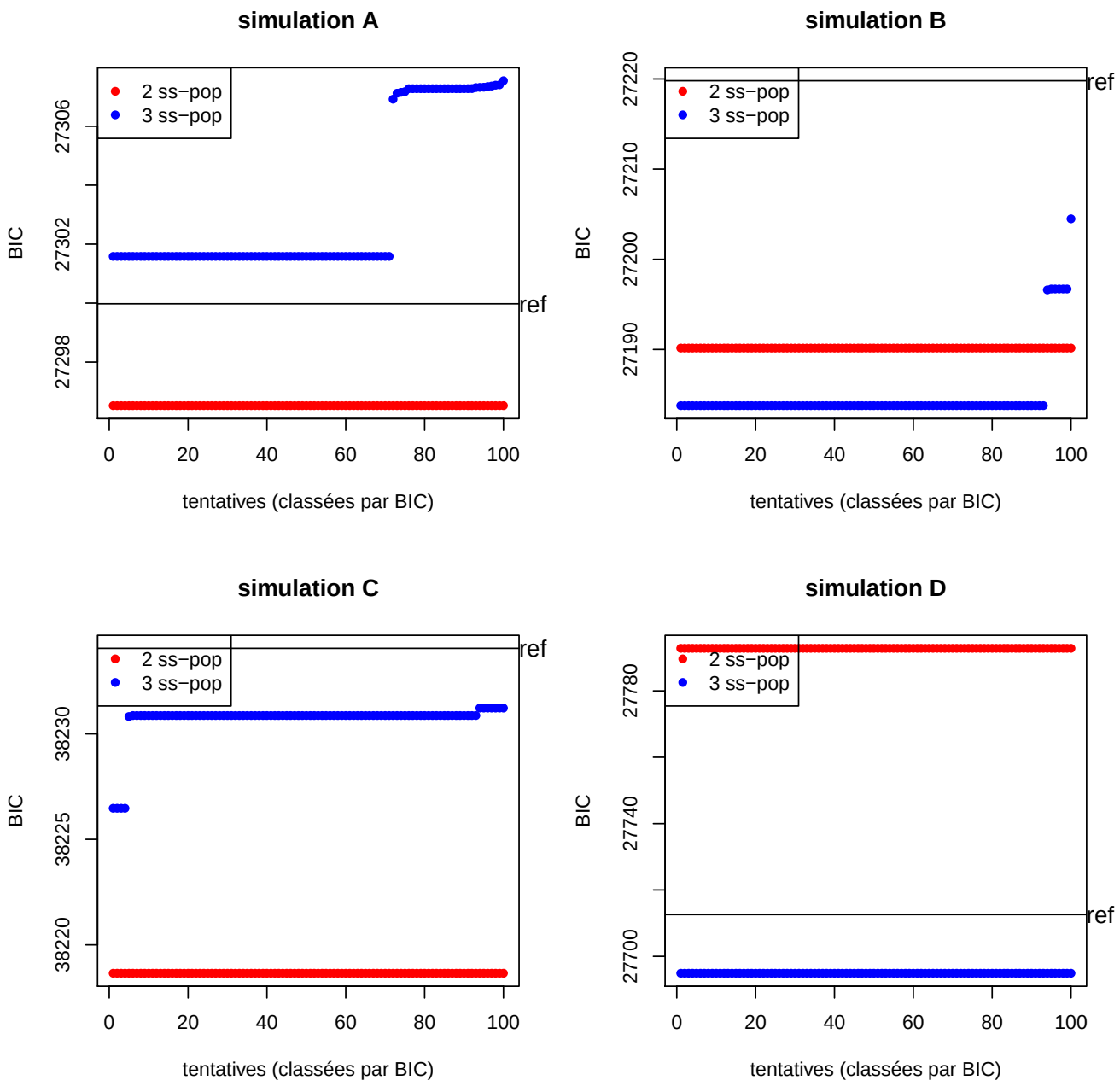
Figure 7: Residual Sum of Squares after 200 iterations using DE algorithm



As predicted by the literature, a population size inferior to 10 times the number of parameters does not allow the algorithm to reach the global optimum (figure 7). However, much larger values do not seem to bring an improvement and moreover quickly slow down the computation. With a reasonible population size (NP=80), the computation time remains good (2 seconds) without jeopardizing the fit, as long as $F > 0.5$ and $CR > 0.9$. The default values seem thus to be satisfying.

The capacity of the DE algorithm to reach the global minimum is very high. As for the SANN technique, all attempts converge, but this time on almost always the same value (figure 8). When the algorithm stays stuck on a local minimum, it is never far from the best solution. For instance, the locla optimum in which the 3-subpopulation model is stuck with the dataset A is only 3 units away from the global minimum, compared to tens or hundreds of units with the other algorithms (figures **??** et 6).

Figure 8: Value of the BIC reached by the DE algorithm on the four artificial datasets

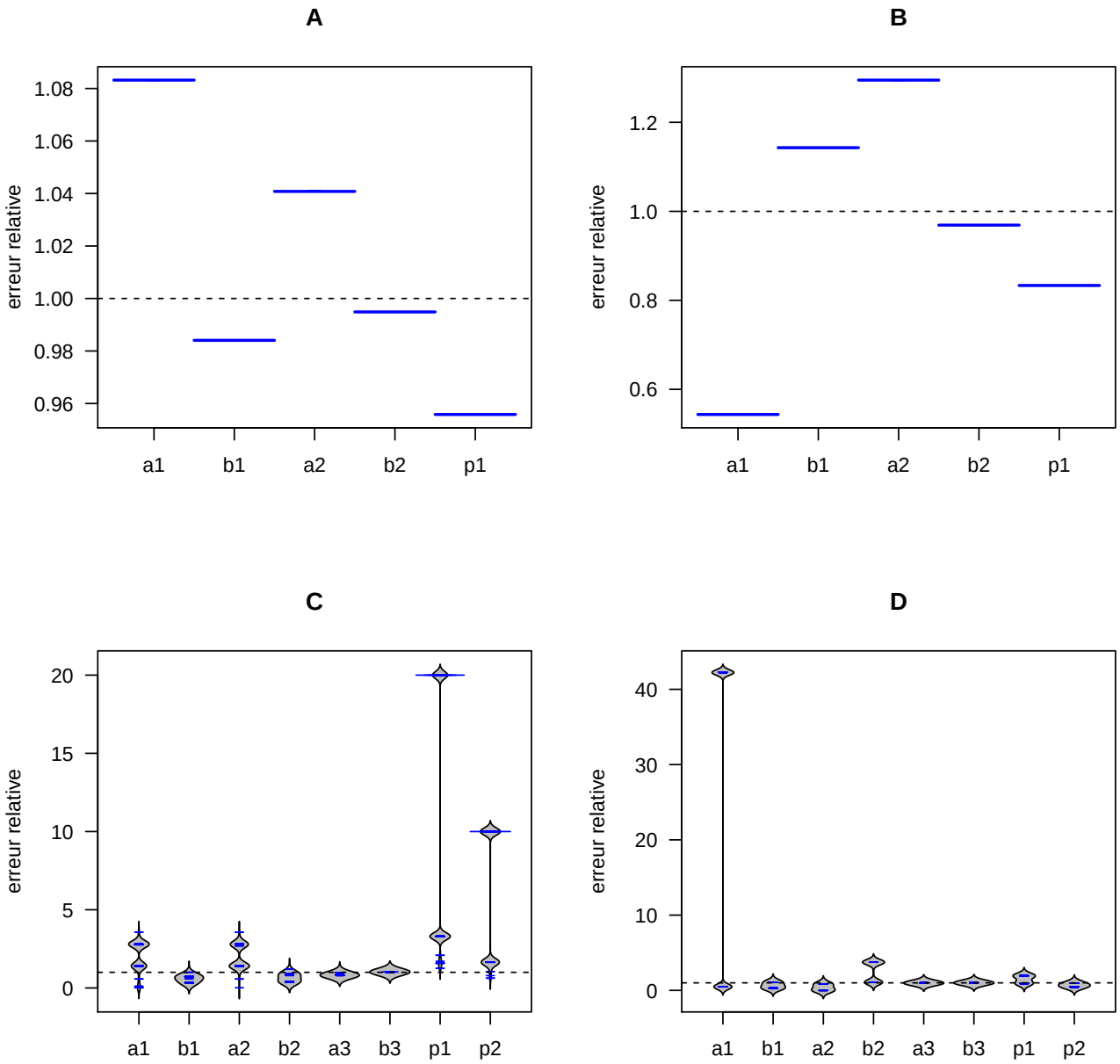


A direct consequence of this ability to systematically reach the global optimum is that it becomes much easier to chose the optimal number of subpopulations. The probability to chose the correct model reaches 100% for the datasets A and D. As for the two other datasets, the two models reach very close results, which are not the anticipated ones. For the dataset B, the 3-subpopulation model reaches a better BIC, even though the simulation used only 2 subpopulations. For the dataset C, this is the opposit.

This calls into question the simple use of the BIC as the criterion to chose the optimal number of subpopulations. One could think of using the AIC, or the AICc, but none of those alternative really change the results. A better idea is to use MCMC simulation, by generating random values of the Mx centered on the observed ones, repeat the estimation for both models, and compare them pairwise. This should allow to eliminate the possibility that a more complex model is favoured only because of its ability to capture stochastic variation. In the case of the dataset B, 10 random datasets each time indicated a better BIC for the 2-subpopulation model. This suggests that the superiority of the more complex model was not justified by the deterministic part of the data.

The parameters are also very well estimated. The relative error is small in each dataset and, more important, all attempts tend to cluster around the same values. What could seem like a large variation for the datasets C and D, is simply the sign that sometimes the two small subpopulations are inverted. Figure 9 clearly shows that the solution cluster around two sets of values that are simply the reciprocal of each other.

Figure 9: Estimation quality of the parameters with the DE algorithm



Finally, the DE algorithm is much quicker that the SANN. Each estimation takes about 10 seconds for the simpler model, and 14 seconds for the complexe one. Although, by decreasing the number of artificial generations, one could expect to bring down those figures to 2 and 5 seconds respectively, without compromising the quality of the fit. These results taken as a whole conclude that the DE algorithm is clearly superior to its alternative in this particular case. It is the only one that is capable of estimating our model.

5 Data

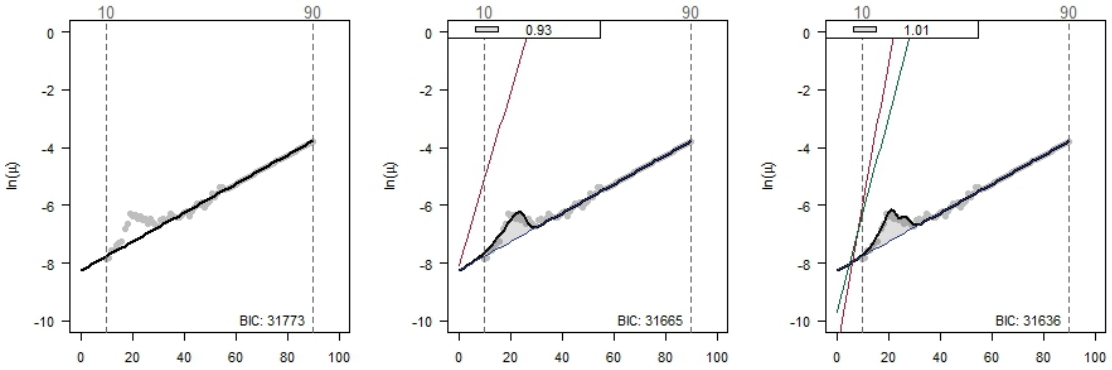
Now that the model has been tested on simulated data, it is possible to apply it on real observations. The following analyses were performed on data collected from the Human Mortality Database. Cohort age-specific exposures and rates by sex were extracted. The observations were left-truncated until the age of the lowest mortality rate in order to avoid the phase of decreasing mortality during childhood, i.e. ontegenescence (Levitis, 2011) [7] The data was also right-censored after age 90 in order to avoid the mortality plateau usually observed at old age (Thatcher et al., 1998).

For unextinct cohorts, unknown mortality rates were extrapolated using a Gompertz law. For cohorts born before 1960, the two parameters α and β were inferred from the known portion of the mortality curve from age 30 and over. For younger cohorts, whose trajectory is too short to infer a reliable rate of ageing for senescence, a semi-fixed method was used by setting $\beta = 0.1$ and estimating α on the three last known values of M_x in order to assure the continuity of the series.

6 Application to the Swiss case

Let us first take one real population and apply the above procedure. The Swiss cohort of males born in 1955 exibit a classical hump pattern, approximately symetrical and centered on age 20. The estimation of a single-population, 2-subpopulation and 3-subpopulation models indicates that the two last ones obtain very close fits 10. The difference in BIC being only of 4 units, it is reasonible to use a MCMC simulation to test if this is really supported by the deterministic part of the data. This procedure shows that 8 times out of 10, the 2-subpopulation models gets out with a lower value of the BIC.

Figure 10: Swiss males born in 1955



The reason why the simpler model is favores is that the deviation from the senescence trend is regular enough to be produced with a single vulnerable subpopulation. Looking closer at the coefficients in the 2-subpopulation model, the vulnerable subpopulation would have represented about 9.9‰of the total population. It would have experienced a baseline mortality $\frac{\alpha_1}{\alpha_2} = 1.2$ times higher, and a rate of ageing $\frac{\beta_1}{\beta_2} = 6$ times stronger than the general population.

This experience confirms that the observed mortality rates of Swiss males born in 1955, between the age of 11 and 90, could have been produced by the existence of two subpopulations, among which no individual experienced a deviation from its rate of ageing during early adulthood. It is therefore not necessary to assume that each and every individual goes through a period of heighten risk of death to observe young adult excess mortality at the aggregated level. Moreover, the share of the population subject to a specific pattern of the force of mortality does not need to be important, in this case less than 1%.

In terms of lost life expectancy, the life expectancy at birth for Swiss males born in 1955 computed on the

[7] Note that this effect could also be estimated by using a third subpopulation representing individuals that are born with extremely high frailty (such as malformations). Such an example can be found in Vaupel Vaupel1985. The ontogenesis as a selection process was suggested in (Levitis, 2011) using a Monte Carlo simulation.

overall estimated force of mortality $\bar{\mu}(x)$ is $\hat{\mathring{e}}_0 = 82.15$ years [8]. Life expectancy for each subpopulation widely differs, with the first one barely reaching 21.5 years, and the second one achieving 83 years. The difference between the overall life expectancy and the one of the main, advantaged, subpopulation is $\hat{\mathring{e}}_{02} - \hat{\mathring{e}}_0 = 0.93$ years. This could be interpreted as the years of life lost to the fact that 1% of the individuals follow the vulnerable path instead of staying in the general population.

Now that we know that the model works on a single population, let us apply it to all Swiss cohorts born from 1876 to 1981 and see how the parameters have evolved over time. Among the 2x106 cohorts, the 2 or 3-population models were always favoured over the single population model. And among the former, the third population sometimes brings a significant improvement over a 2-subpopulation model (table 4). Let us note that without the MCMC procedure, the 3-subpopulation model would have been chosen twice as many times as it was the case here.

Table 4: Number of subpopulations that minimizes the BIC

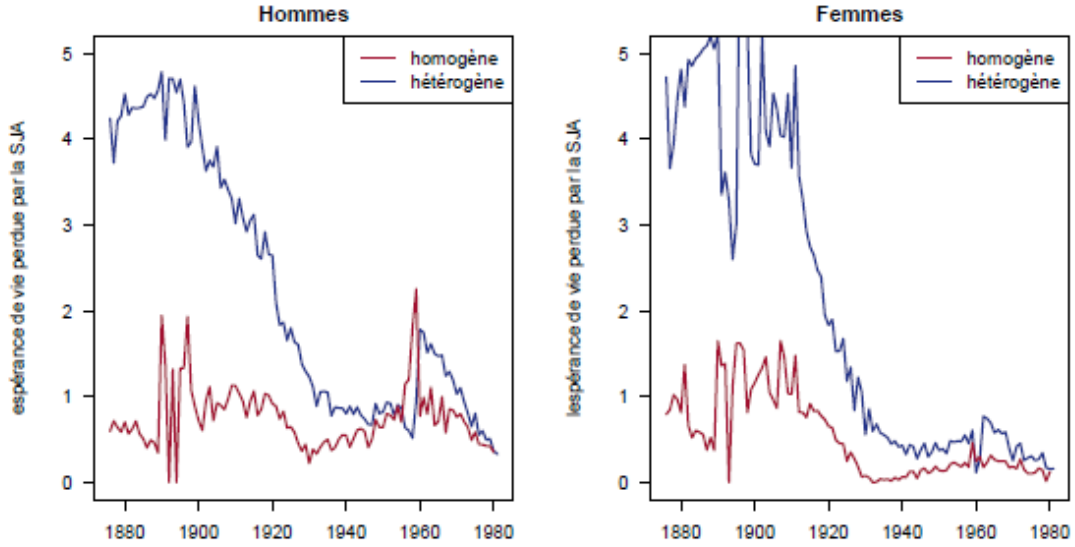
	1	2	3
males	0	73	33
females	0	53	53

The amount of life expectancy lost to the hump was computed twice, first under the homogeneous assumption (figure 1), then under the heterogeneous assumption. In the case where there are three subpopulations, one needs to decide whether the third one should be considered together with the most vulnerable subpopulation, or with the general population. As there is no natural solution, the solution so fat was to consider the third subpopulation as belonging to the vulnerable part of the population if its life expectancy was within 10 years of the lowest one. The difference between both life expectancies corresponds to the loss in life expectancy due to the inclusion of the vulnerable subpopulation.

The comparison of these two measures of *YAEM* is instructive in the sense that, although based on opposite assumptions, they should remain similar. The results show that the two measures are closely related but not systematically (figure 11). Indeed, the measure based on the heterogeneity assumption (blue) is systematically higher than the one based on the homogenous assumption (red). This is particularely true for the begning of the period. However, both series evolve in the same direction most of the time. Both begin at high levels at the end of the nineteenth century, and then drop dramatically in the second half of the twentieth century. This probably corresponds to the disapearance of tuberculosis. The then reach a minimum between the two World Wars, and then climb back in the generations born around 1960. Those are known to be especially concerned with the AIDS epidemics (Valdes, 2011). Then both decrease in the most recent generations.

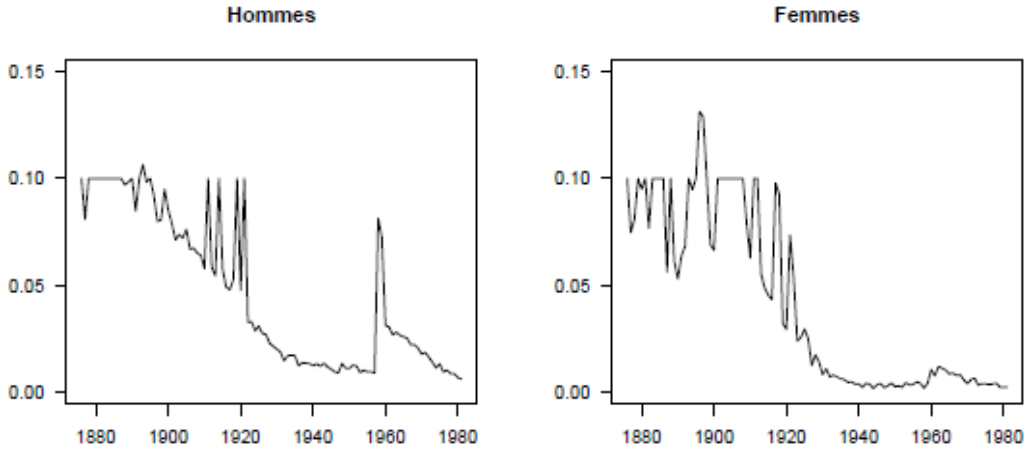
[8] Note that all estimated life expectancies are computed on the same age range (0 - 90) as the original dataset. This is done by projecting the force of mortality outside of the truncated observations, using the estimated parameters.

Figure 11: Evolution of *YAEM* in Swiss cohorts (1876-1981)



In terms of the size of the vulnerable subpopulation required to create the observed hump, the results are also interesting. In both sexes, the proportion has decreased over time, going from about 10% to less than 1% in the last decades. This confirms once again that this subpopulation can be very small and confined to very marginal profiles.

Figure 12: Proportion of individuals in the more vulnerable subpopulation



7 Conclusion and research perspectives

These analyses demonstrate that it is not necessary to postulate, as psychoanalytical and neuro-psychological theories of adolescence do, that each and every individual goes through a phase of 'Storm and Stress' in order to observe young adult excess mortality. Non-observed heterogeneity and selection effects can also produce a hump in the force of mortality even though no individual experiences a period of heighten vulnerability.

Even though a model with 2 subpopulations is sometimes sufficient, the irregularities of the rate of ageing often necessitate a third subpopulation in order to model separately the hump and the bend of the force of mortality. Moreover, the results on Swiss cohorts indicate that the loss of life expectancy due to the presence of a vulnerable subpopulation resembles the loss of life expectancy obtained by interpolating the aggregated force of mortality. For both measures and for both sexes, this loss is maximal in the generations born around 1900, then decreases until the generations born in the middle of the twentieth century. The loss then increases until the 1960 generation, and

then decreases to reach an almost all-time low for the last observed cohort born in 1981.

So far, we have shown that the heterogeneity hypothesis is valid, i.e. that conclusions (the existence of a mortality hump) follow from the premises (the existence of vulnerable subpopulations). What remains to be done is to show that these particular premises are true, and hence have a sound alternative argument about the origins of *YAEM*.

A first good sign is that the vulnerable subpopulation can be small and cover only a marginal portion of the population, subject to very specific conditions such as heavy psychiatric disorders (schizophrenia, bipolar disorder, obsessive-compulsive disorder, etc.), drug addiction, organized crime, or extreme poverty. This cannot however be confirmed with a population-based model.

Moreover, some causes of death known to be particularly associated with young adults' mortality also display a remarkably strong social gradient. It is for instance well documented that tuberculosis is strongly linked with poverty, at the national (Janssens and Rieder, 2008), sub-national (Nagi and Stockwell, 1973), and individual levels (Kitagawa and Hauser, 1968). The same can be said about traffic accidents (Males, 2009b; Hasselberg et al., 2005; Braver, 2003; Cho et al., 2007; Laflamme and Diderichsen, 2000), as well as other types of accidents (Hjern and Bremberg, 2002; Laflamme et al., 2009).

This means that socioeconomic status, together with gender and, more generally, social environment, are serious contenders to justify our hypothesis that some subpopulations follow a steeper force of mortality, explaining the so-far unobserved heterogeneity in rates of ageing. Moreover, in times when economic growth and welfare state provided shelter for the least advantaged of the society, during the 1950s, 1960s and early 1970s, *YAEM* was at its lowest in all Western and Eastern-European countries (Remund, 2012). Globalization, economic stagnation, and the fall of socialist regimes have hit young adults especially hard (Blossfeld, 2008), which could explain the recent increase of *YAEM* that remain high even after the decrease of HIV-related deaths in the Northern hemisphere. The diverging forces of mortality between subpopulations also corresponds to the life course theory of cumulative (dis)advantages, which postulates a "systemic tendency for interindividual divergence in a given characteristic (e.g., money, health, or status) with the passage of time" (Dannefer, 2003)[327].

In order to test this hypothesis we will use partitioning techniques on individual mortality data from the *Swiss National Cohort* (Spoerri et al., 2010).

References

Jeffrey Jensen Arnett. Adolescent storm and stress, reconsidered. *American psychologist*, 54(5):317–326, 1999.

Jeffrey Jensen Arnett. Suffering, selfish, slackers? myths and reality about emerging adults. *Journal of Youth and Adolescence*, 36(1):23–29, 2007.

Judith Bessant and Rob Watts. The mismeasurement of youth: why adolescent brain science is bad science. *Contemporary Social Science*, 7(2):181–196, 2012.

Hans-Peter Blossfeld. *Young workers, globalization and the labor market : comparing early working life in eleven countries*. Edward Elgar, Cheltenham, Glos, UK ; Northampton, MA, 2008.

Elisa R. Braver. Race, hispanic origin, and socioeconomic status in relation to motor vehicle occupant death rates and risk factors among adults. *Accident Analysis & Prevention*, 35(3):295–309, 2003.

David R Brillinger. A biometrics invited paper with discussion: the natural variability of vital rates and associated statistics. *Biometrics*, pages 693–734, 1986.

Hong-Jun Cho, Young-Ho Khang, Seungmi Yang, Sam Harper, and John W Lynch. Socioeconomic differentials in cause-specific mortality among south korean adolescents. *International Journal of Epidemiology*, 36(1):50–57, 2007. doi: 10.1093/ije/dyl239. URL `http://ije.oxfordjournals.org/content/36/1/50.abstract`.

Suparna Choudhury. Culturing the adolescent brain: what can neuroscience learn from anthropology? *Social Cognitive and Affective Neuroscience*, 5(2-3):159–167, 2010.

Dale Dannefer. Cumulative advantage/disadvantage and the life course: Cross-fertilizing age and social science theory. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 58(6):S327–S337, 2003.

Swagatam Das and Ponnuthurai Nagaratnam Suganthan. Differential evolution: A survey of the state-of-the-art. *Evolutionary Computation, IEEE Transactions on*, 15(1):4–31, 2011.

Pierre Dasen. Rapid social change and the turmoil of adolescence: A cross-cultural perspective. *International Journal of Group Tensions*, 29(1-2):17–49, 2000.

Marco Dorigo and Mauro Birattari. *Ant colony optimization*, pages 36–39. Springer, 2010.

Anna Freud. *Adolescence*. American Book, New York, 2d ed. edition, 1968.

Wade Hampton Frost. The age selection of mortality from tuberculosis in successive decades. *American Journal of Epidemiology*, 141(1), 1939.

Jay N Giedd. Structural magnetic resonance imaging of the adolescent brain. *Annals of the New York Academy of Sciences*, 1021(1):77–85, 2004.

Jay N Giedd, Jonathan Blumenthal, Neal O Jeffries, F Xavier Castellanos, Hong Liu, Alex Zijdenbos, Tom Paus, Alan C Evans, and Judith L Rapoport. Brain development during childhood and adolescence: a longitudinal mri study. *Nature neuroscience*, 2(10):861–863, 1999.

Joshua Goldstein. A secular trend toward earlier male sexual maturity: Evidence from shifting ages of male young adult mortality. *PLoS ONE*, 6(8), 2011.

G.S. Hall. *Adolescence: Its psychology and its relations to physiology, anthropology, sociology, sex, crime, religion and education*. D. Appleton and company, New York, 1904.

Marie Hasselberg, Marjan Vaez, and Lucie Laflamme. Socioeconomic aspects of the circumstances and consequences of car crashes among young adults. *Social Science & Medicine*, 60(2):287 – 295, 2005. ISSN 0277-9536. doi: 10.1016/j.socscimed.2004.05.006. URL `http://www.sciencedirect.com/science/article/pii/S0277953604002308`. ¡ce:title¿Equity, Capabilities and Health¡/ce:title¿.

A Hjern and S Bremberg. Social aetiology of violent deaths in swedish children and youth. *Journal of Epidemiology and Community Health*, 56(9):688–692, 2002. doi: 10.1136/jech.56.9.688. URL `http://jech.bmj.com/content/56/9/688.abstract`.

Jean-Paul Janssens and HL Rieder. An ecological analysis of incidence of tuberculosis and per capita gross domestic product. *European Respiratory Journal*, 32(5):1415–1416, 2008.

Evelyn M. Kitagawa and Philip M. Hauser. Education differentials in mortality by cause of death: United states, 1960. *Demography*, 5(1):318–353, 1968.

Anastasia Kostaki. A nineparameter version of the heligmanpollard formula. *Mathematical Population Studies*, 3 (4):277–288, 1992.

L. Laflamme, S. Burrows, and M. Hasselberg. Socioeconomic differences in injury risks. Technical report, Karolinska Institutet, 2009.

Lucie Laflamme and Finn Diderichsen. Social differences in traffic injury risks in childhood and youtha literature review and a research agenda. *Injury Prevention*, 6(4):293–298, 2000.

Rhoshel K. Lenroot and Jay N. Giedd. Brain development in children and adolescents: Insights from anatomical magnetic resonance imaging. *Neuroscience & Biobehavioral Reviews*, 30(6):718–729, 2006.

Daniel A. Levitis. Before senescence: the evolutionary demography of ontogenesis. *Proceedings of the Royal Society B: Biological Sciences*, 278(1707):801–809, 2011.

Michael Males. Does the adolescent brain make risk taking inevitable?: A skeptical appraisal. *Journal of Adolescent Research*, 24(1):3–20, 2009a.

Mike Males. The role of poverty in california teenagers fatal traffic crash risk. *Californian Journal of Health Promotion*, 7(1):1–13, 2009b.

D. Marquardt. An algorithm for least-squares estimation of nonlinear parameters. *Journal of the Society for Industrial and Applied Mathematics*, 11(2):431–441, 1963.

Margaret Mead. *Coming of age in Samoa; a psychological study of primitive youth for western civilisation.* W. Morrow & Company, New York,, 1928.

Seyedali Mirjalili, Seyed Mohammad Mirjalili, and Andrew Lewis. Grey wolf optimizer. *Advances in Engineering Software*, 69(0):46–61, 2014.

Melanie Mitchell. An introduction to genetic algorithms (complex adaptive systems). 1998.

Katharine M. Mullen, David Ardia, David L. Gil, Donald Windover, and James Cline. Deoptim: An r package for global optimization by differential evolution. *Journal of Statistical Software*, 40(6):1–26, 4 2011. ISSN 1548-7660. URL `http://www.jstatsoft.org/v40/i06`.

Mostafa Nagi and Edward Stockwell. Socioeconomic differentials in mortality by cause of death. *Health Services Reports*, 88(5):449–456, 1973.

Daniel Offer and Kimberly A. Schonert-Reichl. Debunking the myths of adolescence: Findings from recent research. *Journal of the American Academy of Child & Adolescent Psychiatry*, 31(6):1003–1014, 1992.

Daniel Offer, Eric Ostrov, and Kenneth I Howard. The mental health professional's concept of the normal adolescent. *Archives of General Psychiatry*, 38(2):149, 1981.

Magnus Erik Hvass Pedersen. *Tuning & Simplifying Heuristical Optimization.* PhD thesis, 2010.

Anne C. Petersen. Presidential address: Creating adolescents: The role of context and process in developmental trajectories. *Journal of Research on Adolescence*, 3(1):1–18, 1993.

DT Pham, Afshin Ghanbarzadeh, E Koc, S Otri, S Rahim, and Mb Zaidi. The bees algorithm. technical note. *Manufacturing Engineering Centre, Cardiff University, UK*, pages 1–57, 2005.

Kenneth V Price, Rainer M Storn, and Jouni A Lampinen. Differential evolution a practical approach to global optimization. 2005.

A. Remund. *eunesse(s) en danger? Approche dmographique de la sant et la mortalit des jeunes adultes en Suisse, 1990-2010.* PhD thesis, University of Geneva, forthcoming.

Adrien Remund. Is young adults' excess mortality a universal phenomenon? Chaire Quetelet, Louvain-la-Neuve, 2012.

Howard Sercombe. The gift and the trap: Working the teen brain into our concept of youth. *Journal of Adolescent Research*, 25(1):31–47, 2010.

L. P. Spear. The adolescent brain and age-related behavioral manifestations. *Neuroscience & Biobehavioral Reviews*, 24(4):417–463, 2000.

Andrej N. Spiess. A better 'nls'?, 2012.

Adrian Spoerri, Marcel Zwahlen, Matthias Bopp, Felix Gutzwiller, and Matthias Egger. Religion and assisted and non-assisted suicide in switzerland: National cohort study. *International Journal of Epidemiology*, 39(6): 1486–1494, 2010.

Laurence Steinberg. Cognitive and affective development in adolescence. *Trends in Cognitive Sciences*, 9(2):69–74, 2005.

Rainer Storn. On the usage of differential evolution for function optimization. In *Fuzzy Information Processing Society, 1996. NAFIPS. 1996 Biennial Conference of the North American*, pages 519–523. IEEE.

Rainer Storn and Kenneth Price. Differential evolutiona simple and efficient heuristic for global optimization over continuous spaces. *Journal of global optimization*, 11(4):341–359, 1997.

A Roger Thatcher, Vin Kannisto, and James W Vaupel. *The force of mortality at ages 80 to 120*, volume 22. Odense University Press Odense, 1998.

Batrice Valdes. La mortalit par sida en suisse: ses caractristiques et son impact sur la mortalit gnrale. *Revue mdicale suisse*, 306(30):1652–1659, 2011.

J. W. Vaupel and Anatoli I. Yashin. *Unobserved popoulation heterogeneity*, volume 1, chapter 21. Elsevier, Amsterdam, 2006.

James W. Vaupel and Anatoli I. Yashin. Heterogeneity's ruses: Some surprising effects of selection on population dynamics. *The American Statistician*, 39(3):176–185, 1985.

James W. Vaupel, Kenneth G. Manton, and Eric Stallard. The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography*, 16(3):439–454, 1979.

James W. Vaupel, Anatoli I. Yashin, and Kenneth G. Manton. Debilitation's aftermath: Stochastic process models of mortality. *Mathematical Population Studies*, 1(1):21–48, 1988.

Xin-She Yang. *A new metaheuristic bat-inspired algorithm*, pages 65–74. Springer, 2010.