Using Information on the Age Distribution of Deaths in Population Reconstruction: An Extension of Inverse Projection with Applications

ALESSANDRO ROSINA *

Istituto di Studi su Popolazione e Territorio, Università Cattolica, Largo Gemelli 1, 20123 Milan, Italy, e-mail: alessandro.rosina@unicatt.it

1 Introduction

The inverse projection technique, developed by Ronald Lee in the late 1960s, is a well-known method used for "inferring levels and changes in fertility, mortality and population age distributions from observed series of births and deaths, along with some information about population size" [5]. It is a simple and flexible method to use. Over time, some extensions to the technique have been proposed to overcome the assumption of migration absence [4] and for the differentiation of population reconstruction by sex [1]. A further extension has been suggested by Rosina and Rossi [6, 9]. It takes into account some information about the age at death, which is usually available in Italian historical sources. The authors named this modified version of Lee's algorithm Differentiated Inverse Projection (IPD).

This paper offers a brief review of the inverse projection methodology of mortality estimation and the generalisation proposed in the differentiated version. The results of a few applications are then presented. These show the importance of considering the age at death on population reconstruction, thus proving the usefulness of the differentiated inverse projection.

2 Mortality Estimation in the Inverse Projection and in the Differentiated Version

Starting from a population of age x at exact time t $(P_{x,t})$, the population of age $x+1$ at exact time $t+1(P_{x+1,t+1})$ can easily be obtained as follows (for simplicity of exposition we assume the absence of migration):

* The author thanks Fiorenzo Rossi and Gianpiero Dalla Zuanna for their valuable comments on an earlier version of this chapter.

$$P_{x+1,t+1} = P_{x,t} - D_{x,t} \qquad x = n, 0, 1, \ldots, \omega - 1;\ P_n = births,$$

where $D_{x,t}$ are the deaths by age occurred during the time interval $(t, t+1)$ (see Figure 1).

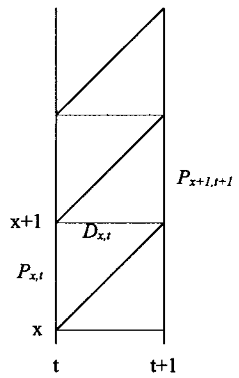

Figure 1. Lexis diagram

Usually, for pre-industrial populations, only the number of total deaths is available. Thus, the reconstruction of these populations necessitates an estimation of the age distribution of deaths. The relationship that links the known data about total deaths and the population by age may be expressed by:

$$D_t = \sum_x P_{x,t} q_{x,t} \qquad x = n, 0, \ldots, \omega - 1, \tag{1}$$

where D_t are the total number of deaths occurred during the time interval $(t, t+1)$, and $q_{x,t}$ is the probability that individuals aged x to $x+1$ at exact time t die before reaching ages $x+1$ to $x+2$ at exact time $t+1$.

The problem is therefore to find that series of q_x (unknown probability of dying by age) that, applied to $P_{x,t}$, gives exactly the observed D_t. A solution is given from the model:

$$q_{x,t} = q_x^0 k_t \qquad x = n, 0, 1, \ldots, \omega - 1. \tag{2}$$

This could be considered a particular form of a proportional hazard model, with baseline function q_x^0 and relative risk k_t. The model assumes that it is possible to decompose the probability of dying in a component which varies by age but is constant over time (q_x^0: age factor of mortality), and a component which is constant by age but varies over time (k_t: period factor of mortality).

The baseline function q_x^0 is an external input that represents the mortality age structure of the population. The period factor k_t can be obtained by plugging formula (2) into formula (1):

$$\mathrm{D_t} = \sum_x P_{x,t} q_{x,t} = k_t \sum_x P_{x,t} q_x^0$$

which, solving for k_t, gives:

$$k_t = \frac{D_t}{\sum P_{x,t} q_x^0} \quad . \tag{3}$$

The following are the main steps for estimating mortality by age:

a) from (3), knowing D_t and $P_{x,t}$, and choosing an appropriate baseline mortality pattern q_x^0, we calculate k_t;
b) from (2) we calculate $q_{x,t}$;
c) the deaths by age now can be easily obtained: $D_{x,t} = P_{x,t} q_{x,t}$.

2.1 Inverse Projection

The solution proposed by Lee [3] is based on two baseline mortality patterns (q_x^a and q_x^b). The proportional hazard assumption becomes less restrictive since it now refers to differences between probabilities:

$$\left(q_{x,t} - q_x^b\right) = k_t \left(q_x^a - q_x^b\right).$$

Solving for k_t, we obtain:

$$k_t = \frac{D_t - \sum P_{x,t} q_x^b}{\sum P_{x,t} q_x^a - \sum P_{x,t} q_x^b} \quad .$$

2.2 Differentiated Inverse Projection

The proportional hazard assumption means that the age mortality structure is the same in every year of the reconstruction period under study. In fact, in model (2) the baseline (q_x^0) is constant over time and the only factor that varies is k_t, which in turn doesn't depend on age. The consequence is that, from one year to the next, only the mortality level varies, while the structure by age remains the same.

This assumption is generally not verified in reality because there could be long-term modifications in the age pattern of mortality. There are also usually short-term variations in mortality incidence by age, mainly as a consequence of epidemics (we know for example that typhus is more intense in the adult ages, and smallpox in the young ages).

A possible solution [6, 10] is to use the information about deaths by age when this is available. The information is easy to find in Italian historical sources [1]. For example, it is not unusual to have enough information that distinguishes at least between deaths in the first five years of life (child mortality) and deaths in other ages (adult mortality). In this case, we can estimate different parameters, k_t, for child mortality (${}_1k_t$) and for adult mortality (${}_2k_t$). Model (2) can now be written as:

$$q_{x,t} = q_x^0 \, {}_1k_t \qquad for\, x = n, 0, 1, \ldots, 4$$

$$q_{x,t} = q_x^0 \, {}_2k_t \qquad for\, x = 5, \ldots, \omega - 1.$$

We can therefore have an independent estimation of k_t in every group of deaths, given by:

$${}_1k_t = \frac{{}_1D_t}{\sum P_{x,t} q_x^0} \qquad for\, x = n, 0, 1, \ldots, 4$$

$${}_2k_t = \frac{{}_2D_t}{\sum P_{x,t} q_x^0} \qquad for\, x = 5, \ldots, \omega - 1,$$

where ${}_1D_t$ and ${}_2D_t$ are respectively child and adult deaths occurred during the time interval $(t, t+1)$.

Generalising the inverse projection model, for every group of deaths g ($g = 1, 2, .., G$), we have:

$$\left(q_{x,t} - q_x^b\right) = {}_gk_t \left(q_x^a - q_x^b\right)$$

and, hence,

$${}_gk_t = \frac{{}_gD_t - \sum P_{x,t} q_x^b}{\sum P_{x,t} q_x^a - \sum P_{x,t} q_x^b}$$

with $x = l(g), \ldots, u(g)$, where $l(g)$ and $u(g)$ are the lower and upper age limits of group g.

We have found very useful to distinguish at least between infant (first year of life) and subsequent mortality. The factors affecting infant mortality are generally different from those affecting adult mortality, and the impact on the demographic system is also generally differentiated. If the information about death by age is more detailed we suggest distinguishing between infant (1^{st} year of life), young (from 2^{nd} to 15^{th} year) and adult (from 15^{th} onwards) mortality. This permits us to appreciate the different incidence of epidemics by age also.

3 Advantages of Using the Differentiated Version

By using more detailed mortality information, we can obtain a more rigorous population reconstruction and better estimations on synthetic mortality measures (e_0). Generally, the improvement, with respect to the classic inverse projection, is relevant only in the occasion of an intense mortality crisis with different incidence by age.

Of greater importance is the possibility of producing more detailed mortality measures. With the classic inverse projection one can only use, as output, summary measures of mortality which represent the level, as life expectancy at birth (e_0), but not the age pattern of mortality. With the differentiated inverse projection, we can obtain more specific outputs by using more detailed data. If for example, we distinguish between infant, young and adult deaths, we can obtain independent measures of mortality in the various groups of deaths, thus appreciating the long-term evolution and short-term variation which are different in infant, young and adult mortality.

4 Some Applications to Italian Data

In this section, we present a few applications to Italian data. These show the importance and usefulness of taking into account age at death with the differentiated inverse projection to study characteristics and dynamics of the pre-industrial populations.

4.1 An Analysis of the Pattern of Mortality by Age

Information about the pattern of mortality by age is usually scarce for past populations. If we use the classic inverse projection, we must assume as an external input a baseline mortality pattern (see section 2). The usual approach is simply to choose model life tables from Coale, Demeny and Vaughan [2]. However, for some historical populations, the mortality pattern could be very different from standard life tables. With the differentiated inverse projection, we overcome this problem if the quality of the information on age at death enables us to distribute the death totals by at least a few suitable age groups.

For example, in the study of the evolution of the population of Militello (a village in the island of Sicily) in the 18^{th} century, we considered the distinction between infant, child (1-14), young and adult (15 and over) deaths. We were able to estimate the surviving probability in these age groups by using the differentiated inverse projection (separately for male and female). We found that, while the probability to survive until 15 was not different for male and female, in the adult ages the male mortality was very high (Figure

2) if compared to the female one of the same population and to male adult mortality in other populations during the same century [8].

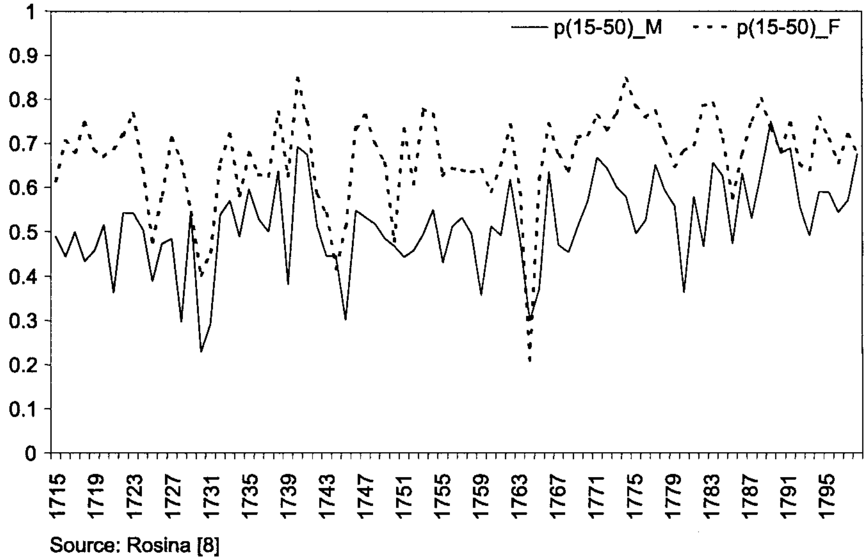

Figure 2. Probability to survive from age 15 to age 50 by sex. Militello, 1715-1799

4.2 An Analysis of the Demographic System

Equation 4 is usually used for representing the demographic system of a population:

$$NRR = GRR\, p_M, \tag{4}$$

where: GRR is the Gross Reproduction Rate, NRR is the Net Reproduction Rate and p_M represents the surviving probability from birth to the mean maternity age (M).

The inverse projection allows the estimation of all these components, so it is possible to see whether the variations observed in the natural growth of the population (NRR) during the reconstructed period are mainly due to fertility variations (GRR) or to mortality variations (p_M).

Looking at Figure 3, we can see that in Veneto (a region in the North of Italy), for example, the population growth in the 19th century is mainly due to the increase of the surviving probability from birth to the mean maternity age: p_{31}.

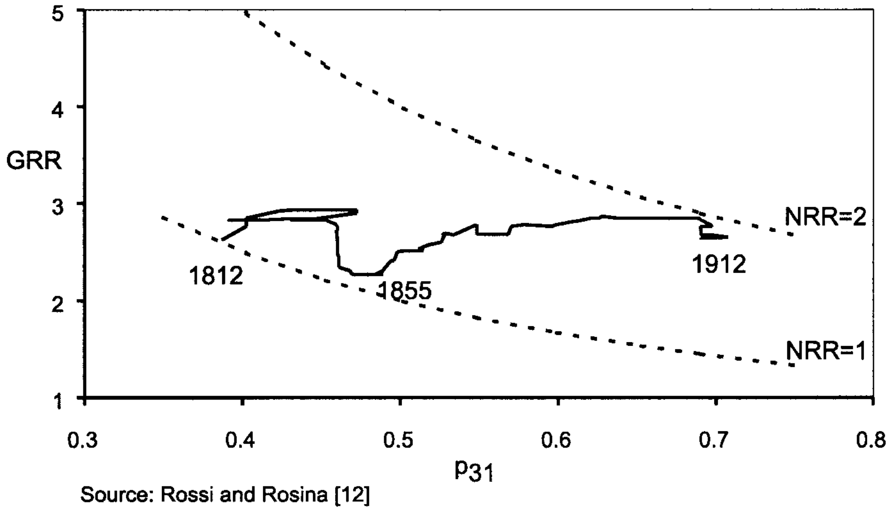

Figure 3. Combination of GRR with p_{31}. Veneto, 1812-1912 (11-value moving median)

With the differentiated version we can also decompose p_M in the contribution given by relevant segments of life, such as infant, young and adult ages:

$$NRR = GRR\,(p_{n-0}\,p_{1-14}\,p_{15-M}). \tag{5}$$

Figure 4 shows, for example, that in Veneto the total increment in the surviving probability from birth to the mean maternity age ($M = 31$) is the result of a differentiated evolution of the surviving probability in the infant ages and in the young-adult ages. While the surviving probability in the first year of life shows a continuous increase, the probability of surviving during the young-adult ages starts to rise continuously only in the second part of the reconstructed period.

From Figures 3 and 4 we can conclude that the considerable population growth in Veneto during the 19th century, at least until the start of the great emigrations during the 1880s, is almost only due to the decline of infant mortality [12].

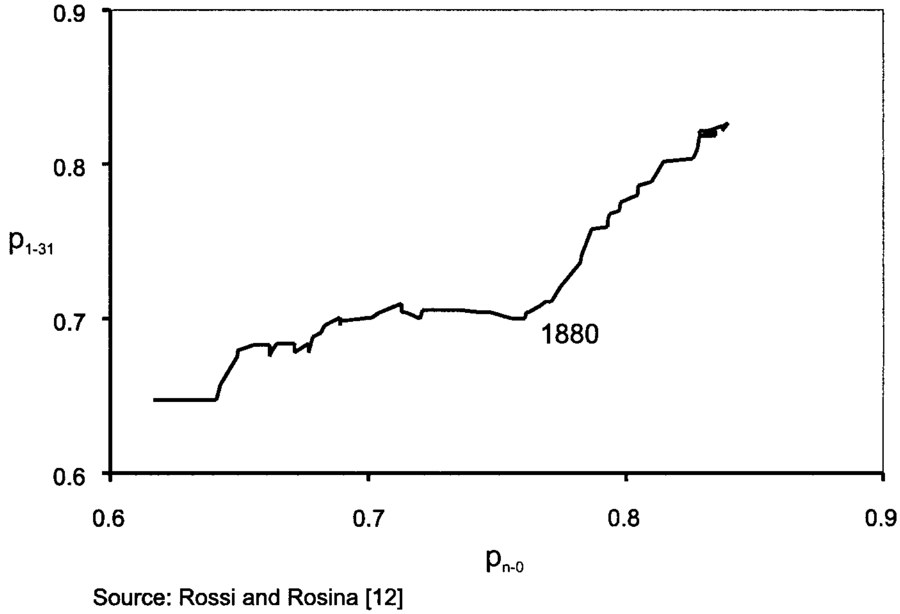

Figure 4. Combination of p_{1-31} with p_{n-0} . Veneto, 1812-1912 (11-value-moving-median)

4.3 An Analysis of the Epidemics Incidence

Figure 5 shows the consequences of the mortality crises[2] according to the relevant age groups obtained from the Chioggia[3] population reconstruction [11]. We can see that the crises can be differentiated by age. For example, it is evident that the years with extra-adult-mortality are not necessarily the same as those with extra-infant-mortality. We can also notice that:

a) relevant rises in adult mortality occur during periods known as famine periods (namely 1648-49 and 1678-79); we may therefore presume the cause to be typhus fever epidemics;
b) we observe now and then an increase in mortality at young ages, which can be attributed to smallpox.

[2] The effect of the crisis (extra-ordinary survival) is computed as the ratio between the probability to survive to age x at time t and the 11-value-moving-median centred on t.

[3] Chioggia is a town near Venice.

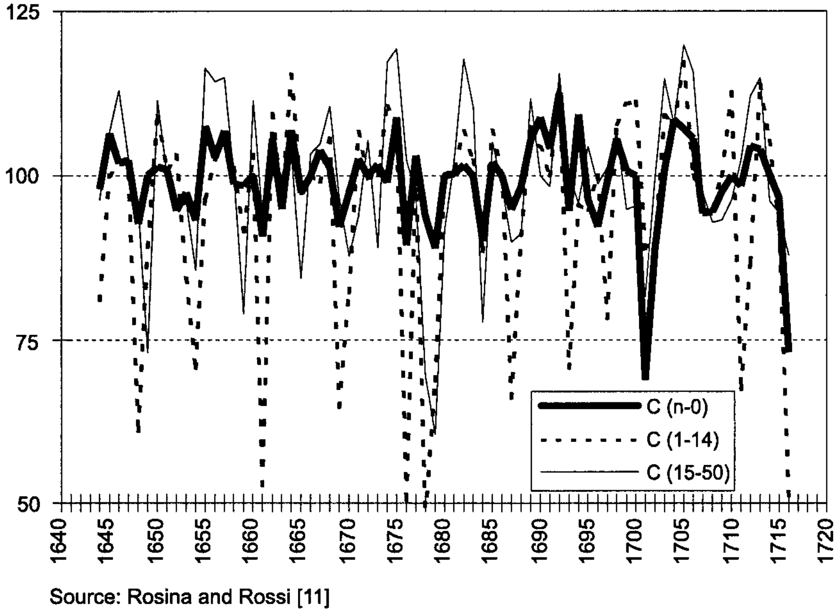

Figure 5. Extra-ordinary survival, $C(x)$, at ages $n-0$, $1-14$, $15-50$. Chioggia, 1640-1720 (11-value-moving-median = 100)

5 Conclusion

We have shown that a simple modification of the inverse projection methodology allows for information on age at death previously unconsidered. This information improves the reconstruction of historical populations, especially to study the age pattern of mortality, its evolution on time and its consequences on the demographic system.

In the Italian sources, age at death is easy to find, and usually the quality of this information is good enough to be used for suitable age groups. Furthermore, the most important distinction is between infant and adult mortality. Having enough information that would help researchers separate deaths at age 0 from other deaths is a quite common situation, not only in Italy but also in other countries with even less detailed historical sources. The differentiated inverse projection can therefore be gainfully used in various situations.[4]

[4] A PC program, called IPD 3.0, is available [7] for population reconstruction by using the differentiated inverse projection technique. The alghoritm is a generalisation of the inverse projection, so it gives the same results as does Lee's procedure when only total deaths are considered as input.

References

1. Breschi M. (1991), Aggregative Methods for Reconstructing Historical Population: Considerations on Their Use in Italy. Paper presented at the seminar on Old and New Methods in Historical Demography, Palma de Mallorca, 12-15 June 1991.
2. Coale A., Demeny P. and Vaughan B. (1983), *Regional Model Life Tables and Stable Populations*. Second edition, Academic Press, New York.
3. Lee R. D. (1974), Estimating series of vital rates and age structure from baptisms and burials: A new technique, with applications to pre-industrial England. *Population Studies* 28(3), 495–512.
4. Lee R. D. (1985), Inverse Projection and Back Projection: A critical Appraisal and Comparative Results for England. *Population Studies* 39(2), 233–248.
5. Lee R. D. (1993), Inverse projection and demographic fluctuations: A critical assessment of new methods. In Reher D. S. and Schofield R. (Eds.), *Old and New Methods in Historical Demography*, Clarendon Press, Oxford, 7–28.
6. Rosina A. (1993), Una generalizzazione dell'Inverse Projection. In *Per una storia della popolazione italiana: problemi di metodo*, Dipartimento di Scienze Statistiche "P. Fortunati", Università degli Studi di Bologna, Bologna, 73–80.
7. Rosina A. (1996), IPD 3.0: applicazione automatica dell'inverse projection differenziata (passo annuale e quinquennale). *Technical Report*, Dipartimento di Scienze Statistiche, Università degli Studi di Padova, Padova.
8. Rosina A. (2000), Le dinamiche evolutive della popolazione: una ricostruzione (1715-1798). In Breschi M. and Santis G. D. (Eds.), *Il quadro macrodemografico di una comunità siciliana nel XVIII secolo: Militello in Val di Catania*, Forum, Udine, 61–86.
9. Rosina A. and Rossi F. (1993), Una estensione dell'Inverse Projection con mortalità differenziata per età. *Statistica* 4, 619–631.
10. Rosina A. and Rossi F. (1994), *Ricostruzioni aggregate dei processi evolutivi delle popolazion.* CLEUP, Padova.
11. Rosina A. and Rossi F. (Eds.) (2000), *Il sistema demografico alla fine delle grandi epidemie. Venezia, il Dogado, Chioggia tra Seicento e Settecento.* CLEUP, Padova.
12. Rossi F. and Rosina A. (1998), Il Veneto tra Sette e Ottocento. *Bollettino di Demografia Storica* 28, 89–114.