Reflections on Inverse Projection: Its Origins, Development, Extensions, and Relation to Forecasting*

RONALD LEE

Department of Demography, University of California, 2232 Piedmont Ave, Berkeley, CA 94720, USA, e-mail: rlee@demog.berkeley.edu

1 Background

Some good ideas just don't work out in practice, while other ideas that seem to be based on very questionable assumptions, surprise us by working better than one could reasonably expect. Inverse Projection (IP) certainly falls in this second category. In this note, I will briefly describe the origin and early development of the idea, then discuss some recent advances, and conclude by suggesting that a related approach may be useful for Third World populations.

I read Tony Wrigley's striking reconstitution study of Colyton, an English parish, while working on my dissertation in the late 1960s. His reconstitution found a certain pattern of change over broad time periods in nuptiality, marital fertility, and life expectancy. Plots of the total number of baptisms and burials each year over three centuries were also shown. Staring at the two different kinds of data, it seemed to me that the trends and fluctuations in the aggregate time series of baptisms and burials were not consistent with estimates of fertility and mortality, and I worried that there was some problem in the reconstitution results. A couple of years earlier, I had taken a course from Nathan Keyfitz, in which he taught us to program computers to simulate demographic processes and do projections, skills that were rare at the time. With this background fresh in my mind, it occurred to me that I could check the consistency of the two kinds of data for Colyton through a kind of inverted projection routine. My idea was to begin with a stable population of appropriate size at the start of the historical period, and then to choose the standard age schedules of fertility and mortality (taken from Coale-Demeny models) that, together with the stable population, would match the totals of baptisms and burials. Then these could be used to project forward the initial stable population, before re-applying this procedure to the next period's data.

* Research for this paper was funded by a grant from NIA, R37-AG11761.

I viewed this as a kind of filtering procedure, to remove the effects of age distribution fluctuations from the variations in the crude birth and death rates. I read the numbers of baptisms and burials off the graph in the paper, and wrote a simple program to carry out the calculations. To my great disappointment, the results matched Wrigley's reconstitution estimates almost exactly, disproving my intuition that the two kinds of data were telling different and inconsistent stories. Only later did I realize that from a different perspective, the outcome was a success.

This work on IP was part of my dissertation which was a study of the broad macro determinants and consequences of aggregate population change in preindustrial England, in a rather Malthusian theoretical framework. The goal was actually quite similar to the project that Palloni outlines in his paper for this volume. One question in my mind was whether different kinds of population cycle, including one-generation classic demographic cycles, or two-generation Easterlin cycles arising from very strong homeostatic control, might occur, again thinking along the same lines as Palloni. For this purpose, I examined the relation of the IP estimates of the Gross Reproduction Rate (GRR) to various IP estimates of population age distribution, such as the ratio of the young working age population to the old as in Easterlin's work. Again I was disappointed to find no association.

After I finished the dissertation, I spent a Postdoctoral year at INED in Paris, under the supervision of Louis Henry. Henry was very cordial, but he was clearly not impressed by my efforts in historical demography. After I returned to the US, however, and eventually published my first article on IP, he wrote me a letter, asking why I didn't make use of the information on the age distribution of deaths. Henry's suggestion is precisely the approach taken in the paper by Rosina in this volume, who calls the method Differentiated Inverse Projection, or IPD. It is certainly a very sensible approach, and indeed one would always want to try to use fully the available data. At one point I wanted to try something of the sort on the French historical data, which for a certain time period distinguished only broad age classes of deaths, as I recall. However, such data are usually not available.

When I finished the dissertation, I sent a copy to Tony Wrigley at the Cambridge Group for the History of Population and Social Structure in England, whose work I much admired. This, and a subsequent meeting with him and Roger Schofield at a workshop in Princeton, led to a plan for collaboration in their reconstruction of English population history based on series of baptisms and burials for a large collection of parishes. Initially, the idea was that I would use IP to reconstruct the population. However, Wrigley and Schofield thought I should modify IP in two ways: first, to go backwards rather than forwards, starting with the census enumerations at the end of the period; and second, to estimate migration rather than to take it as given. However, in my view it was mathematically impossible to estimate migration based only on baptisms and burials and a terminal population, since an infinite range of patterns of migration would be consistent with the data. As for going backward, I saw that a mathematical solution existed, but when I programmed it, I found that the resulting estimates were hopelessly erratic, leading to explosive cycles and negative values for some population elements. This behavior, I realized, was just the mirror image of the ergodicity of population going forward. For these reasons, I declined to develop the program they wanted, so Wrigley and Schofield asked Jim Oeppen to develop a new program which would incorporate these two changes. This became Back Projection. The disagreement with the Cambridge team on these two points continues to this day.

In the 1970s, Helge Brunborg was a doctoral student at the University of Michigan where I was then on the faculty. We developed new versions of the IP program, and worked on incorporating nuptiality and marital fertility. Unfortunately, versions incorporating marriage and marital fertility never performed very satisfactorily. One major problem was that age schedules of marriage rates per unmarried woman (the force of nuptiality) varied in two quite different ways over time. On the one hand, they might shift up or down like age specific death rates, leading indirectly to changes in the mean age at marriage just as changes in the force of mortality lead to changes in life expectancy. But on the other hand, the age schedule might slide toward higher or lower ages reflecting postponement or advancement of marriage, thereby directly affecting the mean age at marriage. While it was straightforward to program these alternatives as switchable options in the IP program, in reality both kinds of changes seemed to occur. Consequently, we were not successful in matching historical changes in first marriage age to changes in the time series of marriages over time. In related work, Brunborg [3] carried out an analysis and comparison of the IP results to the actual population data for Norway, in an unpublished paper that has played an important role in the development of IP, as reflected in many citations to the paper in the present volume.

It was a big step forward for IP when Robert McCaa [6], and McCaa and Brignoli, [7] developed his own program for carrying out IP, in a format which could produce annual estimates. McCaa put his program in the public domain. Although I had been quite willing to share my own software with others from the beginning, mine was difficult to use, input-output operations were idiosyncratic, the programming was a patchwork of contributions by a number of people, and it was poorly documented. McCaa's generous work made the method readily applicable by any interested researcher.

2 New Developments

I have already mentioned two new developments: Back Projection and Differentiated Inverse Projection. Here I will discuss two others: Stochastic Inverse Projection (SIP) and Generalized Inverse Projection (GIP).

Bertino and Sonnino [1, 2] have developed SIP, which uses statistical demography in an ingenious way to enrich the deterministic IP by developing the idea that demographic rates are just probabilities at the individual level. Through this insight, it is seen that there is uncertainty in the distribution by age and sex of the deaths occurring in each year, and also in the age distribution of the mothers of the births occurring in a year. Furthermore, the reconstructed proportional population age-sex distribution (but not, I think, its size) in any year is also uncertain, since the distribution of deaths by age in each earlier year was uncertain, leading to uncertainty in survivors by age-sex. In small populations, this considerably improves the realism of the assumptions. So far as I can see, the mean estimates in forward SIP should be identical to those in IP, although I am not at all certain about this due to the nonlinearities in the procedures. In the papers in this volume that permit comparison, differences in mean values are found. However, with only fifty stochastic realizations of SIP estimates, it seems likely that there will still be a lot of sampling variability in the results. Indeed, faster computers should make it possible to estimate a larger number of SIP trajectories, which would be preferable.

I see two major advantages of SIP. First, it provides an estimate of the uncertainty in the IP results. Second, by introducing increased flexibility in IP, it seems to produce stability in the backwards IP. Let us first consider the uncertainty of IP estimates. Of course, some uncertainty must arise from likely errors in the data, but let us suppose that the data perfectly reflect the actual demography, so that this source of error is absent. Also let us suppose that net migration is known without error. Then, it seems to me, there remain two important sources of error for IP. The first is the one treated elegantly by SIP. The second arises from departures of the true underlying age-sex specific probabilities of birth and death from the model schedules that have been assumed. Such departures will in practice arise for many reasons. For one thing, the age pattern of fertility will depend on the proportions married at each age, and these will vary irregularly over time, which is why it would be useful to incorporate nuptiality in IP. In the case of mortality, since the causes of death vary, and different causes of death entail different age-sex patterns of mortality, the age patterns of mortality are also likely to vary. It would be interesting to learn the relative importance of the two sources of IP error, and it might be possible to do by using known historical populations such as those of Norway or Sweden.

The first source of uncertainty, the one estimated by SIP, is inevitable in finite populations, unlike the second. However, this first source of uncertainty does tend to vanish as the size of the population increases, and I would expect it to be negligible at the level of a national population, or for a sizable aggregation of parish registers such as was used to reconstruct the population of England. I was surprised, therefore, to see in the application to Sweden by Barbi and Oeppen in this volume that the 95% confidence interval on the GRR has a width of 0.31 births per woman in 1790. Some of the uncertainty in the estimate may arise from the relatively small number of fifty simulations in this application. Some may also arise from uncertainty about the base population age distribution to which the age specific probabilities should be applied, since uncertainties arising from the rates as probabilities alone would be only a fraction this large.

Because of the logical link between IP and ordinary population projection, it is interesting to note that the first efforts to assess the uncertainty of population projections were of exactly this sort, based on recognizing that rates were just probabilities at the individual level. Tore Schweder [8] took this approach to assessing the uncertainty in Swedish population projections, but then abandoned it after discovering that the implied forecasting errors were much too small relative to the actual forecasting errors made by statistical agencies. The important source of error, it turns out, is the variation over time in the rates or probabilities themselves. In IP or SIP this time series variation is the object of interest, and is estimated, so it does not compete with the uncertainty identified by SIP.

I was very interested to see that SIP can be run backward in time. However, I am not at all convinced that this is a good thing. After all, it is always possible to use the baptisms and burials, together with the migration assumption, to count back to find the population size at any earlier date. Once this is known, a stable population age distribution can be estimated, and then a forward SIP can be done. With a sufficiently long estimation period, the forward SIP should match the terminal period population age distribution within measurement error (as I found to be the case with IP for England). If this is so, the implication is that there is no further information in the terminal age distribution that can be used to shed light on the initial population age distribution, which therefore might as well be assumed to be stable. Perhaps the backward SIP could be used in a first stage to estimate the general level of fertility and mortality towards the beginning of the period, which could then be used to define the initial stable population age distribution, preparatory to running SIP forward for the final estimates.

Aside from this last possibility, I do not see advantages to running SIP backward, although that appears to be the mode preferred by Bertino and

Sonnino, and it is quite possible that I am misunderstanding something here. First, I note that they do not give the same results. It is true that the comparison plots (Bertino and Sonnino's [2] Figures 12, 13 and 16, for example) for the estimates of Total Fertility Rate (TFR) and life expectancy at birth (e_0) from the two approaches look strikingly similar. However, in their Table 4, forward SIP has the TFR rise by 0.2 births per woman between 1651-1700 and 1851-1870, whereas backward SIP has it rise by 1.3 births per woman over the same time range. Because these differences occur in averages spanning many years, they do not arise from backward SIP's difficulty in capturing annual variations, to which I turn next, but rather from some deeper problem.

Inspection of the plots of estimated probability of dying from birth to 1st birthday (q_0) for males and females (Figures 14 and 15) and comparison to the plotted of numbers of total deaths in the raw data (Figure 3) clearly indicates a problem with the backward estimates. Enormous peaks in the raw death time series at two to three times the normal number are not associated with elevated infant death probabilities from the backward estimates, which cannot be correct. Backward SIP somehow smoothes through annual variations, due to the cohort constraints.

While I am not familiar with the details of Generalized Inverse Projection (GIP), my understanding is that it permits one to specify various demographic data that are available in addition to the time series of births and deaths, and that if these over-identify the system then a solution will be found which minimizes some goodness of fit criterion. This strikes me as an excellent approach to the estimation problems for which IP was designed. I wonder whether GIP could be used to incorporate the information on the age distribution of deaths used by IPD. I would like to see GIP placed in the public domain, where researchers could experiment with it. I am very unclear how it is used to reconstruct populations as in the Barbi-Oeppen paper (in this volume) in the absence of input data on migration, using only a terminal population age distribution. Additional assumptions are used, but I am not clear what they are.

3 Links to Forecasting Mortality and to Stochastic Population Projection

The name "Inverse Projection" invokes the close similarity of the algebra of population projection, and its logical inversion, inverse projection. As it happens, IP has actually led to some useful innovations in the methods of population projection.

A central problem in developing IP was how to represent parsimoniously the age pattern of variations over time in mortality. I experimented first with equal additive changes in age specific mortality, and then, to better capture the actual age pattern of changes, I tried equal multiplicative changes across age. Both these specifications of mortality change were already well-studied in mathematical demography. However, it then occurred to me that linear interpolation between two known age schedules of mortality, and extrapolation outside the range they spanned, would give a very simple, quite flexible, and computationally tractable approach which would easily accommodate the particularities of any population studied. This is the approach I used in my published articles on IP. A variable k_t located the current mortality schedule in relation to the two known schedules chosen as standards, and could be viewed as an index of the force of mortality.

Shortly after I moved to Berkeley, Larry Carter spent a sabbatical term there. We got the idea of using this index k_t to model and forecast mortality. We calculated the index k for a long time series of US mortality data, using it to summarize the changes. To our surprise, the time path of k for the 20[th] century in the US was strikingly linear, which has turned out to be true for the other countries where this has been done as well. We then modeled it as a stochastic time series using standard time series methods, and forecasted it far into the future, obtaining a probability distribution for it as a byproduct of this approach. Given the forecasted k_t, we could then recover the forecasts of the age specific rates for each period, and then the rest of the life table, in the manner familiar from IP. Each forecasted variable then also had a probability distribution, arising from the probability distribution of k.

One problem with this approach quickly became apparent: the forecasts led to negative death rates at younger ages within a few decades. This same problem is noted by McCaa and Barbi, in this volume, for it can crop up in IP as well. The solution we adopted was to switch from linear interpolation to multiplicative interpolation, that is we used linear interpolation for the logs of the death rates. This is the formulation now used in the so-called Lee-Carter method, and I expect it would be useful for IP as well, although the method then becomes more non-linear, and must be solved numerically. Also, whereas with IP the actual age schedules of mortality were not observed and had to be inferred from the number of deaths, in the forecasting context the actual age schedules were typically observed through vital registration data. While the IP approach of interpolating from two known schedules could still be used, it might not give the best fit to the observed series. Consequently, at Ken Wachter's suggestion, we instead used the Singular Value Decomposition (SVD) to find the optimal set of coefficients (a_x, b_x and k_t) for the model. Then, in a second stage using the IP procedure, we recalculated the value of k_t which exactly fit the number of total deaths observed in each historical year, given the first stage estimates of a_x and b_x. As with the contribution by Rosina on IPD, the question arises of using the information on the age distribution of deaths to improve the fit, and John Wilmoth [9] has developed such methods using weighted SVD or Maximum Likelihood. The Lee-Carter method [4], which springs directly from IP, is now used fairly widely for modeling and forecasting mortality in industrial countries, and occasionally for Third World countries as well.

The full IP model requires one parameter indices of the force of mortality (k_t) and the force of fertility (f_t). For a given population age distribution and number of deaths or births, IP then generates an estimate of these indices that exactly matches the number of events and the population age distribution. Suppose, now, that we take trajectories of the indices k_t and f_t as inputs rather than outputs. Given an initial population age distribution, each pair of trajectories for the indices will then generate a full population projection, with fertility, mortality, and population age distribution in each period, assuming the population is closed to migration (which can, of course, be added if desired). Now let these trajectories be stochastic rather than deterministic, and specify their probability distributions. I have already described how this is done for mortality, using the Lee-Carter method. A similar method can be used for fertility. Tuljapurkar and I [5] used this approach to develop stochastic population forecasts for the US. Probability distributions were derived using analytic approximations, and also through stochastic simulation, with the latter approach turning out to be much more tractable. Stochastic projections of this sort are becoming increasingly popular.

4 Toward a More General Model

Arising from the realities of historical data, IP and its variations are designed to use time series of births and deaths, together with at least one measure of the size of the population, or better its age distribution. They exploit this information by employing the basic macro-demographic dynamic accounting identities, and by assuming that the age schedules belong to one parameter families. This strategy for weaving together disparate strands of demographic information into a coherent and consistent whole could be extended to cover situations in which one did not have full time series of births and deaths, and perhaps had more information of other kinds, such as censuses. There is a need for a flexible approach and program of this sort. For example, the United Nations publishes five-yearly data back to 1950 on fertility, life expectancy, and population age distribution, for every country in the world. These data are very valuable, but they have been pieced together from often fragmentary information in a somewhat *ad hoc* way. It would be useful for the UN to have a program which could synthesize these bits of information in a way consistent with the dynamic accounting identities, and I suspect that IP and its variations could be a good starting point. I expect that GIP in particular is already half-way to the desired capability, although in its current form it still requires time series of births and deaths, and in many applications these would not be available. A generalized approach of this sort would also be useful in many historical contexts in which bits of information are available, but not time series of births and deaths. Although in the end the obstacles to creating such a very general approach might prove insurmountable, there is no doubt that the principles could be applied in tailor made way in many particular situations, with specific kinds of data. There are, I believe, benefits for all concerned in a collaboration between those interested in rigorous demographic estimation based on historical data and those interested in estimation in Third World populations with data problems that are often similar.

References

1. Bertino S. and Sonnino E. (1995), La proiezione inversa stocastica: tecnica e applicazione. In *Le Italie Demografiche. Saggi di Demografia Storica*, Dipartimento di Scienze Statistiche, Università degli Studi di Udine, Udine, 99–122.
2. Bertino S. and Sonnino E. (2003), The Stochastic Inverse Projection and the Population of Velletri (1590-1870). *Mathematical Population Studies* 10(1), 41–73.
3. Brunborg H. (1976), The Inverse Projection Method Applied to Norway. Unpublished manuscript.
4. Lee R. D. and Carter L. (1992), Modeling and forecasting the time series of U.S. mortality. *Journal of the American Statistical Association* 87(419), 659–671.
5. Lee R. D. and Tuljapurkar S. (1994), Stochastic population projections for the United States: Beyond high, medium and low. *Journal of the American Statistical Association* 89(428), 1175–1189.
6. McCaa R. (1989), Populate: A Microcomputer Projection Package for Aggregative Data Applied to Norway, 1736-1970. *Annales de Démographie Historique*, 287–298.
7. McCaa R. and Brignoli H. P. (1989), Populate: from births and deaths to the demography of the past, present and future. *Working Paper* 89-0-1, Center for Population Analysis and Policy, University of Minnesota, Minneapolis.
8. Schweder T. (1971), The precision of Population Projections studied by Multiple Predictor Methods. *Demography* 8, 441–450.
9. Wilmoth J. R. (1993), Computational methods for fitting and extrapolating the Lee-Carter model of mortality change. *Technical Report*, Department of Demography, University of California, Berkeley. Downloadable from his web site: `http://www.demog.berkeley.edu/~jrw/vita.html`