APPLICATION OF THE KALMAN FILTER TO MAKE POPULATION ESTIMATES AND PROJECTIONS IN SMALL GEOGRAPHIC AREAS

In demography, it is very common to obtain good-quality population indicators for large aggregates, i.e., at the national level, at the state level, and in major metropolitan areas. But it is difficult to arrive at estimates of reasonable quality in small geographic units due to the high variance evidenced by population figures at those levels, because small changes in demographic components may have a considerable effect upon population indicators. For that reason as well, the population projections made are highly variable. Moreover, in many instances information is not available on demographic components at these geographic levels. The aim of this paper is to make population estimates and predictions in small geographic areas with few inhabitants, using the Kalman filter.

The Kalman filter is a recursive, unbiased estimator using least squares, with a random Gaussian signal. It represents a modern way of discussing the Theory of Least Squares Estimates. The essential difference between the Kalman filter and the conventional linear model is that in the Kalman filter, the state parameter is not constant, but rather may vary in time, i.e., it involves a stochastic process. It is a method that allows us to update and improve population estimates when new information becomes available.

The Kalman filter is comprised of two equations: an observation equation: $Z_k = H_k X_k + V_k$ and a transition equation: $X_k = \phi_k X_{k-1} + W_k$; X_k represents the population at moment k and Z_k is the geographic area, also at moment k. H_k and ϕ_k are parameters that are known at moment k. V_k is the observation error and W_k is the error in the transition equation. Both errors are distributed as normal with mean zero and known variances. In the first equation, a relationship is established between the geographic area and the population for a given moment, while in the second equation, the population at one point in time is associated with the population at a subsequent moment. It is worthwhile to point out that we have two independent sources of information. For one, we have population censuses taken in 1970, 1980, 1990, and the year 2000. And we also have Population Counts from the years 1995 and 2005. Secondly, we have maps for some years, which enable us to analyze how the geographic surface has been expanding. This method combines the censuses with the maps produced by Geographic Information Systems obtained by satellite, making possible to arrive at population estimates and projections in short periods. This method may be applied by utilizing other variables related to demographic dynamics. It is also possible to make estimates of metropolitan areas by taking into account population data and maps. One advantage of the method is that we can estimate the variance of estimate errors.

To conduct this research, we have chosen, as an example, a small locality, Villa Milpa Alta, an area located in the southern part of Mexico City. This is a rural area where the population lives in one-family one- or at the most two-story houses and is small in terms of the number of inhabitants. To make the population estimate for the 1970-2000 period, we started with the population in the year 1970 and used aerial photographs for intermediate years. The observation and transition equations are solved recursively.