Stochastic population forecasting using functional data methods: the case of France

Heather Booth
Sophie Pennec
Rob J Hyndman

This paper applies functional data models and time series methods in forecasting the components of change, mortality, fertility and net international migration, for use in forecasting the population of France. The probabilistic population forecast is compared with the official population projections for France, based on traditional deterministic scenarios.

The use of extrapolative methods for the components presupposes that the trends of the past will be continued into the future. This assumption has often proved to be a better basis for forecasting than either structural modelling involving exogenous variables or methods based on expectation (Booth, 2006). The forecast components are combined using the cohort-component method and Monte Carlo simulation to produce probabilistic population forecasts by age and sex.

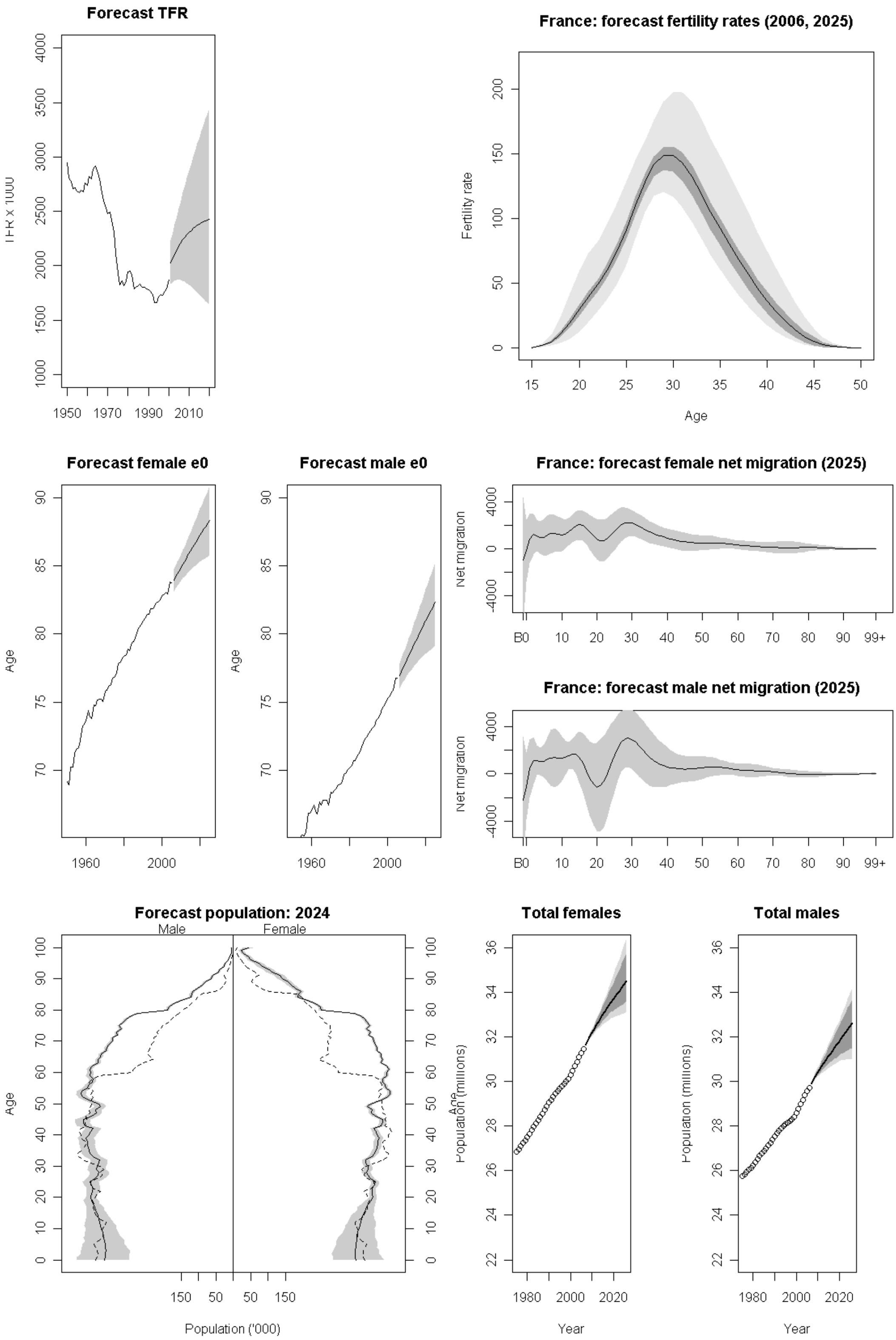
The data are age-time matrices of population and death rates for each sex and fertility rates for France for 1950-2005. Net international migration was estimated by age and sex using the demographic growth-balance equation.

Functional data forecasts of the components are made using functional data models of age-specific fertility and mortality rates and migration numbers (Hyndman & Ullah, 2007). A different Box-Cox transformation is used for each component, selected to minimise out-of-sample forecast error (Hyndman & Booth, 2008). For mortality, the transform is the logarithm; for net migration numbers (which can be negative) it is appropriate to use no transform; and for fertility an intermediate transform is used. The observed rates (or migration numbers) are first modelled using nonparametric regression as a time series of smooth age functions with age and time varying errors. The smooth functions are then modelled as functional data models; the orthogonal basis functions of this model are calculated using a principal components decomposition. The independent coefficients of the basis functions are forecast using time series methods. The procedure is a generalisation of the well-known Lee-Carter (1992) method.

Three sources of randomness are taken into account: the random variation in births, deaths or migrants from the relevant distribution (e.g., Poisson or normal); the residual error in modelling the smooth functions using a finite set of basis functions; and the randomness inherent in the time series model for each coefficient. The variances are adjusted to make sure that one-step forecast variances are equal to those obtained with historical data.

The forecast components and their probability distributions (for preliminary results, see graphs) will be compared with official assumptions. Comparison will also be made between the probabilistic population forecast and the high, medium and low official projections.

Figures: Forecasts of life expectancy, total fertility, age-specific fertility and net migration, with population forecast (preliminary results)



References

Booth, H. (2006). Demographic forecasting: 1980 to 2005 in review. *International Journal of Forecasting*, 22, 547-581.

Hyndman, R.J., & Booth, H. (2008) Stochastic population forecasts using functional data models for mortality, fertility and migration. *International Journal of Forecasting*, 24, 323-342.

Hyndman, R.J., & Ullah, S. (2007). Robust forecasting of mortality and fertility rates: a functional data approach. *Computational Statistics & Data Analysis*, 51, 4942-4956.

Lee, R.D., & Carter, L.R. (1992). Modelling and forecasting U.S. mortality. *Journal of the American Statistical Association*, 87(419), 659-671.