FORECASTING REGIONAL FERTILITY TRENDS: A FUNCTIONAL PCA

Alessandra Carioli, Netherlands Interdisciplinary Demographic Institute and University of Groningen
Prof. L. J. G. van Wissen, Netherlands Interdisciplinary Demographic Institute

Introduction

Spain offers high regional and provincial **variability in fertility trends** over the last two centuries, which **diverge from the main national trends**.

The present work aims to forecast **national and sub-national Spanish fertility time series** to understand **how future fertility trends will develop** and if evidence of geographical heterogeneity will dominate the outline of fertility in the Iberian country. The projection method employs **Functional Principal Component Analysis**. This method has bee applied to **Spain**, **17 Regions** and **50 Provinces** forecasting **TFR** and **ASFR** in the short-run **for 15 years**.

Heterogeneity is present across the sub-national dimension and regions with historical homogeneity in trends, show similar trends in the projections.

Data

Census, Vital Statistics and Municipal Register (Source: Centre d'Estudis Demogràfics, UAB)

1. Age-specific fertility schedules;
2. Female population size by age;
3. Spain, 17 regions (NUTS2) and 50 provinces (NUTS3) for years 1975-2011.

Figure 1: ASFR by age groups

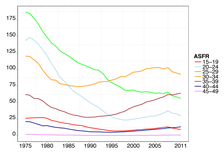
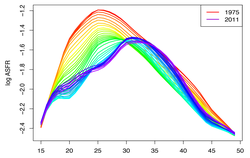


Figure 2: ASFR logs



Method

The approach ideated by Hyndman &Booth(2007), forecasts fertility rates combining non-parametric smoothing, functional data analysis and principal components analysis.

1) Estimate $\gamma_t(x)$ smooth functions using non parametric regression for each year Ediev (2014);
2) Estimate $\mu(x)$ as the mean of the smooth functions;
3) Estimate $\beta_{t,k}$ and $\phi_k(x)$ k=1,…,K using a principal components decomposition of $[y_t(x) - \hat{\mu}(x)]$;
4) Estimate time series models $\beta_{t,k}$ for k=1,…,K (exponential smoothing state space models).

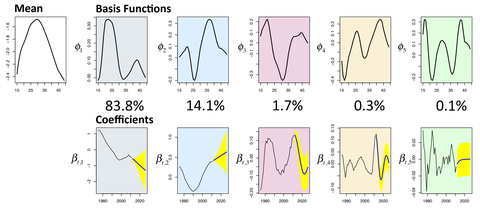
Advantages:

- ✓ **Robust forecasts;**
- ✓ **Easily interpretable of relevant dimensions from PCA;**
- ✓ **TFR and Fertility Schedule projections;**

Forecast Results: Spain

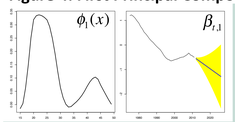
The projection model employs a **functional forecast** using a **K=6 basis functions** for **Spain and the largest regions**, and **K=9 for provinces**, 2012 to 2026.

Figure 3: PCA decomposition for first 5 components



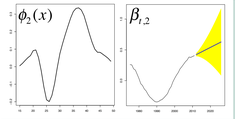
Mean portrays the average of all ASFR over the years, the negative sign underlines the decrease in fertility over the decades;
The **basis functions** model changes in fertility over **age**, while the **coefficients** model fertility changes over **time**.
Each **principal component** explains a share % of the total variation in fertility:

Figure 4: First Principal Component



✓$\beta_{t,1}\phi_1(x)$ **explains the lion's share of the overall change in fertility 83.8%** of the change in fertility and models the **decrease** in **early age fertility** ;
It models early mothers in their 20s and (to a lesser degree) late mothers in their late 30s) occurred since the end of the 1970s until the mid 1990s (lowest-low fertility) and **resumed after the onset of the 2008 economic crisis**

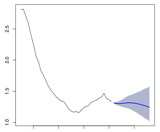
Figure 5: Second Principal Component



✓$\beta_{t,2}\phi_2(x)$ **explains 14.1%** of the change in fertility and **models recuperation of late fertility** (late mothers in their 30s). That is to say, recuperation has been much faster for women aged 30 more. This recuperation is projected to **increase** although its intensity has been lessened by the economic crisis;
✓ $\beta_{t,3}\phi_3(x)$ explains little of the change in fertility, **1.7%**, and models an overall decrease in fertility between 20 and 30 years old coinciding with economic development and crisis years (1993 and 2008);
✓ $\beta_{t,4}\phi_4(x)$ and $\beta_{t,5}\phi_5(x)$ can be interpreted as **postponement fluctuations** over time happening with a **cyclical component affecting determined age groups** (20s and 40s).

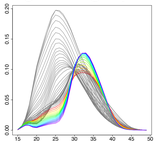
Forecast Results: Spain

Figure 6: Total Fertility Rate Projection



The projected **TFR** suggests that, despite the recuperation of fertility of the late 1990s, the **downward trend** is **likely to continue**. This drop is largely influenced by the economic crisis onset in 2008 from 1.33 (2011) to 1.24 (2026) children per woman.

Figure 7: Age Specific Fertility Schedule Projection



The projected **fertility schedule** (colors) shows how future fertility will **heavily depend on later age births**, while **early age fertility consistently drops**, although the peak age at fertility (around 33 years old) does not increase.

Forecast Results: Check 1

The **first check** implemented to control forecast accuracy is to truncate the time series by 10 years (1975-2000), project it for the next 11 years, up to 2011. The graphs below compare the forecast with the observed data.

The comparison between **observed and forecasted TFR** shows a **robust projection** of the truncated time series, with the exception of the shock created by the economic crisis.
Looking at **ASFR** and **forecast error**, there is an overestimation of early fertility and an underestimation of late fertility. This could be explained by the **stop of recuperation** of early fertility and the **new postponement** driven by the **economic crisis** or by the **decrease in the in-flow of migrants** occurred in the last years.

Figure 8: Test 1 TFR comparison

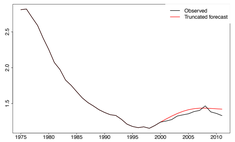


Figure 9: Test 1 Forecast error

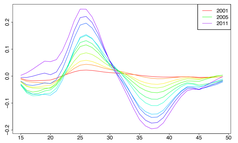
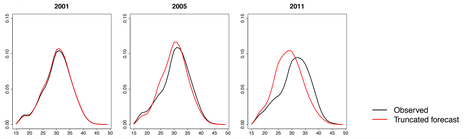
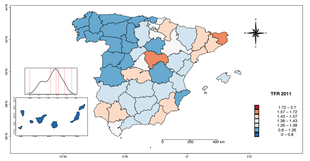


Figure 10: Test 1 Fertility Schedule comparison



Forecast Results: Regions and Provinces

In 2011 there were important high and low fertility clusters (East and North-West). The **projections** at provincial level **confirm** the **presence of patterns diverging from the national path**.

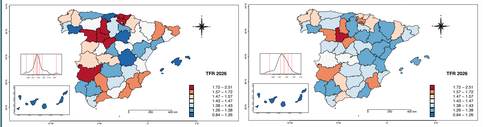


Distortions due to the small population size can bias projections. Thus, we have applied an alternative projection, which **makes use of a top-down approach.**

This method exploit the **mean and basis functions** (similar across provinces) from PCA decomposition and applies it to the various regions and provinces to obtain TFR projections.

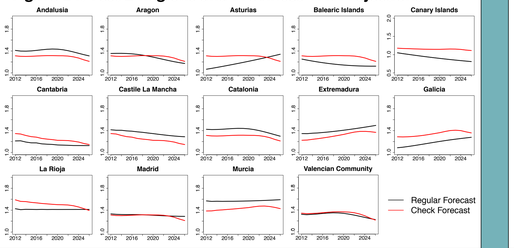
Figure 12 compares Hyndman & Booth (2008) projection and of the top-down approach. Areas with projected higher or lower fertility remain so, and TFR at the extremes of the scale is restrained (seed distribution). Also areas with historically similar fertility pattern keep such consistencies (whole North-West, East, Murcia).

Figure 12: Provincial projection of TFR: regular and Top-Down.



This test highlights the distortion across regions and provinces due to the high variability present in the ASFR obtained through regular projection. Figure 14 compares the forecast model employed for Spain and the top-down check implemented (for regions). The top-down projections adjust but still closely fit Hyndman & Booth (2008) model, confirming the robustness of both approaches.

Figure 14: Test 2 Regions and Provinces TFR Projections



Conclusion

- ✓ The **main hypothesis** of sub-national **heterogeneity** in fertility trends **finds proof in the forecasts of TFR**: although the financial crisis put a stop to the recent increase in TFR, a considerable number of provinces and regions sees an **increase in projected fertility trends**;
- ✓ The **persistent decrease of early age fertility** can explain the lion's share of past and future low fertility levels;
- ✓ Fertility projections for **Spain** depict a **future of low fertility** slightly above lowest-low fertility, 1.3 children per woman
- ✓ Fertility projections seem heavily **influenced by the effect economic crisis** has on fertility rates;
- ✓ The rise in period fertility levels registered during the period 1997-2008, reached a hiatus due to the economic crisis that negatively impacted fertility recuperation everywhere in Spain;
- ✓ PCA suggests that the **rise in fertility** occurred during the early 2000s was mainly due to an increase in **early age** fertility;
- ✓ Postponement of fertility is the **driving force** behind provinces with **projected increasing fertility**. It is likely to continue in the future but to a **much lesser degree than in the past**;
- ✓ Areas showing diverging trends (increase) are geographically closed, such as the North-West and Center West: part of Galicia, Pais Vascos, part of Castilla y Leon;

References

Ediev, D. M. (2013). **Spline interpolation for demographic needs**. (Unpublished manuscript)

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

Hyndman R.J. & Booth H. (2008): **Stochastic population forecasts using functional data models for mortality, fertility and migration**. *International Journal of Forecasting* 2:323-342.

Hyndman, R. J., Booth, H., Tickle, L., & Maindonald, J. (2013, January 27). **Demography: Forecasting mortality, fertility, migration and population data**. R-CRAN.