IOCR 2026 Conference
The 1st International Online Conference on Risks
6–7 July 2026 | Online
MDPI

Dynamic Principal Component Analysis for Syndemic Factors in an Extended Lee–Carter Mortality Model

Maria Carannante[1], Valeria D'Amato[2], Steven Haberman[3], Massimiliano Menzietti[4]
[1]Link Campus University
[2]Sapienza University of Rome
[3]Bayes Business School, City St. George University of London
[4]University of Salerno

INTRODUCTION & AIM

Mortality projections are affected by systematic deviations arising from model misspecification, commonly referred to as model risk. Such deviations may result from period shocks that temporarily alter mortality patterns and reduce forecasting accuracy.

The Lee–Carter (LC) model is one of the most widely used approaches for mortality forecasting thanks to its parsimonious structure, interpretability, and ability to capture long-term longevity improvements. Several studies have extended the LC framework by incorporating external covariates, such as economic, environmental, and health-related indicators, to explain part of the heterogeneity not captured by the traditional model.

Recent research has shown that variables including GDP, public health conditions, climate indicators, and lifestyle factors can significantly improve mortality projections by identifying common trends between mortality and external risk factors. Recently an integrated approach using PCA to combine economic, health, and lifestyle variables as a unique exogenous variable in LC extension is proposed.

The aim of this research is to improve the combined PCA and LC integration, using a dynamic PCA (dPCA), to address multicollinearity and preserve temporal dependence among covariates. In this framework, a syndemic index, derived from multiple socioeconomic and health-related variables is proposed for Italian mortality forecasting.

METHOD

Let $z_t = (z_{1t}, \dots, z_{pt})^\top$ be a vector of covariates observed over time. A lagged data structure is constructed to incorporate temporal information, enabling the extraction of dynamic components that capture both contemporaneous and past behaviour of the variables. The extracted dynamic sindemic index is defined as $f_t = z_t w, t = 1,\dots,T$, where $w \in R_p$ is the loading vector defining a linear combination of the exogenous variables.
To estimate the syndemic index through dPCA, the following inverse problem is defined

$$\max_{w,\widetilde{w}}; w^\top Z_{s+1}^\top Z_s, (\widetilde{w} \otimes w) \qquad \text{s.t.} \quad |w| = 1,\ |\widetilde{w}| = 1$$

Where $\widetilde{w}$ is the temporal weighting vector associated with the lag structure of the model.
The objective function maximizes a lagged cross-covariance between present and future linear projections of the multivariate process, through the iteratively estimation of w and $\widetilde{w}$.
The resulting latent factor $f(t) = \{f_t\}_{t=1}^T$ is the syndemic index to be incorporated into an extended Lee–Carter model as an exogenous driver

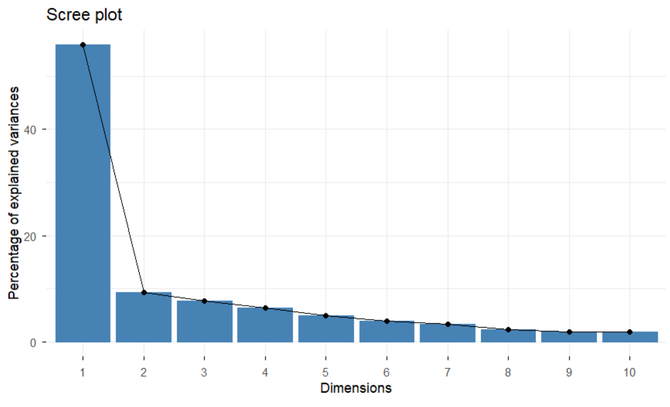
$$\ln \mu\,(x,t) = \alpha(x) + \beta(x)\kappa(t) + \theta(x) f(t) + \varepsilon(x,t)$$

dPCA provides a principled method to reduce the dimensionality of exogenous variables while capturing their dynamic influence on mortality trends.

RESULTS & DISCUSSION

The syndemic index is defined on ISTAT data for Italy and Italian regions from 1995 to 2023.

Description — Acronym
Activity rate — ACT
Alcohol habit rate — ALC
Smoke habit rate — SMK
Obesity rate — OBESE
Per capita nominal GDP — GDP
Average household size — HHSIZE
Average household income — HHINC
No chronic diseases rate — HEALTH
Medical doctors rate — MED
Pediatricians rate — PED
Health expenditure per capita (household) — HHEXP
Public health expenditure per capita — PHEXP
Hospital beds rate — HOSP
Retirement home beds rate — RES



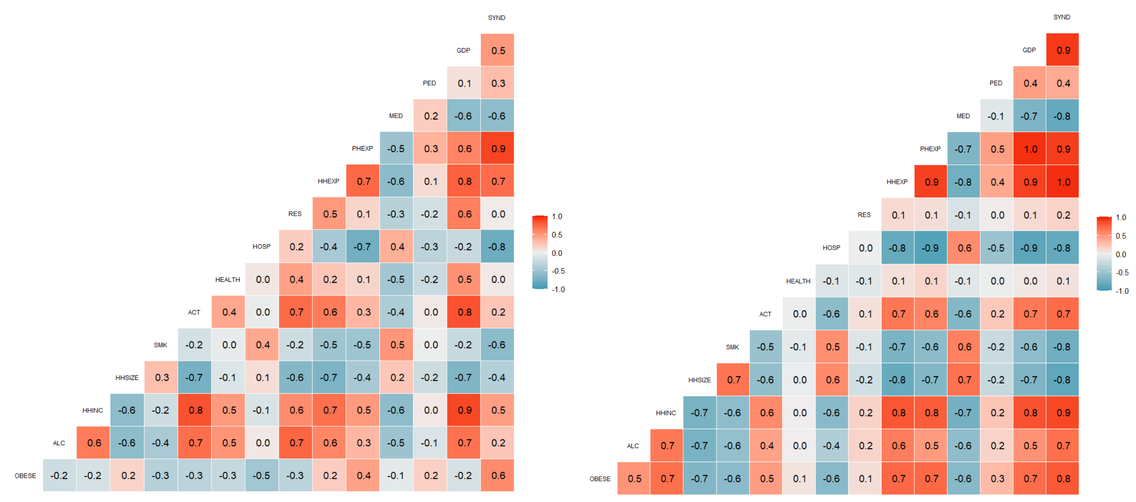
The scree plot shows that selection of variables leads to the definition of a unidimensional indicator, given the significant difference in explained variability between the first and the second latent factors. This result is consistent with the aim of define a unique syndemic index.
Goodness of fit of PCA and dPCA is measured through mean squared reconstruction error (MSRE) and lagged correlation (ρ)

MRSE: PCA = 0.343, dPCA = 0.441. ρ: PCA = 0.738, dPCA = 0.989.

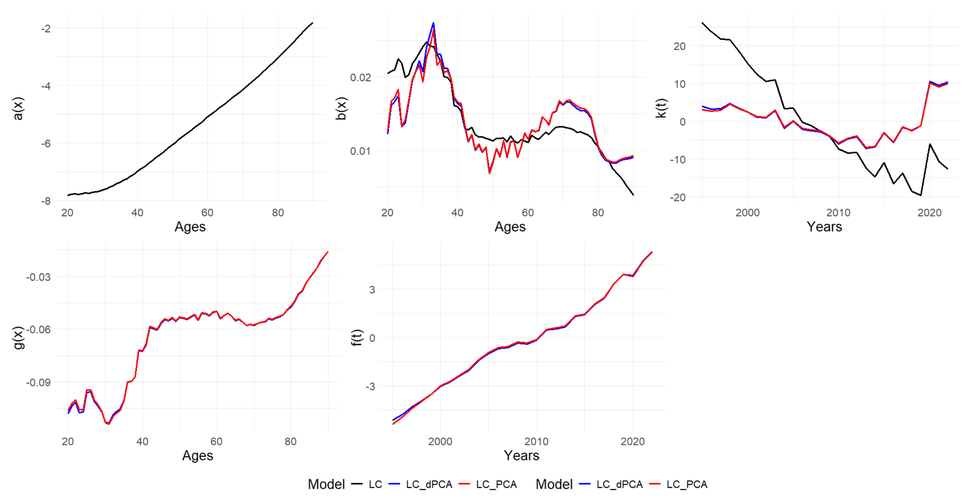
PCA optimise the variability space, while dPCA optimise temporal coherence.

Model deviance: LC: 27628; PCA: 26230; dPCA: 26221

Model comparison via deviance suggests improved fit for both PCA and dPCA extensions, although differences between dynamic and static syndemic specifications are marginal.



Static PCA (left) and dPCA (right) correlation plots exhibit a consistent overall correlation structure, confirming the robustness of the syndemic index (SYND). In both cases, SYND is positively associated with economic and health expenditure variables (GDP, HHINC, HHEXP, PHEXP) and obesity (OBESE), while showing negative correlations with healthcare capacity indicators such as hospital beds (HOSP), medical density (MED), and smoking prevalence (SMK). However, dPCA highlights a stronger separation between these two clusters. The dynamic formulation amplifies correlations, reinforcing the economic–metabolic dimension on one side and the structural healthcare system variables on the other.



LC parameters shows that $a(x)$ is consistent across models; $b(x)$ better capture adult and old-age effects with syndemic index with slight difference for dPCA than PCA; $k(t)$ highlighs a non always decreasing trend as in standard LC, with more shifts for dPCA in the boundary years.

CONCLUSIONS

Using the dPCA model to construct a syndemic index shows its advantages in terms of the model's explanatory ability. While PCA can already improve the definition of mortality trends and age-specific effects by including variables that construct a syndemic index, dPCA allows for a more refined selection of the variability space, while maintaining the dependency structure and therefore, the robustness of the results.
From the perspective of parameter estimation of a LC model, the improvement in index construction is evident in more marked differences, especially for older ages, where it is more difficult to obtain stable estimates of mortality effects.

FUTURE WORK

Future developments focus on two aspects:

Improving variable selection, considering indices already consolidated in the literature, such as the Index of Multiple Deprivation (IMD), in order to consider additional aspects concerning the syndemic phenomenon.

Improve the weight optimization algorithm, based on the temporal relationship conditional on the effect on mortality, using an estimation based on the PLS algorithm.