Frailty heterogeneity and differentials in mortality: a focus on the health insurance market

VALERIA D'AMATO[1]

JOINT RESEARCH WITH MARIA CARANNANTE[1], STEVEN HABERMAN[2], AND MASSIMILIANO MENZIETTI[3]

[1]DEPARTMENT OF PHARMACY, UNIVERSITY OF SALERNO

[2]FACULTY OF ACTUARIAL SCIENCE AND INSURANCE, BAYES BUSINESS SCHOOL, CITY UNIVERSITY OF LONDON

[3]GIOVANNI ANANIA" DEPARTMENT OF ECONOMICS, STATISTICS AND FINANCE, UNIVERSITY OF CALABRIA

Agenda

1. Model risks in projection mortality;
2. Proposal of frailty mortality model;
3. Identification of the factors that allows to quantify frailty.

Introduction and research goals

Model risk

Mortality projections are affected by systematic deviations.

Deviations due to misspecification of mortality model are the **model risk** (Pitacco *et al.* 2009).

Model risk includes shocks caused by **period effects** that temporary change the mortality behaviours.

Frailty

Frailty is the set of unobservable factors that determines the heterogeneity in mortality.

In actuarial literature, frailty is distributed as a non-negative real random variable.

Frailty represents **individual deviations** in mortality from the average behaviour estimated by the model, as analysed by Beard (1959, 1971) and Vaupel *et al.*(1979).

Frailty

Let Z_x the continuous random frailty at age x, with a probability density function, $g_x(z)$.

Let $\mu_x(z)$ the conditional force of mortality for an individual in a population group at age x and with a frailty level z:

$$\mu_x(z) = \lim_{t \to 0} \frac{P(T_x \leq t | Z_x = z)}{t}$$

where T_x being the remaining lifetime.

Note that Z_x is invariant with respect to t.

Frailty

Vaupel *et al.* (1979) define the frailty as a multiplicative factor of the force of mortality:

$$\mu_x(z) = \mu_x \cdot z$$

The survival function of an individual at age 0 considering the frailty is defined as follows:

$$S(x,z) = e^{-\int_1^x \mu_t(Z)dt} = e^{-zH(x)}$$

With $H(x)$the cumulative standard force of mortality in the interval $(0,x)$.

The role of frailty in mortality projections

Improvements in longevity increased mortality heterogeneity due to the onset of **co-morbidities** (Xu *et al.* 2019).

In much of actuarial literature, **co-morbidities, frailty and disability** are often used interchangeably in the identification of the vulnerable elderly (Fried *et al.* 2001, 2004, Jones *et al.* 2004).

The concept of frailty as the onset of a state of **health-related vulnerability to mortality** is inconsistent with the idea that frailty is invariant over time, as defined in most existing models (i.e. Haberman and Butt 2004, Su and Sherris 2012).

Research goals

1. Proposal of a **Lee-Carter family model** that includes a multiplicative parameter of frailty as a factor that determines the mortality by age and time;
2. Definition of a **time-varying** and quantifiable measure of frailty;
3. Detection of the factors that determines the measure of frailty, using the **variable importance** of a tree-based algorithm.

Mortality projection with frailty model

Frailty as a measurable parameter

Vaupel (1979) stresses the need to introduce a frailty parameter within the mortality models.

Frailty was considered difficult to quantify, as it is a latent variable that includes a number of unspecified factors.

This leads to poor specification of the models, with consequent underestimation of the mortality trends.

Frailty as a measurable parameter

Frailty parameter could be included in Lee-Carter family model, thanks to its desiderable properties:

- Few parameters and easy to interpret;
- Allows modeling improvements in longevity;
- Requires a limited number of a priori hypotheses;
- Includes a second stage re-estimation of deaths.

Lee-Carter based frailty model

Let $\mu_{x,t}$ the force of mortality. A Lee-Carter model (Lee and Carter 1992) is defined as follows:

$$y_{x,t} = \log(\mu_{x,t}) = a_x + b_x k_t + \varepsilon_{xt}$$

We define the force of mortality conditional to frailty $\mu_{x,t,z}$ and the relative model:

$$y_{x,t,z} = \log(\mu_{x,t,z}) = a_x + b_x k_t + z_t + \varepsilon_{xt}$$

Where z_t is a time-dependent multiplicative coefficient of the force of mortality.

Lee-Carter based frailty model

We can express the ordinal least squares optimization as minimizing the squared sum of errors:

$$\min_{a_x, b_x, k_t, z_t} \sum_{x,t} \left[y_{x,t,z} - (a_x + b_x k_t + z_t) \right]^2$$

the objective function is minimised by equating to 0 the first derivatives with respect to a_x, b_x, k_t, z_t

Lee-Carter based frailty model

The updating for the parameters can also obtained recursively using normal equation. The required parameters will be obtained numerically according the following algorithm:

$$a_x = \frac{\sum_t D_{xt}\left[y_{xtz} - b_x \widehat{k_t} - \hat{z}_t\right]}{\sum_t D_{xt}}$$

$$b_x = \frac{\sum_t D_{xt}\widehat{k_t}\left[y_{xtz} - \widehat{a_x} - \hat{z}_t\right]}{\sum_t D_{xt}\widehat{k_t}^2}$$

$$k_t = \frac{\sum_t D_{xt}\widehat{b_x}\left[y_{xtz} - \widehat{a_x} - \hat{z}_t\right]}{\sum_t D_{xt}\widehat{b_x}^2}$$

$$z_t = \frac{\sum_t D_{xt}\left[y_{xtz} - \widehat{a_x} - \widehat{b_x k_t}\right]}{\sum_t D_{xt}}$$

Lee-Carter based frailty model

The estimation of a_x, b_x, k_t, z_t is performed in two stages.

The first determines the values of parameters and the second a re-estimation of the deaths by age x and time t.

In this sense, D_{xt} of the second stage re-estimation could be interpreted as the **deaths corrected by frailty**.

Frailty heterogeneity factors identification

Frailty quantification

The heterogeneity in frailty of a demographic population that determines **differentials in mortality**.

The literature shows that neglecting this feature leads to a **bias in projecting the longevity phenomenon**.

To avoid a misrepresentation of the longevity it is necessary to estimate a **frailty score**, using a set of covariates.

Frailty quantification

The idea is to focus on the frailty and detect the covariates to determine its heterogeneity.

The use of variable importance of a Random Forest algorithm allow overcoming the functional form of the model and considering non-linear correlations.

ELSA data

The English Longitudinal Study on Ageing (ELSA) is a longitudinal household survey dataset for the study of health, economic position, and quality of life among the elderly.

The dataset is harmonised with similar ones of many countries.

ELSA data

The dataset is composed of 9 waves from 2002 to 2019.

The starting sample included 11,050 respondents aged 50 and over on March 1, 2002.

The sample is refreshed every two waves, including individuals aged 50 years and over and their partners.

ELSA data

The survey is divided into parts, each of which deals with a different theme in the life of the respondents:

A. DEMOGRAPHICS, IDENTIFIERS, AND WEIGHTS

B. HEALTH

C. INSURANCE

D. COGNITION

E. FINANCIAL AND HOUSING WEALTH

F. INCOME AND CONSUMPTION

G. FAMILY STRUCTURE

H. EMPLOYMENT HISTORY

ELSA data

The survey is divided into parts, each of which deals with a different theme in the life of the respondents:

I. RETIREMENT & EXPECTATIONS

J. PENSION

K. PHYSICAL MEASURES

L. ASSISTANCE AND CAREGIVING

M. STRESS

O. END OF LIFE PLANNING

P. CHILDHOOD

Q. PSYCHOSOCIAL

ELSA data

The nine-waves harmonized dataset (Banks *et al.* 2021) includes any individual interviewed at least once, for a total of 19,802 respondents.

Respondents are individuals who were age-eligible at the time of their first interview, while the unit of observations are: individual, the couple (the respondent and his/her partner) and the household.

Data pre-processing

The original dataset is in the form of a cross-sectional data matrix, with the respondent i on the rows and the variables x_i on the columns, replicating the variables for each wave t.

We obtain a panel data matrix, with the respondent and waves on the rows i, t and the variables $x_{i,t}$ on the columns.

Data pre-processing

To do this we do the following steps for each individual respondent i:

1. We detect the first and the last wave in which the respondent participated;
2. We include in the dataset only the waves included in the first and the last waves;
3. We detect the waves included in between the first and the last waves at which the respondent does not participate;
4. The variables with missing data due to non-response are imputed using the median of the response of the individual in the other waves;
5. Other missing data are imputed using the median of the respondents. They represent about the 1% of the sample.

Data pre-processing

The purpose of the random forest is to identify the variables relevant measure the individual frailty.

Considering the size of the original matrix, it is necessary to make a qualitative selection of the variables before implementing the model.

The target variable is the **health status**, measured on a scale ranging from 1, indicating excellent, to 5, indicating poor health status.

The feature variables are selected from the following sections of the survey: A: Demographics, Identifiers, and Weights; B: Health; C: insurance; F: income and consumption; H: employment history; I: retirement and expectations; L: assistance and caregiving; O: end of life planning for a total of **35 variables.**

Summary statistics

Variable	Category	Relative frequency
gender	male	0.443
	female	0.557
race	white	0.964
	non-white	0.036
education	less than upper secondary	0.313
	upper secondary and vocational trading	0.527
	tertiary	0.160
partner	married or civil partnership	0.670
	partnered	0.041
	separated	0.012
	divorced	0.080
	widowed	0.148
	never married	0.049
birthplace	UK	0.912
	Other	0.088
Birth year	min	1908
	q1	1936
	median	1945
	q3	1951
	max	1988

Socio-demographic variables

Summary statistics

Variable	Category	Relative frequency
hypertension	no	0.590
	yes	0.410
diabetes	no	0.899
	yes	0.101
cancer	no	0.907
	yes	0.093
lung	no	0.940
	yes	0.060
heart	no	0.808
	yes	0.192
stroke	no	0.954
	yes	0.046
psyche	no	0.904
	yes	0.096
arthritis	no	0.643
	yes	0.357
asthmae	no	0.871
	yes	0.129
cataracts	no	0.791
	yes	0.209

Co-morbidities variables

Summary statistics

Variable	Category	Relative frequency
parkinson	no	0.993
	yes	0.007
hipfracture	no	0.984
	yes	0.016
angina	no	0.922
	yes	0.078
heartattack	no	0.948
	yes	0.052
rhythm	no	0.905
	yes	0.095
osteoporosis	no	0.933
	yes	0.067
sight	excellent	0.146
	very good	0.333
	good	0.381
	fair	0.105
	poor	0.031
	blind	0.004
hearing	excellent	0.188
	very good	0.276
	good	0.322
	fair	0.164
	poor	0.049

Co-morbidities variables

Summary statistics

Variable	Category	Relative frequency
health status	excellent	0.134
	very good	0.304
	good	0.318
	fair	0.176
	poor	0.068
adl	0	0.815
	1	0.090
	2	0.042
	3	0.023
	4	0.014
	5	0.010
	6	0.008
mobility	0	0.470
	1	0.165
	2	0.114
	3	0.083
	4	0.066
	5	0.053
	6	0.034
	7	0.015
physical activity	more than once a week	0.772
	once a week	0.094
	one to three times a month	0.032
	hardly ever or never	0.102

Health status and habit variables

Summary statistics

Variable	Category	Relative frequency
physical activity	more than once a week	0.772
	once a week	0.094
	one to three times a month	0.032
	hardly ever or never	0.102
drink	no	0.131
	yes	0.869
smoke	no	0.377
	yes	0.623
social participation	no	0.701
	yes	0.299
informal care	no	0.952
	yes	0.048
formal care	no	0.949
	yes	0.051
professional care	no	0.949
	yes	0.051
Survival probability	min	0
	q1	50
	median	60
	q3	80
	max	100

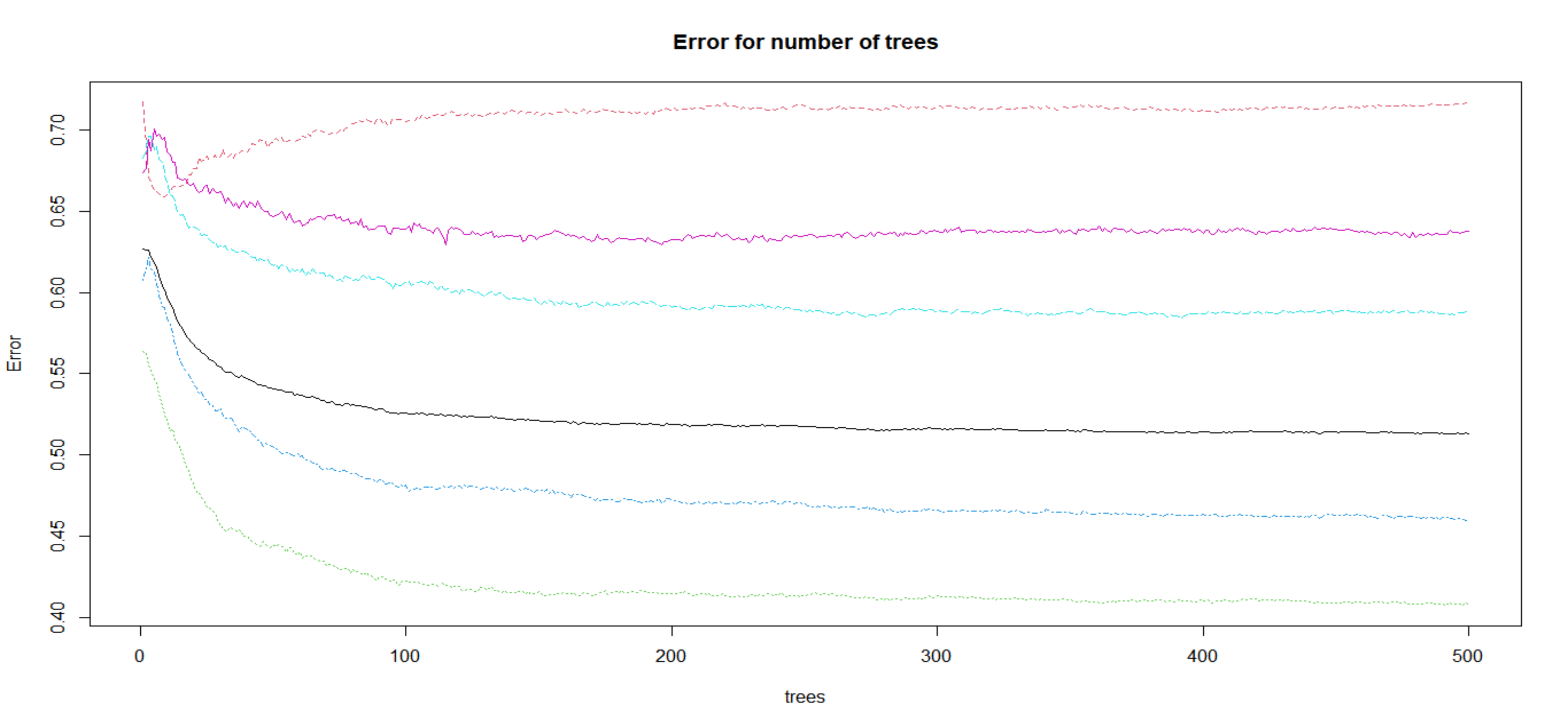
Health status and habit variables

Summary statistics

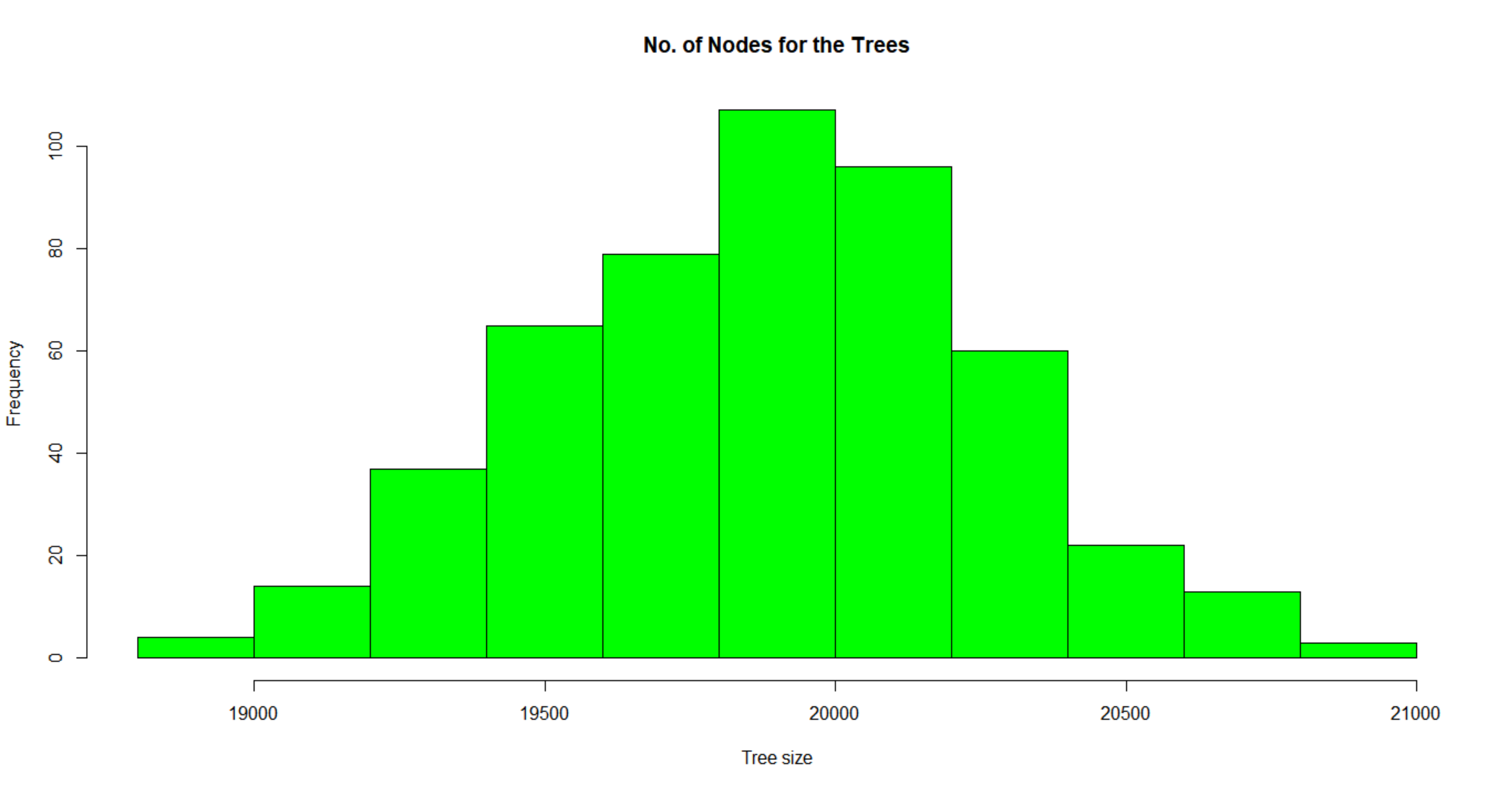
Variable	Category	Relative frequency
labour force status	employed	0.263
	self-employed	0.063
	unemployed	0.011
	partly retired	0.006
	retired	0.539
	disabled	0.051
	looking after home or family	0.067
household income	min	-81174
	q1	12285
	median	20229
	q3	32111
	max	879211
health insurance	no	0.866
	yes	0.134
life insurance	no	0.657
	yes	0.343

Economic variables

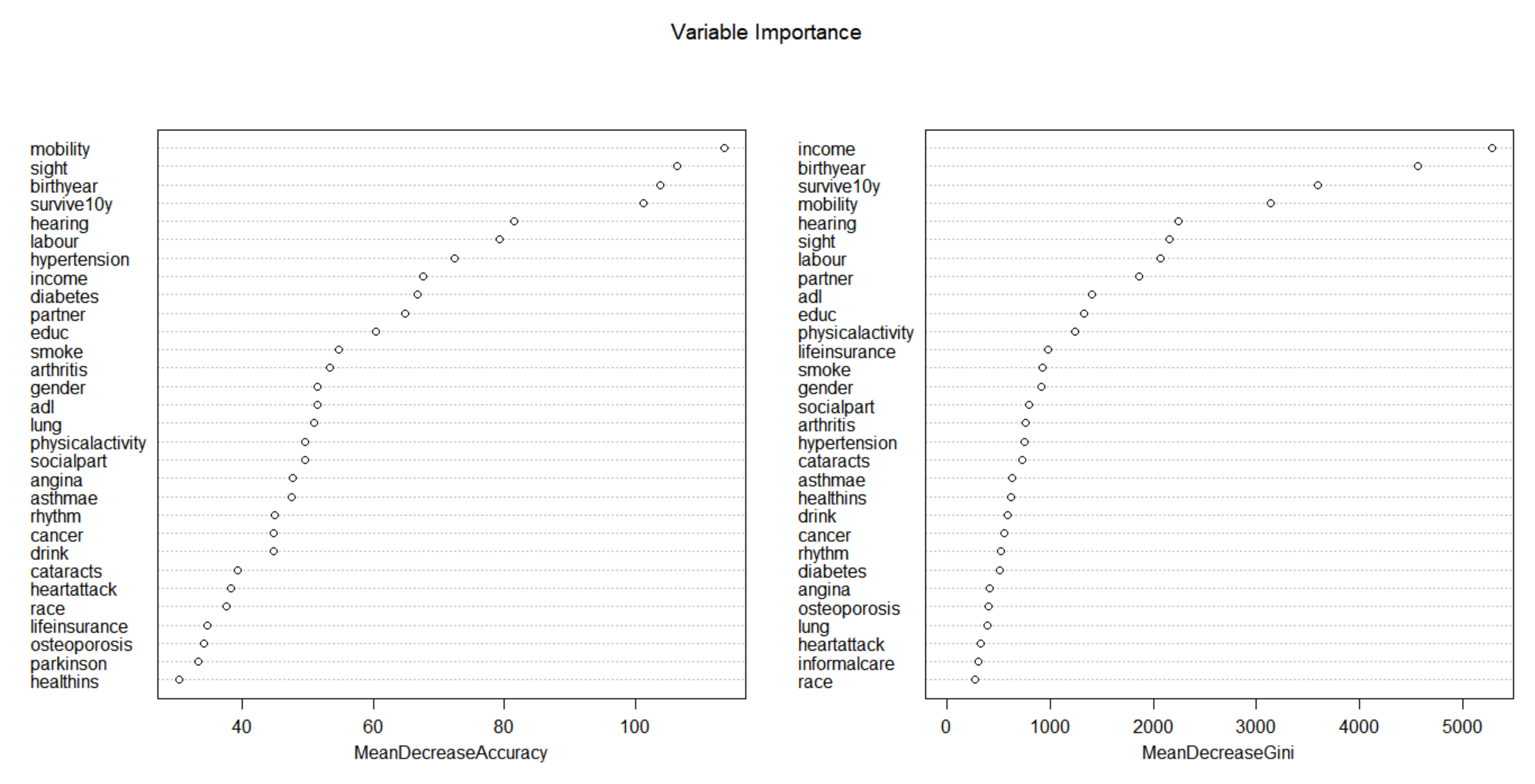
Random forest



Random forest



Random forest



Random forest

Variable importance is considered both as accuracy and leaf homogeneity. In particular:

- For the accuracy, the most important variables are: mobility, sight, birthyear, survive10y, hearing, labour, hypertension, income, diabetes and partner;
- For the leaf homogeneity the most important variables are: income, birthyear, survive10y, mobility, hearing, sight, labour, partner, adl, education.

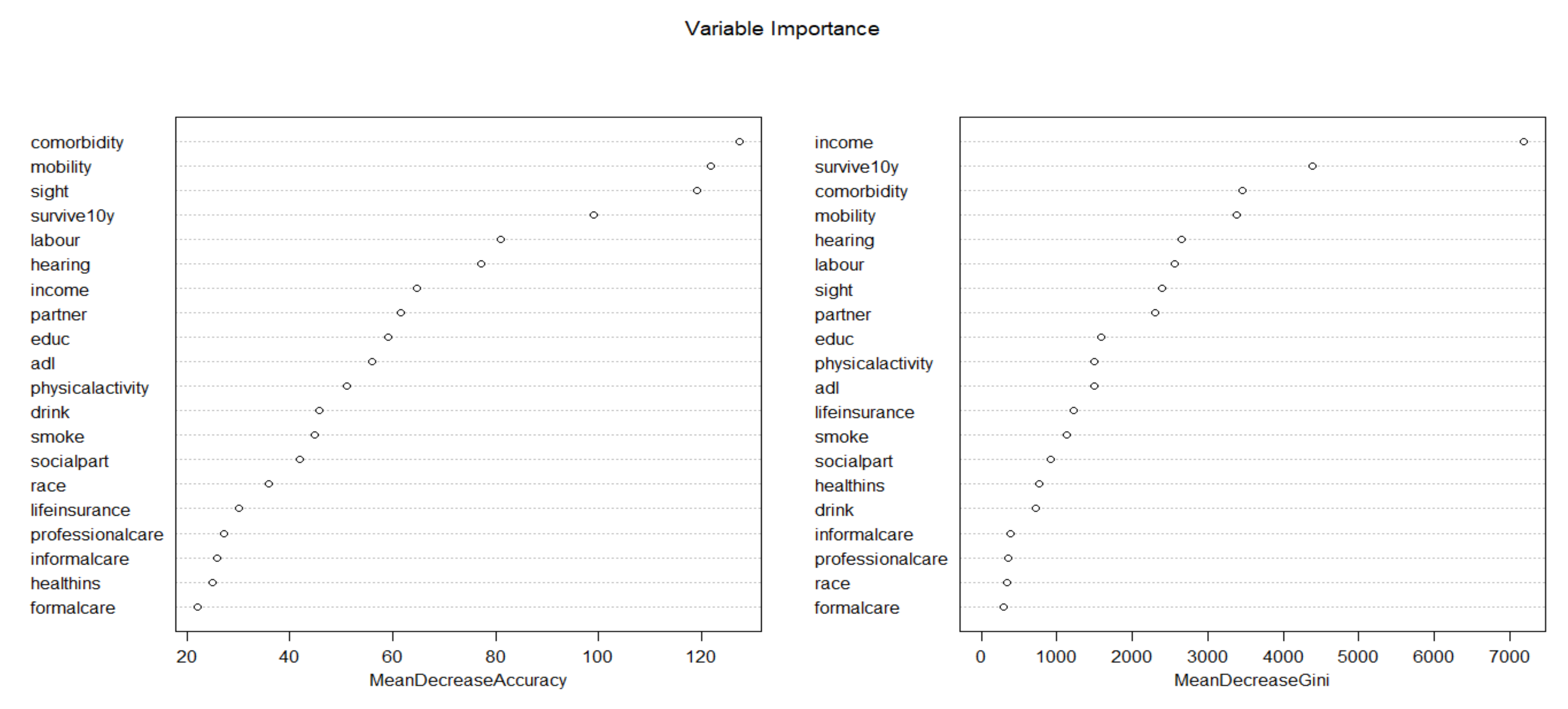
Frailty appears to be linked to an individual's self-sufficiency.

Random forest

Comparing the summary statistics and the variable importance for the co-morbidities variables, we observe the higher the incidence of a disease in the population, the higher its importance. This could lead to an underestimation of the role of comorbidities in the definition of frailty.

For this reason, we again estimate the RF by constructing, in a similar way to the adl, a variable of the **number of co-morbidities** of an individual. Furthermore, the model is estimated without considering the year of birth, being a redundant variable as it is already considered in the mortality models.

Random forest



Random forest

In this formulation, the number of co-morbidities show a significative importance, keeping the importance of the other variables invariant, which suggests that the model is robust.

The RF is used to simulate a frailty variable z for age x and time t, since frailty is not observable for a number of ages and times large enough to allow to estimate a mortality model.

Conclusions

Heterogeneity in frailty causes mortality differentials;

Neglecting the frailty in mortality projections involves dangerous bias;

To capture the mortality differentials, we provide a stochastic mortality model in LC setting including frailty;

Random forest shows the most important variables determining frailty, that is co-morbidity.

Further research concerns investigating the hypotheses on frailty distribution.

Reference

Banks, J., Batty, G. David, Breedvelt, J., Coughlin, K., Crawford, R., Marmot, M., Nazroo, J., Oldfield, Z., Steel, N., Steptoe, A., Wood, M., Zaninotto, P. (2021). English Longitudinal Study of Ageing: Waves 0-9, 1998-2019. [data collection]. 37th Edition. UK Data Service. SN: 5050, DOI: 10.5255/UKDA-SN-5050-24

Beard, R. E. (1959). Note on some mathematical mortality models. In CIBA Foundation Colloquia on Ageing (ed. C. E. W. Wolstenholme and M. O. Connor) 5, Boston: 302–311

Beard, R. E. (1971). Some aspects of theories of mortality, cause of death analysis, forecasting and stochastic processes. In Biological aspects of demography (ed. W. Brass), pp. 57–68. Taylor & Francis, London

Fried, L.P., Ferrucci, L., Darer, J., Williamson, J.D., Anderson, G. (2004) Untangling the concepts of disability, frailty, and comorbidity: implications for improved targeting and care, The Journals of Gerontology: Series A, Biological sciences and medical sciences 59, 255–263

Fried L.P,. Tangen C.M., Walston J. (2001) Frailty in older adults: evidence for a phenotype, The Journals of Gerontology: Series A, Biological sciences and medical sciences 56, 146–156

Jones D.M., Song X., Rockwood K. (2004) Operationalizing a frailty index from a standardized comprehensive geriatric assessment, Journal of the American Geriatrics Society 52, 1929–1933

Haberman S, Butt, Z (2004), Application of frailty-based mortality models using generalized linear models. ASTIN Bulletin, 34(1), 175-197

Pitacco E., Denuit M., Haberman S., Olivieri A., (2009), Modeling Longevity Dynamics for Pensions and Annuity Business, Oxford University Press, Oxford: United Kingdom

Su, S. and Sherris, M. (2012). Heterogeneity of Australian population mortality and implications for a viable life annuity market, Insurance: Mathematics and Economics 51, 322–332

Vaupel, J. W., K. G. Manton, and E. Stallard (1979) The impact of heterogeneity in individual frailty on the dynamics of mortality, Demography 16(3): 439–54

Thanks for your attention!

vdamato@unisa.it