UNIVERSIDAD CARLOS III DE MADRID
working papers

UC3M Working Papers
Statistics and Econometrics
17-14
ISSN 2387-0303
Junio 2017

Departamento de Estadística
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (34) 91 624-98-48

Estimating life expectancy free of dependency: Group characterization through the proximity to the deepest dependency path

Irene Albarrán[a], Pablo J. Alonso-González[b] and Aurea Grané[a]

Abstract

The aging of population is perhaps the most important problem that developed countries must face in the near future. Dependency can be seen as a consequence of the process of gradual aging. In a health context, this contingency is defined as a lack of autonomy in performing basic activities of daily living that requires the care of another person or significant help. In Europe in general and in Spain in particular this phenomena represents a problem with economic, political, social and demographic implications. The prevalence of dependency in the population, as well as its intensity and its evolution over the course of a person's life are issues of greatest importance that should be addressed. The aim of this work is to estimate life expectancy free of dependency (LEFD) using categorical data and individual dependency trajectories that are obtained using the whole medical history concerning the dependency situation of each individual from birth up to 2008, contained in database EDAD 2008. In particular, we estimate LEFD in several scenarios attending to gender, proximity-group and dependency degree. Proximity-groups are established according to an L2-type distance from the dependency trajectories to a central trend within each age-gender group, using functional data techniques. The main findings are: First, the estimated LEFD curves reach higher values for women than for men; Second, their decreasing rate is higher (and more abrupt) for men than for women; Third, the more the dependency trajectories depart from the central trend, the more the gap between the LEFD for major dependency and the other dependency situations widens; Finally, we show evidence that to estimate LEFD ignoring the partition by proximity-groups may lead to non-representative LEFD estimates.

Keywords: ADL, Cox Regression, Dependency, Disability, Functional Data

[a] Department of Statistics, Universidad Carlos III de Madrid.
[b] Economics Department, Universidad de Alcalá.

Financial support from research project MTM2014-56535-R by the Spanish Ministry of Economy and Competitiveness.

Estimating life expectancy free of dependency: Group characterization through the proximity to the deepest dependency path

Irene Albarrán[(1)], Pablo J. Alonso-González[(2)], Aurea Grané[(1)]

(1) Statistics Department. Universidad Carlos III de Madrid.
(2) Economics Department. Universidad de Alcalá. [*†‡]

Abstract

The aging of population is perhaps the most important problem that developed countries must face in the near future. Dependency can be seen as a consequence of the process of gradual aging. In a health context, this contingency is defined as a lack of autonomy in performing basic activities of daily living that requires the care of another person or significant help. In Europe in general and in Spain in particular this phenomena represents a problem with economic, political, social and demographic implications. The prevalence of dependency in the population, as well as its intensity and its evolution over the course of a person's life are issues of greatest importance that should be addressed. The aim of this work is to estimate life expectancy free of dependency (LEFD) using categorical data and individual dependency trajectories that are obtained using the whole medical history concerning the dependency situation of each individual from birth up to 2008, contained in database EDAD 2008. In particular, we estimate LEFD in several scenarios attending to gender, proximity-group and dependency degree. Proximity-groups are established according to an L^2-type distance from the dependency trajectories to a central trend within each age-gender group, using functional data techniques. The main findings are: First, the estimated LEFD curves reach higher values for women than for men; Second, their decreasing rate is higher (and more abrupt) for men than for women; Third, the more the dependency trajectories depart from the central trend, the more the gap between the LEFD for major dependency and the other dependency situations widens; Finally, we show evidence that to estimate LEFD ignoring the partition by proximity-groups may lead to non-representative LEFD estimates.

Keywords: ADL, Cox Regression, Dependency, Disability, Functional Data

AMS Subject Classification: 62-07, 62-09, 62H20, 62H99, 62P05.

[*] Financial support from research project MTM2014-56535-R by the Spanish Ministry of Economy and Competitiveness.

[†] Authors' address: (1) Statistics Department, Universidad Carlos III de Madrid, C/ Madrid 126, 28903 Getafe, Spain. (2) Statistics Department, Universidad de Alcalá, Plaza de la Victoria 2, 28802 Alcalá de Henares, Spain. E–mails: I. Albarrán, `irene.albarran@uc3m.es`, P. Alonso, `pablo.alonsog@uah.es`, A. Grané, `aurea.grane@uc3m.es`

[‡] Corresponding author: A. Grané; Date: June 14, 2017.

1 Introduction

Population aging is an ongoing global phenomenon and a powerful and transforming demographic force. Several reports have warned about the need of evidence-based policies sustained on rigourous research and on the importance to prioritize healthy aging and well-being (see for instance, WHO 2011a, WHO 2011b, Lloyd-Sherlock et al. 2012). In particular, one of the eight tackling societal challenges of the European program Horizon 2020 is concerned with these issues.

The decreasing mortality and increasing life expectancy in most Western European countries during the last decades are well documented (Eurostat 2009). A key issue is to find out whether the increased life expectancy is healthy or either associated with an increase or decrease in disability (Fries 1983).

Life expectancy is one of the most used indicators in order to measure quantity of life. However, if the aim is to measure quality of life, indicators related to healthy life expectancy should be used. These kind of indicators introduce health status (morbility or disability) of the individual (Robine et al. 2003, Robine and Ritche 1991). For instance, Sanderson and Scherbov (2010) propose disability-free life expectancy as a consistent disability aging measure for many countries in order to provide better tools for policy makers. In this paper we are interested in dependency, which is a more restrictive concept than disability. Therefore, our indicator will be life expectancy free of dependency (LEFD) (see Martel and Bélanger 2000).

Other authors have addressed disability mainly with two different approaches: The first one, considers longitudinal data whereas the second one consists only in considering particular calendar years. In the former, Manton and Land (2000) construct a panel from the National Long Term Care Survey of elderly persons in the years 1982 to 1996, and considering 5-year-cohorts from 65 to 85 years old estimate the active life expectancy for male and female in the U.S. elderly population. Crimmins et al. (2009) use two longitudinal surveys of the U.S. community-dwelling population collected from 1984 to 2000 (LSOA I and LSOA II) to estimate disability-free life expectancy for Americans 70 years old and older and compare the transition probabilities for nondisabled, IADL disabled and ADL disabled people[1]. In the latter approach, Cambois et al. (2001) use 1980 and 1991 Surveys on Health and Medical Care (INSEE) to estimate the disability-free life expectancy in the French male population by occupational groups and compare them for two particular ages, 35 and 60 years old. Yi et al. (2004) use the Chinese Longitudinal Healthy Longevity survey collected in 1998 and 2000 to estimate disabled life expectancy for oldest-old Chinese population aged 80-105 for several morbidity profiles.

In the Spanish case, three surveys about disability have been undertaken by the Spanish Statistical Office (INE) during the last 30 years. The first one was conducted in 1986 and was the Survey about Disabilities, Impairments and Handicaps. Then came the Survey about Disabilities, Impairments and Health Status, that was prepared using data from 1999. Finally, the last one was EDAD in 2008. Although all of them talk about disabilities, it is not possible to track this phenomenon in a homogeneous way along the years because the definition of that concept changed through the years depending on the classification that was used to prepare the survey. In

[1] IADL stands for Instrumental Activities of Daily Living and ADL stands for Activities of Daily Living

fact, EDAD 2008 was the first Spanish survey that used the internationally accepted measures established by the 'International classification of functioning, disability and health'.[2] Following the World Health Organization recommendations, this survey is based on the concept of self perceived disability and, despite its drawbacks, the main advantage is that it focuses the attention on the daily activities of the individuals and the problems they may have while doing them, with no consideration of medical matters. Moreover, EDAD 2008 is the first time that the survey includes information useful for studying the dependency phenomenon, such as the average hours per week of special care received by the dependent person.

What is the definition of dependency? Resolution R(98) of the Council of Europe defines dependency as "such state in which people, whom for reason connected to the lack or loss of physical, mental or intellectual autonomy, require assistance and/or extensive help in order to carry out common everyday actions". This definition has been translated into national legislations in a heterogeneous way (Kamette 2011). In Section 2 we particularize to the Spanish case.

The aim of this work is to estimate life expectancy free of dependency (LEFD), that is, the expected number of years that a person can live free of this contingency based on mortality and morbility conditions. The evolution of dependency in the Spanish population will be studied through a pseudo panel constructed from EDAD 2008, in the lack of longitudinal studies or the possibility to link different cross-sectional surveys.

Our main contribution is the estimation of LEFD not only by gender or dependency degree (moderate, severe and major), but also by partitioning the individuals in homogeneous groups with a similar dependency pattern. The characterization of homogeneous groups of individuals is obtained through the proximity of the dependency trajectories (that are obtained from the whole medical history concerning the dependency situation of each individual from birth up to 2008) to a central trend within each age-gender group. These central trends are obtained via functional data techniques. Once these proximity-groups are established, via Cox regression model, we obtain the survival probabilities (in fact, the staying free of dependency probability at a given age given that a person is alive at that age). Then marginal probabilities are calculated by multiplying these estimates by survival probabilities given by the Spanish disabled pensioners' mortality table. Finally, we obtain the LEFD for Spanish population within homogenous groups considering gender, dependency degree and ages from 30 to 100. As far as we know, this is the first time that the dependency evolution is used to estimate life expectancy. Other recent studies on dependency are Albarrán-Lozano et al. (2017), Albarrán et al. (2014) and also Albarrán et al. (2015) regarding dependent children.

The main findings are: Firstly, the estimated LEFD curves reach higher values for women than for men; Secondly, their decreasing rate is higher (and more abrupt) for men than for women; Thirdly, as dependency trajectories depart from the deepest

[2] In 2001, the World Health Organization (2001a) established a framework for measuring health and disability at both individual and population levels, which was known as the 'International classification of functioning (ICF), disability and health'. The ICF tries to establish a consensus in its understanding, by establishing a difference between the basic activities of living daily (ADL) and the instrumental ADL. The basic activities are defined as those activities which are essential for an independent life.

path, the gap between the LEFD for major dependency and the other dependency situations widens; Fourthly, the relative errors of the LEFD calculated using the partition by proximity-groups versus the global LEFD show evidence that the global LEFD may not be representative of the Spanish population. From economic and demographic points of view, this is a relevant finding, since the expected dependent population would demand care services (health care, pensions and other services) that should be covered and related expenditures should be financed; Finally, the Gini index computed on the dependency paths within proximity-groups reveals that the dependency phenomenon tends to a unique dependency extreme-pattern. Additionally, this index always reaches higher values in women than in men, suggesting that women have a wider range of dependency patterns than men.
The paper proceeds as follows. Section 2 contains the definition of dependency and its graduation according to the Spanish legislation. Also some information about survey EDAD 2008 is presented. Section 3 is devoted to explain the construction of the dependency trajectories from a pseudo panel coming from EDAD 2008, a description of the functional data techniques that we are going to use and the proximity measure that will help to characterize groups with homogeneous dependency trajectories. In Section 4 we propose the methodology to estimate the LEFD and analyze the main results. Finally we conclude in Section 5.

2 Dependency situation in Spain: legislation and dataset EDAD 2008

2.1 Spanish legislation on dependency

When talking about dependency two fundamental aspects must be considered. Firstly, the *definition* itself. In the Spanish case, article 2 of Act 39/2006, of 14^{th} December, on the Promotion, Personal Autonomy and care for Dependent persons states that dependency is a 'permanent state in which persons that for reasons derived from age, illness or disability and linked to the lack or loss of physical, mental, intellectual or sensorial autonomy require the care of another person/other people or significant help in order to perform basic activities of daily living or, in the case of people with mental disabilities or illness, other support for personal autonomy'.
Secondly, the *assessment of dependency*, which is usually solved using specific dependency rating scales that take into account the disabilities suffered by the person jointly with their intensity. Royal Decree 504/2007 rules the evaluation of dependency in Spain. The Spanish dependency rating scale goes from 0 to 100 points and it is categorized in four degrees (non dependant, I-moderate, II-severe, III-major). Table 1 contains the dependency graduation according to Spanish legislation.
To acknowledge the entitlement to the benefits of the System a person must reach at least the moderate degree, that is, at least 25 points are needed to be considered dependant in Spain. According to the dependency rating scale value or score reached by an individual, the Spanish legislation establishes a minimum level of protection, which is defined and financially guaranteed by the General State Administration.

Table 1: Dependency graduation according to Spanish legislation

Dependency	Degree	Level	Rating Scale	Dependency	Degree	Level	Rating Scale
Non dependant	-	-	$[0, 25)$	Severe	II	1	$[50, 65)$
					II	2	$[65, 75)$
Moderate	I	1	$[25, 40)$	Major	III	1	$[75, 90)$
	I	2	$[40, 50)$		III	2	$[90, 100]$

Moderate dependency	The person needs help in order to perform various basic ADL[(1)] at least once a day or the person needs intermittent or limited support for his/her personal autonomy.
Severe dependency	The person needs help in order to perform various basic ADL two or three times a day, but he/she does not want the permanent support of a carer or he/she needs extensive support for his/her personal autonomy.
Major dependency	The person needs help in order to perform various basic ADL several times a day or he/she needs the indispensable and continuous support of another person or he/she needs generalised support for his/her personal autonomy.

(1) ADL stands for Activities of Daily Living.

2.2 EDAD 2008 survey

In order to provide reliable estimates at the national level, the EDAD 2008 survey was performed around the country using sampling. In particular, a two-stage sampling was performed, stratified and proportional to the size of the Spanish autonomous regions (with stratified sampling distribution proportional to population size in stratum, within each Spanish province). Therefore, each individual in EDAD 2008 is associated to a weight reflecting the population group that represents. See INE (2010) for more details on the sampling methodology.

EDAD 2008 gives information about people with disabilities that were living either at home or in institutions. In the first case, the survey was prepared interviewing 260,000 people who were living in 96,000 different houses whereas for institutionalized people, 11,000 people in 800 centers were asked about their situation. Interviewed people are not only those suffering disabilities, but also their relatives and/or carers. This survey is based on the concept of self perceived disability, in accordance with the recommendations of the World Health Organization. So, the target people is identified through a set of questions about the possible difficulties they can find in doing some specific activities. Despite its drawbacks, the main advantage of this strategy is that it is focused in the daily activities of the individuals and the problems they may have while doing them, with no consideration of medical matters. That is, it puts the attention of both interviewer and interviewed in functional affairs since they are key aspects when talking about disability (Jiménez and Huete 2010).

In 2008 the Spanish population ascend to 46.66 million people (23.10 million men and 23.56 million women). According to EDAD 2008, there are more than 4.1 million Spanish people suffering at least one kind of disability. Although the global prevalence rate is situated between 8.2%-8.6% with a 95% of confidence, in the case of people living at home, this rate is lower than that for people living in institutions (8.4% and

17.7%, respectively). Disability is related to two main factors: gender and age; until 45 years old, the male prevalence is statistically significant greater than the female one. After that age, the relative incidence is greater for women. In general terms, more than 57% people with this problem are at least 65 years old, being most of them women. Table 2 contains an estimation (derived from the weighted survey data) of children and adult population with disability living at home.

Table 2: Estimation of children and adult population with disability living at home: 95% confidence intervals for number and prevalence rate

	Disabled people (in thousands) and prevalence rate (in %)		
Age (in years)	Total	Male	Female
Under 6	53-67.8 (1.8%-2.3%)	30.9-41.9 (2.1%-2.8%)	24.4-27.6 (1.1%-1.7%)
Between 6 and 44	576.8-648.4 (2.3%-2.7%)	316-374.2 (2.5%-3.0%)	240.5-285.5 (2.0%-2.4%)
Between 45 and 64	897.7-1005.9 (8.0%-9.0%)	379.4-437.6 (6.9%-8.0%)	505.2-580.4 (8.9%-10.3%)
Between 65 and 79	1138-1264.6 (20.8%-23.1%)	422.4-487.2 (17.1%-19.8%)	703.5-789.5 (23.4%-26.3%)
80 or more	971.8-1079.8 (45.8%-50.9%)	277.6-326.2 (36.6%-43.0%)	683.1-764.7 (50.0%-56.0%)
Total	3740.4-3955.2 (8.2%-8.6%)	1488.9-1605.5 (6.6%-7.1%)	2226.6-2374.6 (9.3%-10.2%)

Source: INE elaboration. Results derived from the weighted survey data.

The sample selected for the present study is formed by 7,446 individuals and represents 2.35% of the Spanish population in 2008, that is more than one million people (325,253 men and 773,079 women). We remind that each individual in the sample has a weight reflecting the population group that represents. These weights have been taken into account in all the computations of this paper. We give more details about the selected sample in the next section.

3 From a pseudo panel to dependency trajectories

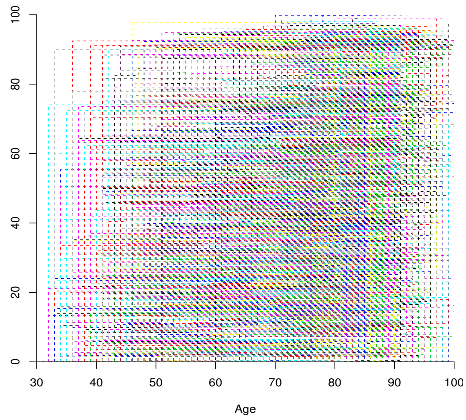
The data-set obtained from EDAD 2008 contain, among many other variables, the ages at which each person in the sample has suffered a change in his/her health condition leading to a jump in his/her dependency score, together with his/her current age (for details on how these scores where calculated see Albarrán and Alonso 2009). Although the survey includes the term 'dependence' in its title, the questionnaire does not consider any question on this topic. So, the dependency score is not reported in the survey but is computed from the information provided in it and applying the Spanish legislation (Act 39/2006 and Royal Decree 504/2007), that is, taking into account the disabilities suffered jointly with their severity and the average hours per week of special care received.

The aim of this section is to construct a dependency trajectory for each individual in the sample, that is, a curve describing the evolution of his/her dependency situation over time, and to use functional data techniques to analyze the database. Indeed, in functional data analysis, individual observations are real functions of time, observed at discrete time points. Each curve provides the evolution of a certain process for a given individual (see Ramsay and Silverman 2005 for an overview). In our case, the process of interest is the evolution of dependency.

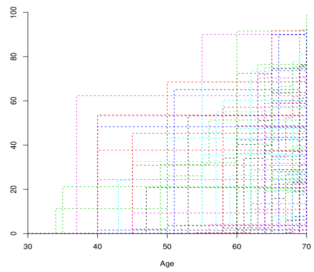
Notice that even if the available data come from a one-time survey, individuals were asked about their whole medical history, so we have information concerning their

dependency situation/score from birth up to 2008. Then, for the i-th individual we observe $(t_{i1}, y_{i1}), \ldots, (t_{in_i}, y_{in_i})$, the ages when changes occur and the dependency scores at these ages, and a_i, the current age (at 2008). From these data, in order to stress the step character of these curves, we add a first point $(0, 0)$ (only if $t_{i1} > 0$), intermediate points $(t_{ih} - \delta, y_{ih-1})$ between (t_{ih-1}, y_{ih-1}) and (t_{ih}, y_{ih}), where δ is a chosen short period of time, and a final point (a_i, y_{in_i}) (only if $t_{in_i} < a_i$). Indeed, even if the person's health/dependency condition can be seen as a smooth process, the dependency score function is piecewise constant since changes in score only take place once some particular disability status has been reached. These transformed sequences will make up our set of observations from now on. For the sake of simplicity, we will still refer to them as $(t_{ih}, y_{ih})_{h=1,\ldots,n_i}$, $i = 1, \ldots, n$. Thus, we have n discretely observed curves $y_1, \ldots, y_n$ defined in different time intervals $[0, a_i]$, $i = 1, \ldots, n$.
As an illustrative example, in Figure 1 we depict the dependency trajectories for women with dependency score of zero at the age of 30 and aged (a) from 50 to 100 in 2008 (that is, $50 \leq a_i \leq 100$), (b) 70 in 2008 (that is, $a_i = 70$) and (c) 80 in 2008 (that is, $a_i = 80$).

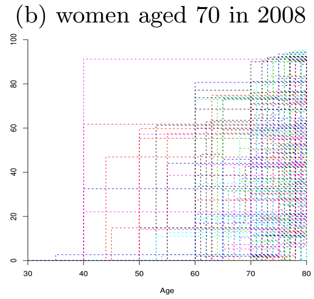
Figure 1: Dependency trajectories for women



(a) women aged from 50 to 100 in 2008



(b) women aged 70 in 2008



(c) women aged 80 in 2008

In order to apply any functional data analysis technique, we need functions that are defined over the same interval. One idea would be to consider the different cohorts present in the sample and analyse the dependency trajectories within each cohort. However, this may lead to many different under-represented cohorts, since the age range of the individuals in the sample is large. Instead of that, we consider disjoint groups of people in age intervals of 10 years. Within each age interval $[A, A + 10)$ we truncate individual curves to get them defined in $[0, A]$. Then, given a starting age A_1, we have the following k groups of individuals and curves

$$\mathcal{I}_{A_j} = \{i \mid 1 \leq i \leq n,\, a_i \in [A_j, A_j + 1)\},\ \mathcal{C}_{A_j} = \{y_i(t),\, t \in [0, A_j] \mid i \in \mathcal{I}_{A_j}\},\ j = 1, \ldots, k,$$

where $A_j = A_1 + 10(j-1)$, except for the last age interval which is defined as $[A_k, \infty)$. For all the analysis performed in this article the starting age A_1 was taken equal to 50 and the number of groups was set to $k = 5$. That is, the first age interval was $[50, 60)$ and the last one $[90, \infty)$. Moreover, we are particularly interested in those people with a dependency score of 0 at the age of 30, and from now on, they will be grouped in 10 age-gender intervals (5 groups per gender) according to their current age at 2008. Notice that we consider such a group of 30-year-old non dependent people in order to obtain LEFD estimates for different dependency scenarios, which are useful for health, economic, demographic and insurance contexts.
One of the objectives of the paper is to search for different patterns within each age-gender group, that is, we are interested in identifying dependency trajectories that lie close/far/very far from the central trend of the group. The reason why is that, as we will see in Section 4.2, life expectancy free of dependency can experiment huge variations as dependency trajectories depart from the central trend of the corresponding age-gender group. The central trend of each age-gender group will be obtained by using functional data techniques, that we describe in the following.

3.1 Estimating the central trend

Providing a measure of centrality when dealing with functional data is not an straightforward task. Indeed, not only the levels of the curves matter, but also their shapes, whose information is more difficult to incorporate to any numerical summary. The problem aggravates if we consider curves for which the main features are not aligned. It is well known that in this context the sample point-wise or cross-sectional mean is a poor estimator of the mean behaviour (Gasser et al. 1984, Kneip and Gasser 1992, Gasser and Kneip 1995). A very simple example of this is to consider two bell-shaped curves, $y_1(t)$ and $y_2(t)$, with different and distant modes. The point-wise or cross-sectional mean of these two curves, that is, $\bar{y}(t) = 0.5(y_1(t) + y_2(t))$, will probably present two modes, and then will not look like any of the two curves, in terms of shape.
In this context, it is extremely important to use measures of centrality that can take into account the misalignment between the curves of the sample. Indeed, in the particular case of the dependency evolution curves that we study in this work, it is very natural to consider that the evolution of dependency may present a common pattern which is accelerated or delayed in some individuals with respect to others. Then, it is useful to consider the following *time warping* model for the generation of the observed curves:

$$y_i(t) = x \circ h_i^{-1}(t) \quad t \in [0, A], \quad i = 1, \dots, n, \tag{1}$$

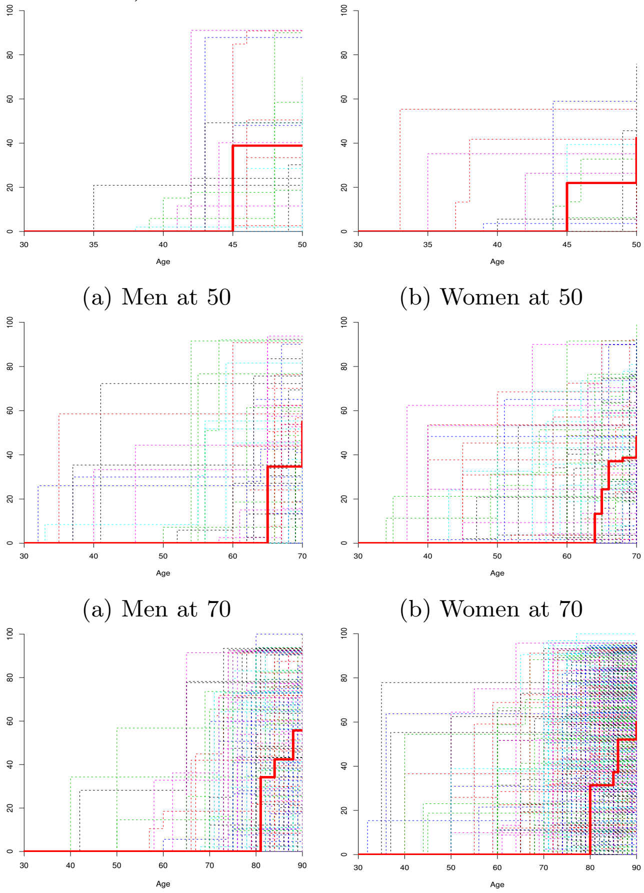
where x is an unknown deterministic function representing the process of interest and h_i^{-1} i.i.d. realization of the so-called *warping* process, H, that represents individual time distortion. These are strictly increasing bijections defined on the observation time interval. A common identifiability condition on the warping process is that $E[H(t)] = t \ \forall t$, meaning that we assume that some of the curves of the sample are accelerated and some others are delayed with respect to x. In the time warping model, two approaches to estimate the central trend or mean behaviour of the data are possible: (i) to align or register the curves, that is, to estimate h_i, and to compute

any desired sample statistic on the registered sample, $y_1(\hat{h}_1), \ldots, y_n(\hat{h}_n)$; and (ii) to define appropriate estimators directly on the observed sample, taking into account the nature of the data. For the analysis of the dependency data-set we will consider an estimator of the second kind that we describe in the following.
Before that, we would like to remark that the time warping model defined in (1) provides a general and flexible framework for the modelling of the dependency trajectories, since it includes any kind of parametric model in which the individual parameters allow for variations in scale and phase with respect to some given functional form, such as growth models, and also semi-parametric models in which this functional form is unknown and estimated from the data, such as shape-invariant models (see Wang and Gasser 1997 for details). Also, notice that we can assume that observations are free of measurement error since they correspond to the evaluation, on an official numerical scale, of the particular conditions suffered by each individual at each moment.

Deepest curve

The literature on estimators directly defined on the unregistered sample is relatively small. Dupuy et al. (2011), Liu and Müller (2004) or Arribas-Gil and Romo (2012) are works which are particularly concerned by the definition of suitable population centrality measures, and their corresponding sample statistics, in the time warping model. However, there might be curves with a typical shape but taking atypical values (abnormally high or low at some locations) and, in this case, a registration procedure would neutralise the effect of those curves with an atypical shape (due to the fact that they may be delayed or accelerated with respect to the rest). Therefore, for the analysis of the dependency data-set we will consider the approach of Arribas-Gil and Romo (2012) since it provides a robust estimator of the central trend for a set of curves.
A way to provide a centrality measure that is robust against the two types of atypical curves is to use functional depth. Indeed, the deepest curve of a sample in terms of modified band depth (López-Pintado and Romo 2009) has been proven to be an accurate and robust estimator of the central pattern of a sample of curves in the time warping model (Arribas-Gil and Romo 2012). It can be understood as a generalization of the median to functional data because, intuitively, it is the curve that is most surrounded by other curves. Therefore, it provides an accurate measure of centrality since: (i) it is a curve geometrically located in the center of the sample and (ii) it presents a typical shape because it is one of the observed curves. These properties make it a robust estimator, against the two types of functional atypical observations described above, even when computed on an un-registered sample.
As it was mentioned before, we are interested in estimating the central trend of each age-gender group. Therefore, for each one of these 10 groups we compute the deepest curve in terms of modified band depth using the `roahd` package in R by Tarabelloni et al. (2016). As an example, in Figure 2 we depict the dependency trajectories with the corresponding deepest curve for several age-gender groups, where we observe that at the same age the first score value reached by these deepest curves is lower for women than for men.

Figure 2: Dependency trajectories for men and women with their corresponding deepest curves (in bold red)



(a) Men at 50 (b) Women at 50

(a) Men at 70 (b) Women at 70

(a) Men at 90 (b) Women at 90

3.2 Distance to the deepest curve

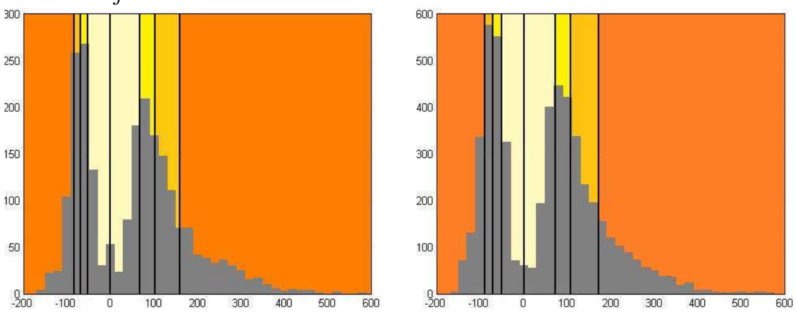
Once we have estimated the central trend within each age-gender group, we propose to search for different patterns of trajectories within each group by computing a proximity measure to its deepest curve.
Therefore, within each group we compute the L^2-distance of each trajectory to the corresponding deepest one multiplied by 1 (or -1) if the trajectory is most of the time above (or below) the deepest curve. Indeed, if we note $y_{ij}(\cdot)$ the dependency curve of individual i in group j, and $m_j(\cdot)$ the deepest curve of group j, we obtain

$$d_j(i) = \sqrt{\sum_t (y_{ij}(t) - m_j(t))^2} \cdot \text{sign}\left(\sum_t (y_{ij}(t) - m_j(t))\right), \quad i = 1, \dots, n_j, j = 1, \dots, 10, \tag{2}$$

where n_j is the number of individuals in group j. This yields a numerical summary for each one of the trajectories that can be used to establish different patterns. As we will see later, these patterns will exhibit quite different life expectancies.

In Figure 3 we depict the histograms for the proximity measure d_j defined in (2) computed over all the trajectories by gender. Notice that the sign of d_j indicates whether the trajectory is below or above the deepest curve within its age-gender group. In particular, negative values of d_j correspond to trajectories below the deepest curve and, therefore, to individuals with lower dependency scores than those of the central trend of their age-gender groups. In fact, the best situations are expected for the left-tail values of d_j. On the other hand, positive values of d_j correspond to trajectories above the deepest curve and, hence, to individuals with higher dependency scores than those of the central trend of their age-gender groups. In this case, the worst situations are expected for large values of d_j.

Figure 3: Histogram for d_j for men and women. Vertical lines stand for the quartiles of the sets $\{d_j < 0\}$ and $\{d_j \geq 0\}$, respectively. Different color indicates the degree of departure from $d_j = 0$



(a) Men (b) Women

Since each set of values $\{d_j < 0\}$ and $\{d_j \geq 0\}$ has a different meaning, we propose to compute the LEFD for the groups of individuals established by the quartiles in each set, yielding to eight groups for each gender, that we call proximity-groups (see Figure 3 where the regions defined by the quartiles are in different color). Remind that, we are also interested in estimating LEFD for the three dependency degrees (moderate, severe, major) within each proximity-group. Depending on the researcher, other partitions or sets of particular interest, such as extreme observations, can be considered.

4 Estimating life expectancy free of dependency

Two methods have been commonly used to estimate life expectancy: Sullivan's method (Sullivan 1971) and Cox regression model (Cox 1972, Cox and Oakes 1984). Sullivan's is the easiest method to estimate life expectancy. However, it presents several disadvantages. For example, it fails to account for possible changes in health conditions. It also fails in measuring the effect of covariates in the estimation of life expectancy. Another disadvantage is that it is not possible to consider the evolution of a morbility or disability for each individual (Andrae et al. 2011). On the other hand, Cox regression model allows to explore the determinants of life expectancy (or survival probability) and to estimate hazard ratios of the covariates included in the model, such as gender, disabilities, etc.

In our case, the event of interest is not 'survival' itself, but 'staying free of dependency at a given age'. We propose to use Cox regression model to obtain those probabilities and the estimation of LEFD is then straightforward. However, it must be pointed out that EDAD 2008 only contains records of alive people at 2008, hence the effect of death is ignored. That is, the estimated staying free of dependency probability at a given age is in fact the staying free of dependency probability at a given age given that a person is alive at that age. Then marginal probabilities are obtained by multiplying these estimates by survival probabilities given by the Spanish disabled pensioners' mortality table (BOE-A-2005-21310 2005).

4.1 Searching for relevant variables

To estimate LEFD we consider different scenarios attending to gender, eight proximity-groups (IV-I for $\{d_j < 0\}$ and I-IV for $\{d_j \geq 0\}$) and three dependency degrees (I-moderate, II-severe, III-major) given by the Spanish legislation. Indeed, for a non dependent person we calculate three different LEFDs, which are the expected number of years that a person can live out of each one of the three dependency degrees. Notice that the dependency history of an individual may not reach all the states, that is, the first score reached by an individual can be greater than 50 or 75 points. This is the reason why, in the following tables and for the sake of simplicity, we call 'degree I' to the expected number of years that a person can live out of any dependency degree (score under 25 points); 'degree II' stands for the expected number of years that a person can live out of severe or major dependency (score under 50 points); 'degree III' stands for the expected number of years that a person can live out of major dependency (score under 75 points).
We include in the analysis people that in 2008 were between 50 and 100 years old, with a dependency score of 0 at the age of 30. We remind that the sample is formed by 7,446 individuals (2,230 men and 5,216 women), that represent more than one million Spanish people, according to INE. As mentioned above, the variables included in the model are the age-gender group and several disabilities suffered. In particular, the considered disabilities are to present difficulties in performing postural changes, washing/bathing, relieving themselves, conducting household life, maintaining interaction and interpersonal relationships, following medical treatments and mobility difficulties (inside and outside the house). We estimate and validate Cox Proportional-Hazards model with `survival` package in R. Table 3 contains an scheme of the main effects derived from Cox regression results for men and women estimated for each proximity-group and dependency-degree (see Table 5 in the Appendix for detailed results).
In the following we interpret the main effects on LEFD of the variables included in the model using Table 3 and also considering the detailed numerically results contained in Table 5 in the Appendix.

Individuals within set $\{d_j < 0\}$: Remind that all individuals within this set have dependency paths below the corresponding deepest curve, that is, all individuals have lower dependency scores than the corresponding central trend. Regarding proximity-groups IV-I (columns 2 to 13 of Table 5 in the Appendix) we observe several variables with zero coefficient, meaning that their impact on LEFD is null. This behavior only happens on dependency degree III groups and the affected variables for men are

Table 3: Main effects derived from Cox regression results (Table 5) in Appendix)

Men	Set $\{d_j < 0\}$				Set $\{d_j \geq 0\}$			
Proximity-group	IV	III	II	I	I	II	III	IV
postural changes	-	↗↗	↗	↗	-	↗	-	-
mobility	-	-	-	-	-	-	-	-
washing/bathing	↗	↗	↗	↗	↗	-	↗	-
relieve themselves	↗	-	-	-	-	↗	-	↗
medical treatments	-	-	-	-	↗↗	-	↗↘	↗↘
household life	-	↗	-	-	-	-	-	-
interpers. rel.	-	-	-	-	-	-	-	-

Women	Set $\{d_j < 0\}$				Set $\{d_j \geq 0\}$			
Proximity-group	IV	III	II	I	I	II	III	IV
postural changes	-	-	↗	-	-	-	↗	-
mobility	-	-	-	-	-	-	-	-
washing/bathing	-	-	-	-	-	↗	↗	-
relieve themselves	↗	↗	↗↗	↗	-	-	-	-
medical treatments	-	↗	-	↗	↗↘	↗↘	↗↘	↗↘
household life	↘	↘	-	↘	↘↗	↘↗	-	-
interpers. rel.	-	-	-	-	↗↗	-	-	↗

Legend: non relevant effect -, the effect increases with dependency degree ↗, highest effect by gender and position with respect to the 'central' trend ↗↗, the effect increases with dependency degree ↘.

(in order of irrelevance) washing/bathing, following medical treatments, conducting household life and maintaining interaction and interpersonal relationships; for women those variables are (in order of irrelevance) following medical treatments, conducting household life, washing/bathing and mobility difficulties inside and outside the house. Considering sense and magnitude of the effects, there are two variables whose effects on men's LEFD tend to increase with the dependency degree (performing postural changes and washing/bathing). In general, these variables have the greatest impact on men's LEFD, since they register the highest coefficients within each proximity-group and dependency degree (see Table 5 in the Appendix). Regarding women, the effect of relieve themselves tends to increase with the dependency degree whereas the opposite happens with conducting household life. In this case, the variable with greatest impact on women's LEFD is relieve themselves. On the other hand, mobility difficulties inside and outside the house and maintaining interaction and interpersonal relationships have the lowest impacts on both genders' LEFD.

Individuals within set $\{d_j \geq 0\}$: In this case, all individuals within this set have dependency paths above the corresponding deepest curve, that is, all individuals have higher dependency scores than the corresponding central trend. Considering proximity-groups I-IV (columns 14 to 25 of Table 5 in the Appendix) we observe only one variable with zero coefficient. This is the case for washing/bathing in Proximity-group I. Regarding sense and magnitude of the effects, there are several variables whose effects tend to increase with dependency degree. In particular, washing/bathing for both genders; relieving themselves for men and maintaining interaction and interpersonal relationships and conducting household life for women.

The variables with the greatest impact on men's LEFD are following medical treatments, followed by washing/bathing and relieving themselves. On the other hand, the variable with the greatest impact on women's LEFD is maintaining interaction and interpersonal relationships, followed by washing/bathing and following medical treatments, although the latter has a non monotonic impact on LEFD. As happens for individuals on set $\{d_j < 0\}$, mobility difficulties inside and outside the house registers the lowest impact on both genders' LEFD.

To sum up, using the proximity measure to characterize the dependency evolution, we observe that:

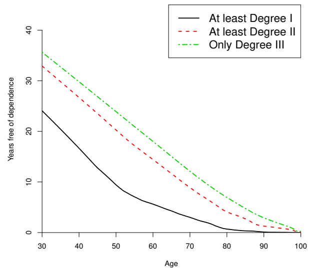
- Following medical treatments begins to impact on men's LEFD when their dependency situation is worse than the central trend; Maintaining interaction and interpersonal relationships plays a similar role on women's LEFD.
- Difficulties in performing postural changes registers the highest impact on men's LEFD when their dependency situation is better than the central trend; however, the impact fades when their dependency situation is worse than the central trend; Relieving themselves plays a similar role on women's LEFD.
- Regarding gender, two variables show an opposite effect on LEFD. In particular, washing/bathing and conducting household life.

4.2 Contribution of the proximity-groups on estimated LEFD

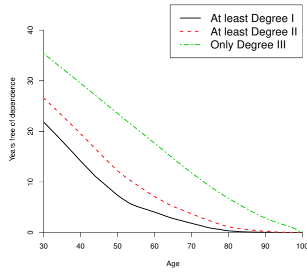
In Figures 4-5 we depict the estimated LEFD for the scenarios considered for the eight proximity-groups for men and women, respectively. In particular, each panel contains three curves corresponding to the evolution of the LEFD along age in three situations (At least Degree I, At least Degree II and Only Degree III).[3] In general, looking at LEFD curves we can observe that, firstly, they reach higher values for women than for men; Secondly, the decreasing rate is higher (and more abrupt) for men than for women; Thirdly, as dependency trajectories depart from the deepest path, the LEFD for 'At least Degree II' gets closer to the LEFD for 'At least Degree I'. This behaviour is observed in both genders, looking at (b)-panels from proximity-group I to IV.
Finally, we focus on (a4) and (b4) panels for both genders, that contain the extreme proximity-groups IV-below the deepest path and IV-above the deepest path, respectively. The former correspond to individuals with the best dependency situation, where we observe that LEFD curves for 'At least Degree I' reach lower values for men than for women, meaning that dependency situation tends to appear earlier in men than in women. Additionally, the estimation of women's LEFD is very similar for the three dependency situations (notice the proximity of the three curves). The latter corresponds to individuals with the worst dependency situation, where we can observe an analogous shape of the LEFD for both genders in the three dependency situations, reaching slightly higher values for women.

[3]'At least Degree I' stands for expected number of years that a person can live out of any dependency degree (score under 25 points); 'At least Degree II' stands for the expected number of years that a person can live out of severe or major dependency (score under 50 points); 'Only Degree III' stands for the expected number of years that a person can live out of major dependency (score under 75 points)

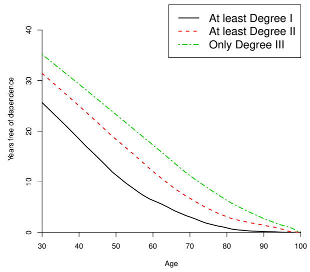
Figure 4: Life expectancy free of dependency for men. Left (right) panels show LEFD for individuals with dependency trajectories below (above) the deepest path



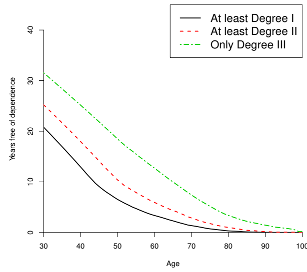
(a1) Proximity-group I; below the deepest path



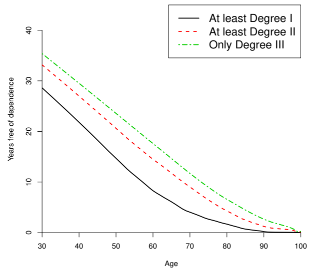
(b1) Proximity-group I; above the deepest path



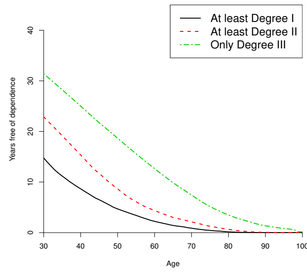
(a2) Proximity-group II; below the deepest path



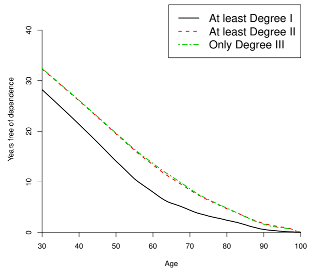
(b2) Proximity-group II; above the deepest path



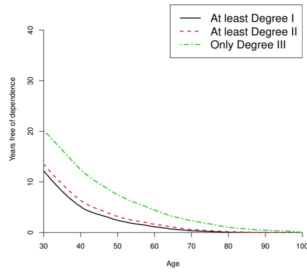
(a3) Proximity-group III; below the deepest pat



(b3) Proximity-group III; above the deepest path



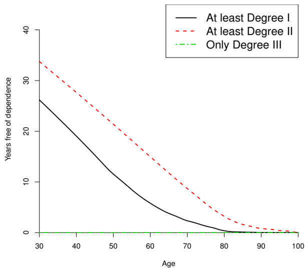
(a4) Proximity-group IV; below the deepest path



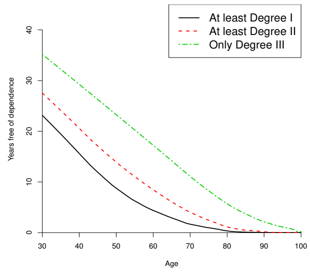
(b4) Proximity-group IV; above the deepest path

As a summary, Table 4 contains the estimated LEFD for men and women at three particular ages jointly with the LEFD calculated without taking into account the partition by proximity-groups (rows LEFD for men and LEFD for women). We may remind that these LEFD estimations are computed from survey EDAD 2008, that contains only dependent people. Therefore, they must interpreted as the 'at least' expected numbers of years free of dependency. Nevertheless, the methodology that

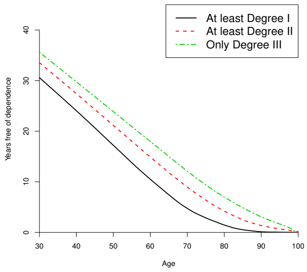
Figure 5: Life expectancy free of dependency for women. Left (right) panels show LEFD for individuals with dependency trajectories below (above) the deepest path



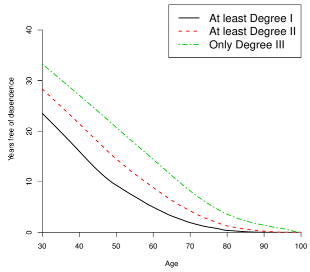
(a1) Proximity-group I; below the deepest path



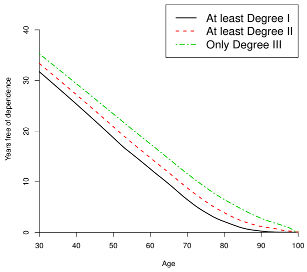
(b1) Proximity-group I; above the deepest path



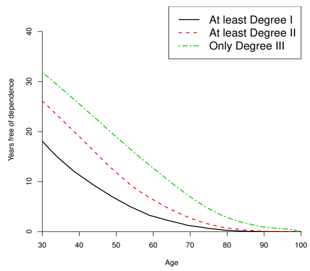
(a2) Proximity-group II; below the deepest path



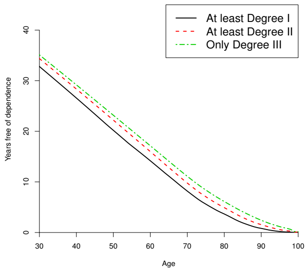
(b2) Proximity-group II; above the deepest path



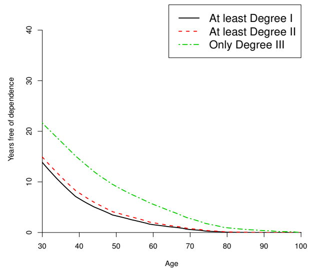
(a3) Proximity-group III; below the deepest path



(b3) Proximity-group III; above the deepest path



(a4) Proximity-group IV; below the deepest path



(b4) Proximity-group IV; above the deepest path

we propose in this paper is not restricted to the database. In Figure 6 we represent the relative error (in %) of the values considered in Table 4 with respect to the global LEFD by gender.

In Table 4 and Figure 6 we observe that the variance of LEFD increases with age and tends to decrease with dependency degree. In general, the variance is greater for women. The relative errors show evidence that the global LEFD by gender, calculated without taking into account the partition by proximity-groups, is far from any of the LEFD values estimated by proximity-groups. This means that the global

Table 4: Summary of LEFD for three particular ages

	Men	30 years old			50 years old			70 years old		
	Degree	I	II	III	I	II	III	I	II	III
$\{d_j < 0\}$	Proximity-group IV	28.234	32.320	32.416	14.117	19.501	19.623	4.417	8.424	8.648
	Proximity-group III	28.616	33.167	35.395	14.701	20.618	23.567	4.040	9.022	11.706
	Proximity-group II	25.664	31.426	35.186	11.352	18.389	23.290	3.065	6.745	11.234
	Proximity-group I	24.069	32.908	35.648	9.424	20.275	23.902	3.004	8.899	12.120
$\{d_j \geq 0\}$	Proximity-group I	21.848	26.600	35.375	7.378	12.228	23.541	1.838	3.711	11.812
	Proximity-group II	20.792	25.213	31.548	6.524	10.375	18.474	1.346	2.857	7.400
	Proximity-group III	14.762	22.884	31.427	4.695	8.647	18.659	0.878	2.127	7.425
	Proximity-group IV	12.191	13.509	20.194	2.426	3.183	7.497	0.359	0.600	2.337
LEFD for men		18.733	23.614	31.868	6.475	10.429	19.391	1.556	3.195	8.583

	Women	30 years old			50 years old			70 years old		
	Degree	I	II	III	I	II	III	I	II	III
$\{d_j < 0\}$	Proximity-group IV	32.805	34.370	35.099	20.195	22.210	23.175	8.187	9.780	11.110
	Proximity-group III	31.714	33.353	35.264	18.694	20.864	23.394	6.465	8.783	11.589
	Proximity-group II	30.578	33.549	35.614	17.215	21.141	23.860	4.753	8.973	12.116
	Proximity-group I	26.172	33.743	-	11.557	21.380	-	2.343	8.699	-
$\{d_j \geq 0\}$	Proximity-group I	23.145	27.523	35.166	8.766	13.904	23.264	1.664	4.017	11.102
	Proximity-group II	23.527	28.316	33.279	9.340	14.536	20.765	1.938	4.283	8.204
	Proximity-group III	18.077	26.109	31.882	6.543	11.858	18.958	1.190	2.745	6.992
	Proximity-group IV	13.852	14.923	21.651	3.288	3.909	9.114	0.579	0.759	2.757
LEFD for women		21.601	26.657	32.860	8.781	13.623	20.577	2.092	4.337	9.045

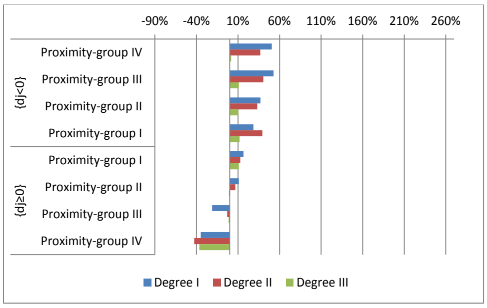
LFED may not be representative of the Spanish dependent population, not even for those individuals within the most central proximity-groups, that is, for those that are the nearest to the corresponding central trend.

To illustrate the contribution of the partition by proximity-groups on the LEFD estimation, we consider the following example. The global LEFD for a 30-year-old women with dependency degree I is 21.6 years, which means that the expected number of years that a 30-year-old woman can live out of any dependency degree is 21.6. In other words, it is expected that a 30-year-old woman can reach 51 years-old out of any dependency degree. However, a more accurate estimation can be obtained by considering proximity-groups, ranging from 32.8 years (proximity-group IV within set $\{d_j < 0\}$) to 13.8 years (proximity-group IV within set $\{d_j \geq 0\}$). That is, in the best situation, a 30-year-old woman can live out of any dependency degree until 62 years old and, in the worst case, a 30-year-old woman becomes dependent at the age of 43 years old. Notice that this difference of around 20 years is relevant, at least, from demographic and economic points of view, in the sense that, the expected dependent population would demand care services (health care, pensions and other services) that should be covered and related expenditures should be financed.

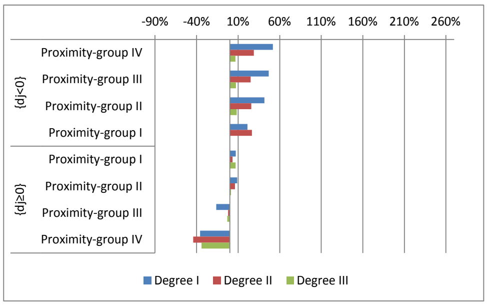
4.3 Do proximity-groups help to characterize different dependency patterns?

In Figures 7-8 we depict individual dependency trajectories for men and women within proximity-groups; (a)-panels contain the individuals whose trajectories are below the deepest path, whereas (b)-panels contain those with trajectories above the deepest path.

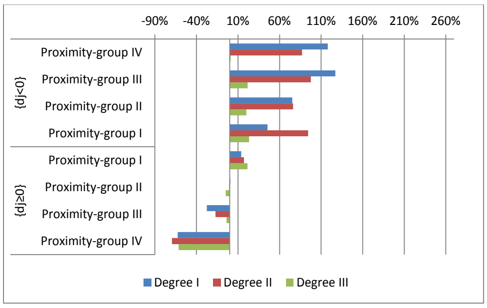
Figure 6: Relative error for LEFD (in %) computed from Table 4 values



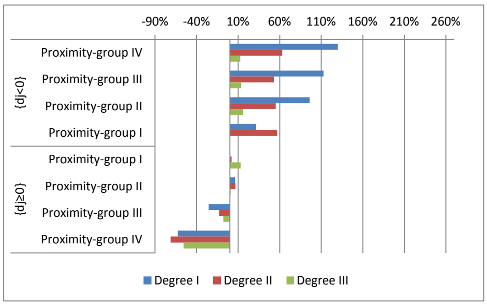
(a1) 30-year-old men



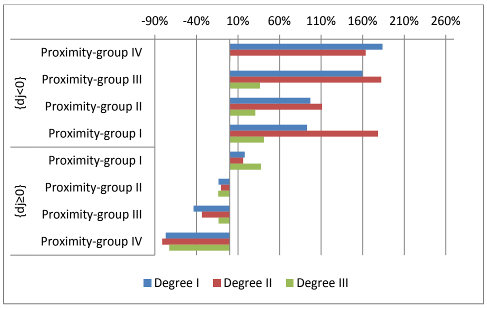
(b1) 30-year-old women



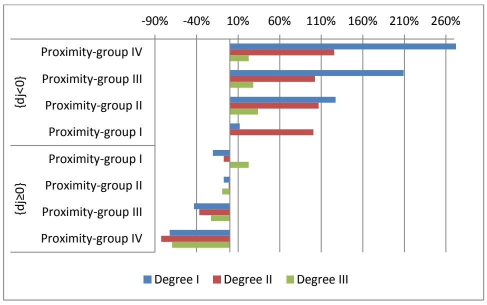
(a2) 50-year-old men



(b2) 50-year-old women



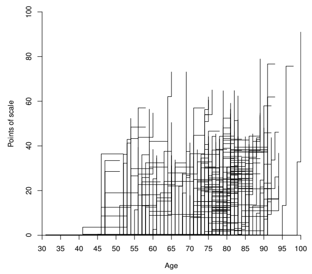
(a3) 70-year-old men



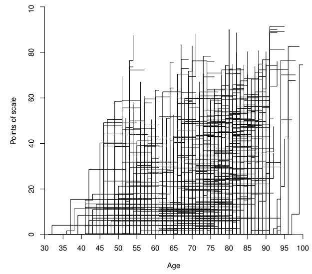
(b3) 70-year-old women

Individuals within set $\{d_j < 0\}$: Focusing in (a)-panels of Figures 7-8 we observe that: First, trajectories tend to reach positive scores at later ages than trajectories in (b)-panels; Second, two patterns can be identified as we depart from the origin (that is, going from proximity-group I to IV): In the first one, the severity of dependency increases with age, that is, the age of the individual maybe the main cause of the evolution of dependency, whereas in the second one, individuals reach dependency scores greater than 25 points (degree I) at a given age. This happens in most of the trajectories in proximity-group II, where scores range from 25 to 75 points, most of the trajectories in proximity-group III with scores going from 50 to 95 points and in proximity-group IV, where scores are over 75 points in most of the cases (degree III). This behaviour tends to appear earlier in men than in women, specially in proximity-group IV.

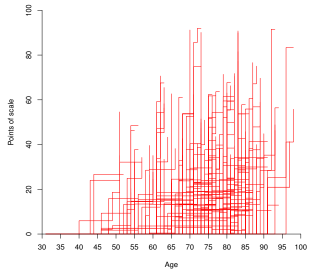
Figure 7: Left (right) panels show individual dependency paths for men with dependency trajectories below (above) the deepest path



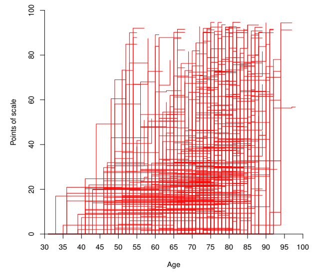
(a1) Proximity-group I; below the deepest path



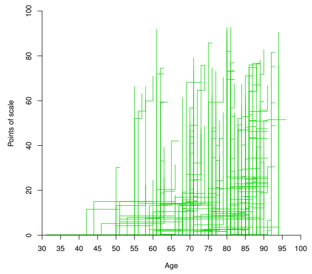
(b1) Proximity-group I; above the deepest path



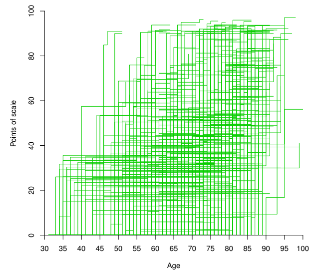
(a2) Proximity-group II; below the deepest path



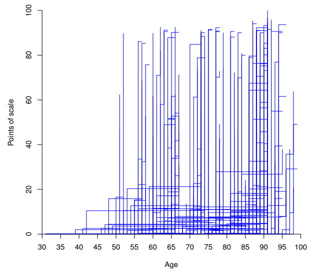
(b2) Proximity-group II; above the deepest path



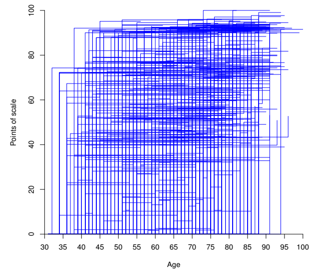
(a3) Proximity-group III; below the deepest pat



(b3) Proximity-group III; above the deepest path



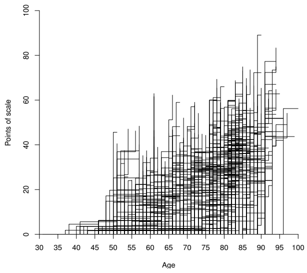
(a4) Proximity-group IV; below the deepest path



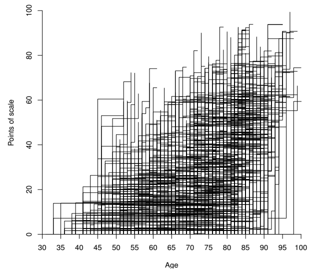
(b4) Proximity-group IV; above the deepest path

Individuals within set $\{d_j \geq 0\}$: Regarding curves in (b)-panels of Figures 7-8, we can say that: First, they tend to reach greater scores at earlier ages than those in (a)-panels; Second, most of the trajectories in proximity-group II reach scores under 50 points, whereas in proximity-group III, most of them have scores over 25 points; Third, the main differences between proximity-groups II and III are related, on the one hand, to trajectories with scores ranging from 25 to 50 points and, on the other

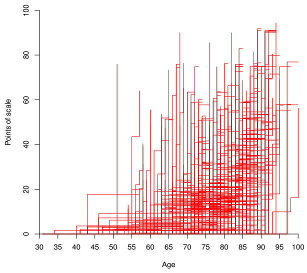
Figure 8: Left (right) panels show individual dependency paths for women with dependency trajectories below (above) the deepest path



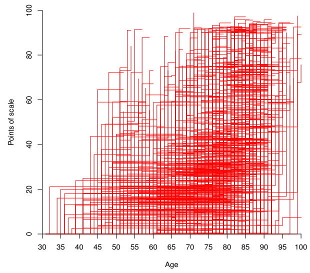
(a1) Proximity-group I; below the deepest path



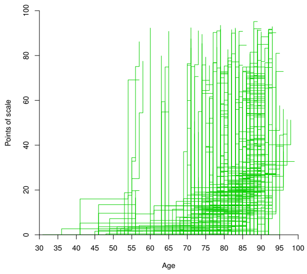
(b1) Proximity-group I; above the deepest path



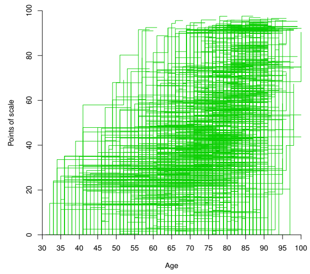
(a2) Proximity-group II; below the deepest path



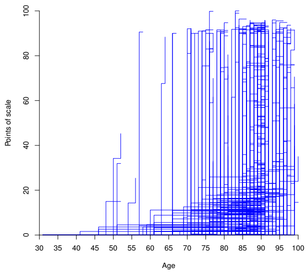
(b2) Proximity-group II; above the deepest path



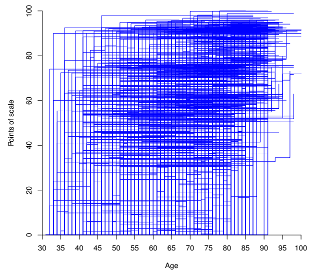
(a3) Proximity-group III; below the deepest path



(b3) Proximity-group III; above the deepest path



(a4) Proximity-group IV; below the deepest path



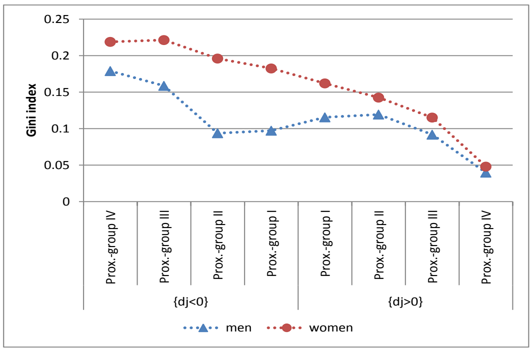
(b4) Proximity-group IV; above the deepest path

hand, with scores greater than 90 points (degree III.2). In particular, in proximity-group II all individuals under 40 years old have scores under 25 points. This is not the case for proximity-group III, where scores over 25 points are reached at 33 years old. Additionally, in proximity-group III we can observe more individuals with scores over 50 points and more individuals with scores over 90 points (degree III.2) than in proximity-group II; Fourth, most of the trajectories in proximity-group IV have scores over 50. Moreover, women reach scores over 90 points before 35 years old, whereas men do it after that age.

As a final comment on Figures 7-8, we can add that the behaviours of trajectories in proximity-groups I (panels a1 and b1) are very similar. This was expected and happens because these groups contain the closest trajectories to the deepest curve (below the central trend in a1-panel and above the central trend in b1-panel).

As far as here, we have seen that proximity-groups characterize different dependency patterns. Additionally, we can use Gini index as a numerical measure to summarize dependency evolution per proximity-group. At this point it is important to remind that dependency curves are associated to individuals that represent population groups of different sizes.[4] Therefore, each individual is associated to a weight. The sum of all of these weights is an estimation of the whole Spanish dependent population.
Two variables are needed in order to calculate the Gini index: p variable for population and q variable for quantities or amounts. In our case, p variable has been computed summing the weights of all individuals with positive score and for each age ranging from 30 to 100 years old. On the other hand, q variable has been computed for each age ranging from 30 to 100 years old as the product of (positive) scores by weights and represents the total amount of score points at each age. We depict the results in Figure 9.

Figure 9: Gini index for men and women within the proximity-groups



We can conclude the following:

- For proximity-groups within set $\{d_j \geq 0\}$, the concentration tends to reduce as d_j increases, which means that dependency paths tend to be more similar as their distance to the deepest path increases.
- The opposite happens for proximity-groups within set $\{d_j < 0\}$, that is, Gini index tends to reduce as d_j gets closer to zero, meaning that dependency paths tend to be less similar as their distance to the deepest path increases.
- As the dependency situation worsens, the Gini index tends to reduce, meaning that the dependency phenomenon tends to a unique dependency extreme-pattern. In women, the Gini index experiments a regular decreasing, reaching always higher values than in men. This may suggest that women have a wider range of dependency patterns than men.

[4] In EDAD 2008 a two-stage sampling was conducted by INE, leading to individuals that represent population groups of different sizes.

5 Conclusions

Dependency, that is, lack of autonomy in performing basic ADL can be seen as a consequence of the process of gradual aging. In Europe in general and in Spain in particular this phenomenon represents a problem with economic, political and social implications. The prevalence of dependency in the population, as well as its intensity and evolution over the course of a person's life are issues of greatest importance that should be addressed.

The aim of this work is to estimate life expectancy free of dependency (LEFD), that is, the expected number of years that a person can live free of this contingency based on mortality and morbidity conditions. The evolution of dependency in the Spanish population is studied through a pseudo panel constructed from EDAD 2008, in the lack of longitudinal studies or the possibility to link different cross-sectional surveys. In particular, individual dependency trajectories are obtained using the whole medical history concerning the dependency situation of each individual from birth up to 2008, contained in EDAD 2008, and applying the Spanish legislation (Act 39/2006 and Royal Decree 504/2007).

The main contribution of this paper is the estimation of LEFD not only by gender or dependency degree (moderate, severe, major), but also by partitioning the individuals in homogeneous groups with a similar dependency pattern. The characterization of homogeneous groups of individuals is obtained through the proximity of the dependency trajectories to a central trend within each age-gender group, with respect to the modified band depth. Once these proximity-groups are established, we use Cox regression model to obtain the 'survival' probabilities (in fact, the staying free of dependency probability at a given age given that a person is alive at that age). Marginal probabilities are calculated by multiplying these estimates by survival probabilities given by the Spanish disabled pensioners' mortality table. Finally, we obtain the LEFD for Spanish population within homogenous groups considering gender, dependency degree and ages from 30 to 100. As far as we know, this is the first time that the dependency evolution is used to estimate life expectancy.

The partition established by the proximity-groups has lead us to identify several variables that seem to be relevant in the estimation of LEFD. In particular, we can conclude that following medical treatments begins to impact on men's LEFD when their dependency situation is worse than the central trend; the same happens for maintaining interaction and interpersonal relationships regarding women's LEFD. On the other hand, difficulties in performing postural changes registers the highest impact on men's LEFD when their dependency situation is better than the central trend, although this impact fades when their dependency situation is worse than the central trend; relieving themselves plays a similar role on women's LEFD.

Looking at the estimated LEFD curves, we observe that, firstly, they reach higher values for women than for men; Secondly, the decreasing rate is higher (and more abrupt) for men than for women; Thirdly, as dependency trajectories depart from the deepest path, the gap between the LEFD for major dependency and the other dependency situations widens.

The relative errors of the LEFD calculated using the partition by proximity-groups versus the global LEFD show evidence that the global LEFD may not be representative of the Spanish population. From economic and demographic points of view, this

is a relevant finding, since the expected dependent population would demand care services (health care, pensions and other services) that should be covered and related expenditures should be financed.
Finally, the Gini index computed on the dependency paths within proximity-groups reveals that the dependency phenomenon tends to a unique dependency extreme-pattern. Additionally, this index always reaches higher values in women than in men, suggesting that women have a wider range of dependency patterns than men.

Appendix

Table 5: Cox regression results for men and women: e^{β}'s and β's standard deviation (within parenthesis)

Men	Set $\{d_j < 0\}$; Individual trajectories below the deepest path												Set $\{d_j \geq 0\}$; Individual trajectories above the deepest path											
	Proximity-group IV			Proximity-group III			Proximity-group II			Proximity-group I			Proximity-group I			Proximity-group II			Proximity-group III			Proximity-group IV		
Degree	I	II	III	I	II	III	I	II	III	I	II	III$^{(*)}$	I	II	III	I	II	III	I	II	III	I	II	III
postural	1.283	1.566	3.107	1.496	2.321	17.951	1.120	3.094	5.323	2.240	4.678	-	1.183	1.547	3.163	1.304	1.580	2.689	1.161	1.320	1.779	0.592	0.816	1.787
changes	(0.02)	(0.026)	(0.047)	(0.014)	(0.02)	(0.068)	(0.016)	(0.031)	(0.102)	(0.015)	(0.029)	-	(0.01)	(0.012)	(0.048)	(0.011)	(0.014)	(0.025)	(0.01)	(0.012)	(0.022)	(0.012)	(0.013)	(0.019)
mobility	1.142	0.862	1.839	1.202	0.225	0.219	1.406	0.416	5.250	1.622	0.310	-	0.674	0.491	0.178	1.057	0.985	0.529	0.789	0.781	1.045	0.711	1.479	0.626
	(0.028)	(0.029)	(0.045)	(0.026)	(0.027)	(0.064)	(0.017)	(0.032)	(0.139)	(0.017)	(0.036)	-	(0.011)	(0.013)	(0.046)	(0.012)	(0.016)	(0.023)	(0.012)	(0.013)	(0.021)	(0.014)	(0.016)	(0.02)
washing/	2.966	4.844	0.000	1.090	7.189	0.000	2.117	3.040	0.000	1.334	2.046	-	1.408	2.657	0.000	1.256	0.829	0.319	1.337	1.361	14.387	1.020	1.002	2.836
bathing	(0.044)	(0.106)	(0.046)	(0.024)	(0.067)	(0)	(0.018)	(0.04)	(0.06)	(0.016)	(0.037)	-	(0.016)	(0.024)	(0)	(0.015)	(0.02)	(0.035)	(0.015)	(0.018)	(0.119)	(0.019)	(0.021)	(0.038)
relieve	0.770	1.548	5.162	0.750	1.287	4.222	0.648	1.105	0.989	1.340	2.845	-	0.523	0.904	2.190	0.544	1.807	5.502	0.386	1.136	5.619	0.437	0.930	4.103
themselves	(0.019)	(0.027)	(0.046)	(0.017)	(0.023)	(0.066)	(0.018)	(0.026)	(0.06)	(0.017)	(0.03)	-	(0.01)	(0.012)	(0.041)	(0.011)	(0.014)	(0.037)	(0.011)	(0.012)	(0.034)	(0.011)	(0.012)	(0.022)
medical	1.471	14.768	0.000	1.359	4.597	0.000	1.225	2.830	0.000	1.419	2.407	-	1.488	1.994	363.577	1.094	1.156	1.780	1.309	2.453	0.520	1.116	3.600	1.022
treatments	(0.025)	(0.086)	(0)	(0.018)	(0.043)	(0)	(0.016)	(0.034)	(0)	(0.016)	(0.037)	-	(0.013)	(0.017)	(0.395)	(0.013)	(0.017)	(0.038)	(0.014)	(0.017)	(0.028)	(0.016)	(0.021)	(0.022)
household	2.116	0.603	0.000	3.164	3.407	6.896	1.835	1.756	0.000	1.292	1.817	-	1.325	1.073	0.539	1.226	1.342	4.731	1.054	1.214	1.759	1.041	0.755	1.130
life	(0.04)	(0.045)	(0)	(0.032)	(0.064)	(0.106)	(0.019)	(0.033)	(0)	(0.017)	(0.037)	-	(0.013)	(0.016)	(0.047)	(0.013)	(0.017)	(0.053)	(0.013)	(0.016)	(0.028)	(0.017)	(0.019)	(0.025)
interpers.	1.189	1.521	0.000	1.555	2.667	3.091	0.797	1.178	0.000	2.134	1.317	-	1.264	2.045	1.901	0.935	1.205	1.872	0.680	0.793	1.818	0.753	0.939	1.995
rel.	(0.017)	(0.02)	(0)	(0.021)	(0.023)	(0.042)	(0.019)	(0.028)	(0)	(0.021)	(0.035)	-	(0.011)	(0.014)	(0.037)	(0.011)	(0.012)	(0.017)	(0.01)	(0.011)	(0.015)	(0.01)	(0.01)	(0.013)

Women	Set $\{d_j < 0\}$; Individual trajectories below the deepest path												Set $\{d_j \geq 0\}$; Individual trajectories above the deepest path											
	Proximity-group IV			Proximity-group III			Proximity-group II			Proximity-group I			Proximity-group I			Proximity-group II			Proximity-group III			Proximity-group IV		
Degree	I	II	III	I	II	III	I	II	III	I	II	III$^{(*)}$	I	II	III	I	II	III	I	II	III	I	II	III
postural	1.069	1.536	3.633	1.709	2.841	2.599	1.027	3.249	11.905	1.032	1.704	-	1.001	1.734	2.640	1.241	1.510	1.366	1.551	1.459	2.161	1.271	1.062	1.902
changes	(0.012)	(0.019)	(0.028)	(0.01)	(0.016)	(0.029)	(0.01)	(0.018)	(0.112)	(0.008)	(0.017)	-	(0.006)	(0.009)	(0.026)	(0.007)	(0.009)	(0.016)	(0.007)	(0.009)	(0.015)	(0.008)	(0.008)	(0.013)
mobility	1.104	0.472	0.349	1.242	0.713	0.648	1.332	0.723	0.000	1.054	0.444	-	0.787	0.644	0.716	0.918	0.933	1.059	0.639	0.902	0.618	0.987	0.930	1.373
	(0.019)	(0.024)	(0.026)	(0.012)	(0.016)	(0.027)	(0.011)	(0.016)	(0)	(0.008)	(0.017)	-	(0.007)	(0.008)	(0.022)	(0.007)	(0.009)	(0.015)	(0.008)	(0.009)	(0.013)	(0.009)	(0.009)	(0.013)
washing/	1.344	2.968	3.718	1.312	1.094	4.165	1.523	2.320	0.000	0.831	1.250	-	1.025	1.461	1.076	1.196	1.373	2.488	0.833	0.698	3.224	0.596	1.324	0.706
bathing	(0.02)	(0.039)	(0.063)	(0.014)	(0.026)	(0.107)	(0.011)	(0.027)	(0)	(0.01)	(0.024)	-	(0.008)	(0.012)	(0.044)	(0.009)	(0.012)	(0.03)	(0.01)	(0.011)	(0.027)	(0.014)	(0.015)	(0.026)
relieve	1.292	5.036	11.112	1.508	2.805	5.735	0.701	1.444	23.578	0.856	1.631	-	0.653	1.012	5.079	0.620	1.712	2.395	0.490	1.356	2.003	0.573	1.108	2.085
themselves	(0.013)	(0.02)	(0.038)	(0.01)	(0.014)	(0.035)	(0.011)	(0.015)	(0.093)	(0.01)	(0.017)	-	(0.006)	(0.008)	(0.029)	(0.007)	(0.009)	(0.018)	(0.007)	(0.008)	(0.015)	(0.007)	(0.008)	(0.011)
medical	1.399	1.255	0.000	0.880	3.403	0.000	0.901	2.711	1.613	0.963	2.802	-	0.973	1.857	0.468	0.910	1.916	1.046	0.859	1.612	0.584	0.969	1.988	0.884
treatments	(0.016)	(0.024)	(0)	(0.011)	(0.023)	(0)	(0.01)	(0.02)	(0.055)	(0.008)	(0.021)	-	(0.007)	(0.009)	(0.027)	(0.008)	(0.012)	(0.017)	(0.008)	(0.01)	(0.015)	(0.01)	(0.01)	(0.013)
household	3.097	2.190	0.000	2.640	1.433	1.793	3.350	1.627	0.000	1.650	0.926	-	1.395	0.722	0.935	1.791	0.708	1.566	1.489	1.304	1.538	1.231	0.928	1.823
life	(0.025)	(0.036)	(0)	(0.018)	(0.028)	(0.079)	(0.018)	(0.029)	(0)	(0.012)	(0.023)	-	(0.01)	(0.011)	(0.039)	(0.01)	(0.012)	(0.031)	(0.01)	(0.011)	(0.022)	(0.012)	(0.011)	(0.021)
interpers.	1.133	1.438	3.128	0.819	1.142	3.460	0.795	1.038	1.122	0.769	1.082	-	0.769	1.308	5.747	0.745	1.048	2.212	0.696	0.890	2.695	0.701	0.760	2.250
rel.	(0.01)	(0.012)	(0.015)	(0.011)	(0.013)	(0.024)	(0.012)	(0.015)	(0.035)	(0.013)	(0.021)	-	(0.008)	(0.009)	(0.022)	(0.007)	(0.008)	(0.011)	(0.007)	(0.007)	(0.01)	(0.006)	(0.006)	(0.008)

(*) The sample size is not enough to obtain significative results.

References

Albarrán, I. and P. Alonso (2009). La población dependiente en España: estimación del número y coste global asociado a su cuidado. *Estudios de Economía 36*(2), 127–163.

Albarrán, I., P. Alonso, A. Arribas-Gil, and A. Grané (2014). Can personal dependency paths help to estimate life expectancy free of dependency? In M. Sibilio and C. Perna (Eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, pp. 1–5. Springer.

Albarrán, I., P. Alonso, and A. Grané (2015). Profile identification via weighted related metric scaling: An application to dependent spanish children. *Journal of the Royal Statistical Society A 178*, 1–26.

Albarrán-Lozano, I., P. Alonso-González, and A. Arribas-Gil (2017). Dependence evolution in the spanish disabled population: a functional data analysis approach. *Journal of the Royal Statistical Society A 180(2)*, 657–677.

Andrae, F., P. Guevara, M. L. ao, Y. Duarte, and J. Santos (2011). Gender differences in life-expectancy and disability-free life-expectancy among older adults in S ao paulo, brazil. *Women's Health Issues 21*, 64–70.

Arribas-Gil, A. and J. Romo (2012). Robust depth-based estimation in the time warping model. *Biostatistics 13*, 398–414.

BOE-A-2005-21310 (2005). Orden TAS/4054/2005. Ministerio de Trabajo y Asuntos Sociales.

Cambois, E., J. M. Robine, and M. D. Hayward (2001). Social inequalities in disability-free life expectancy in the French male population, 1980-1991. *Demography 38*(4), 513–524.

Cox, D. (1972). Regression models and life tables. *Journal of the Royal Statistical Society B 34*, 187–220.

Cox, D. R. and D. Oakes (1984). *Analysis of Survival Data*. London: Chapman & Hall.

Crimmins, E. M., M. D. Hayward, A. Hagedorn, Y. Saito, and N. Brouard (2009). Change in disability-free life expectancy for Americans 70 years old and older. *Demography 46*(3), 627–646.

Dupuy, J., J. Loubes, and E. Maza (2011). Non parametric estimation of the structural expectation of a stochastic increasing function. *Statistics and Computing 21*(1), 121–136.

Eurostat (2009). Health statistics and atlas on mortality in the European Union. Eurostat. Luxembourg.

Fries, J. (1983). The Compression of Morbidity. *The Milbank Memorial Fund Quaterly 61*, 397–419.

Gasser, T. and A. Kneip (1995). Searching for structure in curve samples. *Journal of the American Statistical Association 90*, 1179–1188.

Gasser, T., H. Müller, W. Köhler, L. Molinari, and A. Prader (1984). Nonparametric regression analysis of growth curves. *The Annals of Statistics 12*, 210–229.

INE (2010). Encuesta sobre Discapacidad, Autonomía personal y Situaciones de Dependencia (EDAD), Metodología. Ed. Subdirección General de Estadísticas Sociales Sectoriales (INE), Madrid, España.

Jiménez, A. and A. Huete (2010). Estadísticas y otros registros sobre discapacidad en España. *Política y Sociedad 47*(1), 165–174.

Kamette, F. (2011). Dependency Care in the EU: a comparative analysis. Foundation Robert Schumann. Social Issues European Issue, 196.

Kneip, A. and T. Gasser (1992). Statistical tools to analyze data representing a sample of curves. *The Annals of Statistics 16*, 82–112.

Liu, X. and H. Müller (2004). Functional convex averaging and synchronization for time-warped random curves. *Journal of the American Statistical Association 99*(467), 687–699.

Lloyd-Sherlock, P., M. McKee, S. Ebrahim, M. Gorman, S. Greegross, M. Prince, R. Pruchno, G. Gutman, T. Kirwood, D. O'neill, L. Ferrucci, S. Kritchewski, and B. Vellas (2012). Population ageing and health. *The Lancet 379*, 1295–1296.

López-Pintado, S. and J. Romo (2009). On the cencept of depth for dunctional data. *Journal of the American Statistical Association 114*, 486–503.

Manton, K. G. and K. C. Land (2000). Active life expectancy estimates for the U.S. elderly population: A multidimensional continuous-mixture model of functional change applied to completed cohorts, 1982-1996. *Demography 37*(3), 253–265.

Martel, L. and A. Bélanger (2000). Regression models and life tables. *Canadian Social Trends-Statistics Canada Autumn 2000*, 26–29.

Ramsay, J. and B. Silverman (2005). *Functional Data Analysis* (second ed.). New York: Springer Series in Statistics.

Robine, J., C. Jagger, C. Mathers, E. Krimmins, and R. Suzman (2003). *Determining health expectancies*. Chichester: Wiley.

Robine, J. and K. Ritche (1991). Healthy life expectancy: evaluation of global indicator of change in population health. *BMJ 302*, 457–460.

Sanderson, W. and S. Scherbov (2010). Remeasuring aging. *Science 329*, 1287–1288.

Sullivan, D. (1971). A single index of mortality and morbidity. *HSMSA Health Reports 86*, 347–354.

Tarabelloni, N., A. Arribas-Gil, F. Ieva, A. Paganoni, and J. Romo (2016). roahd: Robust analysis of high dimensional data. Package on CRAN. Version 1.0.4. Published 2016-07-06.

Wang, K. and T. Gasser (1997). Alignment of curves by dynamic time warping. *The Annals of Statistics 25*, 1251–1276.

WHO (2011a). Global health and aging. WHO (World Health Organization). US National Institute of Aging.

WHO (2011b). Global report on disability. WHO (World Health Organization). US National Institute of Aging.

Yi, Z., G. Danan, and K. C. Land (2004). A new method for correcting underestimation of disabiled life expectancy and an application to the Chinese oldest-old. *Demography 41*(2), 335–361.