Dependency evolution in Spanish disabled population: A functional data analysis approach

Irene Albarrán-Lozano

Departamento de Estadística, Universidad Carlos III de Madrid, Getafe, Spain.

Pablo J. Alonso-González

Departamento de Economía, Universidad de Alcalá, Alcalá de Henares, Spain.

Ana Arribas-Gil

Departamento de Estadística, Universidad Carlos III de Madrid, Getafe, Spain.

Abstract.
In a health context, dependency is defined as a lack of autonomy in performing the basic activities of daily living and requiring care-giving or significant help from another person. However, this contingency, if present, changes over one's lifetime. Empirical evidence shows that, once this situation occurs, it is almost impossible to return to the previous state and in most cases increases in intensity. In this article, the evolution of the intensity of this situation is studied for the Spanish population affected by this contingency. Evolution in dependency can be seen as sparsely observed functional data, where we obtain a curve for each individual that is only observed at those points where changes in his/her condition of dependency occur. We use functional data analysis techniques, such as curve registration, functional data depth and distance-based clustering, to analyse this type of data. This approach proves to be useful in this context because it considers the dynamics of the dependency process and provides more meaningful conclusions than simple pointwise or multivariate analysis. We use the sample statistics obtained to predict the future evolution of dependency. The analysed database originates from the Survey on Disability, Personal Autonomy and Dependency Situations, EDAD 2008 (Spanish National Institute of Statistics, 2008). The survey is the largest and most complete survey to be made available in Europe for the study of disability. In addition, the Spanish legislation is one of the most recent in Europe and provides a detailed quantitative scale to assess dependency. In this article, the scale value according to this legislation has been calculated for each individual included in EDAD 2008. Differences between sex, age and the time of first appearance have been considered, and a prediction of the future evolution of dependency is obtained.

Keywords: Chain-ladder, dependency, disability, forecasting, functional data, time warping model

1. Introduction

When we refer to normal living, we usually imagine a situation in which people can do everything that they need or wish to do independently on a daily basis such as dressing and bathing themselves, eating, and drinking. Unfortunately, this is not always possible because of the presence of a disability. This turns into a tougher situation if another person is required to help complete all these activities. In this case, we would be talking about dependency. Traditionally, this problem has been treated as a question of health. However, since the beginning of this century, the social aspects of this problem have also been considered (World Health Organization, 2001). The International Classification of Functioning, Disability and Health (ICF) has attempted to establish a consensus on its understanding. The main objectives of this classification are to provide a common and standardised language as a conceptual framework for describing health and related topics in terms of, for instance, the type of activities in daily life that a person can and cannot do by themselves.

There are many ways of defining *dependency*. One of the most widely accepted definitions is that included in Recommendation No. R(98) 9 of the Council of Europe (Council of Europe, 1998), which 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". Not all dependents suffer this contingency with the same intensity. Therefore, one of the main issues is the measurement of dependency. A scale is typically used to this end. The most standard item to evaluate is the time another person spends helping a dependent person perform certain activities such as dressing or feeding themselves. This is the assessment used by the German and Spanish systems.

Another aspect related to dependency that cannot be forgotten is the increase in intensity throughout one's lifetime. As one can assume, dependency generally increases in intensity with age. This article attempts to analyse the evolution of this suffering over the course of a lifetime once it has been diagnosed. We are particularly interested in estimating the evolution of the personal dependency scores for the period between the time the survey was conducted and the person's death.

The data analysed in this study were collected by EDAD 2008, a recent macro-survey on dependency conducted in Spain. To our knowledge, the database obtained from this survey is the largest and most complete database to be made available in Europe for the study of the phenomenon of disability. It is important to note that Spain is one of the few countries that has collected nationwide information about these contingencies following internationally accepted classifications such as the ICF. In addition, the Spanish legislation is one of the most recent in Europe and provides a highly detailed and quantitative scale to assess dependency. Thus, the Spanish situation provides a unique framework to study disability and dependency for a national scope.

It is important to note the statistical challenges that arise in analysing this type of survey data. Indeed, we will study the evolution of dependency from a one-off survey because there are no longitudinal datasets available to analyse this phenomenon. Moreover, this situation is not likely to be remedied soon because a longitudinal study of this type would require a cohort to be followed over their whole lifetimes or at least for a long enough period of time to obtain sufficient temporal information. In this context, we propose a way to exploit a one-off survey in which historical information is collected to build a pseudo-panel from which dependency evolution trajectories can be extracted. These trajectories correspond to people in different cohorts so that they are defined in different time intervals. We propose to group them by age intervals so that we can apply functional data techniques to the data in each group. As we will see in Section 3, these trajectories are not smooth, and thus, an event-type functional data methodology needs to be applied. In particular, we will see that the presence of large temporal differences among the individual evolution curves makes it necessary to apply a time warping model to remove phase variability before the estimation of a mean trajectory. Moreover, this model would allow one to cluster individuals in terms of their different dependency evolutions and not only in terms of their current dependency situation. Concerning the prediction of future dependency, it is important to note that the evolution trajectories of younger individuals are not long enough to be used for forecasting. In this context, the use of techniques that can borrow historical information from other age groups is desirable. This framework is equivalent to the claim reserving problem in actuarial sciences; therefore, we propose to use the stochastic Chain Ladder method developed in that context to forecast dependency.

It should be noted that this paper is not focused on estimating the global impact of dependency in the general population or the costs for future assistance, which are related but different problems. Indeed, if we wished to obtain an estimation of the future impact of dependency on the whole population at each age, it is necessary to combine the following three elements: an estimation of future dependency scores by age groups, which we provide in this paper; the evolution of the prevalence rate of this contingency; and death probabilities for the dependent population. Given these rates and probabilities for the different age groups, it would be straightforward to obtain an overall weighted average future score. Nevertheless, obtaining these data is an impossible task because there is no statistical information for estimating neither the probabilities to change from an active to dependent state nor death probabilities related to this group of the population. Therefore, we are only able to estimate the future evolution of the personal scores and only when they are referred to people classified as dependent at the time the data were collected. Thus, according to the available statistical information, for the moment, the estimation of the impact of dependency in the general population remains an unsolvable problem. However, these limitations become irrelevant when estimating the future evolution of the scores because the population analysed in this paper has been grouped into similar age intervals; specifically, we have considered cohorts. Indeed, taking an abridged life table, every individual in a certain age interval would be affected by the same death probability during a given time lapse (five years in our case); therefore, the average future score after that time for all individuals in a specific age interval will not be dependent on this set of probabilities. In this setting, we are implicitly assuming that these death probabilities for the dependent population would only depend on the cohort age and calendar year but not on the degree of dependency of each person. This might be regarded as an unrealistic assumption, but it is necessary because of the lack of information about these death probabilities.

The remainder of the paper is organised as follows. In Section 2, we describe the EDAD 2008 survey and present our data. In Section 3, we introduce the time warping model used to analyse the data and to estimate a “mean” evolution curve. We also describe a distance-based cluster strategy for identifying different profiles among individuals. Finally, we present the stochastic Chain Ladder method used to predict future dependency scores from the estimations. Section 4 includes the analysis of the data from EDAD 2008 according to the methodology presented in Section 3, as well as the projected score values with their confidence bands. Finally, the results are discussed in Section 5.

2. Data

The Survey on Disability, Personal Autonomy and Dependency Situations, EDAD 2008, is the most recent large-scale, nation-wide household survey developed in Spain by the Spanish National Institute of Statistics (INE, by its acronym in Spanish) in agreement with the State Department for Social Services, Family and Disability Support (via the Office of Coordination of Sectorial Policies for the Disabled and the Institute for Older Persons and Social Services, IMSERSO) and the ONCE Foundation (the Spanish Organization for Blind People). This survey is the largest survey ever performed in Spain and focused on studying disability among the Spanish population. Data were collected between November 2007 and February 2008 around the country using stratified two-stage sampling. A sample was created for each province. For each of the provinces, the sample includes a uniform part and another part that depends on the size of the province’s population. In the first stage of the survey, more than 260,000 people living in private households were interviewed. In the second stage, more than 11,000 people living in public or private residencies were also included. For those living in households, the sampling units used in the first stage were the existing census sections on April 1^{st}, 2007, whereas the households were the units for the second stage. All members of a selected household were included in the survey; therefore, the sample includes people of all ages. For those living in specialized centres, the sampling units used in the first stage were the centres, whereas the persons living in those centres were the units for the second stage. The number of persons interviewed in each centre was determined based on the type and size of the centre. The sampling design provides a weight associated to each individual in the sample indicating how many people in the population he/she represents (the sum of all the sampling weights equals the population size). See Albarrán and Alonso (2009) and Instituto Nacional de Estadística (2010) for additional details on the sampling methodology.

This survey was based on the concept of self-perceived disability, in accordance with the recommendations of the World Health Organization. All the selected individuals were included in the survey; however, only those acknowledging suffering some type of disability were asked to provide the information contained in the disability questionnaire. Note that this allows one to estimate the disability prevalence (9.1%). This questionnaire included some important questions such as the degree of severity and the starting age of the disabilities. Each person was asked about the history, from birth until the moment the survey was conducted, of all the disabilities they suffered. Despite the fact that the survey includes the term dependency in its title, the questionnaire does not consider any specific question on this topic. Specifically, the intensity of the dependency suffered by each individual is not quantified in any way. The information contained in the survey was used to create a quantitative variable accounting for the degree of dependency of each individual. This variable ranges from 0 to 100 points. The greater the score, the stronger the dependency is, with a value of 0 indicating a non-dependent situation. Note that, according to this, all dependents are disabled, but the opposite is not always true (some disabilities may have an associated score of 0). The sex and age distributions of the dependent individuals (score > 0) included in EDAD 2008 are shown in Figure 1, where we also show mean dependency score values by age and gender.

Using the information about personal evolution contained in the survey, the dependency score could be calculated for each individual at all times at which a new disability appeared, which allowed us to build a pseudo-panel with temporal information. Because the items of the survey did not exactly match the dependency categories of the Spanish legislation, a thorough work on considering all the details in the legislation and a case by case analysis were needed. Additional details on how the scores were obtained are given in Albarrán and Alonso (2009). Note that the disabilities considered in the survey are not supposed to heal or decrease in intensity; therefore, the dependency score is either constant or increasing over time. This pseudo-panel is crucial for studying the evolution of dependency in a context in which longitudinal studies are not available.

Because we are particularly interested in studying the future evolution of personal dependency scores until time of death, we have only considered the dependent population over 50 because the intensity of this contingency increases with age, as has been noted previously. The results have been obtained distinguishing between men and women. The sex and age distributions of the individuals included in the analysis are shown in Table 1. The sampling weights have been used to estimate the corresponding population sizes.

3. Methods

First, we attempt to conduct a descriptive analysis of the dependency trajectories obtained from the database. If we think of the evolution of dependency as a continuous process, we may consider these data as individual realizations of that process, observed only at those moments at which changes in the personal situations occur. Thus, methods for analysing such event-type data in the functional data context are of interest here.

In functional data analysis (FDA), the sample of observations is typically given as a set of curves

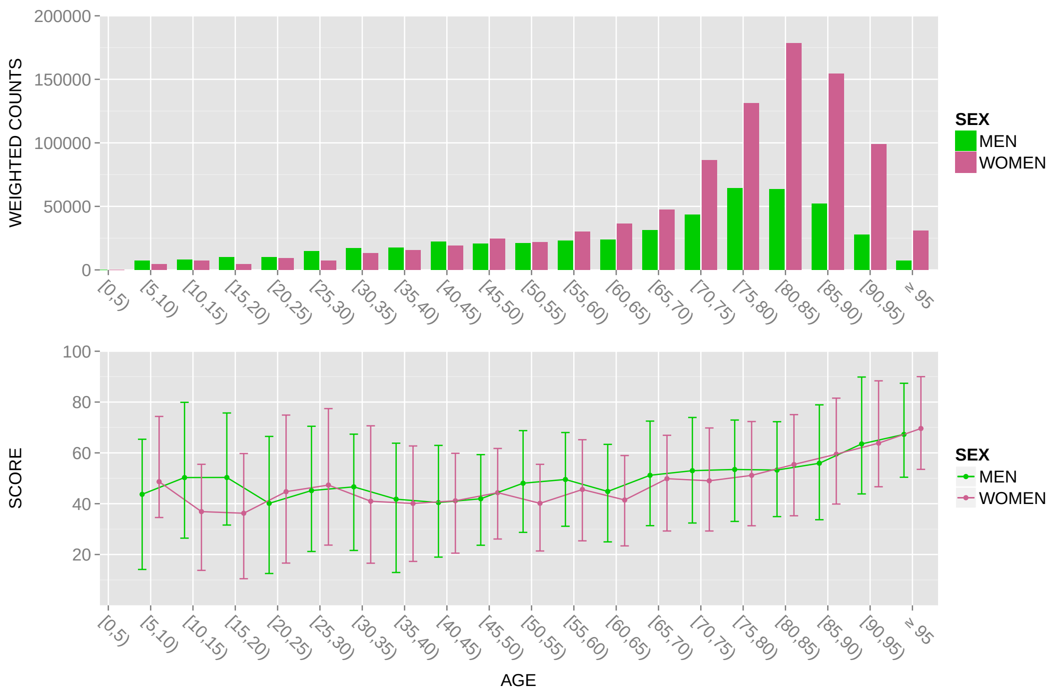

Figure 1. Top: Age distribution of dependent population (score > 0) in the EDAD 2008 sample for men (green) and women (purple). Bottom: Mean dependency score with first and third quartile by age for men (green) and women (purple).

Table 1. Age distribution of the analysed sample (dependent population, score > 0, over 50) for men and women.

	Men		Women	
Group of age	Estimated Population Size	Sample Size	Estimated Population Size	Sample Size
$[50, 55)$	20,816	137	21,297	130
$[55, 60)$	22,038	146	29,335	172
$[60, 65)$	23,314	157	35,429	238
$[65, 70)$	31,163	182	45,812	287
$[70, 75)$	43,437	326	85,418	573
$[75, 80)$	63,677	421	130,041	831
$[80, 85)$	63,549	424	176,779	1,102
$[85, 90)$	51,719	319	153,968	936
$[90, 95)$	27,316	286	98,742	991
≥ 95	7,164	47	30,832	208
Total	354,194	2,445	807,655	5,468

in which each curve represents the behaviour over time of an individual in some process of interest. A common assumption is that individual curves are i.i.d., and the general purpose of the analysis is to extract information about the process. Classical references on this topic include Ramsay and Silverman (2005) and Ferraty and Vieu (2006). Standard FDA techniques assume smoothness of the underlying process so that the observed curves can be expressed as linear combinations of a set of smooth functions. However, in the case of event-type functional data, the smoothness assumption is violated because the process of interest is given by $Y(t) = \sum_{k=1}^{K} S_k \mathbb{1}\{T_k \leq t\}$, $t \in [0,T]$, where K is a random number of events, S_k are i.i.d. random variables describing the severity of the events, T_k are the random times at which the events occur, and $[0,T]$ is the time interval on which the process is observed. In this setting, alternative techniques, such as the ones described in Section 3.1, need to be used to analyse the set of observed curves.

Let us now describe our data as event-type functional observations. The dataset consists of individual pieces of information containing 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. Notice that, even if the data come from a one-time survey, individuals were asked about their whole medical history; therefore, 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 at which changes occur and the dependency scores at these ages, and a_i, the current age (at 2008). From these data, 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 $t_{in_i} < a_i$). These transformed sequences will constitute our set of observations herein. For the sake of simplicity, we will still refer to them as $(t_{ih}, y_{ih})_{h=1,\ldots,n_i}$, $i = 1, \ldots, n$. Indeed, we focus on the event-type character of the data instead of considering any type of smoothing. The reasons for this are the following: First, some changes in the dependency score are indeed true jumps due to specific events (hip fracture, accidents, etc.), and smoothing those data will result in a loss of information. Second, even if dependency also increases gradually, this generally occurs at advanced ages, where the slope of the dependency trajectory is so steep regardless that the difference between smooth and step functions is almost negligible. Finally, the analysis that we undertake focuses on the dependency score, not on the underlying health/dependency condition, which is typically a scale function, that is, a piecewise constant function in which changes only occur once some particular disability/dependency status has been reached. For all these reasons, we analyse our data as event-type functional data and not as smooth functional data. However, we acknowledge that other approaches are also be possible, especially if the focus is on the dependency/disability condition itself and not on the scale that quantifies it.

Thus, we have n discretely observed curves $y_1, \ldots, y_n$ defined in different time intervals $[0, a_i]$, $i = 1, \ldots, n$. However, to apply any functional data analysis techniques, 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 because the age range of the individuals in the sample is large. Instead, we consider disjoint groups of people in age intervals of 5 years. Within each age interval $[A, A+5)$, we truncate individual curves to obtain the curves defined in $[0, A]$. Then, given a starting age A_1, we have the following k groups of individuals and curves

$$\mathcal{I}_{A_j} = \{i \mid 1 \leq i \leq n,\, a_i \in [A_j, A_j + 5)\},\ \mathcal{C}_{A_j} = \{y_i(t),\, t \in [0, A_j] \mid i \in \mathcal{I}_{A_j}\},\ j = 1, \ldots, k, \quad (1)$$

where $A_j = A_1 + 5(j-1)$. The particular values of the age intervals considered for the analysis, that is, the value of the starting age A_1 and the number of groups k, are specified in Section 4.
The idea is now to analyse separately each group of curves. In the following, we describe the

functional data analysis techniques that we will use, and in Section 4, we present the results of the analysis performed on the different groups and the comparisons between them.

3.1. Estimating the central trend

Providing a measure of centrality when addressing functional data is not an straightforward task. Indeed, the levels of the curves as well as their shapes, whose information is more difficult to incorporate into any numerical summary, are both important. The problem is aggravated 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 most likely present two modes and subsequently will be dissimilar to the two curves in terms of shape.
In this context, it is extremely important to use measures of centrality that can consider 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 that 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:

$$y_i(t) = x \circ h_i^{-1}(t) \quad t \in [0, A], \quad i = 1, \ldots, n, \tag{2}$$

where x is an unknown deterministic function representing the process of interest and h_i^{-1} is an 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 that some others are delayed with respect to x. In the time warping model, two approaches used to estimate the central trend or mean behaviour of the data are possible: 1) align or register the curves, that is, estimate h_i and compute any desired sample statistic on the registered sample $y_1(\hat{h}_1), \ldots, y_n(\hat{h}_n)$, or 2) define appropriate estimators directly on the observed sample, considering the nature of the data. For the analysis of the dependency dataset, we will consider one estimator of each type that we now describe.

Notice that the general time warping model (2) includes any type 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). In this sense, the time warping model provides a general and flexible framework for the modelling of the dependency trajectories. In addition, notice that we can assume that observations are free of measurement error because they correspond to the evaluation, on an official numerical scale, of the particular conditions suffered by each individual at each moment.

3.1.1. Cross-sectional mean after registration

Aligning or registering the trajectories consists of estimating the warping functions to obtain $y_i(\hat{h}_i(t))$, $i = 1, \ldots, n$. Many different curve registration methods adapted for different scenarios exist, most of them requiring densely observed data and smoothness (Kneip and Gasser, 1992; Silverman, 1995; Ramsay and Li, 1998; Kneip and Ramsay, 2008; see Ramsay and Silverman (2005), Chapter 7 for an overview). For instance, if one can clearly identify the same common features in all the curves, landmark registration, which consists of estimating the warping functions so that those features are brought together, is the benchmark. Another typical method is to estimate the warping functions by minimising the distance between $y_i(\hat{h}_i(t))$ and some given template to which all the curves are aligned. However, the choice of template is not straightforward and implicitly assumes knowledge about the unknown function x. In addition, in our case, we address event-type data for which landmarks are not clearly identified. Indeed, the dependency trajectories are piecewise constant functions with change points corresponding to the times when changes in the health status of an individual produce a new dependency score according to the official scale. In this sense, most of the curves have a small number of jumps, and the main variability among individuals is driven by the timing of these change points (see Figure 2). In this framework, the method presented in Arribas-Gil and Müller (2014) is specially designed to align curves of this type. The method consists of two steps: finding pairwise alignments between any pair of curves and then constructing a global alignment of the whole sample by averaging the pairwise alignments. This avoids the problem of finding a template to which to align all the curves because in the pairwise alignments one curve serves as a template for the other curve. Indeed, notice that, according to (2), for a pair of curves y_i and y_j, we have $y_i(t) = x(h_i^{-1}(t)) = y_j(h_j(h_i^{-1}(t)))$. Then, aligning y_i towards y_j means estimating the pairwise synchronization $g_{ji}(t) = h_j(h_i^{-1}(t))$. Given some dissimilarity criterion to minimise, this is easily performed through a discrete dynamic programming algorithm because the curves are sparse in nature. Once we have estimates $\hat{g}_{ji}(t)$ and $\hat{g}_{ij}(t)$ for all $i, j = 1, \ldots, n$, we estimate the warping functions as

$$\hat{h}_i^{-1}(t_{ih}) = \frac{1}{n}\sum_{j=1}^{n} \hat{g}_{ji}(t_{ih}), \quad h = 1, \ldots, n_i, \quad i = 1, \ldots, n,$$

because, by the identifiability condition on the warping process, $E[h_j(h_i^{-1}(t))|h_i^{-1}(t)] = h_i^{-1}(t)$. The whole algorithm has a computational complexity of $O(\binom{n}{2}L^2) \approx O((n \cdot L)^2)$ (L standing for some average number of observed points per curve). For more details on the method, see Arribas-Gil and Müller (2014).

Once we have aligned the sample of curves, we can compute any sample statistics such as the cross-sectional mean of the registered sample.

3.1.2. Deepest curve

The literature on estimators of the second kind, namely, those directly defined on the unregistered sample, is relatively small. Liu and Müller (2004), Dupuy *et al.* (2011), or Arribas-Gil and Romo (2012) are works that are particularly concerned with the definition of suitable population centrality measures, and their corresponding sample statistics, in the time warping model.

For the analysis of the dependency dataset, we will consider the approach of Arribas-Gil and Romo (2012) because it provides a robust estimator of the central trend for a set of curves. Indeed, the registration procedure described in Section 3.1.1 neutralises the effect of those curves with an atypical shape (due to the fact that they may be delayed or accelerated with respect to the remaining shapes). However, there might be curves with a typical shape but taking atypical values (abnormally high or low at some locations). One way to provide a centrality measure that is robust against the two types of atypical curves is to use the functional depth. Indeed, the deepest curve of a sample, in terms of band depth (López-Pintado and Romo, 2009), has been proved 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 because (i) it is a curve geometrically located in the center of the sample and (ii) it exhibits 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.

In the analysis of Section 4, we compare, for each group of curves, the sample mean (computed after registration) and the deepest curve (computed directly on the original sample before registration). All of the procedures described above (registration, determination of the deepest curve and sample mean calculation) have been adapted in a straightforward manner to consider the weighted sampling, that is, the fact that each individual in the sample represents a different number of individuals in the population.

3.2. Curve clustering

In the time warping model (2), warping function estimates are useful for individual classification. Indeed, they contain information on how different a curve is with respect to the remaining curves in terms of how accelerated or delayed the process has been registered in that particular individual. Then, following Arribas-Gil and Müller (2014), we will perform distance-based clustering of individual curves using the warping functions to define a distance between individuals. Indeed, we define

$$d(i,j) = \int_0^A \left(\hat{h}_i^{-1}(t) - \hat{h}_j^{-1}(t)\right)^2 dt, \quad i,j = 1\ldots,n. \tag{3}$$

Because warping function estimates are discrete-valued functions, the integral needs to be computed numerically. Instead of using classical Multidimensional Scaling for dimension reduction, we work directly with this distance matrix to obtain groups of individuals with similar profiles. For this, we use a k-means algorithm in which the centroids of the clusters are defined as those individuals who minimise the total sum of square distances to the remaining individuals in the same cluster.

As with any clustering method, there is always the problem of determining the number of clusters. One way to assess cluster quality and choose the number of clusters is to use the *silhouette* coefficient (Kaufman and Rousseeuw, 1990). Given a number of clusters c, and once we run the clustering algorithm to obtain the best configuration with c groups, the *silhouette* coefficient provides a measure of how good that configuration is (in terms of internal cluster cohesion and separation between clusters). The interpretation of this coefficient is two-fold. On the one hand, if the value obtained lies above a certain threshold, we can say that the overall cluster quality is good or, in other words, that there is evidence supporting the existence of c groups. On the other hand, if we compute the coefficient for different values of c, we can choose the number of clusters as the one that maximizes its value. The *silhouette* coefficient is obtained as follows. First, the *silhouette* of any individual i is $sil_c(i) = (b_i - a_i)/\max\{a_i, b_i\}$, where a_i is the mean distance from i to the remainder of the individuals in the same cluster (cohesion measure) and b_i is the mean distance from i to the individuals of the closest cluster (separation measure). Note that if i is correctly assigned to its cluster, one can expect to have a positive value of $sil_c(i)$, with larger values corresponding to better assignments. Finally, the *silhouette* coefficient of a clustering is the mean *silhouette* across individuals $sil_c = \frac{1}{n}\sum_{i=1}^n sil_c(i)$. This coefficient provides a measure of the global clustering quality, with larger values corresponding to appropriate clustering of the data. Therefore, the choice of the number of clusters can be made by maximizing the *silhouette* coefficient on the number of clusters c.

3.3. Forecasting of dependency scores

Once we have a common evolution pattern for each age group, these will be the basis for estimating future scores. Due to the division of the sample into different age groups, the resulting average or deepest curves obtained as explained in Section 3.1 exhibit a special structure. Indeed, we have, for age intervals $[A, A+5]$, a representative score curve S_A defined in $[0, A]$, $A = A_1, \ldots, A_k$. Considering that, in practice, each curve is a series of discretised values, we can summarise all the information in a $k \times k$ table wherein each row represents an age group or cohort and each column stands for a time point. Specifically, row j will contain j values, $S_{A_j}(A_1), \ldots, S_{A_j}(A_j)$, leaving the last $k - j$ columns empty. For the sake of simplicity, we will denote these values as $S_{j,h}$, $h = 1, \ldots, j$, $j = 1, \ldots, k$. The resulting table has a similar structure to that of a run-off triangle. This tool is quite usual in actuarial practice for estimating the level of reserves that will be necessary to face potential claims in the future. Therefore, the projected scores can be obtained using techniques that are appropriate in this context such as the *Chain Ladder* method (see, for instance, Taylor, 1986). In an actuarial context, this method shows how claims are developing over the years using cumulative claim payments (in our case, this variable is replaced by the cumulative score). The method attempts to estimate a rate of change for the cumulative score in any year with respect to the previous year. The increase rate associated to each year is called the development factor $\hat{f}_k$, which can be defined as

$$\sum_{i=k+1}^{I} x_{i,k+1} / \sum_{i=k+1}^{I} x_{i,k} \quad 1 \leq k \leq I - 1,$$

where k is the development year, I is the maximum number of accident years, and x is the cumulative amount of the analysed variable. This method can be estimated in a parametric way using GLM, as can be observed in Kremer (1982) or in Renshaw and Verrall (1994). According to this approach, the logarithm of the incremental claims amount can be considered as the response variable and is regressed on two non-interactive sets of covariates, treated as factors, with a parameter α_i for each accident year i and a parameter β_j for each development year j. However, the disadvantage of this scheme is from the extreme complexity involved in the analytic expressions used to calculate prediction errors. This problem can be avoided using bootstrap techniques. England and Verrall (1999) and Liu and Verrall (2010) showed how to use this methodology to calculate the prediction errors in claims reserving. The key point is to use a residual definition that can be considered appropriate to the model under consideration. In our case, the direct application of the Chain Ladder on scores can lead to non-desirable results because it would be possible to obtain scores greater than 100 points. These results would be inconsistent with the measurement of dependency that we are considering because that level should be the maximum estimated score reached by the model. Due to this restriction, we have used a model based on the Brass logit model (Brass, 1971), as will be explained in Section 4.4.

Thus, the Chain Ladder algorithm has been applied with logits instead of incremental scores, as it were for a traditional run-off triangle. The stochastic character of the estimation has been obtained using the approach proposed by England and Verrall (1999) and Liu and Verrall (2010). This scheme was the one used to estimate the Renshaw and Verrall (1994) model for reserving. However, the scheme followed in this paper differs from that in the definition of the residuals employed for resampling. Because our incremental logits can be negative, it would be impossible to use the unscaled Pearson residuals, as in England and Verrall (1999). Instead, we have used response residuals. Confidence intervals for every forecast have been estimated using bootstrapping.

4. Analysis of dependency evolution data

As explained in Section 3, the first step of the analysis is to divide the individuals of the sample into groups of people in terms of their age. Because we are interested in the older population, we consider the age intervals $[50, 55)$, $[55, 60)$, $[60, 65)$, $[65, 70)$, $[70, 75)$, $[75, 80)$, $[80, 85)$,

$[85, 90)$, $[90, 95)$, and $[95, \infty)$. Specifically, we define a collection of 10 groups of individuals $\{\mathcal{I}_A\}_{A\in\mathcal{A}}$ and their corresponding groups of dependency evolution curves $\{\mathcal{C}_A\}_{A\in\mathcal{A}}$, with $\mathcal{A} = \{50, 55, 60, 65, 70, 75, 80, 85, 90, 95\}$. In Figure 2, we present the curves for the different groups of ages before and after alignment by the technique described in Section 3.1. Although at first glance it may seem that all the trajectories are shifted right, that is, all the jumps in the dependency score are shifted as late in life as possible, this is only true for the (few) trajectories with jumps at young ages (the most visible ones), whereas the majority of the trajectories (the trajectories with very few jumps at the end of the time interval) are slightly shifted left so that, overall, the registration process is "centred".

4.1. Data analysis of the whole dataset

We have computed the cross-sectional mean, the cross-sectional mean of the registered curves and the deepest curve for each of the age intervals previously defined. These results are displayed in Figure 2. It can be observed that, in every subsample, the cross-sectional mean before registration obtains higher scores at younger ages than the cross-sectional mean after registration. This is due to the presence of some atypical individuals for which high scores are reached at early ages. Their influence is reduced by aligning the curves. However, there is another type of atypical individual: those who reach very high scores at typical ages. Their effect cannot be attenuated by the registration process because their temporal behaviour is standard. This point can be addressed using a robust measure such as the deepest curve in each subsample. Indeed, we observe that, for every age interval, the deepest curve is systematically lower (during the whole time interval) than any of the cross-sectional means. This indicates that the distribution of the dependency evolution might present a slight positive asymmetry. However, the difference between the deepest curve and the mean of the registered trajectories is small and almost negligible for many of the age groups. Therefore, herein, and for the sake of interpretability, all our analyses will be based on the mean after registration.

4.2. Differences by gender

The analysed dataset (dependent individuals 50 years of age or older) is composed of 5,468 women and 2,445 men, who represent 807,655 women and 354,194 men in the Spanish population. We now consider them separately to study the differences in their evolution profiles. We repeat the analysis performed over the whole sample, now applied to the gender groups. After defining, for men and women, the different age groups, we obtain the mean curves shown in Figure 3. Figure 3 shows that the mean dependency score at the beginning of each age interval increases with age. However, the estimated mean dependency score at any particular age, say, 70 years, varies according to the age of the participant when the survey was completed. For example, the mean dependency score at age 70 is over 40 points for a man who was aged [70,75) at the time of the survey but almost zero for a man aged [80, 85) at the time of the survey. This is a consequence of the survey design and the assumption that the score cannot decrease after the dependency occurs. The rate at which dependency increases with age intensifies with age category: the [50,55) age group increases its score by approximately 30 points in 15 years, the [70,75) age group increases by approximately 40 points in 10 years, and the increase is approximately 60 points for the oldest group in just 7 years.

As for differences between men and women, we can say that, in the first age groups (until the age of 70), men evolve earlier than women in their dependency situation and even have slightly higher final scores, with the exception of the group of people between 60 and 65 years of age, in which this situation is reversed. However, for people older than 75, the situation is quite the opposite: men deteriorate more slowly and reach lower final scores than women. Regarding the [60,65) group, an atypically low presence of men with high scores at early ages, presumably due to sampling effects,

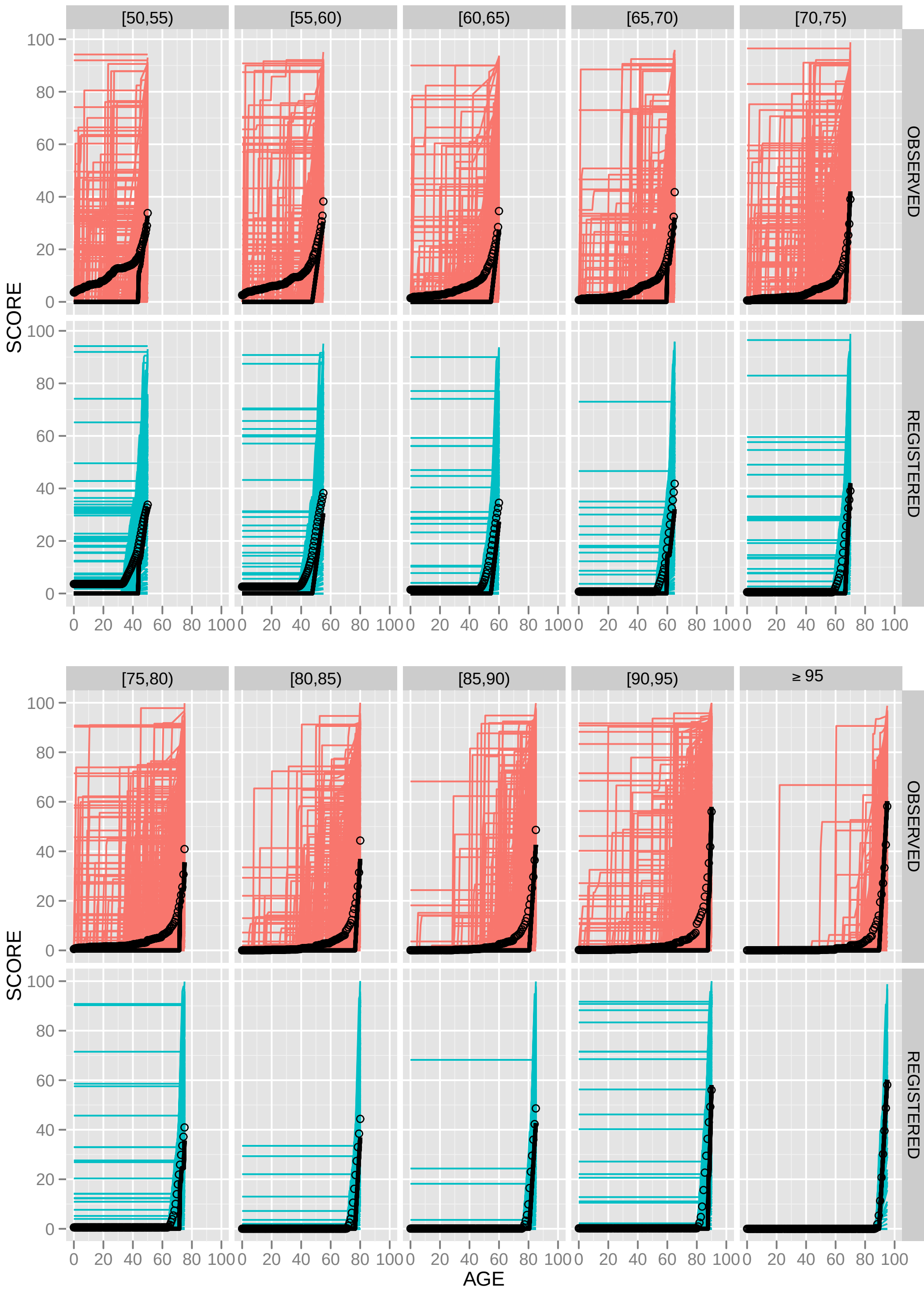

Figure 2. Top (first and third rows): Dependency evolution curves by age intervals. Black circles represent the cross-sectional mean of the observed curves, and the solid black line indicates the deepest curve of the sample. Bottom (second and fourth rows): Aligned dependency evolution curves by age intervals. Black circles represent the cross-sectional mean of the registered curves, and the solid black line indicates the deepest curve of the sample.

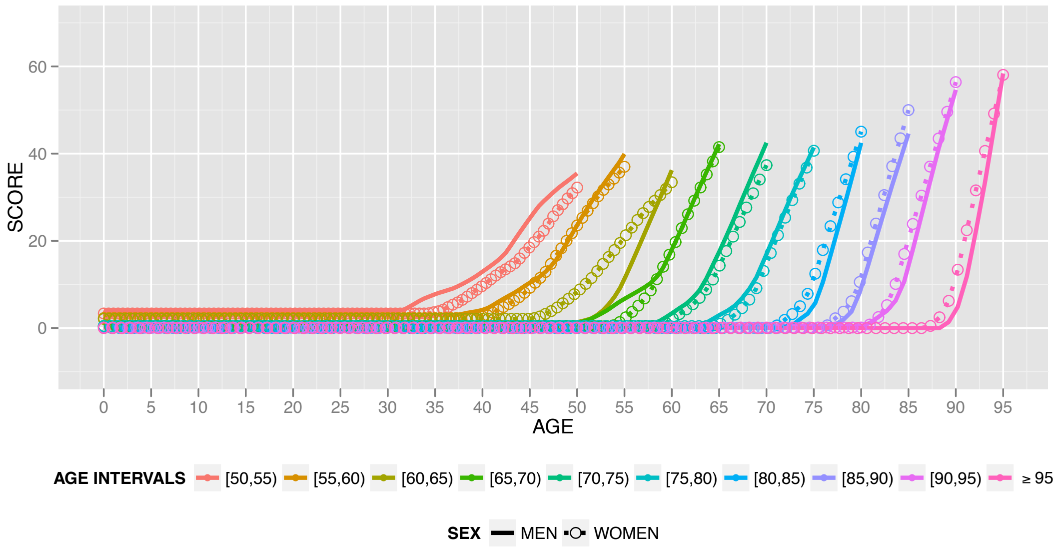

Figure 3. Cross-sectional means after registration for men (solid lines) and women (circles) for the 10 different age groups.

might be the cause of the discrepancy between average curves for men and women (this phenomenon can be noticed in the top-left panel of Figure 7, where the [60,65) group curve takes on very low values).

The mean for men and women has been calculated by registering the two sets of curves within each age group separately. Specifically, we have considered the original trajectories observed for men and women for each age interval and aligned them separately, therein obtaining two sets of registered curves, and have then computed the cross-sectional mean of each of these sets. This is different from jointly aligning all the curves (as shown in Figure 2) and then averaging the registered curves corresponding to men and to women separately. Indeed, the registration process eliminates time distortion among individuals. However, if the distortion is due to the presence of different subpopulations and not only to sample variability, we will be interested in retaining those differences. Therefore, we need to register the different subpopulations separately because the global registration procedure will produce an undesirable homogenization effect.

4.3. Identification of profiles

With the aim of identifying different profiles in the dependency evolution of the individuals of the sample, we have performed a warping-based clustering analysis, as explained in Section 3.2. For this, we use the warping function estimates obtained for each of the original age intervals $[50, 55)$, $[55, 60)$, $[60, 65)$, $[65, 70)$, $[70, 75)$, $[75, 80)$, $[80, 85)$, $[85, 90)$, $[90, 95)$, and $[95, \infty)$, that is, without considering any differentiation in terms of gender or the age of first occurrence. We then apply a k-means type of clustering algorithm to the distance matrix $D = (d(i, j))_{i,j}$. We can clearly identify two clusters of dependency curves for each age interval, which are shown in Figure 4. We have performed the analysis with a varying number of clusters, $c = 2, 3$ and 4, and the maximum mean silhouette coefficient was systematically obtained for $c = 2$. Moreover, this value was found to be greater than 0.82 in every age interval (all the obtained values are presented in Table 2). According to Kaufman and Rousseeuw (1990) (page 88), a value of this coefficient greater than 0.71 indicates

Table 2. Mean silhouette coefficient for the cluster analysis within each age interval with different numbers of clusters $c = 2, 3, 4$.

sil_c / Age group	$c = 2$	$c = 3$	$c = 4$
$[50, 55)$	0.8295	0.7535	0.7052
$[55, 60)$	0.8799	0.8150	0.7938
$[60, 65)$	0.8560	0.7949	0.6811
$[65, 70)$	0.8961	0.8190	0.6179
$[70, 75)$	0.9061	0.8495	0.7668
$[75, 80)$	0.9242	0.8589	0.7599
$[80, 85)$	0.8996	0.8596	0.8130
$[85, 90)$	0.9051	0.8437	0.8138
$[90, 95)$	0.9507	0.8757	0.7892
≥ 95	0.9348	0.8242	0.7719

that a strong structure has been found in the data. Therefore, there is evidence that the two identified groups correspond to two differentiated profiles. The conclusions are similar for any of the age groups: The first, and less numerous, cluster corresponds to individuals with *early-onset* dependency and many jumps homogeneously distributed throughout their lifetimes. The second cluster contains the most common profile, which consists of individuals with *regular* dependency or *late-onset* dependency but with very few jumps, concentrated at the end of their life. Specifically, the two clusters do not exactly correspond to *early-onset* and *regular* dependency evolution; rather, they correspond to individuals experiencing continuous worsening and individuals experiencing deterioration mostly concentrated at the end of their lives, which represent the majority of the population.

If we now consider, within each age group, the two clusters, we can analyse the differences between them. The means after registration for the two profiles are presented in Figure 5. As we explained above for the means by gender, these means are obtained by registering separately the trajectories of cluster 1 and cluster 2 within each age group. We can see how the mean trajectories for the first cluster exhibit a faster increase, obtaining higher scores much earlier in time than those for the second cluster. However, for the groups of older individuals (over the age of 80), the mean final scores in both clusters tend to become closer. Specifically, it seems that the differences between the two profiles decease as individuals become older. An exception to this appears in the last age group, namely, the age group with individuals older than 95, in which the difference between the two clusters is very important. However, this fact should considered with caution because the first cluster in this group contains only eight individuals.

4.4. Forecasting of dependency scores

As was explained in Section 3.3, forecasting will be achieved by the stochastic Chain Ladder method together with bootstrapping to build confidence intervals. The first step is to obtain the logit of each known average score $S_{j,h}$, where *j* refers to the age group and *h* stands for time. We define $l_{j,h} = log((S_{max} - S_{j,h})/(S_{j,h} - S_{min}))$, where S_{max} is equal to 100 and S_{min} is equal to $S_{j,h-1}$, that is, the previous average score in the same age group, with $j = 50, 55, \ldots, 95$ and $h = j, j+5, \ldots, 95$. Using this setting, we ensure that every future score is at least equal to the previous score.

Because the chain ladder will be used on logits, the whole set of information required to compute them will include 65 average score values, and the set of logits on which the technique will be implemented is composed of 55 values. Once the future logits have been estimated, the projected scores are obtained as $\hat{S}_{j,h} = (S_{max} + S_{min}\exp(\hat{l}_{j,h}))/(1 + \exp(\hat{l}_{j,h}))$.

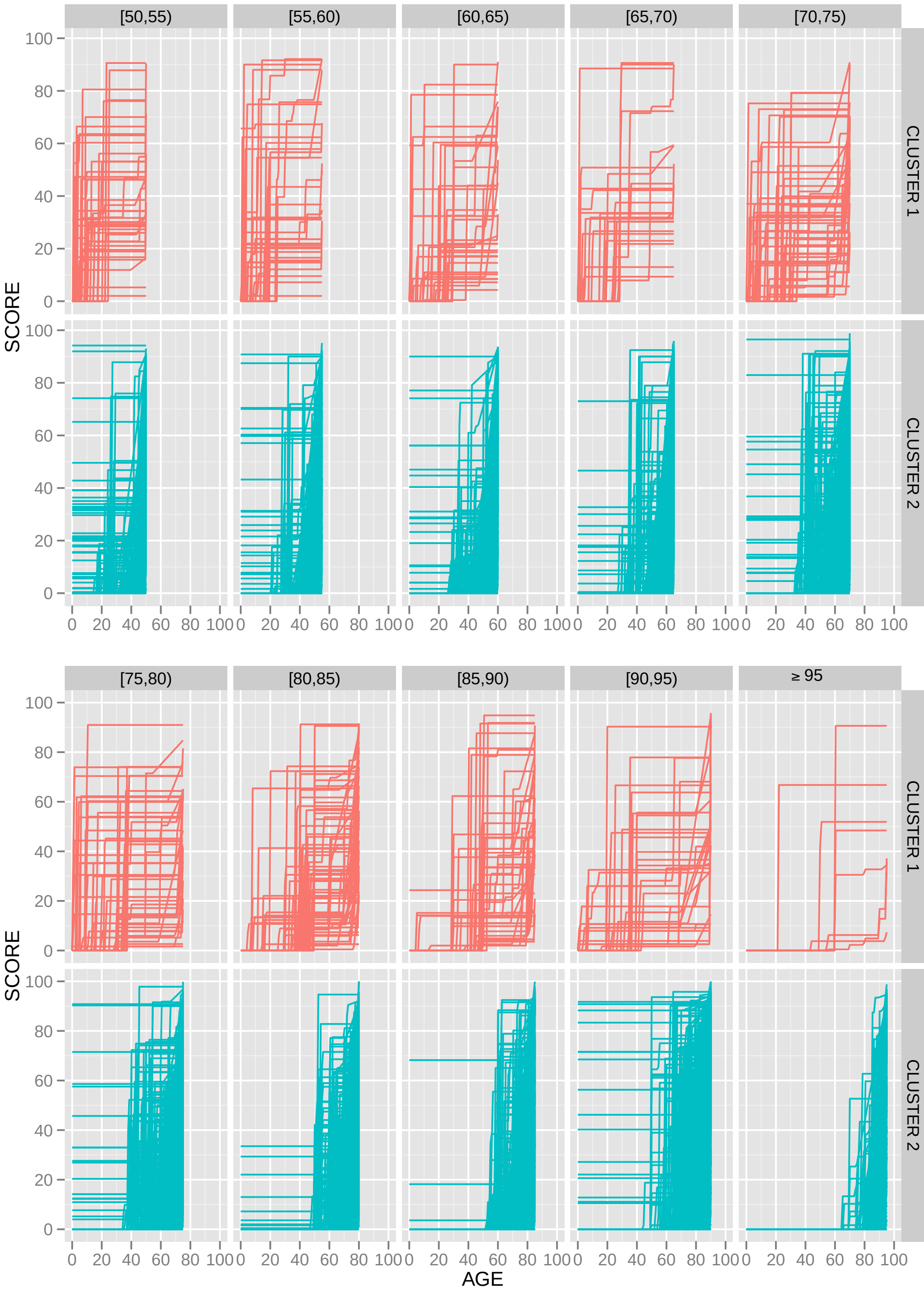

Figure 4. Dependency evolution curves in the two clusters by age intervals.

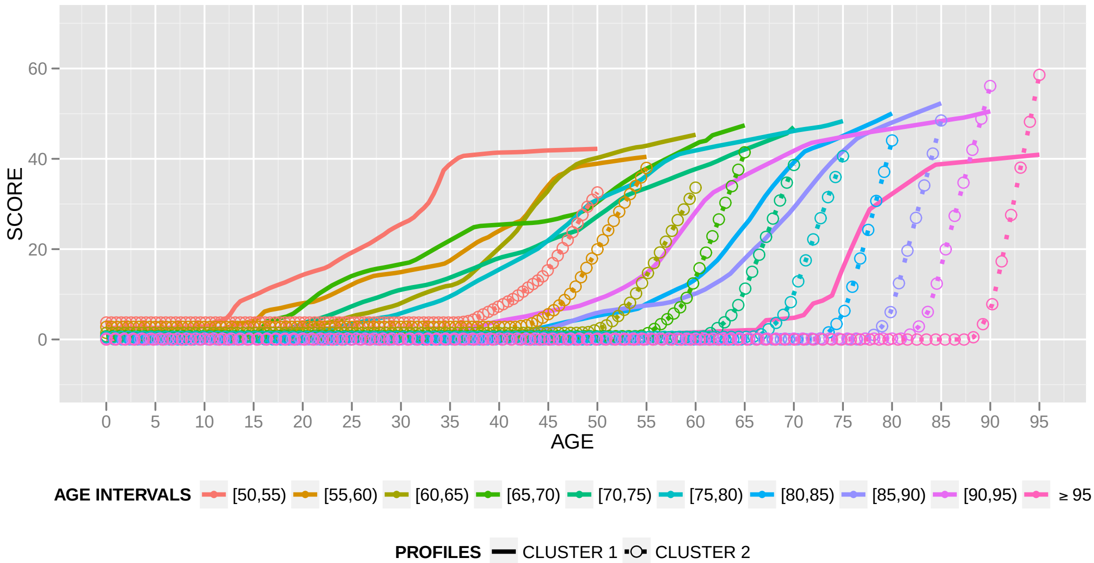

Figure 5. Cross-sectional means after registration for individuals in the first (solid lines) and second cluster (circles) for the different age intervals.

In Figures 6 and 7, we present the results by cluster and by gender and cluster. The actual predicted values and confidence intervals used in these graphics are given in Tables 2, 3 and 4 of the Supplementary Materials.

Looking at the different curves in Figures 6 and 7 (or at the different rows of the corresponding tables), for any given cohort, the forecasted scores increase with age as expected. If we now focus on any given future age (any column in the tables), in general, the scores become smaller as the cohorts get closer to that age. In other words, the younger a person becomes dependent, the higher the scores they will obtain in the future. This is true in almost all the cases for the second cluster but does not hold in general in the first cluster, in which the observed trajectories (Figure 5) did not previously follow this pattern. By profiles, individuals in the second cluster (the cluster representing a slow evolution of dependency) will obtain higher scores than those in the first cluster, especially for the oldest cohorts.
Combining sex and profiles, men behave similarly to the the global population, that is, higher scores are expected in the second cluster, except for the youngest cohort. For women, the differences between the forecasted scores in the two clusters are smaller; however, the scores also tend to be higher in the second cluster.
Comparing only gender, in the two clusters, the forecasted scores for women are almost always higher than those for men in any given cohort, especially for predictions at very advanced ages.

In both figures, we can appreciate how the confidence bands are systematically wider for the second cluster, which is because the bootstrap standard deviations $\hat{\sigma}^B_{j,h}$ are always higher in the second cluster. This is explained by the fact that the observed trajectories are steeper in the second cluster than in the first cluster, which results in higher variability for the estimates of the second cluster. It is also important to note that, given the restrictions of monotonicity and upper score limit

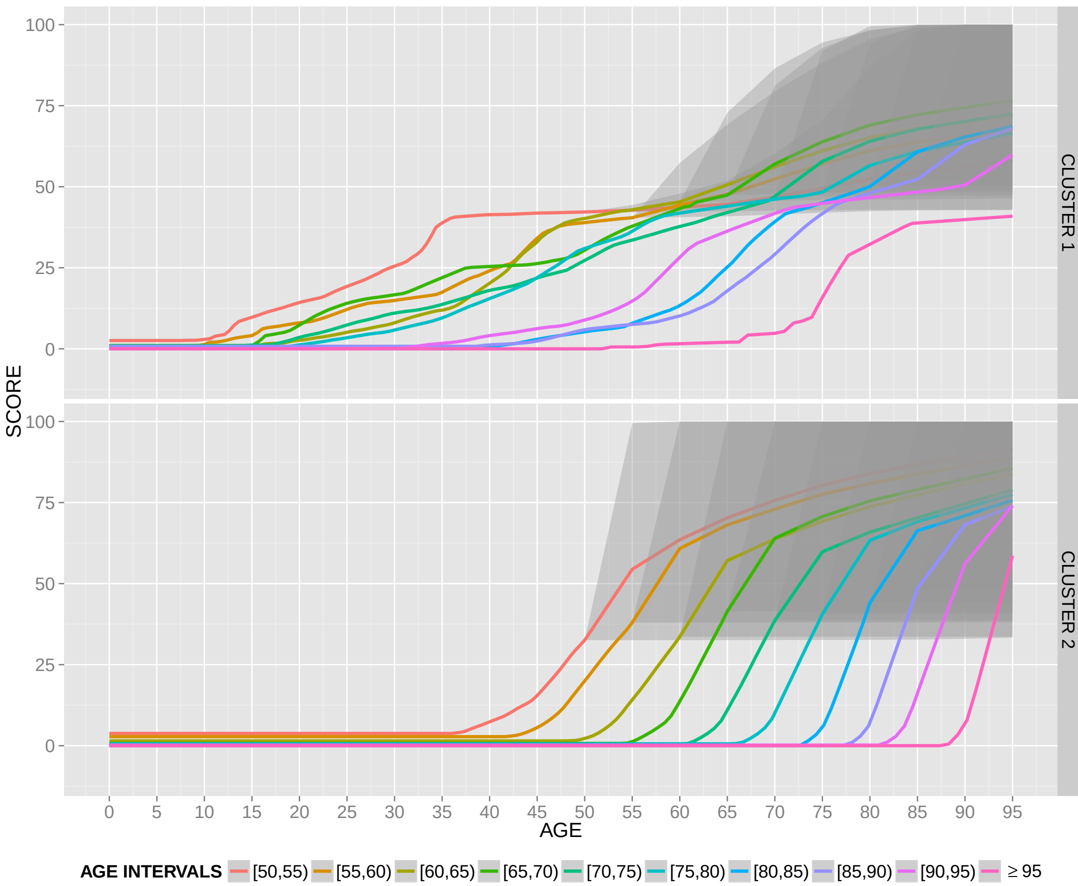

Figure 6. Observed mean trajectories and forecasted scores by cluster together with 95% confidence bands.

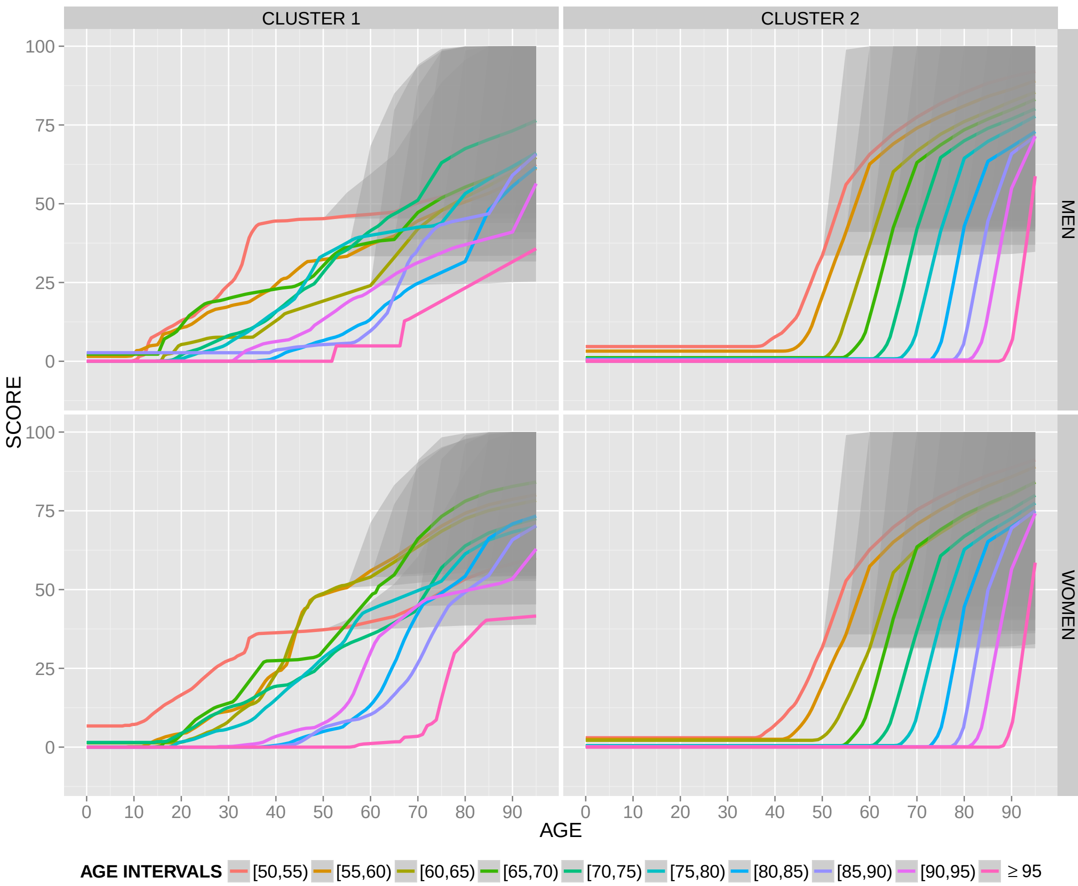

Figure 7. Observed mean trajectories and forecasted scores by sex and cluster together with 95% confidence bands.

included in the confidence intervals, these are not symmetric with respect to the forecasted score.

5. Discussion

Interest related to disability and dependency issues has increased in recent years not only in terms of their medical aspects but also, and specifically, in terms of aspects related to social and economic matters. Deteriorating physical conditions due to aging, worsened in some cases by physical and/or mental limitations, represent a huge problem for a continually aging society. This needs to be considered if we wish to find possible solutions to these issues. The time to address this problem is already at hand, and the projected age structure for the population suggests that it will become an even more serious issue in the coming decades.

Therefore, it seems appropriate to address the future situation of those affected by the contingency of dependency. The first step in this process is to obtain good statistical support to help us describe the present situation of individuals in both situations: disability and/or dependency. The reader is reminded that every dependent person suffers several types of disability (which may be specifically age related), whereas the opposite is not always true; not all disabled individuals are dependent. In this paper, we analyse a recent survey, EDAD 2008, on disability and dependency in Spain because it is the largest and most complete survey that has ever been made available in Europe to study these contingencies. However, the statistical methods and techniques applied are not specific to this study and can be applied to any other dataset of the same type. In this article, we focus on dependency rather than disability because the Spanish legislation allows us to quantify this variable (according to an official scale), which is crucial for the study of its evolution over time. Specifically, if the aim is to study dependent populations over time, there is the problem of longitudinal information being unavailable. Therefore, the only thing that can be done is to prepare a pseudopanel using the data included in EDAD 2008, as explained in Section 3. However, we are aware of the potential biases induced by this approach. Indeed, it should be noted that one of the drawbacks in the design of the questionnaires is how the events have been recorded: all the history declared by and registered for each individual is based on his/her memories of his/her life. It is then possible that there are important differences between the ages declared by interviewees as representing the beginning of a disability and the real such ages. This may be especially true for individuals at very advanced ages, who may have trouble remembering the dates, but also for general individuals, who may have tendencies to round ages. As an example, 15.6% of all disabilities recorded have been declared to begin at an age multiple of 10. This percentage becomes 24.7% for multiples of 5. Both percentages are far from expectations in the age distribution (10% and 20%, respectively). In addition to this recall bias, we note that the dependency scores were obtained by linking the disability variables included in the survey with the (later) Spanish legislation, which was not established in the exact same terms, thereby possibly introducing some impreciseness in the data.

There are several statistical challenges related to this study. First, the longitudinal trajectories obtained from the pseudo-panel cannot be seen as observations of a standard functional data sample. On the one hand, they are not defined over the same time interval, and on the other hand, they correspond to realizations not of a smooth process but of a jump process. To address the first point, we have grouped the individuals by cohorts with 5-year spans, allowing us to obtain not only the trajectories of each group defined over the same time interval but also predictions for homogeneous groups of people to which the same survival probabilities could be applied in a future. Indeed, as explained in the Introduction, the 5-year span is relatively small, and thus, we can assume that all individuals in any group are affected by death with the same intensity. In addition to the lack of information about the transition rate from non-dependency to dependency, if mortality tables for the dependent population became available, we could apply them, aggregated by 5-year intervals, to the

predicted scores for the different age groups to obtain an overall average score. This is necessary if the target is to obtain an assessment of costs for future assistance, for which we would also need some evolution of prices for services and care.
To address the fact that the trajectories are not smooth, we have used specific functional data techniques for event-type data. In particular, we have employed a time warping model for this precise type of data to estimate mean evolution trajectories in each age group. The time warping model allows one to reduce phase variability so that mean estimates are not affected by atypical trajectories with very early onset. Moreover, the estimated warping functions, which contain the temporal information of each individual, provide a way of clustering the dependency trajectories in terms of their dynamics and not only in terms of the obtained score values.
With respect to the forecasting of future dependency scores, we utilise the structure of the resulting data in the form of run-off triangles to apply actuarial forecasting techniques, such as the stochastic Chain Ladder method, which benefits from the historical information available in the older age groups to estimate the future evolution of the younger age groups.

To our knowledge, this is the first study on a nationwide scale that makes projections of the individual intensity of dependency. Indeed, available studies worldwide focus on very specific subpopulations within the dependent population and analyse panel data over very shorts period of time (see for instance Hankey *et al.* (2002), Manton *et al.* (2007) or Chou and Leung (2008)). In other cases, dependency is analysed as a categorical rather than a quantitative scale, as it is established by French and German regulations (see for instance Biessy (2015)).

Our study suggests that, at least considering EDAD 2008, two main groups of people can be identified: a group related to individuals with early-onset dependency and another group, the more numerous one, related to individuals with worsening concentrated towards the end of their life (regardless of the moment of the start of the dependency). The main difference between these groups is the rhythm at which the dependency increases over the course of the individuals' lifetimes: continuous worsening vs. a rapid decline at the end of life. This classification into two main groups may be extended to four if gender is considered for each group. According to our projections, the scores are expected to increase with age. However, for a certain age in the future, the scores are expected to be lower for those cohorts who are close to that age at present. In other words, the younger a person becomes dependent, the greater the decline observed in the future. With respect to gender, our projections suggest that women will obtain higher scores in the future in almost all cases, that is, they will present further dependency.

Disability affects health status and quality of life and is a significant public health issue worldwide (Lin and Lave, 2000). Due to the clear impact that disability and dependency have on social and health systems, it is essential to study their prevalence, causes, and effects in order to formulate a suitable plan for public health policy (Chalise *et al.*, 2008). According to the statistical results of this paper, the design of social policies for taking care of these individuals should consider the differences between such individuals. Specifically, the amount of resources allocated to dependents will depend on the age at which this contingency first appeared, the speed of progression and the gender. Concerning this last question, it is interesting to note that a higher life expectancy for women is generally accepted; therefore, their need for specialized and usually expensive care will be longer in duration. All these factors should be considered when planning the allocation of financial resources that a society might reserve for providing adequate attention to these issues not only now but also in a future characterized by an aging population in Western countries.

Acknowledgements

This work has been partially supported by Spanish grant MTM2010-17323 and MTM2014-56535-R (Spanish Ministries of Science and Innovation and Economy and Competitiveness, respectively). In addition, the research of Ana Arribas-Gil was supported by grant ECO2011-25706 of the Spanish Ministry of Science and Innovation. The authors thank the Associate Editor and two anonymous referees whose comments have helped to improve the quality of the paper.

References

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

Arribas-Gil, A. and Müller, H. (2014) Pairwise dynamic time warping for event data. *Comput. Stat. Data Anal.*, **69**, 255–268.

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

Biessy, G. (2015) Long-term care insurance: a multi-state semi-Markov model to describe the dependency process for elderly people. *Bulletin français d'actuariat*, **29**, 41–74.

Brass, W. (1971) On the scale of mortality. In *Biological Aspects of Demography* (ed. W. Brass), 69–110. London: Taylor and Francis.

Chalise, N., Saito, T. and Kai, I. (2008) Functional disability in activities of daily living and instrumental activities of daily living among nepalese newar elderly. *Public Health*, **122**, 394–396.

Chou, K. and Leung, J. (2008) Disability trends in Hong-Kong community-dwelling Chinese older adults. *Journal of Ageing and Health*, **20(4)**, 385–404.

Council of Europe (1998) Recommendation No R (98) 9 on dependence adopted by the Committee of Ministers on 18 September 1998 at the 641^{st} meeting of the Ministers' Deputies. (Available from https://wcd.coe.int/ instranet.InstraServlet?command=com.instranet.CmdBlobGet&InstranetImage=532369 &SecMode=1 &DocId=486242&Usage=2).

Dupuy, J.-F., Loubes, J.-M. and Maza, E. (2011) Non parametric estimation of the structural expectation of a stochastic increasing function. *Stat. Comput.*, **21**, 121–136.

England, P. and Verrall, R. (1999) Analytic and bootstrap estimates of prediction errors in claim reserving. *Insur. Math. Econ.*, **25**, 281–293.

Ferraty, F. and Vieu, P. (2006) *Nonparametric Functional Data Analysis: Theory and Practice*. Springer Series in Statistics. New York: Springer.

Gasser, T. and Kneip, A. (1995) Searching for structure in curve samples. *J. Am. Stat. Assoc.*, **90**, 1179–1188.

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

Hankey, G., Jamrozid, K., Broadhurst, R., Forbes, S. and Anderson, C. (2002) Long-term disability after first-even stroke and related prognostic factors in the Perth community stroke study. *Stroke*, **33**, 1034–1040.

Instituto Nacional de Estadística (2010) Encuesta sobre Discapacidad, Autonomía personal y situaciones de Dependencia. Metodología. (Available from http://www.ine.es/metodologia/t15/t1530418.pdf).

Kaufman, L. and Rousseeuw, P. J. (1990) *Finding Groups in Data: an Introduction to Cluster Analysis*. Wiley-Interscience.

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

Kneip, A. and Ramsay, J. (2008) Combining registration and fitting for functional models. *J. Am. Stat. Assoc.*, **103**, 1115–1165.

Kremer, E. (1982) IBNR claims and the two ways model of ANOVA. *Scand. Actuar. J.*, 47–55.

Lin, C. and Lave, J. (2000) Utilization under children's health insurance programs: children with vs. without chronic conditions. *J. Health Soc. Policy*, **11**, 1–14.

Liu, H. and Verrall, R. (2010) Bootstrap estimation of the predictive distributions of reserves using paid and incurred claims. *Variance*, **4**, 121–135.

Liu, X. and Müller, H. (2004) Functional convex averaging and synchronization for time-warped random curves. *J. Am. Stat. Assoc.*, **99**, 687–699.

López-Pintado, S. and Romo, J. (2009) On the concept of depth for functional data. *J. Am. Stat. Assoc.*, **104**, 718–734.

Manton, K., Lamb, V. and Gu, X. (2007) Medicare cost effects of recent U.S. disability trends in the elderly. *Journal of Ageing and Health*, **19(3)**, 359–381.

Ramsay, J. and Li, X. (1998) Curve registration. *J. R. Stat. Soc. Series B Stat. Methodol.*, **60**, 351–363.

Ramsay, J. and Silverman, B. (2005) *Functional Data Analysis*. New York: Springer.

Renshaw, A. E. and Verrall, P. (1994) A stochastic model underlying the chain ladder technique. In *Proc. XXV ASTIN Colloquium*, 45–69. Cannes, France.

Silverman, B. (1995) Incorporating parametric effects into functional principal components analysis. *J. R. Stat. Soc. Series B Stat. Methodol.*, **57**, 673–689.

Taylor, G. (1986) *Claim Reserving in Non-Life Insurance*. Amsterdam: North-Holland.

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

World Health Organization (2001) *International Classification of Functioning, Disability and Health (ICF)*. Geneva: World Health Organization.