



Article A Stochastic Markov Chain for Estimating New Entrants into Professional Pension Funds

Alessandro Fiori Maccioni ወ

Department of Economic and Business Sciences, University of Cagliari, Via Sant'Ignazio 74, 09123 Cagliari, Italy; alessandro.fiori@unica.it

Abstract: This paper presents a stochastic Markov chain model for estimating new entrants into professional orders and their related pension funds. The model considers the interactions between demographic, socio-economic and regulatory variables. The intuition behind it is that, in the medium term, trends in academic education can anticipate changes in the job market and preferences for highly skilled professions. Similarly, in the long term, fertility trends can anticipate the number of future young adults, thus influencing the overall occupational structure of employment. The model has been formalized mathematically and successfully validated by backtesting over historical data. The model's predictions have been compared with the observed data of new entrants into the Italian order of chartered accountants (CNDCEC) between 2012 and 2021. The related professional pension fund (CNPADC) has also been analyzed under the additional assumption of stochastic returns with an evaluation of the impact of future new chartered accountants on its demographic and financial evolution between 2020 and 2070.

Keywords: Markov chain; stochastic new entrants; demographic risk; professional pension funds; accounting profession

MSC: 60J20; 62P05

JEL Classification: C63; G22; J11

1. Introduction

The uncertain evolution of new contributors is a crucial component of demographic risk in social insurance schemes. Under a pay-as-you-go system, workers' contributions are used to pay the benefits of current pensioners; thus, an adequate number of new contributors is required to guarantee the financial sustainability in the long term. Under a fully funded system, new contributors affect the size of assets under management and, therefore, the scheme's investment strategy. In both cases, when the scheme is subject to a risk-based capital requirement, new entrants can affect the amount of provisions for solvency margins and the redistribution of its burden among all members.

Public pension schemes are generally open to different professional groups to prevent financial unbalances due to a decline in a specific profession. Variations in the number of contributors are influenced mainly by changes in the age structure of the population. On the other hand, private pension schemes can choose to admit only a homogeneous class of people, for example, workers with a given qualification. If they do so, the demographic risk can be more relevant because changes in the job market can have a stronger influence on the number of contributors. In this perspective, the present study addresses the evaluation of new entrants into pension schemes with educational and/or professional entry requirements.

Among the literature on stochastic models for social insurance, there are relatively few studies on new entrants. A common assumption following Dufresne (1988) is that



Citation: Fiori Maccioni, Alessandro. 2023. A Stochastic Markov Chain for Estimating New Entrants into Professional Pension Funds. *Journal* of Risk and Financial Management 16: 276. https://doi.org/10.3390/ jrfm16050276

Academic Editor: Richard J. Cebula

Received: 17 February 2023 Revised: 26 April 2023 Accepted: 9 May 2023 Published: 17 May 2023



Copyright: © 2023 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). the population of active members is stationary, such that every contributor who retires is replaced by a new individual. Bowers et al. (1986) support a closer integration between the stochastic and deterministic approaches, also in relation to new entrants which they model through population dynamics with stochastic rates of change (see also Iyer 2008, 2015). Lee (1993), Lee and Tuljapurkar (1994) and Rosati (1996) propose stochastic fertility models that can be used to assess future workers and, consequently, new contributors to national pension schemes. De Dominicis et al. (1991), Janssen and Manca (1997) and Chang and Cheng (2002) use semi-Markov models to describe the transition pattern of pension scheme members, including new entrants, on the basis of their work status. Mandl and Mazurová (1996) use stationary random sequences to model fluctuations in new entrants and investment returns. Colombo and Haberman (2005) analyze the optimal contribution strategies for minimizing the impact of stochastic new entrants on the imbalance between the expected and actual investment returns. Menoncin (2005) propose a model of asset allocation for a PAYG pension fund, where the total number of contributors follows a Brownian motion, subject to the same risk sources of asset prices. Melis and Trudda (2012) assume that the active population evolves exponentially according to an Ornstein-Uhlenbeck process. He et al. (2020) model new entrants' average salary via a Brownian motion correlated to asset prices.

While these studies are widely applicable to the broad range of pension schemes, the literature lacks a stochastic model that specifically addresses the estimation of new entrants into professional pension funds, admitting only a homogeneous population of academically qualified individuals (such as physicians, lawyers, etc.). The present study aims to fill this gap by presenting a Markov chain model for estimating new entrants into professional orders and their related pension funds. The model considers the effects of trends in populations, education choices and the appeal of a profession. The intuition behind this is that, in the medium term, trends in academic education can anticipate changes in the job market and preferences for highly skilled professions. Similarly, in the long term, fertility trends can anticipate the number of future young adults, thus influencing the overall occupational structure of employment of the population. The proposed approach is inspired by the stage-classified demographic theory based on Markov chain, described by Caswell (2001, chp. 5) and originating in the work on population dynamics by Lefkovitch (1965), who departed from the traditional matrix model (Leslie 1945) by considering stages instead of age classes of population. For a review of this theory, see also Lebreton (2005), Caswell (2009) and Vindenes et al. (2021). The main innovation of this paper is the adoption of success rates in academic and professional education, instead of biological rates typical of their use in ecological models.

The model has been validated by comparing its predictions with the observed data of new enrollments, between 2012 and 2021, into the Italian order of chartered accountants, known as CNDCEC¹. An extension of the model, considering stochastic interest rates, was applied to forecast the impact of new entrants on the demographic and financial dynamics of the related pension fund, known as CNPADC². The choice of analyzing this professional group derives from its propensity to publish data and statistics, which has fostered research (see, for example, the actuarial analyses by Attias et al. 2020; Fiori Maccioni 2011; Melis and Trudda 2012, 2010). In both cases, separate calculations were carried out for women and men to contribute to the ongoing debate on gender issues in accounting (see Haynes 2017).

The paper is organized as follows. Section 2 presents the mathematical formalization of the model. Section 3 presents its validation through backtesting over historical data on the Italian order of chartered accountants (CNDCEC). Section 4 extends the model and applies it to the related professional pension fund (CNPADC). Finally, in Section 5 conclusions are drawn.

2. The Model

The dynamics of employment in a given professional group depend on different variables pertaining to trends in population, in educational choices and in the appeal of the profession. Accordingly, the population of new entrants into the group can be estimated in subsequent steps, outlined in Table 1, by identifying progressively the populations of potential applicants for academic education; students who enroll and, subsequently, graduate in the prerequisite course for entering the profession; graduates who pursue the professional career and, consequently, join its pension fund.

Table 1. Stages for estimating new entrants into a professional order and its pension fund.

Population of potential university students \downarrow	Regarding trends in population
Enrollments in the specific field of study \downarrow Graduations in the specific field of study \downarrow	Regarding trends in education
New enrollments into the professional order \downarrow New contributors to the professional pension fund	Regarding the appeal of the profession

The Markov chain in Figure 1 models the stages that professionals have to leave behind for entering into their specific pension fund. The term $p^{i \rightarrow j}(y)$ is the probability of transitioning from state *i* to state *j* during year *y*.

The time periods h and k are deterministic parameters that represent the length for successfully completing, respectively, the course of study and the required training for entering the professional category (and, thus, joining its pension fund). They are treated deterministically, since it is assumed that they can be reliably estimated from the central tendency of the corresponding historical values.

Each potential future contributor can only be in one state at any given time and can move to a greater state exclusively after a fixed time interval—consisting of 0, h or k years—depending on the state itself.

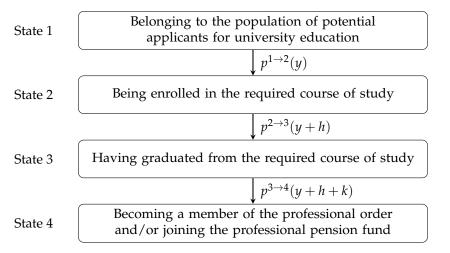


Figure 1. Markov chain for the estimation of future entrants to the pension fund of a professional category.

The stochastic process $NE_s(y)$ represents the number of new entrants into the professional order and/or its pension fund, of sex *s* at time *y*, defined as:

$$NE_{s}(y) = POP_{s}(y-h-k) \cdot p_{s}^{1\to 2}(y-h-k) \cdot p_{s}^{2\to 3}(y-k) \cdot p_{s}^{3\to 4}(y)$$
(1)

where $POP_s(y)$ represents the number of potential applicants for university education at the beginning of year *y* divided by sex while the terms $p_s^{i \to j}(y)$ represent the transition probabilities of the Markov chain in Figure 1. More precisely, $p_s^{1 \to 2}(y)$ represents the

probability that an individual belonging to $POP_s(y)$ is enrolled, in year y, as a first-year student in the prerequisite degree for entering the profession; $p_s^{2\to3}(y)$ represents the probability that an individual, who was enrolled in year y - h as a first-year student holds in year y the prerequisite degree; $p_s^{3\to4}(y)$ represents the probability that an individual, who was enrolled in year y - h as a first-year student holds in year y - h - k as a first-year student and who holds in year y - k the prerequisite degree, joins in year y the professional order and/or its pension fund.

The population $POP_s(y)$ of potential university applicants (that is, the starting population of the Markov chain) is modelled via the stochastic process:

$$POP_{s}(y) = \begin{cases} 0 & if \quad \epsilon_{s} < -\frac{\sum POP_{s,x}(y)}{\sigma_{s}} \\ \sum_{x=m}^{M} \overline{POP}_{s,x}(y) + \sigma_{s} \epsilon_{s} & if \quad \epsilon_{s} \ge -\frac{\sum POP_{s,x}(y)}{\sigma_{s}} \end{cases}$$
(2)

where $\overline{POP}_{s,x}(y)$ is the expected value of the national resident population of sex *s* and age *x* at the beginning of year *y*; σ_s is the standard deviation of $\sum POP_{s,x}(y)$; *m* and *M* are integer numbers indicating, respectively, the minimum and maximum age of the cohorts considered in the potential population; ϵ_s is a standard normal variable.

The transition probabilities of the Markov chain are modelled via the stochastic process:

$$p_{s}^{i \to j}(y) = \begin{cases} 0 & if \quad \epsilon_{s}^{i \to j} \leq -\frac{\overline{p}_{s}^{i \to j}}{\sigma_{s}^{i \to j}} \\ \overline{p}_{s}^{i \to j} + \sigma_{s}^{i \to j} \epsilon_{s}^{i \to j} & if \quad -\frac{\overline{p}_{s}^{i \to j}}{\sigma_{s}^{i \to j}} < \epsilon_{s}^{i \to j} < 1 - \frac{\overline{p}_{s}^{i \to j}}{\sigma_{s}^{i \to j}} \\ 1 & if \quad \epsilon_{s}^{i \to j} \geq 1 - \frac{\overline{p}_{s}^{i \to j}}{\sigma_{s}^{i \to j}}. \end{cases}$$
(3)

The term $\overline{p}_s^{i \to j}$ can be seen as the expected value of the transition probability from state *i* to *j*, while $\sigma_s^{i \to j}$ is its standard deviation. The term $\epsilon_s^{i \to j}$ is a standard normal variable with distribution $N \sim (0,1)$.

By Equations (2) and (3), it is possible to estimate the number of new entrants into the professional pension fund as:

$$NE_{s}(y) \approx \left[\sum_{x=m}^{M} \overline{POP}_{s,x}(y-h-k) + \sigma_{s} \epsilon_{s}\right] \cdot \left[\overline{p}_{s}^{1 \to 2}(y-h-k) + \sigma_{s}^{1 \to 2} \epsilon_{s}^{1 \to 2}\right]$$
$$\cdot \left[\overline{p}_{s}^{2 \to 3}(y-k) + \sigma_{s}^{2 \to 3} \epsilon_{s}^{2 \to 3}\right] \cdot \left[\overline{p}_{s}^{3 \to 4}(y) + \sigma_{s}^{3 \to 4} \epsilon_{s}^{3 \to 4}\right].$$
(4)

The model assumes that each standard normal variable $\epsilon_s^{i \to j}$ is dependent on its homologous variable of the opposite sex, with:

$$\epsilon_{s} = \rho \epsilon_{\bar{s}} + \sqrt{(1 - \rho^{2})}\epsilon \tag{5}$$

and:

$$\epsilon_s^{i \to j} = \rho_{ij} \epsilon_{\bar{s}}^{i \to j} + \sqrt{(1 - \rho_{ij}^2)} \epsilon_{ij}, \quad \forall \, i, j \tag{6}$$

where ρ is the correlation coefficient between ϵ_s and $\epsilon_{\bar{s}}$; ρ_{ij} is the correlation coefficient between $\epsilon_s^{i \to j}$ and $\epsilon_{\bar{s}}^{i \to j}$; ϵ and ϵ_{ij} are independent standard normal variables. On the other hand, for a given sex *s*, no additional assumption has been made regarding the dependency of the variables $\epsilon_s^{i \to j}$ among themselves, which can be the object of further research should superior data become available.

3. Model Validation

The Markov chain model has been validated by comparing its predictions with the observed data of new enrollments into the Italian order of chartered accountants in the

decade from 2012 to 2021. An additional baseline scenario, used to benchmark the model accuracy, has been calculated via the ordinary least squares (OLS) method.

This section is organized as follows. Section 3.1 introduces the professional order and its entrance regulations. Section 3.2 describes the parameter estimation of the Markov chain model and of the baseline scenario. Section 3.3 presents the results of the simulations and discusses the ability of the model to reproduce the observed data.

3.1. The CNDCEC Professional Order

The Italian Order of Chartered Accountants and Accounting Experts, known as CND-CEC, was established in 2008 after the merger of the two pre-existing orders of chartered and public accountants, which represented this profession in Italy up until then.³ Any qualified individual who wants to exercise the accounting profession in a liberal form should register with the CNDCEC and become its member. The steps for qualifying as a chartered accountant are, currently, the achievement of a master's degree (*Laurea magistrale*) in economics and business sciences, eighteen months of professional practicum (which can be partly performed while still attending university) and the passing of a state examination. The steps for qualifying as an accounting expert are the same, except for a bachelor's degree (*Laurea triennale*) instead of a master's degree.

Table 2 shows the member total and the annual new enrollments of the CNDCEC since its foundation. Accounting experts have constituted, approximately, only one out of every twenty new members, the remaining being chartered accountants. Notably, the two groups must enroll into different pension funds: chartered accountants into the CNPADC (as described in Section 4.1), and accounting experts into the CNPR⁴.

3.2. Parameter Estimation

The parameters of the Markov chain model, used for simulating new enrollments into the CNDCEC in the decade from 2012 to 2021, are shown in Table 3. The input data used for the parameter estimation were taken from the following sources. Data on Italian resident population were retrieved from the official portal of the Italian National Institute of Statistics.⁵ Data on annual enrollments⁶ and graduations⁷ in the prerequisite degrees have been retrieved from the official portal of the Italian Ministry of University. Data on the CNDCEC annual new members, shown in Table 2, were kindly provided by the Research Foundation of Italian Chartered Accountants.

The population of potential university applicants, $POP_s(y)$, was assumed to be equal to the population of Italian residents aged 19 at the beginning of each year from 2003 to 2012, divided by sex. Since this data are already known at the beginning of the forecasting period, the standard deviation of the population (that is, the parameter σ_s) was assumed to be equal to zero.

The parameters $\overline{p}_s^{1\to 2}$, $\sigma_s^{1\to 2}$ and ρ_{12} were assumed to be equal to, respectively, the average, the standard deviation and the inter-sex correlation coefficient of the ratios between the annual enrollments in the prerequisite bachelor's degrees and the population of Italian residents aged 19 at the beginning of each year. The period considered is from 2000 to 2011. Indeed, these bachelor's degrees were implemented in the Italian universities only from the academic year 2000/01 by the reform introducing the so-called '3 + 2' system.⁸

The parameter *h*, representing the time for completing the academic studies, was assumed to be equal to seven years. This value is consistent with the average time-to-degree in economics and business studies for graduates in Italian universities, shown in Figure 2, according to the annual surveys conducted between 2007 and 2021 by the inter-university consortium AlmaLaurea (2023).⁹ Accordingly, the parameters $\overline{p}_s^{2\rightarrow3}$, $\sigma_s^{2\rightarrow3}$ and ρ_{23} were assumed to be, respectively, the average, the standard deviation and the inter-sex correlation coefficient of the ratios between the annual graduations in the prerequisite master's degrees in each year from 2007 to 2011 and the annual enrollments in the corresponding prerequisite bachelor's degrees in the seventh preceding year (e.g., graduates in the calendar year 2007 were divided by first-year students in the academic year 2000/01).

		Registered Me	nbers at 1 Januar	У	
Year	Fema	iles	Mal	es	Total
2008	30,100	28%	77,399	72%	107,499
2009	32,842	30%	76,632	70%	109,474
2010	32,793	30%	77,994	70%	110,787
2011	32,528	29%	79,636	71%	112,164
2012	34,650	31%	75,585	69%	110,235
2013	34,326	30%	79,715	70%	114,041
2014	36,131	31%	78,936	69%	115,067
2015	36,187	32%	78,330	68%	114,517
2016	37,553	32%	79,799	68%	117,352
2017	38,087	32%	79,829	68%	117,916
2018	39,997	34%	78,336	66%	118,333
2019	38,914	33%	79,725	67%	118,639
2020	39,315	33%	79,460	67%	118,775
2021	39,726	33%	79,572	67%	119,298
2022	41,733	35%	78,536	65%	120,269
	N	ew Members by	y Year of Enrollm	ent	
Year	Fema	iles	Mal	es	Total
2008	1235	40%	1856	60%	3091
2009	1133	40%	1727	60%	2860
2010	1156	40%	1756	60%	2912
2011	1113	41%	1581	59%	2694
2012	1072	43%	1420	57%	2492
2013	1143	43%	1495	57%	2638
2014	1246	44%	1574	56%	2820
2015	1235	45%	1506	55%	2741
2016	1097	44%	1389	56%	2486
2017	1021	46%	1221	54%	2242
2018	985	44%	1265	56%	2250
2019	1007	45%	1220	55%	2227
2020	993	40%	1478	60%	2471
2021	1256	43%	1664	57%	2920

Table 2. Distributions of members of the CNDCEC professional order since its foundation.

Source: Research Foundation of Italian Chartered Accountants (FNC).

Table 3. Parameters for the transition probabilities of the Markov chain used in the Monte Carlo simulations of new entrants into the CNDCEC professional order in 2012–2021.

	$\overline{p}_{s}^{1 ightarrow 2}$	$\sigma_s^{1\to2}$	$ ho_{12}$	$\overline{p}_s^{2 \to 3}$	$\sigma_s^{2 ightarrow 3}$	ρ_{23}	$\overline{p}_s^{3 ightarrow 4}$	$\sigma_s^{3 ightarrow 4}$	$ ho_{34}$
Females Males	0.0699 0.0757	0.0052 0.0048	0.6312	0.3817 0.3109	0.0146 0.0260	0.5891	0.1399 0.2003	0.0183 0.0301	0.9740

The parameter *k*, representing the time for completing the professional training and qualification, was assumed to be three years, corresponding to the then-in-effect legal duration of the professional practicum. Accordingly, the parameters $\overline{p}_s^{3\to4}$, $\sigma_s^{3\to4}$ and ρ_{34} were assumed to be, respectively, the average, the standard deviation and the inter-sex correlation coefficient of the ratios between the annual enrollments into the CNDCEC in each year from 2008 to 2011 and the graduations in the prerequisite master's degrees in the third preceding year.

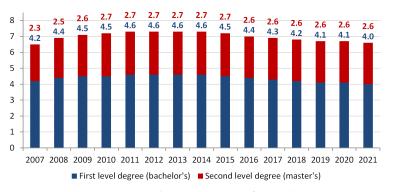


Figure 2. Average time-to-degree (in years) for graduates in economics and business sciences in Italian universities. Source: AlmaLaurea (2023).

The standard normal variables $\epsilon_s^{1\to2}$, $\epsilon_s^{2\to3}$, and $\epsilon_s^{3\to4}$ were assumed to be dependent with their homologous variable of the opposite sex (with correlation coefficients ρ_{12} , ρ_{23} and ρ_{34} , respectively) and independent between themselves.

In the baseline scenario, new enrollments into the professional order between 2012 and 2021 were estimated using ordinary least squares (OLS) regression on time series data of annual new entrants into the CNDCEC between 2008 and 2011 (see also Table 2). The equations of the regression lines obtained are:

$NE_T(y) = -113.9(y - 2011) + 3174$	for total new entrants	(7)
$NE_F(y) = -34.3(y - 2011) + 1245$	for females	(8)
$NE_M(y) = -79.6(y - 2011) + 1929$	for males.	(9)

Following Maccheroni and Nocito (2017), the goodness of fit to the historical time series was assessed using as an index the root-mean-square deviation (RMSD) calculated as:

ŀ

$$RMSD_s = \sqrt{\frac{\sum_{y=2012}^{2021} [NE_s^O(y) - NE_s^P(y)]^2}{10}},$$
(10)

where $NE_s^O(y)$ and $NE_s^P(y)$ represent, respectively, the observed and predicted new entrants into the CNDCEC in year *y* for a given sex *s*. This index was normalized by the historical mean, to facilitate the comparative assessment of the various estimates. The resulting normalized root-mean-square deviation (NRMSD) was calculated as:

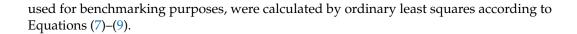
$$NRMSD_s = \frac{RMSD_s}{\overline{NE}_s^O},\tag{11}$$

where \overline{NE}_{s}^{O} represents the mean of the observed values of new entrants into the CNDCEC between 2012 and 2021 for a given sex *s*.

3.3. Results

The observed historical data of new enrollments into the CNDCEC between 2012 and 2021, divided by sex, were compared to the model predictions and to the baseline scenario. Both the annual and cumulative values were compared and the results are presented, respectively, in Figures 3 and 4.

The Markov chain model was used separately under stochastic and deterministic assumptions. In the stochastic case, new entrants were estimated via 20,000 Monte Carlo simulations based on Equations (3)–(6). In the deterministic case, new entrants—corresponding to the model expected values—were estimated via Equation (4) under the assumption that all the dispersion parameters $\sigma_s^{i \rightarrow j}$ of the random variables are to be zero. The descriptive statistics for both cases are shown in Table 4. On the other hand, the baseline predictions,



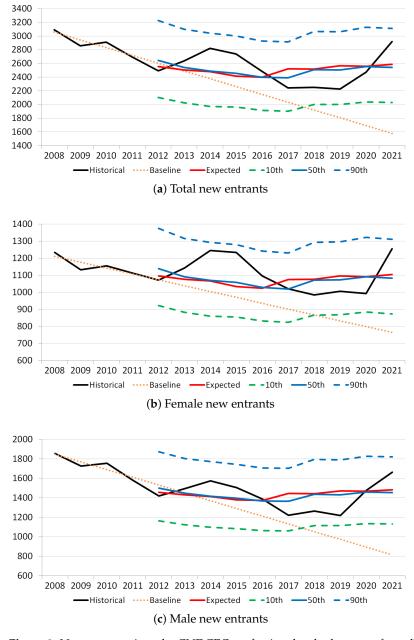


Figure 3. New entrants into the CNDCEC professional order by year of enrollment. Comparison between annual values of observed historical data; baseline predictions using OLS regression; expected predictions using the proposed Markov chain model; and 10th, 50th (median) and 90th percentiles of the frequency distributions obtained in the Monte Carlo simulations using the proposed Markov chain model.

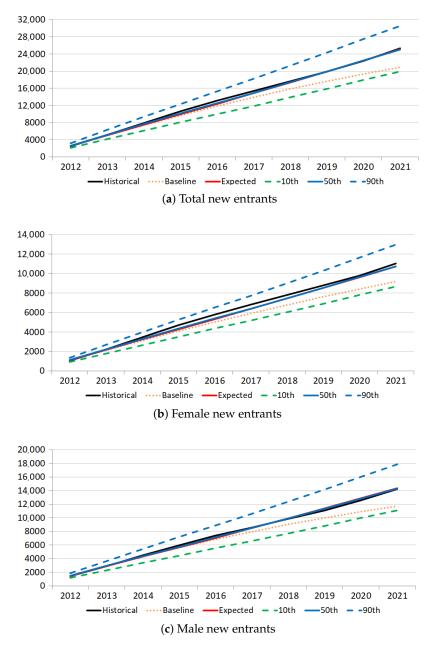


Figure 4. Time-cumulative values of new entrants into the CNDCEC professional order from 2012 onwards. Comparison between time-cumulative values of observed historical data; baseline predictions using OLS regression; expected predictions using the proposed Markov chain model; and 10th, 50th (median) and 90th percentiles of the frequency distributions obtained in the Monte Carlo simulations using the proposed Markov chain model.

The proposed model has proven to be highly accurate in replicating the observed cumulative data over the analyzed period (that is, the sum of new entrants from 2012 onwards). Namely, the cumulative expected and median values achieved an accuracy of up to 99% in predicting the cumulative historical data, while the accuracy of the baseline predictions is less than 80% (see Figure 4). These results confirm the ability of the model to capture the demographic dynamics of new entrants in the medium-long term.

The model provided an adequate level of fit also on an annual basis, as suggested by the goodness-of-fit measures shown in Table 5. For example, the normalized root-mean-squared deviation (NRMSD) between the predicted and observed total new entrants is,

respectively, 9.98% and 9.25% for the expected and median values of the model, a more accurate measure than the 22.73% of the baseline scenario.

Table 4. Statistics of the Monte Carlo experiment for new entrants into the CNDCEC order, obtained from 20,000 simulations using the proposed Markov chain model. Expected values obtained under the assumption that the dispersion parameters σ_s and $\sigma_s^{i \rightarrow j}$ in Equation (4) are equal to zero.

	Total New Entrants										
	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	
Expected	2554	2507	2480	2414	2400	2522	2519	2568	2560	2587	
Mean	2657	2556	2502	2473	2413	2400	2523	2523	2569	2559	
Median	2645	2544	2490	2457	2398	2389	2511	2507	2553	2543	
Minimum	1202	1204	993	1130	1021	939	980	1040	1110	1149	
Maximum	5054	4319	4200	4383	4078	4253	4614	4319	4483	4671	
Q1	2353	2264	2214	2193	2140	2128	2237	2240	2277	2265	
Q3	2943	2829	2773	2736	2672	2654	2793	2796	2848	2836	
Std. Dev.	438	421	415	408	397	393	415	414	424	424	
Coeff.Var.	16.5%	16.5%	16.6%	16.5%	16.5%	16.4%	16.4%	16.4%	16.5%	16.6%	
Skewness	0.21	0.21	0.20	0.23	0.22	0.21	0.17	0.19	0.22	0.25	
Kurtosis	0.10	0.11	0.03	0.13	0.12	0.03	0.06	0.03	0.07	0.17	
Female New Entrants											
	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	
Expected	1096	1076	1067	1034	1025	1075	1077	1097	1091	1104	
Mean	1144	1097	1074	1064	1034	1025	1076	1078	1098	1090	
Median	1139	1091	1070	1059	1029	1020	1071	1073	1092	1083	
Min	546	523	480	502	453	443	482	506	498	516	
Max	2017	1853	1784	1792	1838	1804	1880	1756	1792	1918	
Q1	1022	980	958	949	924	915	961	963	979	973	
Q3	1260	1208	1184	1173	1139	1129	1186	1189	1211	1203	
Std. Dev.	178	170	168	165	161	158	167	167	171	171	
Coeff.Var.	15.5%	15.5%	15.6%	15.6%	15.6%	15.4%	15.5%	15.5%	15.5%	15.7%	
Skewness	0.21	0.20	0.19	0.20	0.22	0.21	0.17	0.19	0.19	0.24	
Kurtosis	0.10	0.06	0.02	0.07	0.12	0.03	0.06	0.03	0.00	0.10	
				Male N	ew Entra	ints					
	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	
Expected	1458	1431	1413	1380	1375	1447	1442	1471	1469	1483	
Mean	1513	1459	1428	1409	1379	1375	1447	1445	1471	1469	
Median	1501	1449	1417	1398	1367	1365	1436	1432	1459	1455	
Min	616	611	509	571	551	465	498	534	551	601	
Max	3037	2678	2639	2728	2538	2756	2734	2616	2691	3054	
Q1	1319	1273	1244	1231	1204	1200	1262	1260	1284	1283	
Q3	1694	1632	1597	1574	1542	1537	1619	1618	1647	1644	
Std. Dev.	278	267	264	259	252	251	265	264	271	270	
Coeff.Var.	18.3%	18.3%	18.5%	18.4%	18.3%	18.2%	18.3%	18.2%	18.4%	18.4%	
Skewness	0.26	0.27	0.26	0.29	0.28	0.26	0.23	0.24	0.29	0.30	
Kurtosis	0.13	0.15	0.08	0.21	0.15	0.08	0.07	0.07	0.14	0.23	

The historical data in Figure 3 show a cyclic and counter-cyclic trend over the observation period. This peculiar behavior depends on the reform in the duration of the professional practicum, which has decreased from three years to eighteen months since 2013.¹⁰ Thanks to this unexpected change, trainees who fulfilled the new requirement could anticipate their entrance into the professional order. As a result, the observed new entrants exceeded the annual model predictions between 2013 and 2015, while they fell below by a similar amount between 2017 and 2019. This fact can explain the higher accuracy of the model in predicting the cumulative (rather than the annual) historical values over the analyzed period. The predictive accuracy of the model, despite the aforementioned reform of the professional practicum, does not seem to diminish significantly over time. This is particularly noteworthy since the last two years of the analyzed period were affected by the COVID-19 pandemic. On the other hand, starting from the third year of the forecast, the baseline projections exhibit a systematic underestimation of the historical data, which intensifies in 2020 and 2021 (that is, in the two pandemic years).

Table 5. Goodness-of-fit measures of the predicted distributions of new entrants into the CNDCEC professional order by year of enrollment.

Total New Entrants								
	Baseline Scenario	Expected Values	Median Values					
RMSD	575	252	234					
NRMSD	22.73%	9.98%	9.25%					
	Female	New Entrants						
	Baseline Scenario	Expected Values	Median Values					
RMSD	224	116	112					
NRMSD	20.26%	10.51%	10.12%					
	Male N	lew Entrants						
	Baseline Scenario	Expected Values	Median Values					
RMSD	362	150	136					
NRMSD	25.41%	10.56%	9.54%					

Note: Root-mean-squared deviation (RMSD) and normalized root-mean-squared deviation (NRMSD) between observed $NE_s^O(y)$ and predicted $NE_s^P(y)$.

4. An Application to the CNPADC Fund

This section aims at extending the model to estimate the impact of the stochastic variable 'new entrants' on the future demographic and financial dynamics of the national pension fund of Italian chartered accountants, known as CNPADC.

The focus is on presenting the methodology, rather than on measuring precisely all the characteristics of the fund. Therefore, the fund's regulations were simplified to improve clarity. Namely, new benefits are only awarded for old-age pensions (while disability pension and survivors' pension are not considered, except for the pre-existing ones). Unless otherwise specified, data used in the calculations are taken or derived from the actuarial balance sheet by Annibali (2020), to which the interested reader is referred for further information on the CNPADC fund.

This section is organized as follows. Section 4.1 presents the pension fund, its historical development and its main entrance regulations. Section 4.2 presents the method used for mortality projections. Section 4.3 deals with the estimation of new entrants via the Markov chain model. Sections 4.4 and 4.5 describe the cohort-component method used for estimating the demographic evolution of contributors and pensioners. Section 4.6 describes the method used for estimating the financial evolution of the fund, including the stochastic process used for modelling asset returns. Demographic and financial results are presented, respectively, in Sections 4.7 and 4.8.

4.1. The CNPADC Pension Fund

The CNPADC was established in 1963 as a non-profit public law institution to ensure social security and welfare provisions to chartered accountants and their families.¹¹ Any chartered accountant who is a member of the Italian professional order and practices the profession must enroll and pay contributions to the fund. The fund initially provided retirement benefits under a funded defined-contribution model, but then in 1986 it shifted to a pay-as-you-go defined-benefit model.¹² In 1994, all pension funds of Italian professional orders, including the CNPADC, were privatized, thus becoming autonomous non-subsidized

private institutions compliant with national regulations and supervised by the Italian ministries of Labour and Economy.¹³ In 2004, to ensure its long-term financial sustainability, the fund adopted a notional defined-contribution model, which is still in place today.

The mandatory qualification requirements in Italy for becoming a chartered accountant have significantly changed over time. Until 1996, it was necessary to have a bachelor's degree in economics and business sciences and to pass a qualifying examination. Then, a new requirement was introduced consisting of a three-year unpaid practicum in the employ of a senior chartered accountant.¹⁴ Then, bachelor graduates rushed to become chartered accountants prior to its implementation, so as to take the qualifying examination without any additional requirements. As a result, the new enrollments to the CNPADC almost doubled in 1996 with respect to the previous year, and quadrupled with respect to 1992, while they fell abruptly in 1997 and 1998. This effect is evident in Figure 5, which shows the distribution of new contributors to the fund by year of enrollment. Since 2008, the bachelor's degree and the practicum are no longer sufficient for qualifying as a chartered accountant, but a master's degree in economics and business sciences is additionally required (as well as passing the qualifying examination). Since 2013, the duration of the professional practicum was reduced from three years to eighteen months, as described in Section 3.3.

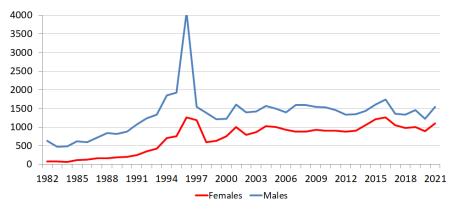


Figure 5. New entrants into the CNPADC fund by year of enrollment. Source: official data from the CNPADC fund (except years 2007–2015, which are the author's estimates).

On 31 December 2019, the demographic composition of the fund consisted of 65,261 active contributors (42,874 males and 22,387 females) and 8491 pensioners (5759 males and 2732 females); their distribution by age is shown in Figure 6. The biggest cohort of active members consisted of individuals aged 46–55 (thus born in the decade 1964–1973), who made up nearly 40% of active contributors (for comparison, those born in the previous decade made up 16.5%).

4.2. Mortality Assumptions

Experienced mortality, measured by annual life tables, can be affected by occasional and transitory effects. The most important in recent years was the outbreak of the COVID-19 pandemic. This effect is considered in the present study by analyzing past mortality trends through parabolic interpolation of death rates over a number of consecutive years.

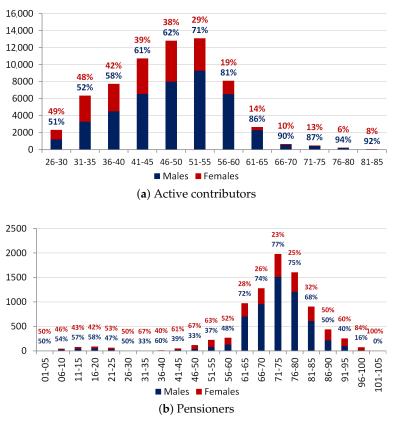


Figure 6. Distribution by age of the insured population of the CNPADC fund on 31 December 2019 (distribution by sex for each age group shown in column labels). Source: Annibali (2020).

Let $q_{sx}(y)$ be the mortality rate in year y for a cohort of sex s and age x. The adopted technique consists of interpolating a second-degree curve with a series of seven consecutive historical values:

$$q_{sx}(y-3), q_{sx}(y-2), q_{sx}(y-1), q_{sx}(y), q_{sx}(y+1), q_{sx}(y+2), q_{sx}(y+3),$$

and determining the numerical values of the parameters by the ordinary least squares method, given the abscissa values +3, +2, +1, 0, -1, -2, -3. More precisely, the technique consists of finding the scalars *a*, *b* and *c* that solve the optimization problem:

$$\min \sum_{i=-3}^{3} [ax_i^2 + bx_i + c - q_{sx}(y_i)]^2$$
(12)

The solution depends on the system of partial derivatives:

$$\begin{cases} f'_{a} = 2\sum[ax_{i}^{2} + bx_{i} + c - q_{sx}(y_{i})]x_{i}^{2} = 0\\ f'_{b} = 2\sum[ax_{i}^{2} + bx_{i} + c - q_{sx}(y_{i})]x_{i} = 0\\ f'_{c} = 2\sum[ax_{i}^{2} + bx_{i} + c - q_{sx}(y_{i})] = 0 \end{cases}$$
(13)

Recalling that there are seven terms, the above system becomes:

$$196a + 28c = 9[q_{sx}(y-3) + q_{sx}(y+3)] + 4[q_{sx}(y-2) + q_{sx}(y+2)] + q_{sx}(y-1) + q_{sx}(y+1)$$

$$28b = -3q_{sx}(y-3) - 2q_{sx}(y-2) - q_{sx}(y-1) + q_{sx}(y+1) + 2q_{sx}(y+2) + 3q_{sx}(y+3)$$

$$28a + 7c = q_{sx}(y-3) + q_{sx}(y-2) + q_{sx}(y-1) + q_{sx}(y) + q_{sx}(y+1) + q_{sx}(y+2) + q_{sx}(y+3)$$
(14)

Let $\hat{q}_{sx}(y)$ be the central value $f(x_i)$ for $x_i = 0$. Then, by solving the system, one obtains:

$$\hat{q}_{sx}(y) = \frac{1}{21} \cdot \{7q_{sx}(y) + 6[q_{sx}(y-1) + q_{sx}(y+1)] + 3[q_{sx}(y-2) + q_{sx}(y+2)] - 2[q_{sx}(y-3) + q_{sx}(y+3)]\}$$
(15)

The above formula was used to estimate a 'theoretical' mortality table for the year 2018, by parabolic interpolation of the historical values between 2015 and 2021, thus including two years of the COVID-19 pandemic. The historical values of $q_{sx}(y)$ were taken from the official mortality tables of the Italian population, published by the Italian National Institute of Statistics.¹⁵ Indeed, following Annibali (2020), it was assumed that the accountant population has a behavior similar to the whole national population.

Projected mortality tables in the forecasting period were calculated as follows. For $y \ge 2020$, one obtains:

$$q_{sx}(y) = \hat{q}_{sx}(2018) \cdot (1 + \Delta q_{sx})^{y-2018}$$
(16)

where Δq_{sx} represents the rate of annual variation in mortality. This parameter was estimated based on the geometric average of the theoretical variation over a decade:

$$\Delta q_{sx} = \sqrt[10]{\frac{\hat{q}_{sx}(2018)}{\hat{q}_{sx}(2008)}} - 1.$$
(17)

To smooth the distribution of Δq_{sx} over the ages x, the values obtained via the above equation were interpolated via parabolic arcs following the formulae described by Marsili (2001, pp. 13–14). Namely, for the ages from 5 to 91, the interpolated value $\Delta \bar{q}_{sx}$ is calculated as:

$$\Delta \bar{q}_{sx} = \frac{1}{21} \cdot \left[7\Delta q_{sx} + 6(\Delta q_{sx-1} + \Delta q_{sx+1}) + 3(\Delta q_{sx-2} + \Delta q_{sx+2}) - 2(\Delta q_{sx-3} + \Delta q_{sx+3}) \right]$$
(18)

The above interpolation formula is similar to that in Equation (15) because it is obtained via an analogous technique. On the other hand, the parabolic interpolation formulae used for the ages from 1 to 4 are:

$$\Delta \bar{q}_{s1} = \frac{1}{35} \cdot \left[31 \Delta q_{s1} + 9 \Delta q_{s2} - 3 \Delta q_{s3} - 5 \Delta q_{s4} + 3 \Delta q_{s5} \right]$$
(19)

$$\Delta \bar{q}_{s2} = \frac{1}{280} \cdot \left[79\Delta q_{s1} + 98\Delta q_{s2} + 87\Delta q_{s3} + 46\Delta q_{s4} - 25\Delta q_{s5} - 5\Delta q_{s7} \right]$$
(20)

$$\Delta \bar{q}_{s3} = \frac{1}{140} \cdot \left[-\Delta q_{s1} + 39\Delta q_{s2} + 54\Delta q_{s3} + 44\Delta q_{s4} + 9\Delta q_{s5} + 5\Delta q_{s6} - 10\Delta q_{s7} \right]$$
(21)

$$\Delta \bar{q}_{s4} = \frac{1}{140} \cdot \left[-15\Delta q_{s1} + 21\Delta q_{s2} + 42\Delta q_{s3} + 48\Delta q_{s4} + 39\Delta q_{s5} + 15\Delta q_{s6} - 10\Delta q_{s7} \right]$$
(22)

4.3. New Entrants

The Markov chain model was used to estimate the future new entrants into the CNPADC fund over the years 2020–2070, based on 20,000 Monte Carlo simulations. Table 6 presents the parameters used in the simulations. The data sources used for the parameter estimation are the following. As described in Section 3.2, data on Italian resident population were retrieved from the official portal of the Italian National Institute of Statistics, while data on annual enrollments and graduations in the prerequisite degrees were retrieved from the official portal of the Italian Ministry of University. Data on annual new members of the pension fund come from official publications by the CNPADC (2020, 2021).

Table 6. Parameters for the transition probabilities of the Markov chain used in the Monte Carlo simulations of new entrants into the CNPADC fund, years 2020–2070.

	$\overline{p}_s^{1 \to 2}$	$\sigma_s^{1\to 2}$	ρ_{12}	$\overline{p}_s^{2 \to 3}$	$\sigma_s^{2 ightarrow 3}$	ρ ₂₃	$\overline{p}_s^{3 \to 4}$	$\sigma_s^{3 ightarrow 4}$	$ ho_{34}$
Females Males	0.0681 0.0766	$0.0055 \\ 0.0064$	0.6662	0.4213 0.3615	0.0514 0.0651	0.9713	0.1094 0.1530	0.0095 0.0136	0.9280

The steps for qualifying as a chartered accountant are currently the achievement of a master's degree (*Laurea magistrale*) in economics and/or business subjects, eighteen months of professional practicum and the passing of a qualifying examination. Chartered accountants must then enroll into the pension fund once they start practicing the profession. According to these steps, the parameters of the Markov chain were estimated as follows.

The population of potential university applicants, $POP_s(y)$, was assumed to be the population of Italian residents aged 19 at 1 January of each year from 2000 to 2020. For the years from 2021 onwards, the population was estimated according to Formulaes (2) and (5), with the following assumptions: age of potential university students x = 19 (and thus m = M = 19); expected value of the population $\overline{POP}_{sx}(y)$ equal to the official forecasting of Italian residents aged 19; standard deviation of the population σ_s estimated under normality assumption by the percentile distribution of the official forecasting of the Italian residents aged 19, published by the Italian National Institute of Statistics (See note 15 above); correlation coefficient between the male and female population $\rho = 1$.

The parameters $\overline{p}_s^{1\to 2}$, $\sigma_s^{1\to 2}$ and ρ_{12} were assumed to be, respectively, the average, the standard deviation and the inter-sex correlation coefficient of the ratios between annual enrollments in the prerequisite master's degrees and the population of Italian residents aged 19 in the period 2000–2020.

The parameter *h*, representing the time taken to complete the relevant academic studies, was assumed to be seven years, in accordance with data presented in Section 3.2. Accordingly, the parameters $\overline{p}_s^{2\rightarrow3}$, $\sigma_s^{2\rightarrow3}$ and ρ_{23} were assumed to be, respectively, the average, the standard deviation and the inter-sex correlation coefficient of the ratios between the annual graduations in the prerequisite master's degrees in each year from 2007 to 2019 and the annual enrollments in the corresponding prerequisite bachelor's degrees in the seventh preceding year (for example, graduates in the calendar year 2007 were divided by first-year students in the academic year 2000/01).

As in Section 3.2, the parameter k, representing the time taken to complete the professional training and qualification, was assumed to be three years. Accordingly, the parameters $\overline{p}_s^{3\to4}$, $\sigma_s^{3\to4}$ and ρ_{34} were assumed to be, respectively, the average, the standard deviation and the inter-sex correlation coefficient of the ratios between the annual enrollments into the CNPADC fund in each year from 2017 to 2021 (see also Figure 5), and the graduations in the prerequisite master's degrees in the third preceding year.

The standard normal variables ϵ_s , $\epsilon_s^{1\rightarrow2}$, $\epsilon_s^{2\rightarrow3}$, and $\epsilon_s^{3\rightarrow4}$ are assumed to be dependent on their homologous variable of the opposite sex (with correlation coefficients ρ , ρ_{12} , ρ_{23} and ρ_{34} , respectively) and independent of one another.

4.4. Evolution of Contributors

Let the stochastic process $cPOP_{sxat}(y)$ represent the contributors alive on 31 December of year *y*, of sex *s*, age *x* and working seniority *a*. The starting population $cPOP_{sxa}(2019)$ consists of 65,261 active contributors on 31 December 2019 (42,874 males and 22,387 females; see also Figure 6).

Contributors with exactly one year of working seniority (for whom a = 1) corresponds to the annual new entrants into the pension fund:

$$cPOP_{sxa}(y) = NE_s(y) \cdot f_x, \quad \text{with } a = 1$$
 (23)

where $NE_s(y)$ represents the annual new entrants of sex *s* in year *y*, calculated according to Equation (4) and f_x represents their age distribution. Following Annibali (2020), the following was assumed:

$$f_{x} = \begin{cases} 4\%, & x = 27 \\ 7\%, & x = 28 \\ 11\%, & x = 29 \\ 13\%, & x = 30 \\ 15\%, & x = 31 \\ 15\%, & x = 32 \\ 13\%, & x = 33 \\ 11\%, & x = 34 \\ 7\%, & x = 35 \\ 4\%, & x = 36. \end{cases}$$
(24)

The CNPADC adopted the notional defined-contribution model in 2004. Then, the pre-reform contributors (with working seniority a > y - 2003) can retire with an old-age pension upon reaching the age and working seniority of, respectively, x = 62 and a = 5. On the other hand, the post-reform contributors (with working seniority $a \le y - 2003$) can retire upon reaching the age and working seniority of, respectively, x = 68 and a = 33. Accordingly, the evolution of existing contributors at the beginning of the year (for whom a > 1) was calculated as:

$$cPOP_{sxa}(y) = cPOP_{sx-1a-1}(y-1) \cdot \left[1 - \frac{q_{sx-1}(y-1) + q_{sx}(y-1)}{2}\right],$$

for: {a > y - 2003, x ≤ 62, a ≤ 5}
or {1 < a ≤ y - 2003, x ≤ 68, a ≤ 33}
(25)

4.5. Evolution of Pensioners

Let the stochastic process $pPOP_{sxat}(y)$ represent the pensioners alive on 31 December of year *y*, of sex *s*, age *x* and working seniority *a*, who have fulfilled the pension requirements in year *t*. The starting population $pPOP_{sxat}(2019)$ consists of 8491 pensioners alive on 31 December 2019 (5759 males and 2732 females; see also Figure 6); for this population it is assumed by convention that t = 0 at the beginning of the forecasting period.

The evolution of existing pensioners at the beginning of the year (for whom t < y - 1) was calculated as:

$$pPOP_{sxat}(y) = pPOP_{sx-1a-1t}(y-1) \cdot \left[1 - \frac{q_{sx-1}(y-1) + q_{sx}(y-1)}{2}\right],$$

for: $t < y - 1$ (26)

where $q_{sx}(y)$ is the probability of dying within one year from *y*.

The fund switched from the defined-benefit to the defined-contribution model in 2004; thus, pre- and post-reform contributors have different pension requirements consisting of $\{x = 62, a = 5\}$ and $\{x = 68, a = 33\}$, respectively (see also Section 4.6.4). Accordingly,

new pensioners in year *y*, who fulfilled the requirements in the previous year and for whom t = y - 1, were estimated as follows:

$$pPOP_{sxat}(y) = cPOP_{sx-1a-1}(y-1) \cdot \left[1 - \frac{q_{sx-1}(y-1) + q_{sx}(y-1)}{2}\right],$$

for: { $t = y - 1, a > y - 2003, x \ge 63, a \ge 6$ }
or { $t = y - 1, a \le y - 2003, x \ge 69, a \ge 34$ }. (27)

4.6. Financial Evolution

The net asset value of the fund at the end of each year y was modelled with the following recursive equation:

$$V(y) = V(y-1) \cdot [1+\delta_y] - E(y) + C(y) - P(y)$$
(28)

where δ_y , E(y), C(y) and P(y) represent, respectively, the rate of return, the administrative expenses, the collected contributions and the pension disbursement in the calendar year y. The initial value of the fund on 31 December 2019 is V(2019) = €9,105,140,000. Cash flows are assumed to take place at the end of the year.

4.6.1. Rate of Return

The rate of return is assumed to follow a Gaussian AR(1) process:

$$\delta_{y} = \phi_{0} + \phi_{1}\delta_{y-1} + \epsilon_{y},$$
(29)
with $\epsilon_{y} \sim N(0, \sigma^{2}), \ cov(\epsilon_{y}, \epsilon_{j}) = 0 \ \forall y \neq j, \ cov(\epsilon_{y}, y_{j}) = 0 \ \forall y \neq j.$

The errors ϵ_y are independently and identically distributed as a normal distribution with zero mean and constant variance σ^2 .

The parameters of the AR(1) process were estimated with the standard linear regression model, using the average net annual returns of Italian closed pension funds in the period 2000–2020, published by the Italian supervisory commission COVIP (2021). Accordingly, it was assumed that $\phi_0 = 3.412\%$, $\phi_1 = -11.614\%$ and $\sigma^2 = 0.159\%$. The process is stationary since it results in $|\phi_1| < 1$.

4.6.2. Administrative Expenses

The administrative expenses were estimated to be equal to their initial value in 2019, adjusted for the annual rate of inflation i_y :

...

$$E(y) = E(2019) \cdot \prod_{k=2020}^{y} (1+i_k)$$
(30)

where $E(2019) = \in 18,410,000$ and:

$$i_y = \begin{cases} 0.50\%, & 2020 \le y \le 2023\\ 1.15\%, & 2024 \le y \le 2025\\ 1.94\%, & 2026 \le y \le 2030\\ 2.00\%, & y \ge 2031. \end{cases}$$
(31)

4.6.3. Contributions

Active members of the CNPADC should pay both an income-based and a revenuebased contribution. The first is called *contributo soggettivo* and funds the notional definedcontribution accounts of individuals. The second is called *contributo integrativo*; it has a solidarity function and does not influence the amount of the individual pension.

The total contributions collected each year by the fund are:

$$C(y) = C_{INC}(y) + C_{REV}(y)$$
(32)

where $C_{INC}(y)$ and $C_{REV}(y)$ represent, respectively, the income-based and the revenuebased contributions paid in each year y by the cohort of contributors $cPOP_{sxa}(y)$.

The income-based contributions are calculated as:

$$C_{INC}(y) = \sum_{s} \sum_{x} \sum_{a} 0.127 \cdot INC_{sxa}(y) \cdot cPOP_{sxa}(y)$$
(33)

where $INC_{sxa}(y)$ is the average nominal income of the cohort $cPOP_{sxa}(y)$ in the year y, and the parameter 0.127 is the average contribution rate choosen by CNPADC contributors (the minimum mandatory rate is 12% but contributors are allowed to pay more to increase their future pensions, see also Annibali 2020, p. 54).

The income-based contributions were computed separately for each cohort $cPOP_{sxa}(y)$ and for each year since the introduction of the defined-contribution model in 2004, in order to assess the balance of notional individual accounts at the beginning and throughout the forecasting period.

The revenue-based contributions were calculated as:

$$C_{REV}(y) = \sum_{s} \sum_{x} \sum_{a} 0.04 \cdot REV_{sxa}(y) \cdot cPOP_{sxa}(y)$$
(34)

where $REV_{sxa}(y)$ is the average nominal revenue of the cohort $cPOP_{sxa}(y)$ in the year y, and 0.04 is the standard revenue-based contribution rate.

The nominal values of income and revenue correspond, respectively, to those of year 2019, adjusted to inflation:

$$INC_{sxa}(y) = \begin{cases} INC_{sxa}(2019) \cdot \prod_{k=y+1}^{2019} (1+i_k)^{-1}, & 2004 \le y \le 2018 \\ INC_{sxa}(2019), & y = 2019 \\ INC_{sxa}(2019) \cdot \prod_{k=2020}^{y} (1+i_k), & y \ge 2020, \end{cases}$$
(35)

and:

$$REV_{sxa}(y) = REV_{sxa}(2019) \cdot \prod_{k=2020}^{y} (1+i_k), \quad y \ge 2020.$$
 (36)

The inflation rates i_y for years 2020–2070 are indicated in formula (31), while those for years 2005–2019 were assumed to be equal to the historical annual inflation rates in Italy published by the Italian National Institute of Statistics.¹⁶

4.6.4. Pensions

The total pensions paid each year by the fund are:

$$P(y) = \sum_{s} \sum_{x} \sum_{a} \sum_{t} p_{sxat}(y) \cdot pPOP_{sxat}(y)$$
(37)

where $p_{sxat}(y)$ is the pension paid in year y to an individual of sex s, age x and seniority a, who fulfilled the pension requirements in year t, with:

$$p_{sxat}(y) = \overline{p}_{sxat} \cdot \prod_{k=2020}^{y} (1+i_k), \quad y > t.$$
 (38)

The term \overline{p}_{sxat} represents the initial pension calculated at retirement time *t* (note that the first pension payment occurs at time y = t + 1, since cash flows take place at the end of the year). It has two alternative methods of calculation, depending on whether the retiree

enrolled after the notional defined-contribution reform in 2004. If so, the initial pension is calculated according to the defined-contribution method:

$$\overline{p}_{sxat} = N_{sxa}(t) \cdot \frac{\kappa_x + \kappa_{x+1}}{2}, \quad \text{for } a \le t - 2003$$
(39)

where $N_{sxa}(t)$ is the balance of the individual notional account at retirement, and κ_x is the age-related conversion coefficient.

On the other hand, the accrued rights of pre-reform contributors were protected by providing them with retirement benefits computed according to the pro rata method. Accordingly, their pension consists of two shares calculated proportionally to the years of permanence under the defined-benefit and the defined-contribution models:

$$\overline{p}_{sxat} = \frac{\kappa_x + \kappa_{x+1}}{2} N_{sxa}(t) + 0.02 \sum_{j=t-a+1}^{2003} \prod_{k=y+1}^{t} \left[INC_{s,x-t+j,a-t+j}(1+i_k) \right],$$

for $a > t - 2003.$ (40)

The first summand of the above equation corresponds to the defined-contribution component of the initial pension, calculated as in Formula (39). The second summand corresponds to the defined-benefit component, which is 2% of the average annual income in the pre-reform years, for each year of pre-reform contribution (or, in equivalent terms, 2% of the total pre-reform income).

4.7. Demographic Results

The demographic and financial dynamics of the CNPADC fund were estimated over the years 2020–2070 in a standard Monte Carlo experiment with 20,000 simulations. New entrants were estimated using the Markov chain model proposed in Section 2 with the parameters shown in Table 6. Figure 7 shows the percentiles of their frequency distributions obtained in the experiment, while Table 7 shows their corresponding expected values, calculated deterministically by setting to zero all the dispersion parameters σ_s and $\sigma_s^{i \rightarrow j}$ in Equation (4).

Table 7. Expected new entrants into the CNPADC fund, years 2020–2070.

Year	Fem.	Mal.	Tot.	Year	Fem.	Mal.	Tot.	Year	Fem.	Mal.	Tot.
2020	917	1320	2237	2037	902	1314	2216	2054	726	1078	1804
2021	928	1332	2260	2038	891	1307	2198	2055	731	1083	1814
2022	899	1292	2191	2039	879	1291	2170	2056	731	1084	1815
2023	879	1267	2146	2040	858	1261	2119	2057	734	1086	1820
2024	866	1255	2121	2041	845	1241	2086	2058	735	1087	1822
2025	868	1263	2131	2042	812	1194	2006	2059	737	1090	1827
2026	867	1276	2143	2043	799	1179	1978	2060	739	1091	1830
2027	869	1315	2184	2044	776	1147	1923	2061	740	1093	1833
2028	872	1303	2175	2045	758	1120	1878	2062	742	1094	1836
2029	886	1307	2193	2046	730	1088	1818	2063	743	1096	1839
2030	882	1279	2161	2047	702	1050	1752	2064	744	1096	1840
2031	877	1270	2147	2048	707	1057	1764	2065	745	1096	1841
2032	885	1284	2169	2049	710	1060	1770	2066	744	1096	1840
2033	894	1299	2193	2050	713	1063	1776	2067	744	1094	1838
2034	884	1295	2179	2051	717	1067	1784	2068	742	1091	1833
2035	893	1307	2200	2052	720	1070	1790	2069	740	1087	1827
2036	895	1308	2203	2053	723	1074	1797	2070	737	1083	1820

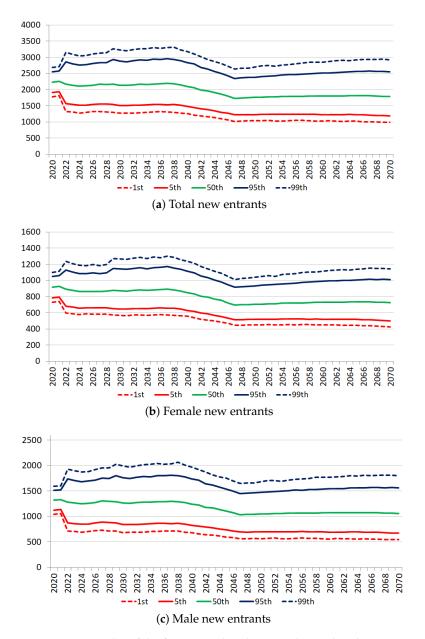


Figure 7. Percentiles of the frequency distributions obtained in the Monte Carlo simulations of new entrants into the CNPADC fund, years 2020–2070.

The simulations reveal a constant trend in the number of annual new entrants in the period 2020–2037, followed by a gradual decrease of around 20% until year 2047, and by a slow recovery in the remaining horizon.

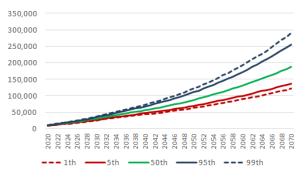
Females are estimated to be 40–41% of annual new entrants over the whole forecasting period. This value is lower than the share of females who joined annually the CNDCEC order since its foundation, which ranged between 40 and 46% (see Table 2). The difference can be explained by a lower propensity of women to practice the profession in a liberal form compared to men. Indeed, chartered accountants working as employees must subscribe to the Italian national institute for social security (INPS), rather than to the CNPADC. This can be interpreted as a hint for the existence of a 'glass ceiling' in the Italian accounting profession, in line with studies on other countries (see Cohen et al. 2020).

An alternative explanation can be that female new members of the CNDCEC are more likely than men to practice in a liberal form as 'accounting experts', thus joining the CNPR pension fund (as described in Section 3.1). However, according to data from Micocci (2022, p. 13), this statement holds true only for those who joined the CNPR before the age of thirty,

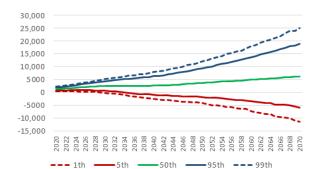
while the contrary holds for the other cohorts. However, the relatively small population involved is not enough to justify the lower-than-expected number of female new entrants into the CNPADC fund, thus supporting the first interpretation.

4.8. Financial Results

Figure 8 shows the percentiles of the frequency distributions of the fund value, the total balance and the pension balance, while Table 8 summarizes the statistical properties of these distributions. The simulations reveal an upward trend of the fund value over the whole forecasting period, driven by a pension balance—the difference between total contributions received and total pensions paid—which is constantly positive with 99% confidence level.



(a) Fund value



(b) Total balance

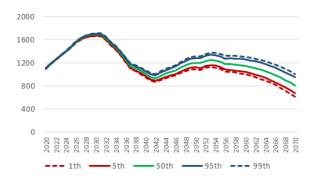




Figure 8. Percentiles of the frequency distributions obtained in the Monte Carlo simulations of the financial dynamics of the CNPADC fund, years 2020–2070. Nominal values in million euros.

The analysis of the pension balance reveals two insights. First, for any given year, its statistical dispersion is significantly lower than that of the total balance, as measured by their standard deviations and coefficients of variation. This result suggests, under the

assumptions of the model, a higher exposition of the fund to the risk of investment returns rather than to the risk of new entrants.

The second insight is about the joint population dynamics of future new entrants and of pre-existing contributors. The pension balance tends to follow two cycles, clearly visible in Figure 8. In the first, the balance tends to increase by about 53% between 2020 and 2030, and to decrease by about 45% up to 2042. The second cycle appears after 2042 and has qualitatively similar trends but of lesser magnitude. It can be inferred that the first increasing phase of the pension balance is largely driven by new entrants, while the following decreasing phase is due to a surge in retirements that occurs when the biggest cohorts of pre-existing contributors meet the pension requirements. This group consists of individuals born in the decade 1964–1973, creating the 'demographic hump' visible in Figure 6.

Table 8. Statistics of the Monte Carlo simulations of the financial dynamics of the CNPADC fund, years 2020–2070. Monetary values in thousand euros.

Fund Value										
	31 December 2020	31 December 2030	31 December 2040	31 December 2050	31 December 2060	31 December 2070				
Mean	10,465,535	30,781,797	56,755,951	88,474,082	132,921,607	190,359,559				
Median	10,469,071	30,685,178	56,389,473	87,548,789	130,975,445	187,380,617				
Minimum	9,043,565	22,151,789	36,511,753	51,510,232	73,064,273	96,183,623				
Maximum	12,006,282	42,087,287	89,061,040	152,796,821	247,860,539	406,040,488				
Q1	10,220,285	29,046,664	52,219,749	79,564,874	117,442,179	164,508,606				
Q3	10,709,630	32,423,164	60,799,819	96,281,546	146,381,647	212,180,472				
Std. Dev.	363,995	2,491,916	6,416,604	12,625,957	22,049,905	36,069,026				
Coeff.Var.	3.48%	8.10%	11.31%	14.27%	16.59%	18.95%				
Skewness	-0.017	0.225	0.356	0.451	0.523	0.578				
Kurtosis	-0.036	0.044	0.222	0.343	0.448	0.605				
Total Balance										
	31 December 2020	31 December 2030	31 December 2040	31 December 2050	31 December 2060	31 December 2070				
Mean	1,360,390	2,526,249	2,667,188	3,766,217	5,075,835	6,391,544				
Median	1,363,883	2,535,672	2,649,323	3,718,917	5,012,783	6,268,094				
Minimum	-61,575	-1,968,612	-6,494,394	-11,211,975	-17,111,129	-25,955,836				
Maximum	2,901,142	7,236,884	11,974,155	18,324,708	28,185,383	50,976,979				
Q1	1,115,145	1,755,227	1,198,679	1,488,966	1,634,153	1,382,552				
Q3	1,604,490	3,291,193	4,105,184	6,023,213	8,481,429	11,195,194				
Std. Dev.	363,995	1,137,049	2,184,903	3,422,921	5,209,454	7,583,524				
Coeff.Var.	26.76%	45.01%	81.92%	90.88%	102.63%	118.65%				
Skewness	-0.017	0.002	0.053	0.081	0.074	0.121				
Kurtosis	0.036	0.019	0.172	0.222	0.336	0.402				
			Pension Balar	ice						
	31 December 2020	31 December 2030	31 December 2040	31 December 2050	31 December 2060	31 December 2070				
Mean	1,105,065	1,690,800	1,015,567	1,205,474	1,151,399	810,395				
Median	1,105,070	1,690,750	1,015,623	1,205,399	1,150,332	809,279				
Minimum	1,102,064	1,655,495	916,977	1,032,131	921,384	502,193				
Maximum	1,108,428	1,729,781	1,131,774	1,425,088	1,429,819	1,134,733				
Q1	1,104,505	1,684,154	997,888	1,172,422	1,106,533	753,614				
Q3	1,105,618	1,697,382	1,032,917	1,238,137	1,194,992	867,278				
Std. Dev.	825	9878	25,778	48,802	65,559	84,502				
Coeff.Var.	0.07%	0.58%	2.54%	4.05%	5.69%	10.43%				
Skewness	0.007	0.055	0.035	0.029	0.077	0.038				
Kurtosis	0.013	0.031	-0.017	-0.022	-0.018	-0.006				

5. Conclusions

This study contributes to the literature on stochastic models for new entrants in social insurance schemes by addressing the case of pension funds with professional entry requirements. The paper proposed a Markov chain model for estimating new entrants, with stochastic coefficients related to trends in population, educational choices and appeal of the profession. The topic addressed is relevant, since the uncertainty in new entrants can pose a risk to pension funds. This is especially true for those financed by a pay-as-you-go system, where contributions of current active members are used to pay benefits of pensioners. Thus, a lacking number of new entrants can lead to financial imbalances and threaten the long-term financial sustainability of the system.

The model was successfully validated through backtesting over historical data of new entrants into the Italian order of chartered accountants (known as CNDCEC). The results, presented in Section 3.3, demonstrate the ability of the model to predict new entrants into a professional group, both in the short and in the long run. As shown in Figure 4, the model reached an accuracy of up to 99% in predicting the cumulative value of observed new entrants into the CNDCEC between 2012 and 2021, compared to the accuracy of the baseline model (using OLS regression) of less than 80%. This result is particularly noteworthy because it was obtained despite significant unexpected events that occurred during this interval, namely, a reform of the professional practicum and the outbreak of the COVID-19 pandemic.

The paper also presents an application to the pension fund of Italian chartered accountants (known as CNPADC). The model was expanded to account for the demographic and financial evolution of the fund, under the additional assumption of stochastic returns. The estimates were made for the years 2020-2070 by Monte Carlo simulations, and the results are presented, respectively, in Sections 4.7 and 4.8. Although the focus was on presenting the methodology, three noteworthy findings were observed. First, a comparative analysis revealed that the annual share of female new entrants into the CNPADC tended to be lower than that into the CNDCEC order. The proposed explanation was a lower propensity of female chartered accountants (with respect to males) to practice the profession in a liberal form. This may suggest the existence, within the Italian accounting profession, of a 'glass ceiling' effect already observed in other countries (see Cohen et al. 2020). Second, under the assumptions of the model, the pension balance was significantly more clustered than the total balance, thus suggesting a higher exposition of the fund to the risk of investment returns rather than to the risk of new entrants. This conclusion is supported by the relative stability in the population of Italian chartered accountants, as measured via the proposed Markov chain model. Third and finally, the pension balance was found to follow a cyclic pattern in which the first increasing phase was driven by new entrants, while the following decreasing phase was due to to the retirement of the bulk of pre-existing contributors, born in the decade 1964–1973. These joint population dynamics and their financial impact on the fund were captured and properly quantified via the model.

The present study makes a significant contribution to the existing literature by providing a new method for actuaries to assess future new entrants into professional pension funds. The traditional methods are mostly based on either heuristic assumptions (for example, a constant population structure), regression techniques or rate of growth of population. This paper brings a different approach inspired by the stage-classified demographic theory based on Markov chain, typical of ecological and population studies, with the innovation of replacing its traditional biological rates with success rates in academic and professional education. The proposed method permits to investigate, more effectively than its traditional alternatives, the structural connections between demographic, socio-economic and regulatory variables affecting the employment dynamics of professional groups. Although the primary purpose of the model is for actuarial valuations, it can also be used to support policy analysis of employment and labour participation. Its main shortcoming is that it requires relatively long time series of a different nature: national population, educational statistics and professional enrollments. This amount of data may be unavailable for various professions or countries. Future developments of this research will aim at the joint stochastic modelling of new entrants, investment returns and future mortality rates, thus adding longevity risk to the topics addressed in this paper.

Funding: This work was supported by the Fondazione di Sardegna (research grant with CUP number F73C22001420007, Bando FdS 2021). The funding sponsors had no role in the design of the study; in the collection, analyses or interpretation of data; in the writing of the manuscript; and in the decision to publish the results.

Data Availability Statement: The datasets generated and/or analyzed during the study are available from the author on reasonable request.

Acknowledgments: The author wishes to thank the Research Foundation of Italian Chartered Accountants, Tommaso Di Nardo, Lorenzo Magrassi, Gianluca Scardocci and Tiziana Mastrogiacomo for useful data and information. The author thanks also Katia Colaneri, Giovanni Masala, Laura Casula, Aldona Skučaitė, Kristina Šutienė and the two anonymous reviewers for useful comments and suggestions. Any errors are the responsibility of the author.

Conflicts of Interest: The author declares no conflict of interest.

Notes

- ¹ Acronym for Consiglio Nazionale dei Dottori Commercialisti e degli Esperti Contabili.
- ² Acronym for Cassa Nazionale di Previdenza e Assistenza dei Dottori Commercialisti.
- ³ The merger was established by the Italian Legislative Decree 139/2005 entered into force on 1 January 2008. The pre-existing orders were that of *Dottori Commercialisti* (chartered accountants) and that of *Ragionieri* (public accountants). For further details, see also Costa et al. (2013), Veneziani et al. (2016), Lucianelli and Citro (2018).
- ⁴ Acronym for Cassa Nazionale di Previdenza Ragionieri.
- ⁵ Source: https://demo.istat.it/app/?i=RIC&l=it (last retrieved 15 May 2023).
- ⁶ Source: http://dati.ustat.miur.it/dataset/immatricolati (last retrieved 15 May 2023).
- ⁷ Source: http://dati.ustat.miur.it/dataset/laureati (last retrieved 15 May 2023).
- ⁸ See the Italian Ministerial Decree 509/1999.
- ⁹ The average time-to-degree exceeds the legal duration of '3 + 2' years because of those students who are unable to complete exams within the set time period. For more on this issue, see Aina et al. (2011) and Licursi (2018).
- ¹⁰ See Italian Presidential Decree 137/2012.
- ¹¹ See Italian Law 100/1963.
- ¹² See Italian Law 21/1986.
- ¹³ See Italian Legislative Decree 509/1994.
- ¹⁴ The reform was approved in principle by Italian Law 206/1992 and implemented by Ministerial Decree 327/1995.
- ¹⁵ Source: http://dati.istat.it/Index.aspx? (last retrieved 26 April 2023).
- ¹⁶ Source: https://www.istat.it/it/prezzi?dati (last retrieved 26 April 2023).

References

- Aina, Carmen, Eliana Baici, and Giorgia Casalone. 2011. Time to degree: Students' abilities, university characteristics or something else? Evidence from Italy. *Education Economics* 19: 311–25. [CrossRef]
- AlmaLaurea. 2023. Profilo dei Laureati. Available online: https://www2.almalaurea.it/cgi-php/universita/statistiche/tendine.php? config=profilo (accessed on 26 April 2023).
- Annibali, Antonio. 2020. Bilancio Tecnico Della Cassa Nazionale di Previdenza e Assistenza a Favore dei Dottori Commercialisti. Proiezioni: 2020–2069. Rome, December. Available online: https://www.cnpadc.it/sites/default/files/allegati-bilancio/2021/0 4/BT%202020-2069.pdf (accessed on 26 April 2023).
- Attias, Anna, Simona Ciavalini, Carla Morrone, and Daniela Saitta. 2020. An actuarial mathematical model for a new pension philosophy. An application to the accountant pension fund. *Pure Mathematics and Applications* 28: 1–32. [CrossRef]
- Bowers, Newton L., Hans U. Gerber, James C. Hickman, Donald A. Jones, and Cecil J. Nesbitt. 1986. *Actuarial Mathematics*. Schaumburg: The Society of Actuaries.
- Caswell, Hal. 2001. Matrix Population Models. Construction, Analysis and Interpretation. Sunderland: Sinauer Associates.
- Caswell, Hal. 2009. Stage, age and individual stochasticity in demography. Oikos 118: 1763-82. [CrossRef]
- Chang, Shih-Chieh, and Hsin-Yi Cheng. 2002. Pension valuation under uncertainties: Implementation of a stochastic and dynamic monitoring system. *Journal of Risk and Insurance* 69: 171–92. [CrossRef]

- CNPADC. 2020. CDC Reputational Report 2020. Rome. Available online: https://www.cnpadc.it/la-cassa/i-numeri-della-cassa/ anno-2020.html (accessed on 26 April 2023).
- CNPADC. 2021. CDC Reputational Report 2021. Rome. Available online: https://www.cnpadc.it/la-cassa/i-numeri-della-cassa/ anno-2021.html (accessed on 26 April 2023).
- Cohen, Jeffrey R., Derek W. Dalton, Lori L. Holder-Webb, and Jeffrey J. McMillan. 2020. An analysis of glass ceiling perceptions in the accounting profession. *Journal of Business Ethics* 164: 17–38. [CrossRef]
- Colombo, Luigi, and Steven Haberman. 2005. Optimal contributions in a defined benefit pension scheme with stochastic new entrants. *Insurance: Mathematics and Economics* 37: 335–54. [CrossRef]
- Costa, Massimo, Patrizia Torrecchia, and Alan Sangster. 2013. Accounting education for accounting educators evidence from Italy -The case of SISSIS/TFA (2002–2013). *Procedia-Social and Behavioral Sciences* 106: 295–307. [CrossRef]
- COVIP. 2021. Relazione per l'anno 2020. Rome, June. Available online: https://www.covip.it/la-covip-e-la-sua-attivita/pubblicazionistatistiche/relazioni-annuali/relazione-annuale-2020 (accessed on 26 April 2023).
- De Dominicis, Rodolfo, Raimondo Manca, and Laura Granata. 1991. The dynamics of pension funds in a stochastic environment. *Scandinavian Actuarial Journal* 1991: 118–28. [CrossRef]
- Dufresne, Daniel. 1988. Movements of pension contributions and fund levels when rates of return are random. *Journal of the Institute of Actuaries* 115: 535–44. [CrossRef]
- Fiori Maccioni, Alessandro. 2011. A stochastic model for the analysis of demographic risk in pay-as-you-go pension funds. *arXiv* arXiv:1106.5081.
- Haynes, Kathryn. 2017. Accounting as gendering and gendered: A review of 25 years of critical accounting research on gender. *Critical Perspectives on Accounting* 43: 110–24. [CrossRef]
- He, Lin, Zongxia Liang, and Fengyi Yuan. 2020. Optimal DB-PAYGO pension management towards a habitual contribution rate. *Insurance: Mathematics and Economics* 94: 125–41. [CrossRef]
- Iyer, Subramaniam. 2008. Stochastic actuarial modelling of a defined-benefit social security pension scheme: An analytical approach. *Annals of Actuarial Science* 3: 127–85. [CrossRef]
- Iyer, Subramaniam. 2015. Social insurance pension schemes: Stochastic actuarial valuation using an analytical model. *Asia-Pacific Journal of Risk and Insurance* 9: 261–301. [CrossRef]
- Janssen, Jacques, and Raimondo Manca. 1997. A realistic non-homogeneous stochastic pension fund model on scenario basis. Scandinavian Actuarial Journal 1997: 113–37. [CrossRef]
- Lebreton, Jean-Dominique. 2005. Age, stages and the role of generation time in matrix models. *Ecological Modelling* 188: 22–29. [CrossRef]
- Lee, Ronald D. 1993. Modeling and forecasting the time series of US fertility: Age distribution, range and ultimate level. *International Journal of Forecasting* 9: 187–202. [CrossRef]
- Lee, Ronald D., and Shripad Tuljapurkar. 1994. Stochastic population forecasts for the United States: Beyond high, medium and low. *Journal of the American Statistical Association* 89: 1175–89. [CrossRef]
- Lefkovitch, Leonard P. 1965. The study of population growth in organisms grouped by stages. Biometrics 21: 1–18. [CrossRef]
- Leslie, Patrick H. 1945. On the use of matrices in certain population mathematics. Biometrika 33: 183–212. [CrossRef]
- Licursi, Sabina. 2018. Graduation after time and peer tutoring. Italian Journal of Sociology of Education 10: 90–109.
- Lucianelli, Giovanna, and Francesca Citro. 2018. Accounting education for professional accountants: evidence from Italy. *International Journal of Business and Management* 13: 1–15. [CrossRef]
- Maccheroni, Carlo, and Samuel Nocito. 2017. Backtesting the Lee-Carter and the Cairns-Blake-Dowd stochastic mortality models on Italian death rates. *Risks* 5: 34. [CrossRef]
- Mandl, Petr, and Lucie Mazurová. 1996. Harmonic analysis of pension funding methods. *Insurance: Mathematics and Economics* 17: 203–214. [CrossRef]
- Marsili, Marco. 2001. Perequazione dei tassi grezzi di mortalità. In *Tavole Della Popolazione Italiana per Provincia e Regione di Residenza*. Informazioni: Istat, pp. 13–14.
- Melis, Roberta, and Alessandro Trudda. 2010. Demographic risk indicators in pay-as-you-go pension funds. *Problems and Perspectives in Management* 8: 117–26.
- Melis, Roberta, and Alessandro Trudda. 2012. Solvency indicators for partially unfunded pension funds. *Investment Management and Financial Innovations* 9: 71–77.
- Menoncin, Francesco. 2005. Cyclical risk exposure of pension funds: A theoretical framework. *Insurance: Mathematics and Economics* 36: 469–84.
- Micocci, Marco. 2022. Bilancio Tecnico Attuariale al 31.12.2021 CNPR–Cassa nazionale di Previdenza e Assistenza a favore dei Ragionieri e Periti Commerciali. Rome, 18 October 2022. Available online: https://www.cassaragionieri.it/bilanci (accessed on 26 April 2023).
- Rosati, Furio Camillo. 1996. Social security in a non-altruistic model with uncertainty and endogenous fertility. *Journal of Public Economics* 60: 283–94. [CrossRef]

Veneziani, Monica, Claudio Teodori, and Giulia Bendotti. 2016. The role of university in the education of professional accountants: degree of IES 2 application. *Journal of Modern Accounting and Auditing* 12: 127–46. [CrossRef]

Vindenes, Yngvild, Christie Le Coeur, and Hal Caswell. 2021. Introduction to matrix population models. In *Demographic Methods across the Tree of Life*. New York: Oxford University Press, pp. 163–81.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.