Extended Abstract

Projecting Gender Inequalities in Work-Life Trajectories and Retirement Expectations at Age 55: Estimates for Germany

Linda Vecgaile (The Max Planck Institute for Demographic Research), **Emilio Zagheni** (The Max Planck Institute for Demographic Research), **Luca Badolato** (The Ohio State University, Columbus, USA); **Luiz Felipe Vecchietti** (Institute for Basic Science, Daejeon, Korea, South)

Research Question

This extended abstract presents a study aimed at predicting work-life trajectories, retirement timing, and the nuanced gendered experiences that accompany the aging process, utilizing cutting-edge machine learning techniques. In response to demographic changes, including an aging population and decreasing working-age individuals, governments worldwide are grappling with the need to adapt pension systems and labor markets. Increasing the retirement age is a common policy response, but it must be implemented carefully to avoid prolonged involuntary unemployment among vulnerable groups.

Older individuals, both men and women, encounter distinctive challenges along their life course. These encompass the necessity to acquire new skills, navigate health-related issues, confront age-based discrimination, and overcome other obstacles to remain employed. In the context of aging and retirement, gendered experiences play a pivotal role. Prolonged unemployment, exacerbated by gendered factors, such as disparities in career opportunities, caregiving responsibilities, and wage gaps, can ultimately steer individuals toward early retirement, as re-entering the workforce becomes increasingly difficult.

Traditional research in this area has primarily focused on predicting work-life expectancy, providing essential baseline data for policy decisions. In this study, we extend this research by adopting a life course approach, recognizing that retirement is a complex, gradual process marked by multiple transitions. Moreover, it takes into account the unique gendered experiences that shape individuals' trajectories as they progress through the different stages of their working lives and aging processes. This holistic perspective enhances our understanding of the multifaceted factors influencing retirement decisions, enabling more inclusive and equitable policy formulation.

Literature: As the population ages, the importance of increasing workforce participation among older individuals becomes evident. This study draws from previous research that has highlighted various factors influencing retirement timing, including demographics, health, social influences, job-related characteristics, financial considerations, retirement preferences, and macroeconomic determinants. These factors can vary significantly from one country to another, making it essential to tailor policies to specific contexts. While some studies have explored the impact of certain factors on

retirement decisions, there remains a gap in understanding how complex life course sequences, such as combinations of unemployment, work training, and career changes, affect retirement timing.

Methodology: To address this gap, our study relies on machine learning techniques: LSTM (Long Short-Term Memory) and Transformer (Encoder-Decoder), to analyze and predict work-life trajectories between ages 56 and 65. Both Transformers and LSTMs are used for sequence modeling tasks, where the goal is to understand and capture patterns in sequences of data. Life course sequences are inherently sequential data, with each year's state being dependent on the previous year's state. LSTM and Transformer models are well-suited for handling sequential data. They have the ability to capture dependencies and patterns over time, making them suitable for modeling life course sequences where employment statuses can evolve over time. These machine learning techniques are employed to provide both aggregate and individual-level predictions.

Data: In this study we employ The Scientific Use File SUFVVL2020 (Vollendeten Versichertenleben, VVL) which is a dataset provided by the Deutsche Rentenversicherung (German Pension Insurance) in Germany. This dataset contains information about individuals who have completed their insured lives and those who are actively contributing to the pension system. The SUFVVL2020 dataset represents comprehensive details regarding various aspects related to pensions. These details encompass contribution histories, pension entitlements, and demographic characteristics. Specifically, the SUFVVL2020 dataset consists of a 25% sample of pension entry cases, which amounts to a total of 218,907 cases. In constructing our model, we specifically focus on individuals for whom comprehensive life course data is accessible spanning ages 18 to 65. This subgroup consists of individuals born between 1940 and 1955, meaning that the observation period encompasses the years from 1958 to 2020. Within this cohort, we identify 28,999 cases distinguished by any missing values within the age range of 18 to 65. Additionally, we select individuals who reached the age of 50-55 in 2020, which marks the final year of our observations. This particular group forms the dataset for our future predictions and consists of individuals born between 1965 and 1970.

Preliminary Results: Both LSTM and Transformer-based models exhibited comparable performance, with around 70% accuracy in predicting social employment statuses across diverse age groups.

Furthermore, we scrutinized the distribution of social employment states in the predictions compared to the observed statuses. The graphical representations in Figure 1 comprises a significant level of prediction accuracy on the aggregate level. Prediction models utilize the first 15 years of an individual's observed life as input data to predict their life course sequences, therefore observations in the plot start from the age 33. Additional analysis demonstrated consistency in prediction accuracy also across genders.

Figure 1. Distribution of Actual and Predicted Social Employment States by Age of Individuals Born between 1940-1955
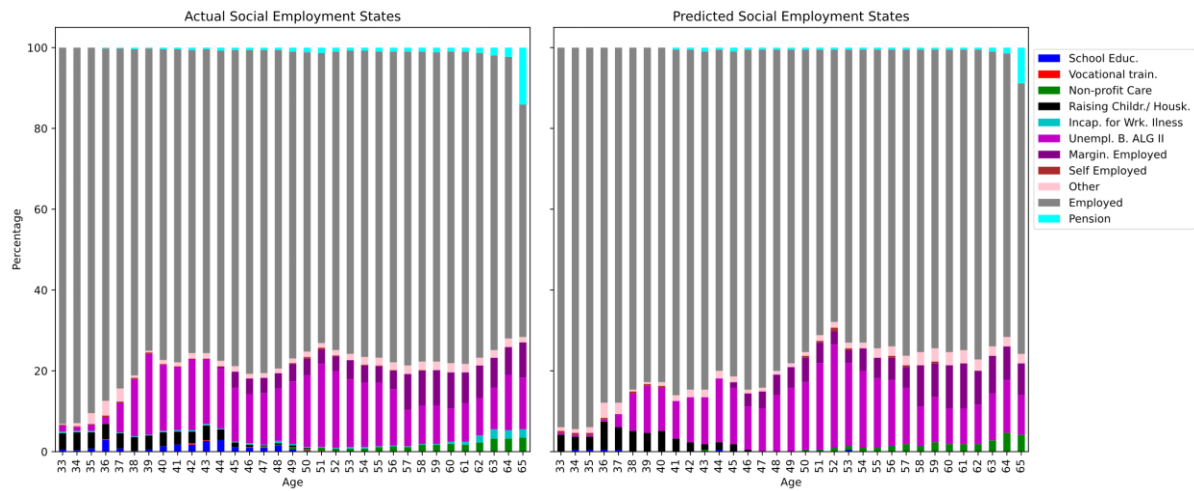
Furthermore, we applied the developed model to forecast social employment statuses at age 56-65 for men and women aged 50-55 in 2020, which represents the extent of our available observations. The predictions depicted in Figure 2 reveal that the frequency of unemployment states is quite similar between men and women. Interestingly, women are projected to engage in marginal employment at a significantly higher rate than men just before entering retirement.

Figure 2. Distribution of Predicted Social Employment States by Age and by Sex of Individuals Born Between 1965-1970
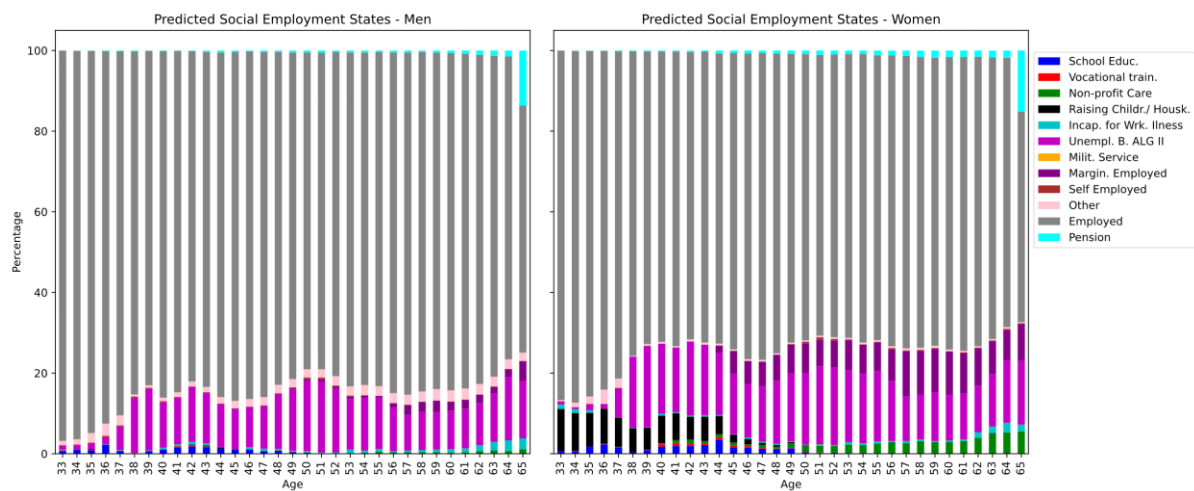
A more in-depth analysis revealed that individuals at risk of experiencing extended periods of unemployment (3 years or more) between the ages of 56 and 65 are primarily associated with distinct occupational categories as presented in Figure 3. Among men, these categories encompass individuals in simple manual work, agricultural occupations, skilled manual work and simple services.

Among women, these include simple manual work, simple services, agricultural occupations and qualified services.

Figure 3. Percentage Distribution of Occupations among Individuals Born Between 1965-1970 Predicted to Experience Extended Periods of Unemployment (3 years or more) between Ages 56 and 65
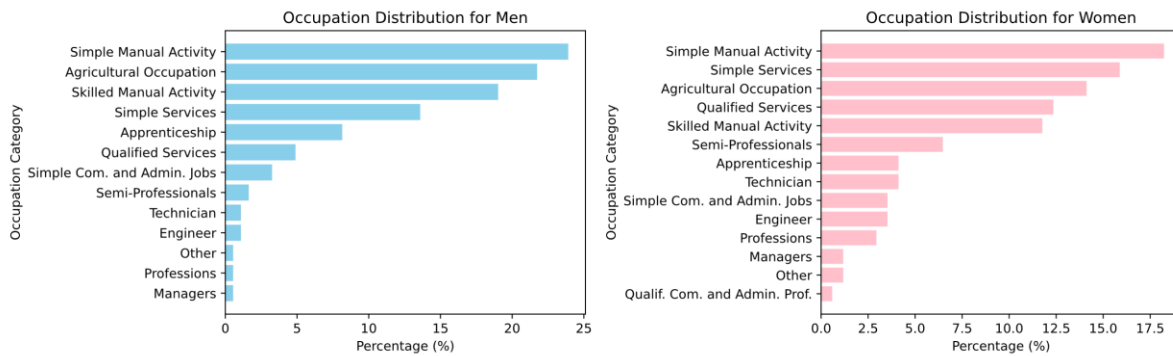
Additionally, our analysis also focused on individuals who are expected to experience early retirement, defined as retirement occurring more than 3 years before the mandatory retirement age. As depicted in figure 4, among men, professions in which individuals are predicted to retire earlier include simple manual activity, agricultural occupation, and skilled manual work. Among women, the predominant occupational categories are simple manual work, simple services, and semi-professionals.

Figure 4. Percentage Distribution of Occupations among Individuals Born Between 1965-1970 Predicted to Experience Early Retirement between Ages 56 and 65
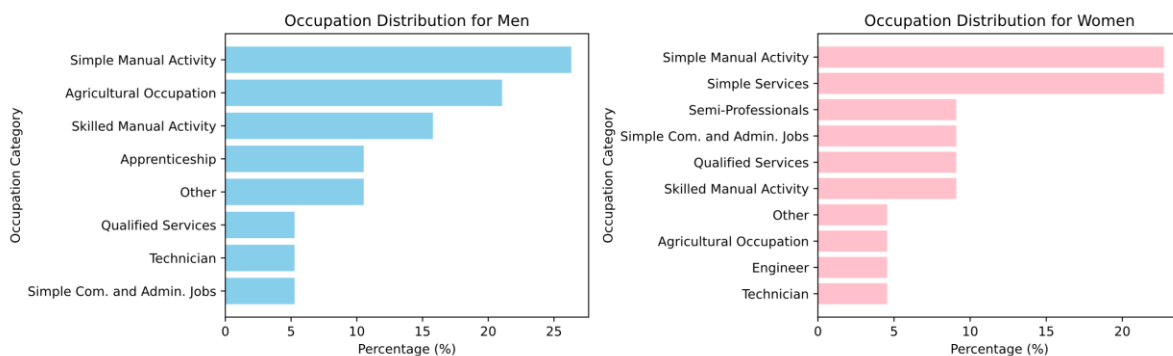
Moving forward, our next analytical step will involve sequence analysis to investigate whether individuals at risk of enduring over 3 years of unemployment during the 10-year span between ages 56 and 65, or those heading for early retirement, share common life course sequences and whether these differ between men and women.