Applying Transformers to Predict Life Course Sequences

Work in progress

Linda Vecgaile (The Max Planck Institute for Demographic Research), Emilio Zagheni (The Max Planck Institute for Demographic Research), Luiz Felipe Vecchietti (Institute for Basic Science, Daejeon, Korea, South), Alessandro Spata (Independent Researcher)

Key words: life course, sequence prediction, transformers, machine learning, labor market transitions

Abstract

Building upon life course theory, which emphasizes the interconnectedness and cumulative effects of life events, this study aims to advance predictive modeling of life course sequences. Our approach investigates whether past sequences of life events (ages 18-55) can effectively predict future sequences of life events (e.g., for ages 56-60), such as transitions into employment, unemployment, or retirement. Utilizing the Transformer encoder-decoder framework—renowned for its ability to analyze sequential data in natural language processing—we develop a customized model that treats life events over time as analogous to words in a sentence, capturing underlying patterns, relationships, and temporal dependencies within life course trajectories. We test the approach on data from the German Pension Insurance, which provides a minimal set of inputs—11 distinct social employment states along with basic demographic information. Our analysis reveals that the transformer model achieves a high predictive overall accuracy of 85.5%. It performs particularly well in predicting outcomes for individuals with stable life paths, who make up the majority of the dataset, but it also captures future states that deviate from recent patterns, demonstrating its ability to account for earlier life experiences. As we refine our models, we expect to be able to provide insights into what makes certain life course trajectories predictable or unpredictable and how machine learning techniques can help us identify patterns early in life that may be conducive to precarious employment circumstances later in life, during a critical stage of the life course.

Introduction

Life course transitions—such as the shift from school to work, retirement, prolonged employment, childcare leave or periods of unemployment—are often pivotal moments that significantly contribute to the (re)production and accumulation of social disparities. Life course theory suggests that these events carry both temporal and contextual significance, influencing individuals' future trajectories through their cumulative effects (Elder, 1998). For instance, transitions from paid to unpaid work, such as childcare, can significantly impact labor market trajectories, as demonstrated in the literature on motherhood and its effects on labor market outcomes (Blau & Kahn, 2017; Budig & England, 2001; Waldfogel, 1998). Similarly, early exposure to unemployment during the transition from school to work can have a lasting impact on future employment prospects, potentially shaping one's entire career path trajectories (Arulampalam, 2001; Biewen & Steffes, 2010; Gregg, 2001; Nilsen & Reiso, 2014). These early experiences often set a course that can be difficult to alter, as the opportunities and limitations faced by individuals early in life can have long-lasting consequences.

However, traditional research in sociology and demographics has largely focused on predicting life course transitions as discrete events—typically emphasizing the movement from one specific state to another (Bolano, 2015). Similarly, studies that aim to predict employment status or fertility often yield binary outcomes, such as whether an individual will enter or exit employment or whether they will have a child in the next period (Buh, 2023). While these approaches are valuable, they do not fully embrace the life course perspective, which recognizes the interconnectedness of various life events over time. In this paper, we take a different approach. Rather than concentrating on individual transitions, our goal is to determine whether sequences of past life events can be used to predict sequences of future life events. This shift in focus is crucial as it lays the groundwork for future research that could explore the mechanisms by which past life course sequences shape future trajectories or determine which life courses are more or less predictable.

To achieve this, we utilize data from the German Pension Insurance and employ the Transformer encoder-decoder framework—a powerful tool in natural language processing—to model and predict life trajectories. Drawing from life course theory (Elder, 1998), which views an individual's life as a sequence of interconnected events, we treat social employment states in each year of an individual's life as analogous to words within a sentence. By doing so, we aim to capture the underlying patterns and relationships among life events over time and assess whether these patterns can be used to forecast future sequences of events. Our model is both innovative and minimalistic. It relies on just 11 distinct social employment states—ranging from school education to retirement, including periods of unemployment, childcare, vocational training, and marginal employment—along with basic demographic information such as gender and year of observation. The rationale for using such limited information is twofold. First, by simplifying the input data, we can test the robustness of predictive models and assess whether core life course dynamics can be captured with minimal yet essential information. Second, this approach increases the generalizability and applicability of the model across different contexts and datasets, where detailed information may not be available. If we can demonstrate that meaningful predictions are possible with these sparse inputs, it suggests that fundamental life course patterns are inherently strong and can be detected even with limited data, making the model useful in a wider array of scenarios.

We specifically focus on predicting the transition period between the ages of 56 and 60. Utilizing data from the German Pension Insurance, which includes individual biographies up until 2020, this age range is particularly relevant. By focusing on people born up until 1960, who turned 60 by 2020, we can fully leverage the available data to maximize our training potential. Importantly, ages 56 to 60 represent a critical phase where individuals are increasingly at risk of unemployment, marking the beginning of potential economic vulnerability. This period is crucial as it often signals the onset of retirement insecurity and social inequality. By predicting life trajectories during this phase, we seek to understand the predictive power of past life events in forecasting future states, which could inform future analyses aimed at explaining the underlying causes and effects.

Literature Review

Rooted in the Life Course Theory (LCT), which posits that individuals' experiences unfold as a series of interconnected events, our lives are shaped by the timing, duration, and order of preceding states, as well as the contextual status such as surrounding cohorts and their demographic composition, historical context, institutional structures including political climates, and prevailing systems of education and social welfare (Hutchison, 2010). Within this framework, transitions in the life course—such as moving from education and training to employment or entering parenthood—play a crucial role in shaping one’s life trajectory. Additionally, life events like periods of work disability, or engaging in further education exert significant influence on employment patterns (Udayar et al., 2021).

The life course approach in analyzing development of life trajectories initiated by Glen H. Elder (Elder, 1994, 1998) who explored how the economic upheavals of the Great Depression influenced family stability (Elder, 2018). Elder's observations revealed consistent life patterns among groups encountering comparable economic upheavals and their subsequent recoveries. Central to his investigations were work and employment, which he regarded as pivotal factors shaping social mobility and societal conduct. Life course concepts were further developed by adding the dimension of age and temporality, illustrating how the timing of life events shape individuals' life trajectories. For instance, early career decisions regarding employment and income significantly influence one's entire professional life. Numerous studies have shown empirical evidence of the adverse long-term impact that unemployment and inactivity early in life can have on future opportunities in the labor market, often referred to as the "scarring effect" (Arulampalam, 2001; Biewen & Steffes, 2010; Gregg, 2001; Nilsen & Reiso, 2014). From this emerged the "early timing hypothesis," positing that early institutional involvement in a young individual's life, such as military service, on-the-job training, or vocational education, enhances their occupational standing and economic prosperity [reference].

Studies have also shown that belonging to a cohort, a collective of individuals who share the same birth period and undergo similar societal shifts within a specific cultural context, progressing through them in parallel at corresponding ages, can predict fertility (Buhr & Huinink, 2014; Mishra et al., 2010), and retirement income (Rasner, 2014). Additionally, encountering joblessness, particularly between the ages of mid-20s to late-40s, has been shown to correlate with poorer physical and mental well-being by the age of 50 (Arpino et al., 2018; Frech et al., 2022).

Traditional approaches to forecasting life course events, such as labor market outcomes, often neglect the significant impact that the timing and sequence of events have on an individual's life trajectory. Typically, these methods focus on discrete transitions, such as moving from unemployment to employment or from work to retirement, without accounting for the preceding sequence of life events that contribute to these changes. This narrow focus overlooks the complex interplay and timing of events within the broader context of an individual's life course. A notable exception to this traditional approach is the emergence of studies using language models to predict mortality based on the sequence of preceding life events (Savcisens et al., 2024).

By adopting a holistic approach to predicting transitions—one that incorporates the historical sequence of events and accounts for prior transitions in both the labor market and family life, such as shifts from paid work to work disability, retirement, or the return from childcare to employment—we can enhance the predictive accuracy of discrete events. Additionally, by forecasting entire sequences of events, this approach provides valuable material for further analysis, offering deeper insights into the life course patterns that lead to the anticipated life events.

Method

In the analysis of life course data, where individuals traverse different states such as education, employment, and marriage over time, researchers have turned to Markov chains (Bolano, 2015, 2017). Markov chains represent a stochastic process, where the likelihood of transitioning to a future state depends solely on the current state, adhering to the Markov property. Transition probabilities between states are typically organized in a transition matrix, which illustrates the probability of moving from one state to another [reference] (Bolano, 2017). These transition probabilities can be derived from longitudinal data tracking individuals' life trajectories (Bolano, 2015, 2017). However, relying solely on transition matrices where a future state depends solely on the current state while ignoring the cumulative effects of the past events in the preceding sequence may limit the prediction accuracy of the future event and limit our understanding of the labor market transitions that lead to the future event. Moreover, life courses have been growing in complexity, de-standardization, and the emergence of gender-related differences (reference). These factors imply that life trajectories might not strictly adhere to the assumptions of Markov chains, which assume stationary transition probabilities over time. Therefore, while Markov chains offer a useful framework for employing life course data and predicting transitions, they may need to be supplemented with additional methods to capture the full complexity of life courses.

To address the limitations of relying solely on transition matrices and Markov chains in understanding the intricate dynamics of life course sequences, a more advanced approach is warranted. One such method is the Transformers encoder-decoder architecture, which has demonstrated remarkable capabilities in capturing complex temporal dependencies in various sequential data tasks [reference].

Transformer (Encoder-Decoder) is 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. Transformer models are well-suited for handling sequential data. They have the ability to capture dependencies and patterns over time (reference), making them suitable for modeling life course sequences where employment statuses can evolve over time. Transformers use a self-attention mechanism that allows them to weigh the importance of different elements in a sequence (Vaswani et al. 2017). This makes them effective at capturing both short-term and long-term dependencies in the data. For predicting life course sequences, this is beneficial as it helps in understanding how different social employment statuses in various years influence each other.

Transformers have been successfully applied to a wide range of sequence modeling tasks, including natural language processing (NLP) tasks like language modeling and machine translation. This demonstrates their versatility and ability to capture complex patterns in sequential data, which is relevant when predicting life course sequences that may involve intricate social and employment transitions. The Transformer's encoder-decoder architecture is particularly useful when the goal is to predict future states based on past states. The encoder can capture patterns in historical data, and the decoder can generate predictions for future years based on that information (Vaswani et al. 2017). This aligns well with the objective of predicting future employment statuses based on past ones in a life course sequence. Additional strength of Transformers is their parallelizability, which makes them suitable for handling large datasets efficiently given that life course data can span many years and involve a significant amount of data (Vaswani et al. 2017). The ability to scale up the model's capacity is crucial for accurate predictions.

By adopting Transformers architecture, past sequences of life course events can be utilized alongside the current state, potentially improving prediction accuracy and offering more comprehensive analyses of life course transitions.

Data and Variables

This study relies on The Scientific Use Files VVL2020 (Vollendeten Versichertenleben, VVL) and VSKT2020 which are datasets provided by the Deutsche Rentenversicherung (German Pension Insurance) in Germany. The VVL2020 dataset comprises individuals residing in Germany, born between 1940 and 1995, who completed their insured lives by 2020. They are observed from the age of 14 until they transition to retirement in 2020. The VSKT2020 dataset includes individuals living in Germany, born between 1953 and 1990, who are actively contributing to the pension system. These individuals are observed from the age of 14 until their age in 2020. The data is based on monthly records.

Importantly, these datasets are not representative of the entire population but rather represents a subset of data. This subset is derived from administrative records and offers comprehensive details regarding various aspects related to pensions. These details encompass contribution histories, pension entitlements, and demographic characteristics. Specifically, the VVL2020 dataset consists of a 25%

subsample of pension entry cases, which amounts to a total of 218,907 cases while VSKT2020 contains 106,553 cases. These samples offer valuable insights into a subset of the population consisting of pension beneficiaries and individuals who are actively contributing to the statutory pension insurance in Germany. This implies that individual life biographies are available, providing sequences of social employment states that define the accumulated statutory pension insurance.

Combining VVL2020 and VSKT2020 results in a dataset containing a total 325,460 individuals. Considering the relatively large number of individuals observed and the frequency of monthly records, significant amounts of data are generated. To streamline the model and decrease complexity, the monthly records have been aggregated into yearly data. VVL2020 and VSKT2020 provide monthly data on social employment status, detailing 15 distinct states (outlined in the Appendix). To slightly improve the balance in the distribution of these states, we aggregate some of the less common states into 10 broader categories: School Education, Vocational Training, Childcare or Non-Profit Care, Unemployed or Incapacitated, Military Service, Marginal Employment, Other, Employed or Self-Employed, and Pension. While this aggregation does not fully resolve the imbalance, it helps to reduce the impact of states with very small shares. We introduced an additional category labeled "no contribution," which includes periods when individuals were not making contributions to the German statutory pension insurance scheme captured in the VVL2020 and VSKT2020 data. This indicates that during these periods, individuals may have changed their employer and been covered by insurance other than the statutory pension insurance scheme. For example, Government employee pension schemes (Beamtenversorgung) are designed for employees of government agencies, public sector organizations, or governmental bodies in Germany. This category serves to capture instances where individuals are not actively contributing to the state pension scheme, thereby providing a more comprehensive understanding of their employment and insurance status over time.

These distinct categories, when combined, provide a comprehensive characterization of individuals' life courses spanning ages 14 to 65. However, considering that only a very small proportion are involved in contributing to the statutory pension insurance scheme at the age of 14, we limit the observation period to start from age 18.

In addition to the social employment states, we differentiate individuals by gender. Finally, we also include a variable that identifies a person’s age and the specific year of that age. The latter variable serves to capture the potential dynamics of individuals' life courses over time, such as the destandardization of individual life courses.

Model Training: Methodology and Approach

To prepare the data for Transformer-based encoder-decoder models, we transform individuals' life events and states from ages 18 to 65 into a text dataset, ensuring the preservation of chronological order. We create sequences for each individual, representing their social employment states over this time span. To adapt this structured data for text-based encoder-decoder models, we encode categorical variables, like event types or states, as text labels or embeddings. Considering that individuals enter retirement at varying ages and we observe individuals born in different years, the sequences of their life events vary in length. Therefore, we utilize tokenization to divide the text into manageable units (social employment states) for model processing.

For the encoder-decoder architecture, we create sequence pairs. The encoder receives the input sequence, and the decoder predicts the next sequence in chronological order. These paired sequences are used for training. We build a token vocabulary or dictionary that maps tokens (social employment states) to unique integer IDs. This vocabulary is used during tokenization and for generating predictions during inference. The text-based data is encoded into numerical form using the token vocabulary. Each token in the text is replaced with its corresponding integer ID. To efficiently train the model, we organize the data into batches. Batches consist of multiple sequence pairs and are fed into the model during training. Finally, we train the encoder-decoder model using the prepared dataset. These batches comprise multiple sequence pairs and are supplied to the model during training. This batching approach helps streamline the learning process and enhance the model's overall performance.

Model Training: Effect of Sequence Length on Prediction Accuracy

One aspect of developing our model involves the input sequence length. In our context, the 'input sequence' refers to a series of data points or tokens that serve as the initial information given to the encoder-decoder model. These tokens typically represent various aspects of an individual's life course, such as education, employment, or other relevant factors. The purpose of this input sequence is to provide context and information to the model, enabling it to make predictions about future sequences. The 'length' here refers to the number of tokens or data points contained within the sequence. It's essential to recognize that input sequences can vary in length, meaning they can be short or long, depending on the specific application or dataset.

In our study, the goal is to determine whether at the age of 55 by providing a preceding life course sequence from age 18 can predict a sequence of the following five years. Therefore, we use the input sequence lengths of 38 tokens (age 18-55).

Additionally, we examine the use of only 1 input state which allows us to assess the applicability of life course theory. In other words, we investigate whether prediction accuracy improves with longer sequence information compared to providing only one current state to predict the next state.
In the VVL2020 and VSKT2020 datasets, individuals are observed up until the year 2020. To train our model to predict social employment states between the ages of 56 and 60, we require individuals who have reached at least age 60. Therefore, we focus on the biographies of individuals born between 1940 and 1960 by 2020.

Assessing the Algorithm's Performance

In the process of building and training our encoder-decoder model, we evaluate it on the testing dataset to assess its ability to generalize to unseen data. A common data split ratio is employed, with 70% of the data used for training, 15% for validation, and the remaining 15% reserved for testing. This division ensures that the model generalizes well to new data. When training and testing the model, we follow this general rule. First, we compile a dataset comprising all individuals from the VVL2020 dataset, which includes those who retired in 2020 and their social employment biographies up to that point. Additionally, we select 70% of individuals from the VSKT2020 dataset who are actively contributing to the statutory pension insurance scheme in 2020, along with their social employment biographies up to 2020. From the remaining 30% of individuals in the VSKT2020 dataset, we select those born between 1961 and 1964. Since these individuals have not yet reached the age of 60 by 2020, we utilize the model to predict their future social employment states until they reach 60 years old.

Next, we combine the VVL2020 and selected VSKT2020 datasets to create training (70%), validation (15%) and testing (15%) datasets for the model. Since capturing the dynamics of life course changes over time is one of the model's objectives, we focus the training dataset on individuals born between 1940 and 1957. Subsequently, we evaluate the model's performance on individuals born within this range as well as those born in years not included in the training dataset, specifically 1958-1960.

In addition, we also compare our Transformer (Encoder-Decoder) model's performance against other relevant methods used in the social sciences to predict life course sequences: Markov chains and a Last Observed State Projection method. The Last Observed State Projection method assumes that an individual's most recent employment state (e.g., at age 55) will persist into the future, serving as a baseline prediction approach. We aim to investigate whether the transformer encoder-decoder architecture yields superior prediction accuracy compared to both Markov chains and this projection method. As outlined in an earlier section addressing methods for predicting employment transitions, Markov chains constitute a stochastic process wherein the probability of transitioning to a future state relies solely on the current state.

In evaluating the accuracy of sequence predictions within our testing dataset, we adopt a binary system to distinguish between correct and incorrect predictions (0 for correct, 1 for incorrect). Additionally, we calculate the proportion of correctly predicted social employment states at each age for each model with varying numbers of input states. This assessment provides a general estimation of overall prediction accuracy.

Moreover, we conduct separate calculations for each social employment state to discern which state tends to be better predicted. Beyond examining the impact of sequence input length on prediction accuracy, our investigation delved into the influence of various included variables within the sequence. Initially, we incorporated age and social employment state status. To test, whether the model can be further advanced, we introduced a temporal indicator – the year of observation. Subsequently, we additionally integrated birth year and sex variables.

Furthermore, we consolidate all observations and create chronograms that illustrate the distribution of social employment states by age for 10-year birth cohorts. These visual representations allow us to compare the model's predictions with the actual ground truth sequences, enabling us to visualize the model's aggregate-level prediction performance and identify patterns in social employment states across birth cohorts.

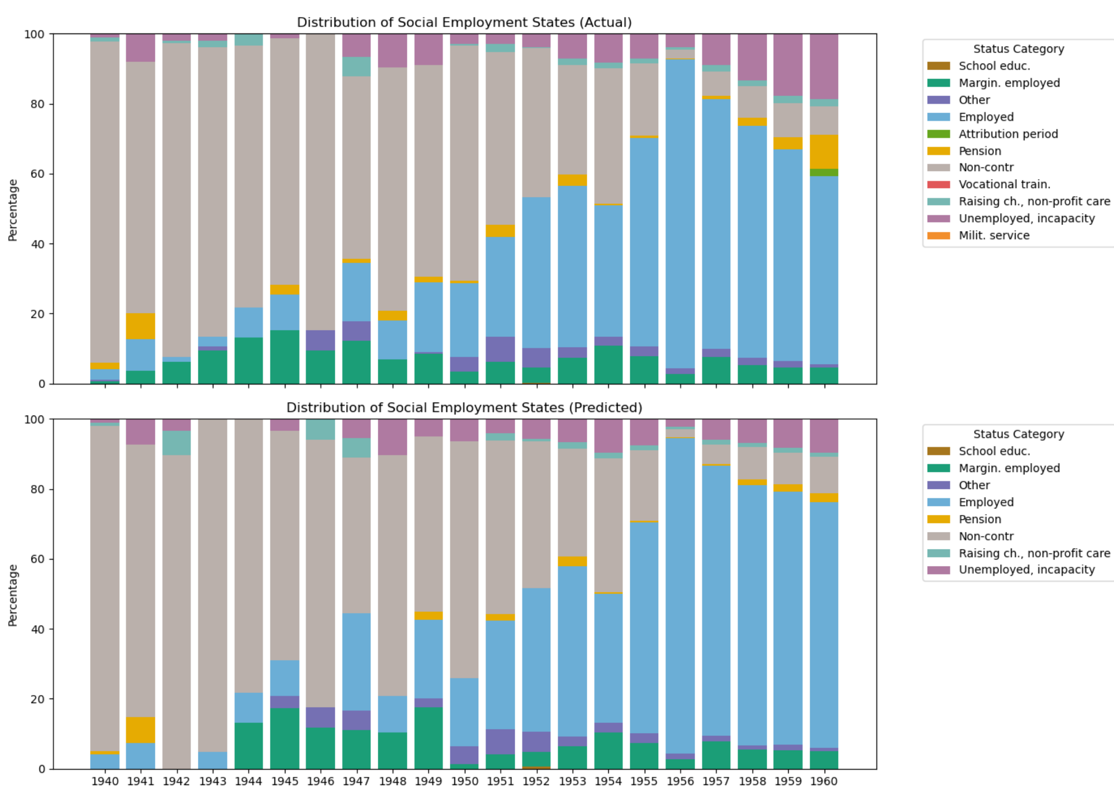
The approach described above not only aids in evaluating the model's prediction accuracy at both the individual and aggregate levels, but also helps determine which model, with how many input states, performs the best for both individual and aggregate predictions, as well as for each social employment state separately.

We also conduct sequence analysis to identify the types of individuals for whom the model excels at predicting ages 56-60, as well as those for whom it underperforms. To achieve this, we group individuals into three categories based on their prediction accuracy: individuals with perfect prediction, those with partial accuracy (greater than 0% but less than 100%), and those with 0% accuracy. Within each category, we further analyze the five largest groups of individuals, determined by their sequence similarity during ages 56-60. Sequence similarity is measured using Levenshtein distance (Yujian & Bo, 2007), allowing for a maximum of two insertions or deletions.

Preliminary Results

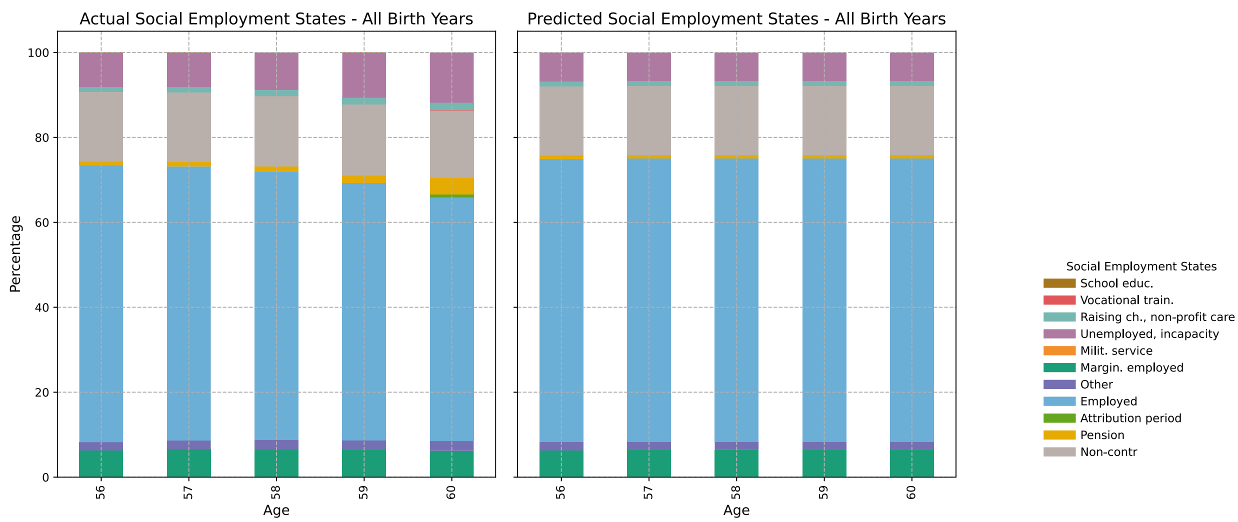
As shown in Figure 1, the model effectively recreates the distribution of social employment states, accurately predicting marginal employment, missing values, and non-profit care. However, it tends to underpredict transitions to retirement for individuals born in several years such as 1945, 1948, 1959 and 1960 and unemployment for those born between 1957 and 1960. This underprediction is offset by an overestimation of the employment state.

Figure 1. Distribution of Actual vs. Predicted Social Employment States by Birth Year.



We further analyze the distribution of social employment states by age, comparing the actual and predicted distributions for ages 56-60. Figure 2 below shows that the model tends to predict social employment states for age 56 consistent across subsequent ages. This trend is also observed when examining distributions by age for each birth year individually (in Appendix A).

Figure 2. Distribution of Actual vs. Predicted Social Employment States by Age.



We use F1 scores to assess the model's accuracy in predicting social employment states for each target age group (56-60) by birth year. As detailed in Table 1, the overall F1 scores range from 0.67 to 0.95. When evaluating the F1 scores for each age individually, the model's accuracy is notably lower for the last two ages (59 and 60). This trend—higher accuracy for the earlier ages and lower accuracy for the later ones—aligns with expectations due to the iterative nature of sequence prediction. As predictions progress, errors can accumulate, resulting in a gradual decline in overall accuracy. Notably, F1 scores for individuals born in years not included in the training set show relatively high prediction accuracy, indicating that the model generalizes well, with the exception of age 60, where the F1 score drops significantly, approaching or falling below 0.5 for individuals born in 1959 and 1960.

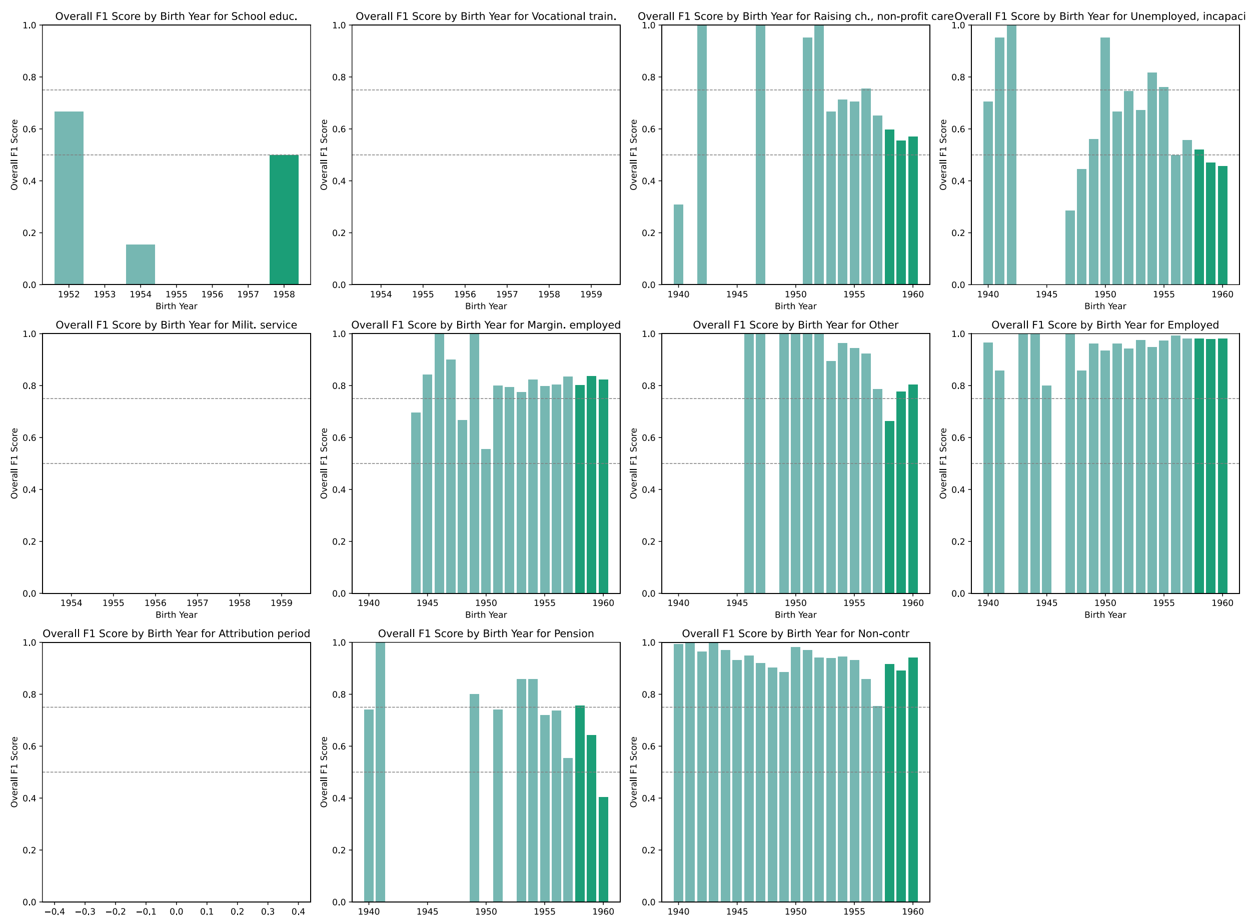
Table 1. F1 Score for Predicted Social Employment States by Age and Birth Year.

Age	**1940**	**1941**	**1942**	**1943**	**1944**	**1945**	**1946**
Overall F1 Score	*0.95*	*0.91*	*0.85*	*0.79*	*0.84*	*0.81*	*0.95*
56	0.97	1.00	0.93	0.86	0.95	0.86	0.97
57	0.96	1.00	0.78	0.79	0.91	0.81	0.97
58	0.96	0.85	0.86	0.72	0.88	0.80	0.97
59	0.94	0.85	0.86	0.81	0.74	0.81	0.92
60	0.91	0.87	0.86	0.81	0.74	0.76	0.92
Age	**1947**	**1948**	**1949**	**1950**	**1951**	**1952**	**1953**
Overall F1 Score	*0.83*	*0.71*	*0.81*	*0.91*	*0.90*	*0.88*	*0.85*
56	0.88	0.90	0.87	0.95	0.94	0.94	0.92
57	0.84	0.76	0.81	0.95	0.90	0.92	0.89
58	0.84	0.73	0.83	0.92	0.89	0.86	0.85
59	0.84	0.63	0.80	0.88	0.89	0.85	0.82
60	0.79	0.58	0.77	0.86	0.88	0.85	0.80
Age	**1954**	**1955**	**1956**	**1957**	**1958**	**1959**	**1960**
Overall F1 Score	*0.86*	*0.88*	*0.93*	*0.83*	*0.81*	*0.76*	*0.67*
56	0.92	0.92	0.96	0.92	0.92	0.90	0.91
57	0.87	0.89	0.94	0.88	0.87	0.86	0.84
58	0.85	0.87	0.93	0.84	0.83	0.79	0.76
59	0.83	0.86	0.92	0.80	0.76	0.71	0.57
60	0.81	0.84	0.91	0.73	0.67	0.54	0.32

Additionally, we calculated F1 Scores for each social employment state separately (Figure 3). (For a detailed breakdown of F1 Scores by social employment state, birth year and age, please refer to the tables in the Appendix B.) As previously noted when analyzing the distributions of actual versus predicted social employment states, the model performs best in predicting transitions to employment and marginal employment. It also accurately predicts periods of missing contributions (non-contributing) and unspecified activities (other). However, the model's predictions are weaker for unemployment, particularly among younger cohorts. Similarly, it struggles to predict involvement in non-profit care, especially for younger cohorts. The lowest accuracy is observed for predicting school education, vocational training, and military service, as these states are infrequent in the dataset, particularly within the relevant age range.

For individuals born in years not included in the model training set (1958-1960), marked in darker green, the F1 scores across each social employment state category reflect favorable results, despite these birth years not being part of the training data. However, a notable decline in prediction accuracy begins already before 1958. This trend suggests that the reduced F1 Scores for these out-of-sample birth years are not solely attributable to the model's generalization limitations. Instead, it indicates the likely presence of confounding factors that may impact prediction accuracy for certain birth cohorts. The empty plots indicate that the respective social employment states were not predicted for the age range of 56-60. *(the X axis for "Attribution period" needs to be fixed)*

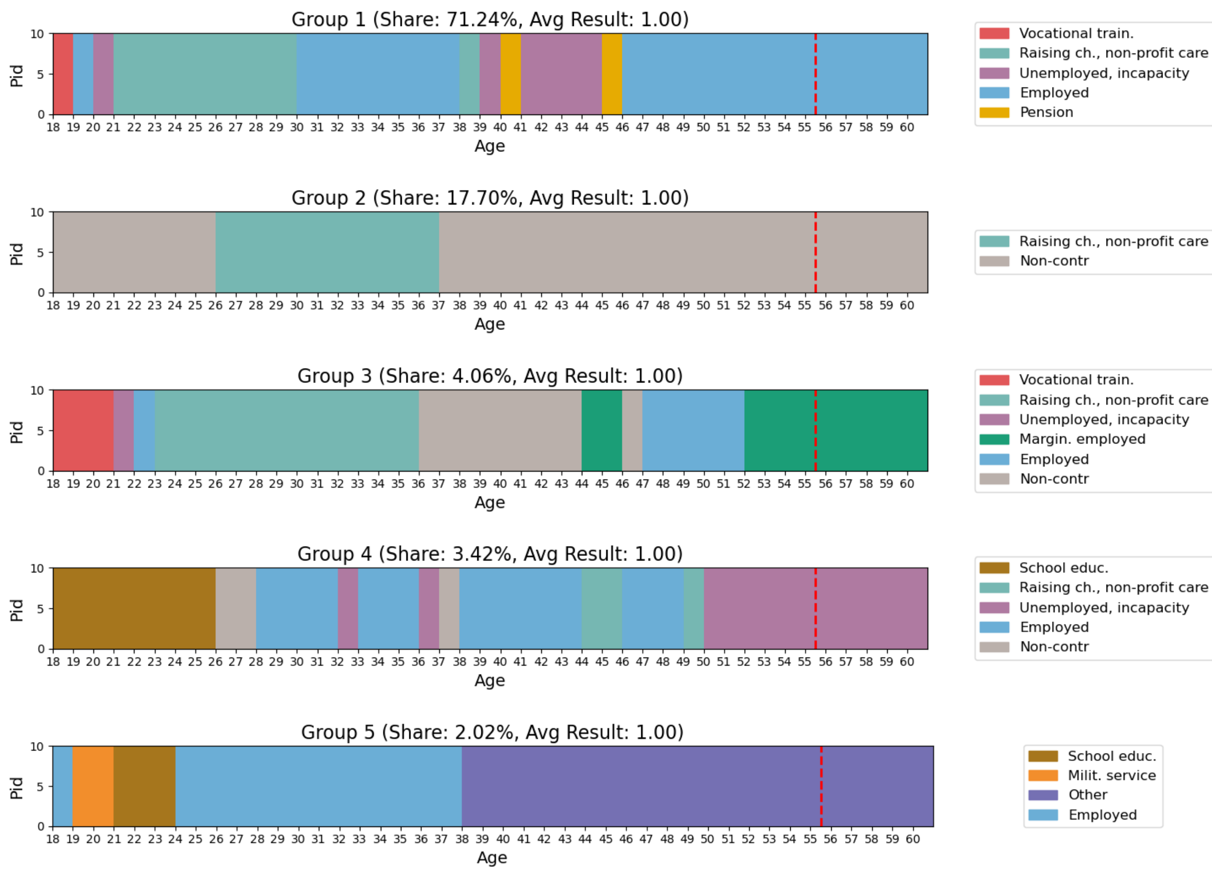
Figure 3. F1 Score by Social Employment State and Birth Year.



Additionally, we aim to identify the types of sequences that are predicted with 100% accuracy, along with the preceding life course patterns that define them. We also investigate sequences predicted with less than 100% accuracy and those predicted with 0% accuracy. To do this, we group sequences in each of the three categories based on Levenshtein distance and present the five largest identified groups within each accuracy category.

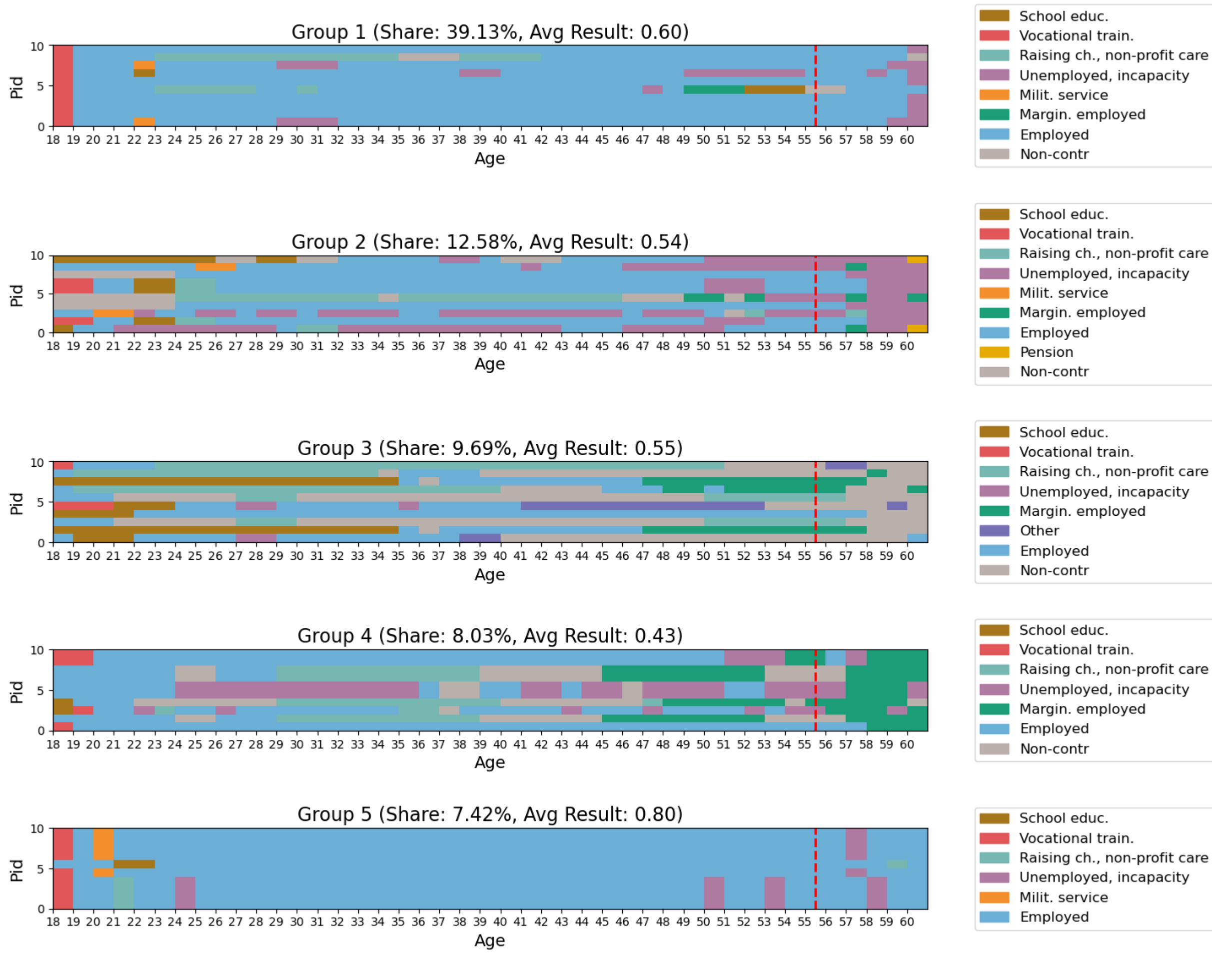
The Figure 4 below illustrates that sequences predicted with complete accuracy (100%) typically exhibit stability, with the last five or more preceding states remaining constant. These life courses show minimal transitions between states, reflecting a stable pattern without much fluctuation. This category dominates the dataset, accounting for 72.96% of all individuals.

Figure 4. Sequences Predicted with Complete Accuracy: Representing 72.96% of All Individuals.



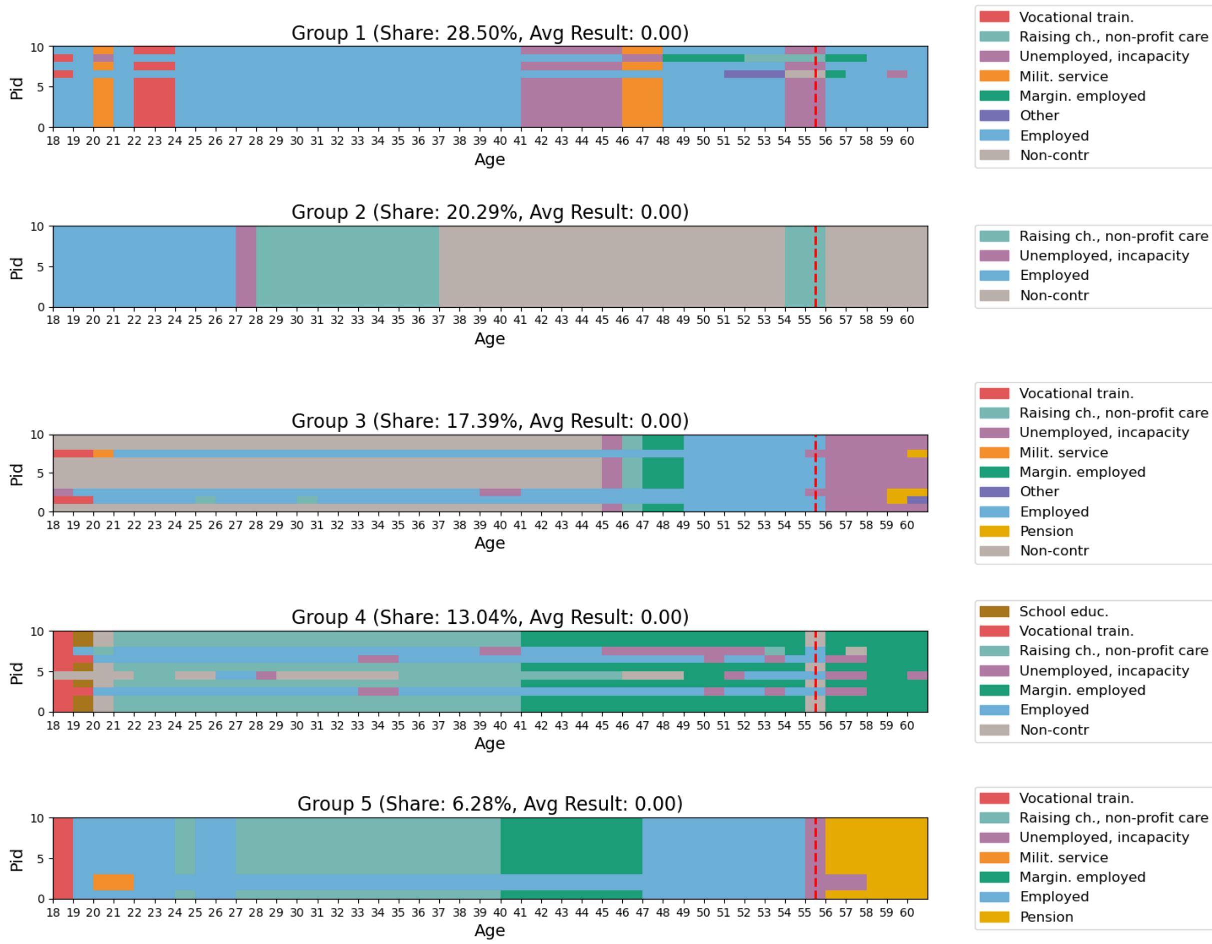
Sequences that are predicted with partial accuracy—less than 100% but greater than 0%—are shown in the next Figure 5. This group represents 22.9% of the total population. These sequences show some variation in the last five years observed, although there is often a dominant state that prevails. The model tends to predict this dominant state for individuals in this category. Despite the variability seen in the years leading up to the prediction, the model successfully identifies the prevailing state in most cases, demonstrating its ability to capture the dominant trend in the final years.

Figure 5. Sequences Predicted with Partial Accuracy (Less than 100% but Greater than 0%): Representing 22.9% of All Individuals.



Lastly, sequences with the least predictive accuracy exhibit characteristics that combine features from both of the previous groups. This is the smallest category, comprising 4.14% of the population. In some cases, these sequences show frequent transitions between states, with high variability both in the years to be predicted and in the preceding years. However, other sequences within this group exhibit more stability. The challenge for the model in this category arises from the greater unpredictability and inconsistency in life course patterns.

Figure 6. Sequences with Zero Predictive Accuracy (0%): Representing 4.14% of All Individuals.



Discussion

The results of this study provide valuable insights into the predictive accuracy of the Transformer model in forecasting life course sequences during the critical ages of 56 to 60. The model demonstrates high overall accuracy, with F1 scores ranging from 0.67 to 0.95. It performs particularly well in predicting stable life trajectories, where individuals' employment states show little variation over time. These findings align with life course theory, which emphasizes the cumulative nature of life events and the tendency for stability in later life stages.

One of the key observations from the results is the model's ability to accurately predict transitions into employment, marginal employment, and missing contributions. This reinforces the model's strength in capturing patterns in stable life courses, where transitions are predictable and consistent with prior states. However, the model shows limitations in predicting transitions to retirement and unemployment, especially for individuals born in specific years (e.g., 1945, 1948, 1959, 1960). This underprediction is balanced by an overestimation of employment states, which suggests that the model may have difficulty in detecting abrupt changes in life courses, particularly those related to retirement or unemployment.

The analysis of F1 scores by age reveals an interesting trend: the model's predictive accuracy declines progressively for older ages (59 and 60). This is likely due to the cumulative nature of prediction errors in sequence-based models, where early inaccuracies compound over time. The lower F1 scores for these later ages suggest that while the model is effective in capturing general patterns, it struggles with the increased uncertainty that comes with predicting transitions further into the future. This trend is especially pronounced for individuals born in the late 1950s and 1960s, where the F1 scores drop significantly, particularly at age 60.

Moreover, when examining the F1 scores for specific social employment states, it becomes clear that the model performs best in predicting transitions to employment, marginal employment, and non-contributing periods (e.g., missing contributions or unspecified activities). However, it faces challenges in predicting less frequent transitions, such as unemployment and non-profit care, particularly for younger cohorts. The model's difficulty in predicting transitions into vocational training, school education, and military service can be attributed to the rarity of these states in the dataset, which limits the model's ability to learn their patterns effectively.

Notably, the model's generalization capacity is reflected in its relatively high F1 scores for individuals born in years not included in the training data (1958-1960). Despite the absence of these cohorts during training, the model achieves reasonable accuracy, indicating that it can generalize well to new data. However, the notable decline in F1 scores for these birth years—particularly for age 60—suggests the presence of confounding factors that may not be adequately captured by the model. This drop in accuracy likely points to structural changes in employment patterns or policy shifts that disproportionately affect these cohorts.

The results also highlight distinct patterns in sequences predicted with 100% accuracy, partial accuracy, and no accuracy. Sequences predicted with complete accuracy tend to exhibit stability, with little variability in the preceding five years, aligning with the model's strengths in handling stable life trajectories. Conversely, sequences predicted with partial accuracy show some variability in the last five years, though a dominant state typically emerges. The model effectively captures this dominant state, despite the complexity in preceding years. Lastly, sequences with the least predictive accuracy are characterized by significant variability, frequent transitions, and less predictable patterns. These sequences, representing the smallest group, pose the greatest challenge for the model due to their dynamic nature.

Conclusions

This study seeks to determine whether past sequences of life events can effectively predict future sequences using minimal information. This research serves as a foundational step toward more comprehensive models that can later be used to explore how these past events shape future trajectories. Through our innovative methodology, we hope to provide a new perspective on life course research and open the door to future studies that will analyze the complex dynamics of life trajectories in greater depth.

This study contributes to life course theory by demonstrating that past life events are significant predictors of future states, particularly during the critical transition period between ages 56 and 60. The model's strong performance in predicting stable life trajectories aligns with the principles of path dependency central to life course theory. Path dependency suggests that early life experiences and accumulated transitions have a lasting influence on future outcomes, making individuals' trajectories more predictable as they age (Hanger-Kopp et al., 2022). Our findings reinforce this concept, showing that individuals with stable life courses exhibit highly predictable transitions in later life, with little deviation from their prior states.

The results also highlight the cumulative effects of life events, a key tenet of life course theory, which emphasizes that each life transition builds upon previous ones (Elder, 1994). This is particularly evident in the model's accurate predictions for sequences characterized by consistent employment or marginal employment over time. Conversely, the model struggles with individuals whose trajectories show more variability, as frequent transitions between different states introduce unpredictability. This demonstrates how variability earlier in life creates complexity in forecasting future states, making transitions to unemployment, retirement, or other states more challenging to predict.

The observed patterns contribute to the broader understanding of how life events cluster over time and provide empirical support for the idea that early life experiences set individuals on paths that are difficult to alter, especially as they approach older ages. These findings extend life course theory by quantifying the extent to which stability and variability in life course sequences affect predictability, offering new evidence on how path dependency operates in the context of employment and retirement transitions.

Further Improvement in Model Accuracy

While the model achieves high overall accuracy, particularly for stable trajectories, the inclusion of additional individual-level data could further enhance its predictive performance. Currently, the model relies on a minimal set of social employment states and basic demographic information. However, life course transitions—especially into unemployment or early retirement—are influenced by a broader range of factors that are not captured by the current model.

Incorporating variables such as health status would likely improve accuracy, as poor health is a significant predictor of early retirement and prolonged unemployment (Hasselhorn et al., 2022; Rice et al., 2011; Toczek et al., 2022). Professional background and education are also crucial, as certain professions and levels of education are associated with different risks of unemployment and career longevity. For example, individuals in physically demanding jobs or those with lower levels of education may face earlier retirement or more frequent transitions into unemployment due to the challenges posed by job demands or market changes (Andrasfay et al., 2021).

Additionally, family status could provide valuable insights, as marital status and caregiving responsibilities (e.g., caring for a spouse or children) can significantly affect employment stability and the timing of retirement (Schmauk, 2024). Moreover, technological changes and shifts in the labor market, particularly with the increasing prevalence of automation and digitalization, are likely to play a growing role in determining career trajectories and the risk of unemployment (Casas & Román, 2023). By including variables that capture regional economic conditions or industry-specific changes, the model could better account for the external factors that influence individual employment patterns.

Other key factors that could enhance prediction include access to social welfare programs, pension system specifics, and policy changes related to retirement age and unemployment benefits (Arranz & García-Serrano, 2023). These structural factors can alter life course trajectories by providing safety nets or incentives for individuals to remain in the workforce longer or to exit earlier.

It is reasonable to speculate that with the inclusion of these additional variables, the model's prediction accuracy—particularly for individuals with less stable life courses—would improve significantly. The currently observed decline in accuracy for individuals nearing retirement age or experiencing frequent transitions could be mitigated, as these added dimensions would enable the model to better capture the nuanced and multifaceted nature of life course transitions.

In sum, while this study demonstrates the predictive power of past life events using a simplified model, incorporating a more comprehensive set of factors could further enhance the model's performance and provide deeper insights into the complex interplay of life events that shape individual trajectories. These improvements would not only bolster the model's applicability but also contribute to the ongoing refinement of life course theory by highlighting the diverse range of factors that influence long-term life outcomes.

Bibliography

Andrasfay, T., Raymo, N., Goldman, N., & Pebley, A. R. (2021). Physical work conditions and disparities in later life functioning: Potential pathways. *SSM-Population Health*, *16*, 100990.

Arpino, B., Gumà, J., & Julià, A. (2018). Early-life conditions and health at older ages: The mediating role of educational attainment, family and employment trajectories. *PloS One*, *13*(4), e0195320.

Arranz, J. M., & García-Serrano, C. (2023). Assistance benefits and unemployment outflows of the elderly unemployed: The impact of a law change. *The Journal of the Economics of Ageing*, *26*, 100466.

Arulampalam, W. (2001). Is unemployment really scarring? Effects of unemployment experiences on wages. *The Economic Journal*, *111*(475), F585–F606.

Biewen, M., & Steffes, S. (2010). Unemployment persistence: Is there evidence for stigma effects? *Economics Letters*, *106*(3), 188–190.

Blau, F. D., & Kahn, L. M. (2017). The gender wage gap: Extent, trends, and explanations. *Journal of Economic Literature*, *55*(3), 789–865.

Bolano, D. (2015). *Markovian modelling of life course data* [PhD Thesis, University of Geneva, Switzerland]. https://access.archive-ouverte.unige.ch/access/metadata/64cd9c76-c9ec-4479-9b37-1f6d3b8b5c4d/download

Bolano, D. (2017). Using hidden Markov models to model life course trajectories. *Proceedings of the Population Association of America (PAA) Annual Meeting*, 27–29. https://paa.confex.com/paa/2017/mediafile/ExtendedAbstract/Paper10800/Bolano_Markov%20Models_PAA.pdf

Budig, M. J., & England, P. (2001). The wage penalty for motherhood. *American Sociological Review*, *66*(2), 204–225.

Buh, B. (2023). Measuring the effect of employment uncertainty on fertility in low-fertility contexts: An overview of existing measures. *Genus*, *79*(1), 4. https://doi.org/10.1186/s41118-023-00185-x

Buhr, P., & Huinink, J. (2014). Fertility analysis from a life course perspective. In *Advances in Life Course Research* (Vol. 21, pp. 1–9). Elsevier. https://www.sciencedirect.com/science/article/pii/S1040260814000112

Casas, P., & Román, C. (2023). Early retired or automatized? Evidence from the survey of health, ageing and retirement in Europe. *The Journal of the Economics of Ageing*, *24*, 100443.

Elder, G. H. (1994). Time, human agency, and social change: Perspectives on the life course. *Social Psychology Quarterly*, 4–15.

Elder, G. H. (1998). The Life Course as Developmental Theory. *Child Development*, *69*(1), 1–12. https://doi.org/10.1111/j.1467-8624.1998.tb06128.x

Elder, G. H. (2018). *Children of the great depression*. Routledge. https://books.google.com/books?hl=en&lr=&id=uOlgDwAAQBAJ&oi=fnd&pg=PP1&dq=Elder,+G.,+Jr.+(1974).+Children+of+the+Great+Depression.+Chicago:+University+of+Chicago+Press.&ots=rG0DiK2RgF&sig=q5h5J1xlywg8P94ngY26-IKbHv4

Frech, A., Damaske, S., & Ohler, A. (2022). The Life Course of Unemployment and Midlife Health. *Journal of Aging and Health*, *34*(6–8), 1081–1091. https://doi.org/10.1177/08982643221091775

Gregg, P. (2001). The Impact of Youth Unemployment on Adult Unemployment in the NCDS. *The Economic Journal*, *111*(475), F626–F653. https://doi.org/10.1111/1468-0297.00666

Hanger-Kopp, S., Thaler, T., Seebauer, S., Schinko, T., & Clar, C. (2022). Defining and operationalizing path dependency for the development and monitoring of adaptation pathways. *Global Environmental Change*, *72*, 102425.

Hasselhorn, H. M., Leinonen, T., Bültmann, U., Mehlum, I. S., du Prel, J.-B., Kiran, S., Majery, N., Solovieva, S., & de Wind, A. (2022). The differentiated roles of health in the transition from work to retirement–conceptual and methodological challenges and avenues for future research. *Scandinavian Journal of Work, Environment & Health*, *48*(4), 312.

Hutchison, E. D. (2010). A life course perspective. *Dimensions of Human Behavior: The Changing Life Course*, *4*, 1–38.

Mishra, G. D., Cooper, R., & Kuh, D. (2010). A life course approach to reproductive health: Theory and methods. *Maturitas*, *65*(2), 92–97. https://doi.org/10.1016/j.maturitas.2009.12.009

Nilsen, Ø. A., & Reiso, K. H. (2014). Scarring effects of early-career unemployment. *Nordic Economic Policy Review*, *1*(2014), 13–46.

Rasner, A. (2014). Gender pension gap in Eastern and Western Germany. *DIW Economic Bulletin*, *4*(11), 42–50.

Rice, N. E., Lang, I. A., Henley, W., & Melzer, D. (2011). Common health predictors of early retirement: Findings from the English Longitudinal Study of Ageing. *Age and Ageing*, *40*(1), 54–61.

Savcisens, G., Eliassi-Rad, T., Hansen, L. K., Mortensen, L. H., Lilleholt, L., Rogers, A., Zettler, I., & Lehmann, S. (2024). Using sequences of life-events to predict human lives. *Nature Computational Science*, *4*(1), 43–56.

Schmauk, S. (2024). Pathways to retirement in West Germany: Does divorce matter? *Advances in Life Course Research*, *60*, 100595.

Toczek, L., Bosma, H., & Peter, R. (2022). Early retirement intentions: The impact of employment biographies, work stress and health among a baby-boomer generation. *European Journal of Ageing*, *19*(4), 1479–1491. https://doi.org/10.1007/s10433-022-00731-0

Udayar, S., Canzio, L. I., Urbanaviciute, I., Masdonati, J., & Rossier, J. (2021). Significant life events and career sustainability: A three-wave study. *Sustainability*, *13*(23), 13129.

Waldfogel, J. (1998). Understanding the “Family Gap” in Pay for Women with Children. *The Journal of Economic Perspectives*, *12*(1), 137–156.

Yujian, L., & Bo, L. (2007). A normalized Levenshtein distance metric. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *29*(6), 1091–1095.

Appendix A

Distribution of Actual and Predicted Social Employment States by Age and Birth Year.

byear_1940

Category	56_Status	57_Status	58_Status	59_Status	60_Status	56_Prediction	57_Prediction	58_Prediction	59_Prediction	60_Prediction
10	0	0	0	1.02	2.04					
12	0.51	0.51	0.51	0.51	0					
13	4.08	3.57	3.06	3.06	2.04	4.08	4.08	4.08	4.08	4.08
15	1.02	1.53	1.53	1.53	3.06	1.02	1.02	1.02	1.02	1.02
16	91.84	91.84	92.86	91.84	90.82	92.86	92.86	92.86	92.86	92.86
3	1.02	1.53	1.02	1.02	1.02	1.02	1.02	1.02	1.02	1.02
6	1.53	1.02	1.02	1.02	1.02	1.02	1.02	1.02	1.02	1.02

byear_1941

Category	56_Status	57_Status	58_Status	59_Status	60_Status	56_Prediction	57_Prediction	58_Prediction	59_Prediction	60_Prediction
10	0	0	7.41	7.41	3.7					
13	7.41	7.41	11.11	11.11	7.41	7.41	7.41	7.41	7.41	7.41
15	7.41	7.41	7.41	7.41	7.41	7.41	7.41	7.41	7.41	7.41
16	77.78	77.78	66.67	66.67	70.37	77.78	77.78	77.78	77.78	77.78
6	7.41	7.41	7.41	7.41	11.11	7.41	7.41	7.41	7.41	7.41

byear_1942

Category	56_Status	57_Status	58_Status	59_Status	60_Status	56_Prediction	57_Prediction	58_Prediction	59_Prediction	60_Prediction
10	0	10.34	6.9	6.9	6.9					
13	3.45	3.45	0	0	0					

16	96.55	82.76	89.66	89.66	89.66	89.66	89.66	89.66	89.66	89.66
3	0	3.45	0	0	0	6.9	6.9	6.9	6.9	6.9
6	0	0	3.45	3.45	3.45	3.45	3.45	3.45	3.45	3.45

byear_1943

Category	56_Status	57_Status	58_Status	59_Status	60_Status	56_Prediction	57_Prediction	58_Prediction	59_Prediction	60_Prediction
10	9.52	9.52	9.52	9.52	9.52					
12	0	0	4.76	0	0					
13	4.76	4.76	4.76	0	0	4.76	4.76	4.76	4.76	4.76
16	85.71	80.95	76.19	85.71	85.71	95.24	95.24	95.24	95.24	95.24
3	0	4.76	4.76	0	0					
6	0	0	0	4.76	4.76					

byear_1944

Category	56_Status	57_Status	58_Status	59_Status	60_Status	56_Prediction	57_Prediction	58_Prediction	59_Prediction	60_Prediction
10	17.39	13.04	8.7	13.04	13.04	13.04	13.04	13.04	13.04	13.04
13	8.7	8.7	8.7	8.7	8.7	8.7	8.7	8.7	8.7	8.7
16	73.91	78.26	82.61	69.57	69.57	78.26	78.26	78.26	78.26	78.26
3	0	0	0	8.7	8.7					

byear_1945

Category	56_Status	57_Status	58_Status	59_Status	60_Status	56_Prediction	57_Prediction	58_Prediction	59_Prediction	60_Prediction
10	17.24	13.79	17.24	13.79	13.79	17.24	17.24	17.24	17.24	17.24
13	13.79	10.34	6.9	10.34	10.34	10.34	10.34	10.34	10.34	10.34

15	0	3.45	3.45	3.45	3.45					
16	68.97	72.41	68.97	72.41	68.97	65.52	65.52	65.52	65.52	65.52
6	0	0	3.45	0	3.45	3.45	3.45	3.45	3.45	3.45
12						3.45	3.45	3.45	3.45	3.45

byear_1946

Category	56_Status	57_Status	58_Status	59_Status	60_Status	56_Prediction	57_Prediction	58_Prediction	59_Prediction	60_Prediction
10	11.76	11.76	11.76	5.88	5.88	11.76	11.76	11.76	11.76	11.76
12	5.88	5.88	5.88	5.88	5.88	5.88	5.88	5.88	5.88	5.88
16	82.35	82.35	82.35	88.24	88.24	76.47	76.47	76.47	76.47	76.47
3						5.88	5.88	5.88	5.88	5.88

byear_1947

Category	56_Status	57_Status	58_Status	59_Status	60_Status	56_Prediction	57_Prediction	58_Prediction	59_Prediction	60_Prediction
10	11.11	11.11	11.11	16.67	11.11	11.11	11.11	11.11	11.11	11.11
12	5.56	5.56	5.56	5.56	5.56	5.56	5.56	5.56	5.56	5.56
13	16.67	16.67	16.67	16.67	16.67	27.78	27.78	27.78	27.78	27.78
15	0	0	0	0	5.56					
16	44.44	55.56	55.56	50	55.56	44.44	44.44	44.44	44.44	44.44
3	5.56	5.56	5.56	5.56	5.56	5.56	5.56	5.56	5.56	5.56
6	16.67	5.56	5.56	5.56	0	5.56	5.56	5.56	5.56	5.56

byear_1948

Category	56_Status	57_Status	58_Status	59_Status	60_Status	56_Prediction	57_Prediction	58_Prediction	59_Prediction	60_Prediction

10	10.34	10.34	6.9	3.45	3.45	10.34	10.34	10.34	10.34	10.34
13	13.79	13.79	13.79	10.34	3.45	10.34	10.34	10.34	10.34	10.34
15	0	0	0	3.45	10.34					
16	68.97	65.52	68.97	72.41	72.41	68.97	68.97	68.97	68.97	68.97
6	6.9	10.34	10.34	10.34	10.34	10.34	10.34	10.34	10.34	10.34

byear_1949

Category	56_Status	57_Status	58_Status	59_Status	60_Status	56_Prediction	57_Prediction	58_Prediction	59_Prediction	60_Prediction
10	12.5	7.5	7.5	7.5	7.5	17.5	17.5	17.5	17.5	17.5
12	2.5	0	0	0	0	2.5	2.5	2.5	2.5	2.5
13	22.5	17.5	20	22.5	17.5	22.5	22.5	22.5	22.5	22.5
15	2.5	2.5	0	0	2.5	2.5	2.5	2.5	2.5	2.5
16	52.5	60	62.5	62.5	65	50	50	50	50	50
6	7.5	12.5	10	7.5	7.5	5	5	5	5	5

byear_1950

Category	56_Status	57_Status	58_Status	59_Status	60_Status	56_Prediction	57_Prediction	58_Prediction	59_Prediction	60_Prediction
10	2.6	1.3	2.6	5.19	5.19	1.3	1.3	1.3	1.3	1.3
12	5.19	3.9	3.9	3.9	3.9	5.19	5.19	5.19	5.19	5.19
13	18.18	20.78	19.48	22.08	24.68	19.48	19.48	19.48	19.48	19.48
15	0	0	1.3	1.3	1.3					
16	67.53	68.83	68.83	66.23	64.94	67.53	67.53	67.53	67.53	67.53
3	1.3	1.3	0	0	0					
6	5.19	3.9	3.9	1.3	0	6.49	6.49	6.49	6.49	6.49

byear_1951

Category	56_Status	57_Status	58_Status	59_Status	60_Status	56_Prediction	57_Prediction	58_Prediction	59_Prediction	60_Prediction
10	6.19	5.15	6.19	6.19	7.22	4.12	4.12	4.12	4.12	4.12
12	7.22	7.22	7.22	7.22	7.22	7.22	7.22	7.22	7.22	7.22
13	30.93	27.84	27.84	27.84	27.84	30.93	30.93	30.93	30.93	30.93
15	2.06	3.09	4.12	4.12	4.12	2.06	2.06	2.06	2.06	2.06
16	48.45	51.55	50.52	49.48	47.42	49.48	49.48	49.48	49.48	49.48
3	2.06	2.06	2.06	2.06	3.09	2.06	2.06	2.06	2.06	2.06
6	3.09	3.09	2.06	3.09	3.09	4.12	4.12	4.12	4.12	4.12

byear_1952

Category	56_Status	57_Status	58_Status	59_Status	60_Status	56_Prediction	57_Prediction	58_Prediction	59_Prediction	60_Prediction
1	0.53	0	0.53	0	0	0.53	0.53	0.53	0.53	0.53
10	4.74	4.74	2.63	3.68	5.79	4.21	4.21	4.21	4.21	4.21
12	5.79	5.79	5.79	5.79	4.74	5.79	5.79	5.79	5.79	5.79
13	41.58	43.16	44.74	43.16	42.63	41.05	41.05	41.05	41.05	41.05
16	43.16	41.58	42.63	42.63	43.68	42.11	42.11	42.11	42.11	42.11
3	0.53	0.53	0	0	0	0.53	0.53	0.53	0.53	0.53
6	3.68	4.21	3.68	4.74	3.16	5.79	5.79	5.79	5.79	5.79

byear_1953

Category	56_Status	57_Status	58_Status	59_Status	60_Status	56_Prediction	57_Prediction	58_Prediction	59_Prediction	60_Prediction
10	6.61	6.9	7.42	7.65	7.94	6.31	6.31	6.46	6.46	6.46
12	2.97	3.12	3.19	2.97	3.04	2.82	2.82	2.82	2.82	2.82
13	46.62	46.84	46.18	46.25	44.77	48.7	48.7	48.7	48.7	48.7

15	2.75	3.12	3.34	3.41	3.86	2.67	2.67	2.67	2.67	2.67
16	31.33	31.4	31.55	31.25	30.88	31.03	31.03	31.03	31.03	31.03
3	1.56	1.63	2	2.08	2.23	1.71	1.71	1.71	1.71	1.71
6	8.17	6.98	6.31	6.38	7.28	6.76	6.76	6.61	6.61	6.61

byear_1954

Category	56_Status	57_Status	58_Status	59_Status	60_Status	56_Prediction	57_Prediction	58_Prediction	59_Prediction	60_Prediction
1	0.01	0.02	0.02	0.01	0.01	0.03	0.03	0.03	0.03	0.03
10	9.81	11.33	11.04	10.86	10.58	10.13	10.4	10.4	10.4	10.4
12	2.72	2.68	2.66	2.67	2.7	2.8	2.8	2.8	2.8	2.8
13	37.3	37.57	37.75	37.63	37.69	36.91	36.91	36.91	36.91	36.91
15	0.35	0.35	0.39	0.42	0.5	0.35	0.35	0.35	0.35	0.35
16	38.56	38.61	38.87	38.76	38.81	38.37	38.37	38.37	38.37	38.37
2	0.01	0.01	0.01	0.01	0					
3	1.57	1.52	1.43	1.49	1.47	1.59	1.59	1.59	1.59	1.59
6	9.68	7.91	7.84	8.15	8.25	9.82	9.54	9.54	9.54	9.54
9	0	0.01	0	0	0					

byear_1955

Category	56_Status	57_Status	58_Status	59_Status	60_Status	56_Prediction	57_Prediction	58_Prediction	59_Prediction	60_Prediction
10	7.94	7.87	8.09	7.88	7.3	7.37	7.37	7.37	7.37	7.37
12	2.76	2.72	2.78	2.77	2.81	2.76	2.76	2.76	2.76	2.76
13	59.7	59.63	59.56	59.42	59.12	60.22	60.22	60.22	60.22	60.22
15	0.7	0.68	0.67	0.75	0.92	0.47	0.47	0.47	0.47	0.47
16	20.38	20.7	20.58	20.77	20.9	20.21	20.24	20.24	20.24	20.24
2	0.01	0	0	0	0					

3	1.47	1.4	1.4	1.41	1.43	1.44	1.44	1.44	1.44	1.44
6	7.03	7	6.92	7	7.53	7.51	7.48	7.48	7.48	7.48
1						0.01	0.01	0.01	0.01	0.01

byear_1956

Category	56_Status	57_Status	58_Status	59_Status	60_Status	56_Prediction	57_Prediction	58_Prediction	59_Prediction	60_Prediction
1	0	0	0.01	0.01	0					
10	2.59	2.65	2.72	2.63	2.78	2.6	2.6	2.6	2.6	2.6
12	1.65	1.68	1.7	1.79	1.83	1.66	1.66	1.66	1.66	1.66
13	89.01	88.59	88.39	87.99	87.24	90.33	90.33	90.33	90.33	90.33
15	0.27	0.3	0.34	0.4	0.47	0.23	0.23	0.23	0.23	0.23
16	2.31	2.41	2.4	2.4	2.49	2.32	2.33	2.33	2.33	2.33
2	0.01	0.01	0.01	0.01	0.01					
3	0.75	0.72	0.77	0.77	0.72	0.74	0.74	0.74	0.74	0.74
6	3.41	3.63	3.68	4.02	4.46	2.12	2.11	2.11	2.11	2.11

byear_1957

Category	56_Status	57_Status	58_Status	59_Status	60_Status	56_Prediction	57_Prediction	58_Prediction	59_Prediction	60_Prediction
1	0.01	0.03	0.01	0	0.01	0.02	0.02	0.02	0.02	0.02
10	7.78	7.72	7.79	7.67	7.16	7.69	7.69	7.69	7.69	7.69
12	1.86	2.04	2.41	2.64	2.79	1.82	1.83	1.83	1.83	1.83
13	75.1	73.68	71.85	69.68	66.07	77.1	77.1	77.1	77.1	77.1
15	0.58	0.82	1.02	1.24	1.58	0.46	0.46	0.46	0.46	0.46
16	5.79	6.26	6.55	7.4	7.81	5.57	5.57	5.57	5.57	5.57
2	0.01	0	0	0	0					
3	1.62	1.66	1.71	1.8	2.76	1.5	1.5	1.5	1.5	1.5

6	7.25	7.8	8.65	9.57	11.82	5.84	5.83	5.83	5.83	5.83
9	0.01	0	0.01	0.01	0					

byear_1958

Category	56_Status	57_Status	58_Status	59_Status	60_Status	56_Prediction	57_Prediction	58_Prediction	59_Prediction	60_Prediction
1	0.04	0	0	0	0	0.01	0.01	0.01	0.01	0.01
10	5.39	5.26	5.41	5.12	4.84	5.45	5.45	5.45	5.45	5.45
12	1.17	1.6	2.4	2.53	2.62	1.16	1.16	1.16	1.16	1.16
13	71.62	70.73	68.12	63.64	57.21	74.41	74.41	74.41	74.41	74.41
15	1.85	2.12	2.3	2.8	3.45	1.69	1.69	1.69	1.69	1.69
16	9.29	8.89	9	9.04	9.02	9.3	9.3	9.3	9.3	9.3
3	1.03	1.11	1.28	2.05	2.15	1.1	1.1	1.1	1.1	1.1
6	9.61	10.29	11.46	14.82	20.71	6.87	6.87	6.87	6.87	6.87
9	0	0.01	0.01	0	0					

byear_1959

Category	56_Status	57_Status	58_Status	59_Status	60_Status	56_Prediction	57_Prediction	58_Prediction	59_Prediction	60_Prediction
10	4.88	4.62	4.57	4.45	4.39	5.36	5.36	5.36	5.36	5.36
12	1.56	1.68	1.91	2.1	2.07	1.45	1.45	1.45	1.45	1.45
13	68.43	67.49	63.31	57.6	45.37	72.43	72.43	72.43	72.43	72.43
15	2.28	2.42	2.81	3.64	6.43	1.97	1.97	1.97	1.97	1.97
16	9.25	9.61	9.7	10.12	10.39	9.23	9.23	9.25	9.25	9.25
2	0.01	0.01	0	0	0					
3	1.35	1.4	2.2	2.51	2.69	1.25	1.25	1.25	1.25	1.25
6	12.24	12.76	15.48	19.57	28.67	8.31	8.31	8.31	8.31	8.31
9	0	0.01	0.01	0	0					

1						0.01	0.01	0	0	0

byear_1960

Category	56_Status	57_Status	58_Status	59_Status	60_Status	56_Prediction	57_Prediction	58_Prediction	59_Prediction	60_Prediction
10	4.78	4.72	4.56	4.26	4.07	5.13	5.13	5.13	5.13	5.13
12	1.09	1.15	1.15	1.17	0.9	0.91	0.91	0.91	0.91	0.91
13	67.44	63.77	58.31	44.29	34.92	70.22	70.22	70.22	70.22	70.22
14	0	0	0	0	10.33					
15	3.09	3.42	4.61	8.79	28.3	2.62	2.62	2.62	2.62	2.62
16	10.08	10.17	10.21	10.22	0.27	10.27	10.27	10.27	10.27	10.27
3	1.25	2.05	2.03	2.21	2.51	1.31	1.31	1.31	1.31	1.31
6	12.27	14.72	19.13	29.07	18.71	9.54	9.54	9.54	9.54	9.54