Aix-Marseille School of Economics

Apprenticeship Report M2 Econometrics, Statistics and Big data

Coraline Best

Apprenticeship conducted from September, 2023 to September, 2024

Apprenticeship Supervisor:	Mr. Raphaël Tremoulu
Academic Advisor:	Mr. Badih Ghattas
Educational Institution:	Aix-Marseille School of Economics
Host Company:	La direction de la Recherche, des Études, de l'Évaluation et des Statistiques (Drees)

Non-plagiarism undertaking

I, Coraline Best declare being fully aware that plagiarism by copying documents or a portion of a document published in all forms and media, including online publications, constitutes copyright infringement and related rights, as well as outright fraud. Consequently, I hereby undertake to quote all sources and authors that I've used to write my apprenticeship report and its appendices.

Date: 5/06/2024

Signature:

Acknowledgments

I would like to express my sincere gratitude to all those who have contributed to the completion of this thesis.

First and foremost, I am deeply thankful to my university advisor, Mr. Badih Ghattas, for his invaluable guidance and support. My sincere appreciation also goes to my mentors, Mrs. Viard-Guillot Louise, Mr. Benjamin Vignolles, and Mr. Raphaël Tremoulu, for their unwavering support and expertise throughout this research journey. Their insights and encouragement have been instrumental in shaping this work, and I am profoundly grateful for the knowledge they have shared with me.

I also want to extend my gratitude to Mr. Léo Quennesson, an outstanding internship supervisor from whom I gained invaluable knowledge and experience. I would also like to express my appreciation to my colleagues Mrs. Dumont Gwénnaëlle, Mrs.Ramuzat Lauriane, M. Rousset Clément , Mrs. Dufour Camille, and all the others for their immense support and expertise. It was a pleasure working and learning with them.

A special thanks to my family and loved ones, especially my parents, who have taught me to persevere through challenges and never give up.

Finally, I want to acknowledge the numerous researchers, authors, and institutions whose work has been an essential source of information and inspiration.

This thesis would not have been possible without the support and contributions of all those mentioned above. Thank you for being a part of this journey.

Coraline Best
September, 2024
Aix-Marseille School of Economics
DREES (direction de la Recherche, des Etudes, de l'Evaluation et des Statistiques)

Abstract

This master thesis explores the complexities of career trajectories in the French labor market and examines the impact of potential modifications in unemployment policy on the disposable income of various case-types.

The study employs sequence analysis to uncover and categorize distinctive patterns of career paths, leveraging an extensive dataset of longitudinal job records. Multiple dissimilarity criteria were employed to assess these sequences, leading to the identification of seven optimal clusters. Each cluster reflects a unique job pathway that is defined by specific combinations of employment, unemployment, and education-related statuses. The research highlights significant disparities in career trajectories, with noteworthy differences in the distribution of employment status among men and women.

A dynamic model was developed to provide a comprehensive analysis of the impacts of unemployment reforms. This model accurately computes the social benefits of hypothetical households in both pre- and post-reform scenarios. The study evaluates the effects of unemployment policy modifications on different socio-economic groups within the French population. After analyzing the potential unemployment reform, a reduction of 92.4 euros in disposable income following the adoption of the policy. However, the study also demonstrates how social benefits, such as RSA (Revenu de Solidarité Active) and AL (Aides au Logement), mitigate the financial harm.

Abstract

Ce rapport d'alternance explore les complexités des trajectoires professionnelles sur le marché du travail français et examine l'impact des modifications potentielles de la politique de chômage sur le revenu disponible de différents types de cas.

L'étude utilise l'analyse de séquences pour découvrir et catégoriser des typologies de parcours professionnels, en s'appuyant sur un vaste ensemble de données. Plusieurs mesures de distance ont été employées pour évaluer ces séquences, ce qui a conduit à l'identification de sept groupes optimaux. Chaque cluster reflète un parcours professionnel majoritaire défini par des combinaisons spécifiques de statuts d'emploi, de chômage et d'éducation. Ce rapport met en évidence des disparités significatives dans les trajectoires de carrière, avec des différences notables dans la répartition des statuts d'emploi entre les hommes et les femmes.

Un modèle dynamique a été développé pour fournir une analyse complète des impacts des réformes de l'assurance chômage. Ce modèle calcule avec précision les bénéfices sociaux des ménages hypothétiques dans les scénarios pré et post-réforme. L'étude évalue les effets des modifications de la politique de chômage sur différents groupes socio-économiques au sein de la population française. Après avoir analysé la réforme potentielle de l'emploi, une réduction de 92,4 euros du revenu disponible suite à l'adoption de la politique. Cependant, l'étude montre également comment les avantages sociaux, tels que RSA (Revenu de Solidarit e Active) et AL (Aides au Logement), atténuent les pertes de revenu disponible.

Table of contents

1 Introduction **4**
1.1 Context . . . 4
1.2 Objectives and missions of the apprenticeship . . . 5
1.3 Background . . . 6
1.3.1 Mapping career paths: analyzing job trajectories . . . 6
1.3.2 Social Benefits Overview . . . 7
1.3.3 Unemployment Insurance in France: the 2024 changes in policy parameters . . . 8

2 Literature review **9**

3 Data **10**
3.1 Tax and Social Income Survey (ERFS) . . . 10
3.1.1 Data sources and sample . . . 10
3.1.2 Descriptive statistics . . . 14
3.2 Dynamic case-type model description . . . 17
3.3 Sample for Evaluating the Unemployment Reform . . . 19

4 Methodology **20**
4.1 Sequence analysis . . . 20
4.1.1 Dissimilarity measures . . . 21
4.1.2 Cluster Analysis . . . 22
4.1.3 Quality Assessment . . . 23
4.1.4 Tree-structured discrepancy analysis . . . 24
4.2 Impact analysis of unemployment benefit policy changes on case-types . . . 25

5 Results **26**
5.1 Sequence analysis . . . 26
5.1.1 Typologies of job trajectories . . . 26
5.1.2 Parametric bootstrap . . . 30
5.1.3 Tree-discrepancy analysis . . . 33
5.2 Result of the analysis of the reform . . . 34

6 Conclusion **36**
6.1 Conclusion on the main results . . . 36
6.2 Improvement and limitations . . . 36

7 Bibliographic References **38**

References **38**

8 Glossary **40**

9 Appendix **41**

1 Introduction

The purpose of this master thesis is to present the main assignments I was given during my apprenticeship period as part of my Master 2 in Econometrics, Big Data and Statistics at the Aix-Marseille School of Economics.

1.1 Context

I completed a one year apprenticeship at the Directorate of Research, Studies, Evaluation, and Statistics (DREES), a directorate of the French central administration of the Health and Social ministries, established in 1998. DREES is a component of the public statistical service, which includes INSEE and 16 ministerial statistical services. DREES is an integral part of the ministerial statistical service for the health and social sectors and operates under the supervision of the Ministry of Labour, Health, and Social Affairs, as well as the Ministry of Economy.

Drees carries out various important tasks to fulfill its position in the public statistical service, which include:

Statistical production: this entails the systematic arrangement, analysis, and dissemination of data. The organization also conducts comprehensive work such as the National Health Accounts, the Social Protection Accounts, and the Health Status of the Population in France.

Public disclosure: DREES disseminates the studies and research it carries out, guaranteeing that crucial information is readily available to the public.

Conducting research: is another crucial objective of the institution. The institution collaborates with the ministry in charge of research and other specialized institutions to provide guidance on research policy and contribute to the development and promotion of research findings. Additionally, it is accountable for overseeing the organization and control of data, algorithms, and codes within the ministry.

DREES has a crucial role in generating and examining statistical data pertaining to health, solidarity, and social protection. By engaging in various activities, it actively enhances public policies by offering unbiased information, thorough evaluations, and well-informed recommendations to educate decision-making and drive reforms.

The organization is structured into several key divisions and offices, each responsible for specific areas within health and social affairs. Within the institution there are three sub-departments: the Sub-Directorate for Health and Health Insurance Observation; the Sub-Directorate for Solidarity Observation and the Sub-Directorate for Summaries, Economic Studies, and Evaluation.

I worked within the Sub-Directorate of Summaries, Economic Studies, and Evaluation (SEE) which includes:

- The redistribution and evaluation office (BRE)
- The social accounts analysis office (BACS)
- The research mission (MiRe)
- The international relations and studies (MREI)

During my apprenticeship, I was affiliated in the Redistribution and Evaluation Office, specifically within the microsimulation unit. This unit comprises six agents who collaboratively manage the Ines microsimulation model in partnership with INSEE[1] and CNAF[2].

The microsimulation unit creates publications about the redistribution of wealth and social policies. Additionally, it generates reports on social and fiscal measures in response to specific requests from ministerial cabinets. The Ines model enables the simulation of different socio-fiscal legislation elements on a representative sample of the population. It estimates the various components of household disposable income and evaluates the specific effects of each measure on public finances, income inequalities, and poverty. The model's source data, the INSEE survey on tax and social incomes (ERFS), includes socio-demographic information (family configuration, employment status) and economic information (levels of activity and replacement income) received by households, necessary for implementing simulations. The model is employed in detailed studies to contribute to economic and social discussions concerning issues such as monetary redistribution, taxation, and social protection.

Finally the Ines model is requested by:

- Ministers overseeing the DREES, as a decision-making tool for calibrating reforms;
- Various high councils, as a support tool for their deliberations (e.g., High Council for the Family, High Council for the Financing of Social Protection, Council of Mandatory Levies);
- Control agencies for evaluation purposes (Cour des Comptes)[3].

Due to the complexity and specificity of certain households in the ERFS, which feeds the Ines microsimulation model, they also use a static model that simulates the current socio-fiscal legislation on fictional case-types. This model provides exportable outputs in the form of tables and graphs that present the stylized redistributive effects of the socio-fiscal system. The model serves several purposes. One use is to understand the redistributive effects of the French socio-fiscal system. Another use is to compare the redistributive effects on different types of households.

1.2 Objectives and missions of the apprenticeship

Throughout my apprenticeship, my primary focus was on examining the correlation between labor market transitions and the redistribution carried out by the French socio-fiscal system. I conducted an analysis on the consequences of the unemployment insurance reform, which was planned to be implemented through a decree on July 1st. This change would specifically impact persons who became unemployed after December 1st. Although the decree has been suspended due to the legislative election, I will still evaluate the potential effects of the policy change.

One of the goals of my apprenticeship was to create a typology of typical job seeker trajectories in France. To achieve this, I used a statistical information system that integrates comprehensive individual administrative data on employment trajectories and unemployment benefits with data on the provision of social and family benefits as well as other information such as the age, structure of the family, sex etc. To do the analysis, I used an optimal matching sequence analysis to determine the main trajectories on the sample studied. Income and population makeup were aged to fit the 2023 distribution using weights and average income evolution, to identify the share of individuals broadly stable in work, those following unstable work pathways and those who experience unemployment.

[1] Institut national de la statistique et des études économiques

[2] Caisse nationale d'allocations familiales

[3] Cour des Comptes is a financial jurisdiction of the administrative order in France, primarily responsible for verifying the regularity of public accounts.

Additionally, I was responsible for developing a dynamic case-type model to simulate all mandatory deductions and social benefits as well as unemployment benefits on fictional households. This is a highly innovative project because, until now, the static model could not simulate unemployment insurance for fictional households. With this new feature, it is possible understand the complexity of the interaction between social benefits and unemployment insurance benefits. The dynamic model allows for the evaluation of the redistribution carried out by all social benefits and their incentivizing effect on employment take-up over time. This model was designed to incorporate changes in unemployment benefits due to reforms and to simulate future reforms. Developed in R, this model is based on a monthly structure adapted from the DGEFP[4], the central administration of the Ministry of Labor, allowing for the analysis of activity and salary trajectories, and showing how socio-fiscal redistribution adjusts over time to smooth the impact of labour market transitions on individuals' income. It shows the monthly Disposable income or Standard of living (after benefits received and taxes paid) of a typical household as a function of its wage and over time. The tool simulates typical trajectories for hypothetical households, where it is possible to choose as input, variables such as age, marital status, income level, number and age of children and the income trajectories of households.

This report focuses primarily on analyzing the trajectories of jobseekers using optimal matching analysis in France. Following this, I perform simulation within the dynamic model I developed using representative cases-types. The primary goal is to estimate the impact of changes in unemployment policy parameters, which was set to affect individuals losing their jobs after December 2024, on their disposable income.

The study examines individual data focusing on trajectories in the French labor market, representative of 2023 households, using sequence techniques. It addresses the following research question:

- **What are the typical trajectories observed in the sample of individuals, and what impact will the potential upcoming unemployment reform have on the disposable income of French individuals experiencing unemployment?**

All of my research aims to document the effects of these reforms on household disposable income based on their configurations. Additionally, it is intended to contribute to enhancing the Ines socio-fiscal microsimulation model to better incorporate these reforms.

1.3 Background

1.3.1 Mapping career paths: analyzing job trajectories

Throughout periods of labor market engagement, a contract might be interrupted or terminated by many occurrences. Various factors such as layoffs, job changes, sick leave, training, or sabbaticals might have various effects on individual income. Transitions are essential components of one's professional life. To examine these inquiries, the study must take into account the personal, economic, and societal aspects that affect persons. Through the creation of typologies, one may effectively represent the primary career paths within the French labor market by grouping individuals who share similar trajectories.

Measuring these transitions will assist the DREES in improving the accuracy of the simulations of career changes for the impacted population. Through the analysis of these transitions, we can gain a deeper comprehension of how compensation fluctuations during various career changes relate to procedures that are intended to reduce and stabilize income losses over a period of time. During these transitional periods, social benefits, unemployment insurance, and taxes and contributions all play a

[4] Délégation générale à l'Emploi et à la Formation professionnelle

role in stabilizing individual incomes.

Gaining a comprehensive understanding of job trajectories is of utmost importance for individuals and policymakers alike. This knowledge offers significant insights into the intricacies of labor markets and employment trends. Through the process of mapping career trajectories, we can examine the progression of individuals as they navigate various positions, industries, and employment statuses during their professional journey. This research facilitates the identification of patterns in job security, professional advancement, and shifts within industries, providing a more distinct understanding of the ways in which different factors impact career progression.

This report will assess the consistency and efficiency of the socio-fiscal system in facilitating professional changes. The evaluation will assess the system's effectiveness in assisting individuals during different career transitions, guaranteeing the preservation of income stability even in the face of disruptions. The results will offer valuable understanding regarding the efficacy of existing policies and propose possible enhancements to bolster the socio-economic assistance provided to those undergoing career changes.

1.3.2 Social Benefits Overview

The social assistance system in France is based on various social benefits, with eligibility criteria, amounts, and income thresholds that vary. These benefits consider different households situations, such as income and family composition. The rules and scales of social benefits, which are often adjusted according to household characteristics, can seem complex. This complexity is partly due to the objectives of the different benefits, which aim to redistribute resources considering the family's burdens while encouraging participation in the labor market.

The following description of social benefits is based on the 2023 legislation.

Unemployment benefits (ARE) represents the primary form of financial support for employees who have lost their jobs. Other types of unemployment benefits will not be studied in this report. Eligible employees who lose their jobs involuntarily can receive an income in the form of an allowance, subject to certain conditions. On average, this allowance represents 72% of the former net salary. In 2023, it can last up to 18 months, or 27 months for those aged 53 and over. To qualify, individuals must register with France Travail and meet the following conditions:

- Job loss must be involuntarily.
- Must have worked for at least 6 months as an employee in the last 24 or 36 months depending on the age of the job seeker.
- Must reside in mainland France, an overseas department or an overseas collectivity covered by the scheme.

The duration of entitlement and the amount of the allowance now depend on the intensity of work during the period prior to entitlement. For those with discontinuous work periods, both the amount and the duration of entitlement are extended, because the periods not worked are taken into account when calculating the reference salary and the entitlement up to a certain period that cannot exceed 75% of the affiliation duration.

The Active Solidarity Income (RSA) is a last resort income guarantee and a differential allowance that supplements a household's initial resources to reach a guaranteed income threshold. The scale varies according to household composition and the number of dependent children. The RSA resource base includes all income received over the past three months, including most social benefits. Additionally, receiving the RSA entitles individuals to an annual 'Christmas bonus' of 156 euros, which is accounted for on a monthly basis in the examples provided. Thus, in July 2023, a single person with no other resources receiving housing assistance would receive an RSA amount of 548 euros per month.

The activity bonus (PA) is an income supplement intended for workers with modest incomes. This bonus is designed to encourage employment and is coordinated with the RSA. In 2023, a single person earning the minimum wage receives an amount of 228 euros in PA.

Housing allowance (AL) helps cover part of a household's housing expenses. The amount is highly adjusted according to income, family size and the area of residence. This benefit has complex calculation rules. For a single person with no resources, who is a tenant (excluding Paris) and whose rent, including charges, is 325 euros or more per month, the housing allowance (AL) amounts to 281 euros. AL is the only social benefit that is computed based on a 12-month continuous income period.

Income tax (IR) is a direct tax that applies to all household incomes. The tax calculation considers the presence of children. Since the income tax scale is progressive, the average tax rate increases with income.

The presence of a children in the family generates costs that the social and fiscal redistribution system mitigates through the payment of family benefits and by modulating other benefits according to household composition. Each benefit is constructed according to its own Equivalence scale, which differs from the one used to calculate the standard of living.

Family allowances (AF) support the standard of living of households with children and are paid to all families with at least two dependent children under 20. The amount paid depends on the number and age of the children; since July 2015, on the parents' income. In 2023, the amount of family allowance is 140 euros for two children.

Family support allowance (ASF) is for single-parent families, paid to those raising at least one child under the age of 20 without the support of one of the parents. It is granted without means-testing. The ASF amount is 195 euros in 2023 per child.

Back-to-school allowance (ARS), means-tested, applies to families with children in school aged 6 to 18. The amount paid depends on the age of the child.

Family supplement (CF) is a means-testing allowance granted to for having at least 3 dependent children aged from 3 to 21.

1.3.3 Unemployment Insurance in France: the 2024 changes in policy parameters

As in many European countries, France's unemployment system imposes a limited period over which unemployed individuals can benefit from unemployment benefits. Unemployment benefits depends on previous income, amounting to 72% of former income. Unemployed individual are required to search actively for a new job, and non-compliance can result in the suspension of benefits. According to 2023 legislation, someone who is unemployed and under 53 can receive unemployment benefit up to 18 months provided that he had worked within the last 24 months. For those over 53, the period extends to 27 months, given they have worked in the last 36 months. When economic conditions are favorable, the duration of compensation for new entrants is reduced by 25%, with a minimum of 6 months of compensation. During economic downturns, an end-of-right supplement can extend entitlements. Additionally, under certain conditions, the unemployment benefits can be decrease from the seventh month for individuals under 57.

The French government introduced a new reform to the unemployment insurance scheme, that intended to affect individuals losing their jobs after December 1, 2024. The Prime Minister announced a significant reduction in compensation entitlements for all workers becoming unemployed after this date, to be enacted by decree. The decree was supposed to be issued at the beginning of July. However, immediately after the results of the first round of the election, the Prime Minister shifted course. The unemployment insurance reform is now suspended, and the current rules will be extended to avoid a legal void. For the purposes of this paper, I will assume that the decree has not been suspended and will evaluate the changes to the policy parameters. This reform is not retroactive and will not affect those already receiving unemployment benefits before December 2024, who will continue to be covered by the old system. Current beneficiaries will maintain their benefits under the existing rules, including the duration of their compensation. To qualify for compensation, jobseekers will now need to work for eight months within a 20-month period, instead of the previous 24 months.

One significant change is the decrease in the maximum length of unemployment benefits to 15 months for persons aged 57 and below, as opposed to the present 18 months for those aged 53 and below. Anticipated savings will be achieved by reducing the maximum period of eligibility, which will in turn limit the number of new claims and shorten the duration of current ones. The second change imposes stricter conditions for entitlement requiring a minimum of eight months of work, compared to six months currently. The third change shortens the reference period for eligibility to 20 months from the current 24 months. The fourth change affects older workers more significantly. Currently, employees aged 53 and 54 can receive benefits for up to 22.5 months, and those aged 55 and over for up to 27 months. The first stage will disappear, and the benefit of longer compensation will be reserved for unemployed people aged 57 and over. To encourage unemployed senior citizens to return to work, the French government has introduced a "senior employment bonus". This bonus supports the return to work by supplementing a lower salary than the previous one for one year. The supplement, paid by the unemployment insurance scheme, compensates for lost earnings on salaries of up to 3,000 euros.

This paper primarily examines the impacts of three significant modifications: the decrease in the maximum length of unemployment benefits, the reduction in the time period utilized to calculate these benefits, and the adjustment to the minimum number of months of employment needed to be eligible for them. The revised policy parameters will be equally applicable to anyone under the age of 57 who experience employment loss on or after December 2024.

I do not estimate the effects on senior workers because simulating the "senior employment bonus" in the dynamic model is not feasible. Therefore I do not invest in analyzing these effects.

2 Literature review

Job trajectories refer to the sequence of job positions an individual occupies over their career. These trajectories can be influenced by various factors including economic conditions, individual skills, educational background, and demographic characteristics. Studies such as those by (Arthur & Rousseau,) and (Hall,) emphasize the shift from traditional, linear career paths to more dynamic, non-linear career trajectories influenced by globalization and technological advancements. Research by (Bidwell & Briscoe,) highlights that career mobility, both within and across organizations, is becoming increasingly common. This is often driven by the need for continuous skill development and adaptation to changing market conditions. Their findings suggest that lateral moves can be as beneficial as upward moves in terms of skill acquisition and career satisfaction.

The literature also points to significant gender differences in job trajectories. Research by (Moen & Sweet,) and (Stone,) discusses how women's career paths are often interrupted or altered by familial responsibilities, leading to non-linear trajectories that differ from their male counterparts. These studies suggest that policies promoting work-life balance and parental leave are crucial for supporting women's career continuity. Economic cycles significantly impact job trajectories. Studies like those by (Kahn,) and (Oreopoulos, von Wachter, & Heisz,) illustrate how entering the job market during a recession can have long-term negative effects on earnings and career progression. These findings underscore the importance of macroeconomic stability for individual career development. In France, the educational system plays a critical role in shaping job trajectories. Research by (Duru-Bellat, Kieffer, & Reimer,) and (Bessin & Legrand,) highlights the strong link between educational attainment and career paths, with higher education often leading to more stable and upwardly mobile trajectories. However, these studies also point to the rigidity of the French labor market, where initial job placement significantly influences future career opportunities. French labor market policies, including strong employment protection legislation and vocational training programs, have a significant impact on job trajectories. Studies by (Gautié & Schmitt,) and (Cahuc & Zylberberg,) discuss how these policies can both support and hinder career mobility, depending on their design and implementation. The transition from education to employment is a critical phase in shaping job trajectories. Research by (Giret, Nohara, & Van de Velde,) and (Dupray & Moullet,) focuses on the challenges faced by French youth, particularly those from disadvantaged backgrounds, in securing stable employment. Their findings suggest that targeted interventions are necessary to facilitate smoother transitions and support long-term career development.

Optimal Matching Analysis (OMA) is a sequence analysis technique originally developed in biology for comparing DNA sequences, but later adapted for social sciences to study life-course trajectories, including job trajectories. It involves computing the cost of transforming one sequence into another through operations like insertions, deletions, and substitutions (Abbott & Tsay,). OMA has been widely used in social science research to study career paths, family formation, and other life-course events. Studies by (Abbott & Hrycak,) and (Brzinsky-Fay,) demonstrate the utility of OMA in identifying common patterns and pathways in large datasets. These applications highlight how OMA can reveal insights into the timing, sequencing, and duration of different life events. The advantages of OMA include its flexibility in handling various types of sequences and its ability to summarize complex patterns in a straightforward manner. However, researchers such as (Wu,) and (Studer & Ritschard,) have pointed out limitations, including the arbitrariness in setting costs for operations and potential sensitivity to these parameters. These critiques have led to the development of alternative methods and enhancements to the traditional OMA approach. Comparative studies using OMA, such as those by (McVicar & Anyadike-Danes,) and (Piccarreta & Lior,), highlight differences in job trajectories across countries and social groups. These studies often combine OMA with other statistical techniques, such as cluster analysis and regression models, to provide a more comprehensive understanding of the factors influencing job trajectories. Studies by (Cousteaux,) and (Rapoport & Le Goff,) apply OMA to analyze the career paths of different cohorts, revealing insights into the impact of economic policies and labor market conditions on career development. These studies contribute to the broader understanding of how structural factors shape individual job trajectories in France.

3 Data

This section describes the samples used in the different analysis and presents some descriptive statistics.

3.1 Tax and Social Income Survey (ERFS)

3.1.1 Data sources and sample

This report uses the Tax and Social Income Survey (ERFS[5]) that is based on the Employment Survey (EEC[6]) of 2021 in order to map job trajectories in France. The Employment Survey is essential for understanding employment and unemployment, serving as the only source that measures these metrics

[5] Enquête revenus fiscaux et sociaux
[6] Enquête Emploi en Continu

according to the standards of the International Labour Organization (BIT[7]). The EEC is a continuous panel survey in which the same household is surveyed every three months for six consecutive quarters. It sheds light on various dimensions of the labor market, such as underemployment, unemployment, and the characteristics of occupied jobs (profession, types of contracts, sectors of activity). Additionally, it encompasses numerous sociodemographic characteristics, including education, country of birth, and family situation. The geographical scope of the Employment Survey covers metropolitan France and the overseas departments (Guadeloupe, Martinique, French Guiana, Réunion) excluding Mayotte. Only individuals aged between 15 and 89 are surveyed. To avoid double counting, such as in cases of multiple residences, individuals are surveyed at their usual residence.

The 2021 ERFS sample consists of respondents from the fourth quarter of the 2021 EEC survey, totaling 93,160 individuals, and includes only those from metropolitan France. A matching process is then carried out with 2021 tax data sources, and this sample provides the highest matching rate. This process involves linking the EEC data with tax records, which include information to identify the declarant as well as details about their income tax. Benefit amounts received by households during the income year in question are collected from the National Family Allowances Fund (CNAF), the National Old-Age Insurance Fund (CNAV[8]), and the Central Fund for Agricultural Social Mutuality (CCMSA[9]). Individuals who were not matched during the initial matching phase, despite being part of a responding household, have their incomes imputed in a subsequent phase of the microsimulation model. After the matching process, over 77,208 individuals were found, representing 82.9% of those aged 18 and older.

The Ines microsimulation model simulates French social and tax legislation. It specifically allows for the evaluation of the budgetary and redistributive effects of various reforms, whether already implemented or under discussion, concerning numerous taxes and social benefits. Based on the ERFS survey, the model simulates the various benefits each household is entitled to, as well as the taxes and levies they must pay. This approach captures the diversity and complexity of real-life situations faced by the French population, using a representative sample of individuals living in ordinary housing in metropolitan France. The sample from the survey, initially representative of ordinary households in France, is aged over time. Certain necessary information for calculating benefits and taxes is imputed. Ines is a static model and does not account for behavioral changes of households, such as labor market participation, which could be influenced by changes in socio-fiscal legislation. The model is now written in R and consists of a sequence of 87 programs. Those programs can be modified in order to simulate reforms, like an increase of RSA for instance. It is then possible to measure for each household the consequence of a reform on the standard of living. Therefore it is possible to identify the potential winners and losers of the simulated reforms.

I specifically focused on the program that reconstructs a monthly activity status calendar for each individual for the year 2021. A filter is applied to the field of individuals over 16 years old to capture only those who can legally be employed. Several imputations are made to complete this calendar, including: retirement assignment for those over 70 years old, considering that they are retired for the entire year; extension of the retirement status when the individual has the retirement status for two consecutive months; filling in sporadically missing months at the end of the calendar (for the last 1 to 3 months of the year); and filling in sporadically missing months at the beginning or middle of the calendar (1 to 3 missing months) and some incomplete calendars are filled with the subsequent/previously recorded status.

To obtain a representative sample for the year 2023, weights calculated in a program of the Ines microsimulation model can be used. This involves "aging" the survey sample, originally representative

[7]Bureau international du travail
[8]Caisse nationale d'assurance vieillesse
[9]Caisse centrale de la mutualité sociale agricole

of ordinary households in metropolitan France in 2021, to reflect these same households in 2023.

This aging process takes two forms:

- 1. Evolution of population characteristics by adjusting weights. For example, if the number of unemployed individuals has increased between year N-2 and year N, the weight of unemployed individuals in the sample is increased. This step is conducted using a "margin calibration."
- 2. Evolution of individual incomes. Differential growth rates are applied to each type of income based on its nature (earnings from employment, from assets, replacement income), using all available economic data. For instance, wage adjustments are made using the Acemo survey from DARES [10], considering each employee's socio-professional category and industry sector.

Once this aging procedure is implemented, certain necessary information for calculating benefits and taxes—such as childcare arrangements for children under six and degree of disability—is imputed, especially where such data is unavailable or incomplete in the ERFS.

After imputations, the monthly activity status calendar consists of 77,206 individuals and includes a total of 24 variables, which are detailed in the table below:

[10] Direction de l'Animation de la Recherche, des Études et des Statistiques

Table 1: Description of variables

Variable	**Description**	**Type**
date	Monthly date	Character
ident	Anonymized housing identifier	Integer
noi	Individual identifier	Integer
rga	Interrogation rank ranging from 1 to 6	Integer
ag	Age of the person	Integer
sexe	Either “men” or “women”	Character
statut	Either “employee”, “independant”, ”unemployed”, ”student”, ”retired”, ”inactivity disability”, ”other inactive”, or ”no information”	Character
typlog5	Either ”Single person”, ”Single-parent family”, ”Couple without children”, ”Couple with children” or ”Other”	Character
ancempl	Length of employment in months (main job)	Integer
ancchom	Length of unemployment	Integer
salaire	Wage	Numeric
chomage	Unemployment and early retirement benefits	Numeric
salaire_independant	Self-employed income	Numeric
retraite	Retirements	Numeric
pension_invalidite	disability pensions	Numeric
retraites_pensions_rentes	Retirement pensions annuities	Numeric
pensions_alimentaires	Amount of child support payments	Numeric
pcstot1	Either ”Farmers”, ”Artisans”, ”Executive”, ”Intermediate occupations”, ”Employees”, ”Laborer”, ”Retired”, ”Student” , ”no information” and ”non coded”	Character
prestations_sociales	Amount of social benefits	Numeric
prestations_fam	Amount of family benefits	Numeric
minimas_sociaux	Amount of minimum social benefits (RSA)	Numeric
minimum_vieilleisse	Amount of minimum old-age pension	Numeric
reg	Region	Character
pauvre60m	Equals 1 when the household’s standard of living is strictly below 60% of the median standard of living of all individuals in this category, and 0 otherwise.	Integer

This dataset is highly comprehensive, offering monthly details on individuals' employment status, income, retirement pensions, social benefits, and unemployment benefits, thereby facilitating in-depth job trajectory analysis.

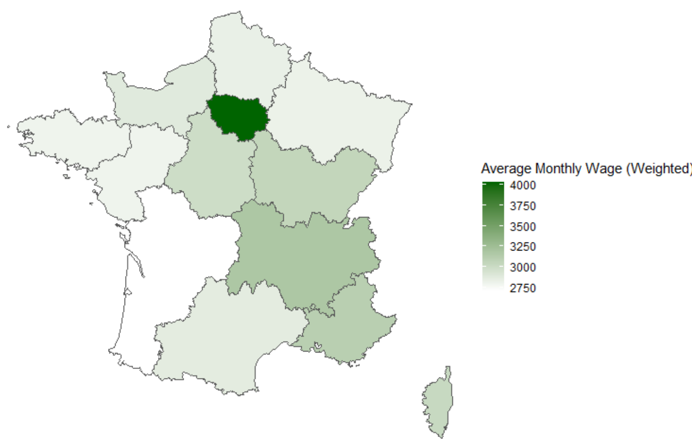
3.1.2 Descriptive statistics

This section begins by outlining descriptive statistics for the ERFS. It aims at providing some visual descriptive statistics of the sample.

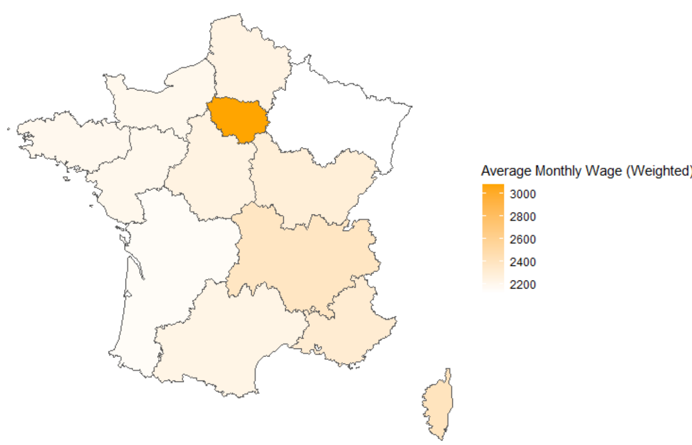
The graph below displays the weighted average monthly wage in France in 2023 by sexe.

Figure 1: Geographical distribution of weighted average wage in France in 2023

(a) Weighted average wage of men



(b) Weighted average wage of women



The geographical distribution of weighted average monthly wages[11] is illustrated using the sample of the ERFS, with weights adjusted to represent the entire French population, focusing exclusively on Metropolitan France. Darker colors represent higher average monthly wages, while lighter colors indicate lower wages. It is evident that the average monthly salary is higher in the Paris region for both men and women. On average, men have a higher monthly salary. Regions such as Grand Est and Nouvelle Aquitaine are depicted in white, indicating lower salaries.

Table 2: Status and weighted counts

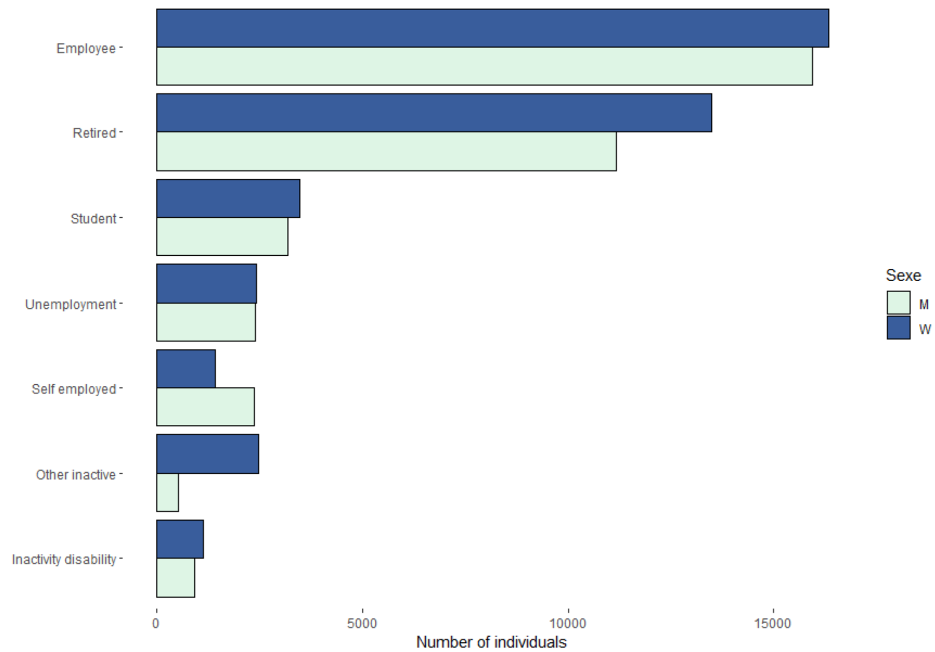
Status	**Weighted Counts**
Employment	23,316,511
Inactivity disability	1,377,222
No information	1,147
Other inactive	2,065,237
Retirement	14,454,919
Self employment	2,735,327
Student	4,809,953
Unemployment	3,577,414

Table 2 presents the weighted counts of individuals across various statuses in France in January 2023. The weights have been applied to reflect the entire French population. The largest group is composed of employed individuals, totaling approximately 23 million. Around 14 million people are retired, 4 million are students, and 3 million are unemployed. Nearly 3 million people in France are self-employed in 2023. This distribution provides a comprehensive overview of the labor market and other activity statuses of the French population during January 2023, illustrating the relative size of each group.

Figure 2 illustrates the number of men and women by status in the tax and social income survey in January 2023.

[11] Wages in this context exclude the self-employed

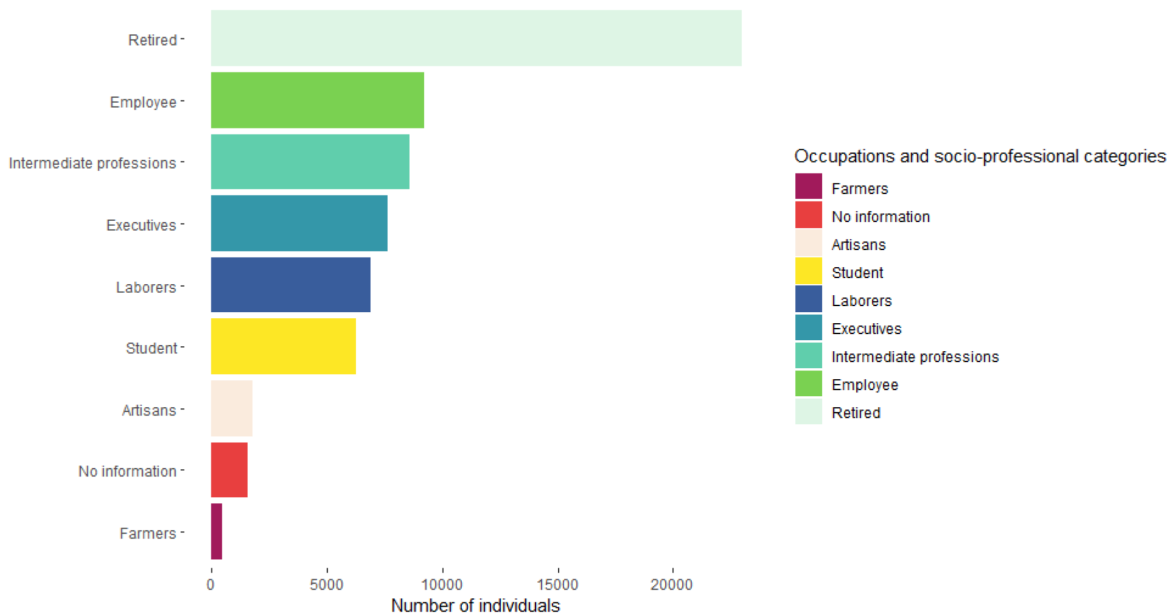
Figure 2: Distribution of employment status by gender



A significant proportion of the surveyed individuals are either employed or retired. The sample shows a slightly higher number of retired women compared to men. Approximately 10,000 students are included in the sample. Nearly 10,000 individuals were unemployed in January 2023. The distribution across genders is fairly balanced for all categories.

Figure 3 illustrates the proportion of individuals across occupations and socio-professional categories who maintained a consistent status throughout 2023. The majority of individuals in the sample did not experience any changes in their status. Specifically, a total of 65,527 individuals maintained a stable status over the year. This trend may be attributed to the limited variability captured within a single year, suggesting potential for greater variability with additional years of data. Notably, retirees comprise the largest group among those with stable status, followed by employees, intermediate professions, and executives.

Figure 3: Proportion of individuals by socio-professional Category with Stable Status



Finally, below is a table that displays some information concerning the minimum, maximum, mean and variance for some variables that were not ploted above.

Table 3: Summary statistics for selected variables

Variable	**Min.**	**Max.**	**Mean**
Age	16	104	52
Salary	0	190,118	1,145
Unemployment	0	80,717	31
Self-employed Income	-15,083	333,849	95
Retirement	0	32,646	597
Disability Pension	0	4,679	14
Pensions and Allowances	0	7,169	2
Social Benefits	0	4,314	187
Family Benefits	0	2,693	72
Social Minimums	0	3,215	82
Old Age Minimum Pension	0	2,046	7
Alimentar Pensions	0	6,047	9

Individuals in the sample are aged from 16 to 104, the average age being 52. The average wage is 1145 in this sample, ranging from people earning 0 to 190,118. It is possible to observe some negative revenue for self-employed individuals, this is particularly due to the fact that they experienced more losses than profits.

3.2 Dynamic case-type model description

The dynamic model of case-types allows for the simulation of the monetary benefits received by a fictional household and the social and tax contributions it pays, based on its family structure and transitions in the labor market. Developed from the DGEFP's dynamic case type model, this monthly step model evaluates how well the socio-fiscal system compensates for income losses following one or more labor market transitions.

This model enables the assessment of redistribution carried out by all social benefits and their incentivizing effects on employment over time. However, this model does not measure behavioral effects. It measures a household's standard of living and disposable income based on its income trajectories, following the 2023 legislation.

Key features of this model include:

- Option to choose the age of the reference person, their spouse, and children, with ages evolving over time.
- Ability to select marital status, number of children, and desired income trajectory for the reference person and spouse.
- Modeling of income neutralization and deduction mechanisms within the RSA, PA, and AL resource bases.

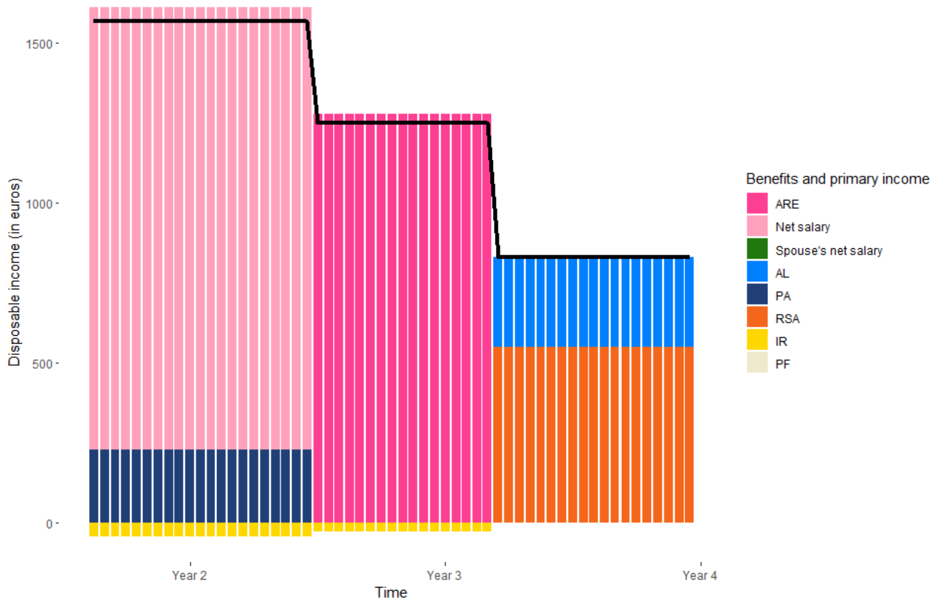
For consistency, the legislation remains stable and unchanged throughout the exercise, and households are assumed to utilize all benefits they are entitled to. It is also assumed that they are renters outside the Paris region. For social contributions, the scale for a non-executive employee on a permanent contract is applied. As mentioned, in the dynamic model, we do not simulate the income of self-employed individuals, nor do we simulate retirement pensions.

A primary use of this model is to illustrate the redistributive effects of the French socio-fiscal system. By configuring household characteristics (such as family structure, housing situation, income trajectory, etc.), one can visualize the taxes paid and social benefits received by the household over time. Additionally, the model allows for the comparison of redistributive effects across different types of households, demonstrating the variability of transfers based on family configuration.

If we take the example of a single 30 year old who works for 2 years at minimum wage and then loses his job, thus relying on unemployment benefits, we can see the different benefits he is entitled to in figure 1. During the first two years of employment, the individual receives an activity bonus (PA) of 228 euros in addition to his net salary of 1383 euros. On the downside, the individual is subject to a small income tax. When the individual loses his job, he receives unemployment benefits amounting to 1275 euros for 17 months. Once the unemployment benefits are exhausted, he receives 548 euros from RSA and 281 euros in housing assistance (AL). This graph illustrates how the social tax system smooths the disposable income of individuals experiencing an income shock.

With this tool, it is possible to conduct a theoretical evaluation of the unemployment insurance affecting individuals losing their job after December 2024. By altering the calculation method of unemployment insurance, we can examine the impact on disposable income for various income trajectories and different family compositions, ensuring a representative sample of the French population. This is considered a theoretical evaluation because it is based on hypothetical households. Consequently, I can analyze how the reform affects different households, specifically by comparing the overall effects on disposable income and the changes in received social benefits.

Figure 4: Composition of disposable income over time for a single person



3.3 Sample for Evaluating the Unemployment Reform

The Ines microsimulation model is a static model, which limits its ability to simulate individual trajectories or assess the impact of unemployment insurance reform. Furthermore, the Ines model does not account for unemployment benefits at the individual level, which is a significant drawback. The ERFS dataset, with its complex household structures, presents challenges in evaluating reforms in a way that is both accurate and representative of the average French household. Additionally, access to other relevant datasets is limited, and it can be difficult to find longitudinal microdata that tracks the same individuals over time. Consequently, I use the dynamic model I developed to theoretically assess policy changes more accurately using case-types.

The impact of unemployment benefit insurance reforms can be measured using several methods. Typically, there are three main approaches: macroeconomic models, microsimulation models, and case-type studies. The objective of using case-types is to maintain a microeconomic approach while providing synthetic information that makes the results easier to understand. In the context of unemployment benefit insurance, knowing the income trajectories and some socio-economic characteristics of individuals allows for the calculation of all social benefits.

Case-types can be either theoretical or representative, with the latter assigned a weight to reflect the number of individuals in the population who share similar characteristics. However, for simplicity, only theoretical case-types will be simulated in this report. The primary objective is to highlight the economic mechanisms underlying the reform. First, I create case-types to thoroughly understand the structure of the reform and identify theoretical winners and losers among the simulated population. The dynamic model then simulates mandatory contributions for hypothetical households based on family configurations and specified income trajectories.

Unlike microsimulation, which retains almost all data, creating case-types requires synthesizing information, resulting in some loss of available data. However, it is important to note that individuals' unemployment behaviors change with reforms. Case-types, with their fixed trajectories, significantly limit their relevance unless a behavioral module is used to adapt their career endings, which is not included in the dynamic model.

The timing of when an individual encounters unemployment at the beginning, middle, or end of their trajectory significantly affects their unemployment entitlements. To address these issues, it is essential to have a sufficient number of case-types that effectively capture the diversity of situations.

The objective is to choose a limited number of case-types that summarize different situations, with different family configurations and income trajectories while maintaining a structure faithful to that of the population. Beyond their educational value, cas-types have the advantage of requiring minimal data collection effort.

I consider a total of 8 case-types in this report, which are as follows:

- A 26-year-old student who works for 7 months with a gross monthly wage of 1700 euros, followed by 1 year of unemployment. For the remaining period, secures a job with a gross wage of 2000 euros per month.
- A 35-year-old single individual without children, employed for 2 years with a gross monthly wage of 2000 euros, then unemployed for 2 years, followed by 1 year of employment at the same wage.
- The same scenario as the previous one, but with two children aged 4 and 7.
- The same situation but for a couple with 2 children, where the reference person follows the same trajectory, while the partner has a stable job earning 2200 euros per month over the 5 years.
- A single seasonal worker who alternates every six months between unemployment and employment. When employed, earns the minimum wage of 1747 euros.
- The same situation as the previous one, but with two children aged 6 and 8.
- The same situation for a couple, with the partner having a stable job earning 2000 euros per month.
- A single parent with a young child aged 3, earning the minimum wage for 1.5 years, then unemployed for 1 year, and earning 1300 euros per month for the remainder of the period.

4 Methodology

This section describes the approach used to generate typologies of typical employment paths in France. I then examine the effect of changes in unemployment compensation policies on various case types.

4.1 Sequence analysis

Sequence analysis was first used in biology research to compare DNA sequence similarities, but it has since gained popularity in the social sciences by (Abbott & Hrycak,). The primary goal of this approach is to quantify the similarity between two sequences by calculating the minimal cost required to transform one sequence into another, and then to construct typologies of typical sequences. A sequence, by definition, is a series of mutually exclusive states ordered in discrete time. In this report, I distinguish seven different categorical states: employee, inactive due to disability, self-employed, retired, student, unemployed, and other inactive. For example, the sequence of someone studying for one year would be represented as S-S-S-S-S-S-S-S-S-S-S-S. Sequence analysis is a useful tool for exploring career trajectories in the labor market.

Sequence analysis typically requires sequences of the same length, and these sequences are often one-dimensional, although this is not always necessary. This method uses sequences of data as input rather than individual data points. Sequences that require fewer operations to be transformed into one another are considered more similar. Conversely, the more operations needed, the higher the cost, and the less similar the sequences are. The operations involved include insertions, deletions (also referred to as 'indels'), and substitutions. Costs are assigned to these operations, so the similarity is calculated by multiplying the number of operations by their respective costs.

Finding the appropriate measure of distance between two sequences is challenging. Various distance measures are available, which will be detailed below. Depending on the specific issue, the dissimilarity measure can account for the order of states and transitions in each sequence, the timing of transitions, and the duration spent in each state. The replacement cost matrix, defined for the different states, results in a square matrix of distances, representing the dissimilarity between each pair of the n sequences.

4.1.1 Dissimilarity measures

Determining the appropriate measure of dissimilarity is a complex task that requires careful parameterization. Although many dissimilarity measures are available, only a few will be presented in this report. Testing all possible measures is unnecessary, as many are not suitable for the analysis I intend to conduct.

Distances between probability distributions can be considered, particularly the distances between state distributions. This method examines the time spent in each state within a sequence. The dissimilarity between sequences is measured by calculating the distance between distribution vectors, using either Euclidean distance or chi-squared distance. The Euclidean distance considers the absolute differences in the proportion of time spent in each state. In contrast, the chi-squared distance assigns weights to the squared differences for each state, based on the inverse of the global proportion of time spent in that state. This approach gives more weight to less frequent states compared to more common ones.

Some other distances are based on counts of common attributes.

Simple Hamming distance uses the number of positions with non-matching states to measure dissimilarity between two sequences. This method applies only to sequences of the same length and is highly sensitive to the timing of mismatches. The Hamming distance does not consider indel costs.

The length of the longest common subsequence (LCS) measures the number of elements in one sequence that can be uniquely matched with elements occurring in the same order in another sequence. The LCS distance can be obtained by computing:

$$d_{\text{LCS}} = A(x,x) + A(y,y) - 2A(x,y) \tag{1}$$

Where $A(x, y)$ represents the number of matching elements between sequences x and y.
Because the LCS distance is not based on positionwise matches, it is less sensitive to timing. However, it relies more on differences in state distribution and order, particularly the sequence of the most common states. The LCS distance uses an indel cost of 1 and a substitution cost of 2.

Optimal matching (OM) has become the most common approach to compute dissimilarities between sequences in social sciences. OM computes the dissimilarity between two sequences using indels and substitutions, with each operation assigned a cost. The aim of the algorithm is to minimize the total cost of transforming one sequence into another. The cost can be modeled in a matrix where the costs are symmetric, fulfill the triangle inequality, and are equal to 0 for the substitution of an element with itself. OM is a highly flexible dissimilarity measure that can be parametrized using different costs. The algorithm combines two approaches: the number of indels and the positionwise mismatches. The positionwise mismatches method focuses on timing differences by counting mismatches at each position. The number of indels method, also known as Levenshtein II, counts the insertions or deletions needed to transform one sequence into the other, making it sensitive to sequence duration and order. Generally, indels have a single, constant cost. Substitution costs can vary, but a popular approach is to derive them from the observed transition rates. The idea is to assign higher costs to substitutions between states when transitions are rare and lower costs when transitions are frequent (?, ?). Another closely related approach considers two states a and b as close when there is a high probability that both states will lead to the same future after q time units. The substitution cost between the two states a and b can be measured by the chi-squared difference between the cross-sectional state distributions:

$$\gamma(a,b) = \left[\frac{\sum_{e \in \mathcal{E}} \{p(e_{t+q} \mid a) - p(e_{t+q} \mid b)\}^2}{\sum_{f \in \mathcal{E}} p(e_{t+q} \mid f)} \right]^{1/2}, \tag{2}$$

Where $p(e_{t+q} \mid f)$ is the probability of transitioning from state f to state e over q time units.

4.1.2 Cluster Analysis

From the dissimilarity matrix, typologies can be created using cluster analysis. Cluster analysis is an unsupervised algorithm that classifies objects into a reduced number of categories, aiming to capture patterns in the data by grouping similar trajectories based on their similarities. This method simplifies a large number of distinct sequences into a few trajectory types, offering a descriptive approach to sequence analysis. Clustering is performed using the dissimilarity matrix, with the goal of minimizing the distance between points within each cluster. Sequences within the same cluster are assumed to follow similar career patterns. However, increasing the number of clusters will reduce the average sample size in each cluster.

There are several types of clustering algorithms, but in this report, I will focus solely on agglomerative hierarchical clustering. As the name suggests, this method constructs clusters based on a hierarchical structure. Hierarchical clustering can be divided into two categories: agglomerative and divisive.
Agglomerative hierarchical clustering treats each observation—in this case, each sequence—as an individual cluster. At each step, the algorithm computes the Euclidean distance between pairs of sequences and merges the most similar clusters. This process is repeated until the optimal number of clusters is reached. Although there are four methods to combine clusters in the agglomerative approach, I have chosen to use Ward's method. Instead of directly measuring the distance between clusters, Ward's method analyzes the variance within clusters. The goal is to minimize the within-cluster discrepancy. Ward's method considers the distance between two clusters as the increase in the sum of squares that occurs when they are merged. Initially, the sum of squares is zero because each sequence is its own cluster; it then evolves as clusters are combined. Ward's method strives to keep the sum of squares as small as possible.

Determining the appropriate number of clusters is crucial. Several methods can be used to choose the optimal number of clusters:

- **Elbow method**: A popular method for selecting the optimal number of clusters. This method computes the within-cluster sum of squares for different numbers of clusters and identifies the point where the within-cluster sum of squares first starts to decrease significantly. This point, known as the "elbow" in the curve, represents the optimal number of clusters.

- **Average silhouette method**: This method assesses whether sequences are correctly assigned to their clusters. The optimal number of clusters is the one that maximizes the average silhouette score. A silhouette coefficient close to 0 indicates that a sequence lies between two clusters; a value close to 1 indicates that the sequence is well-clustered, while a value close to -1 suggests that the sequence may be assigned to the wrong cluster. The silhouette score for the i-th point is calculated as follows:

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \tag{3}$$

  where:

  - $a(i)$ is the average distance from the i-th point to the other points in the same cluster.
  - $b(i)$ is the minimum average distance from the i-th point to points in different clusters, minimized over clusters.

  The average silhouette width for the entire dataset is then given by:

$$\bar{s} = \frac{1}{n} \sum_{i=1}^{n} s(i) \tag{4}$$

- **Dendrogram**: A dendrogram can also be useful for determining the optimal number of clusters. It is a tree-like chart that represents the sequence of cluster merges. When two clusters are merged, the dendrogram joins them in a graph, with the height representing the distance between the clusters. The dendrogram allows visualization of inertia jumps based on the number of chosen clusters.

4.1.3 Quality Assessment

In sequence analysis, various tools have been developed to measure the quality of the typologies obtained. Cluster Quality Indexes (CQIs) are a prominent method for assessing the validity of clustering by evaluating within-cluster homogeneity and between-cluster separation. Despite their utility, CQIs have limitations, highlighting the need for alternative validation methods. Parametric bootstrapping is a useful alternative, serving as a resampling technique to evaluate clustering quality against a baseline of non-clustered data. This approach helps determine whether the observed clustering structure is statistically significant or merely a product of random chance.

The parametric bootstrap procedure operates by repeating the following steps n times:

- Generate similar, but non-clustered, data using a ”null” model.
- Apply clustering to the generated data.
- Compute the value of the CQI of interest. The result of this bootstrap procedure is a set of n CQI values obtained by clustering the non-clustered data. These n values can then be compared with those from our original clustering.

It is possible to assess whether the clustering structure found in the real data is more meaningful than what would be expected from random noise. If the CQI from the actual clustering is significantly higher than the bootstrapped values, it indicates that the clustering captures a meaningful structure within the data. Conversely, if the CQI values are similar, it suggests that the clustering may not reveal a significant structure.

In typical statistical reasoning, these procedures allow us to estimate the null distribution of the CQI, where the null model assumes the absence of any clustering structure in the data. This strategy addresses the challenge of lacking an independence model in cluster analysis, as identified earlier. By following standard statistical reasoning, we can use this information to develop a test-like framework for assessing the clustering structure of the obtained typology.

The framework for validating clustering quality using parametric bootstrapping involves defining a suitable null model and selecting an appropriate Cluster Quality Index (CQI). The null model should generate data similar to the original but without any inherent clustering. It must replicate features such as spell organization and state frequencies, which are common in sequence data but do not indicate clustering. The choice of CQI is also crucial. The Average Silhouette Width (ASW) is one index that combines measures of clustering homogeneity and separation into a single score. ASW assesses clustering quality by comparing the distance of each sequence to the center of its own cluster (reflecting homogeneity) with the distance to the nearest neighboring cluster (indicating separation). While useful, ASW's interpretation thresholds are only indicative and may favor simpler two-group solutions. The Pseudo-$\mathrm{R}^{\mathbf{2}}$ index, on the other hand, measures the proportion of variability explained by clustering but lacks established thresholds. This study introduces thresholds for Pseudo-$\mathrm{R}^{\mathbf{2}}$ to enhance its interpretability.

To account for multiple testing when evaluating clustering quality, this study employs the "Max T" approach. This method involves recording the maximum CQI value across different cluster numbers in bootstrapped samples, thereby addressing multiple comparisons and correlations between tests. Standardizing CQI values using the mean and standard deviation from the null distribution further enhances the reliability of this statistical test.

4.1.4 Tree-structured discrepancy analysis

Sequence analysis is primarily a descriptive method, with limited potential for explaining trajectories. To address this limitation, some researchers have developed methods that combine sequence analysis with tree-discrepancy analysis. This approach enables the study of the relationship between covariates and sequences, providing insights into how different individual characteristics influence job trajectories. This method was first introduced by (Studer & Ritschard,).

The principle of sequence discrepancy analysis involves defining a measure of discrepancy for a set of sequences using any dissimilarity measure. It computes the proportion of the discrepancy accounted for by a given explanatory factor, and it can be extended to regression trees. The discrepancy measures the between-individual variability of trajectories, which may reflect a form of precariousness.

A tree-structured discrepancy analysis can be a robust alternative to clustering analysis. This method displays the relationship between predictor variables and sequences. The tree-structured discrepancy analysis begins by selecting the most important predictor, the one with the highest pseudo R^2, and its most significant values to split the group into two, using a dissimilarity measure, a pseudo R^2, and a pseudo F-test. This process repeats until a stopping rule is reached. It's crucial to manage the number of splits carefully to avoid overfitting. The significance of the split is evaluated using a permutation F-test. The process of growing the tree stops when the chosen split is no longer statistically significant. Using the pseudo-$\mathrm{R}^{\mathbf{2}}$ as the splitting criterion inherently restricts to constructing binary trees. The overall quality of the model can be assessed using the pseudo F-test and the pseudo R^2[12], which provide insights into the statistical significance of the tree and the proportion of total discrepancy

[12] Details regarding the calculation of the sum of squares, pseudo R-squared, and pseudo F-statistic are provided in the appendix.

explained, respectively. An advantage of this method is that it automatically accounts for interactions.

Regression trees start with all individuals grouped in the initial node. Then they recursively divide each node based on the values of a predictor. At each step, the predictor and the split are selected to ensure that the resulting child nodes differ as much as possible from one another, or, equivalently, to minimize the within-group variance. The recursive partitioning generated by a tree offers a clear, interpretable representation of how each selected covariate modifies the effect of those chosen earlier. To achieve this, it's important to display relevant information about the distribution at each node.

The goal of tree-structured discrepancy analysis is to identify the most important covariates and their interactions. It is an iterative algorithm that segments cases using covariate values to form groups as homogeneous as possible.

4.2 Impact analysis of unemployment benefit policy changes on case-types

To evaluate the impact of the unemployment benefit policy change, I simulate the case-types described above under two scenarios: one following the current rules of unemployment (without the reform) and the other applying the rules that would have been in effect if the reform had not been suspended. The dynamic model I created is quite helpful as it enables the implementation of various reforms and the estimation of their effects. By recreating the paths taken by the identical individuals in both scenarios, I can directly compare the results and assess the effect of the change. This methodology successfully mitigates the influence of individual traits, guaranteeing that any observed disparities in outcomes may be directly attributed to the reform itself.

During this approach, I replicate the paths of the case-types in both circumstances and thereafter analyze and contrast the outcomes for each case-type. The outcomes obtained under the existing legislation are considered as the counterfactual, whilst the results obtained under the changed legislation are considered as the treatment scenario. I conduct simulations for each case category over a duration of 5 years. Subsequently, I compute the disparity in expendable earnings prior to and following the alteration in policy for each month. Subsequently, I perform a regression analysis to ascertain the influence by contrasting these simulated results. This methodology for a meticulous examination of alternative scenarios, facilitating accurate measurement of the reform's impact on individuals' disposable income.

Since I lack multi-year monthly empirical data for actual individuals, I cannot apply a quasi-experimental method such as difference-in-differences to estimate the reform's causal effect. Simulating case types would introduce selection bias, compromising the validity of such methods. Therefore, I choose to use a linear regression analysis instead.

The core idea is to quantify the change in the disposable income caused by the intervention by calculating the difference in outcomes for each individual before and after the intervention every months. This method controls for all characteristics that remain constant across the two conditions, effectively isolating the intervention's impact. In my analysis, the outcome of interest is disposable income, and my goal is to assess how this reform affects disposable income.

To apply the regression estimation, I first calculate the difference in disposable income for each case-types before and after the policy change. This difference then serves as the dependent variable in the regression analysis where the only explanatory variable is a constant. The coefficient from this regression provides an estimate of the average treatment effect, reflecting the mean impact of the intervention on the disposable income.

Regression estimation is straightforward and easy to implement, providing a clear measure of the intervention's average effect without the need for complex modeling. By comparing each unit with itself, the method controls for constant characteristics, reducing the risk of bias from unobserved variables.

The method assumes that any observed difference between conditions is solely due to the intervention. If other factors vary between the conditions, the results could be biased. The method relies on differences between conditions and does not account for potential changes over time or other dynamic factors.

5 Results

This section first presents the results of the job trajectory typologies, followed by the findings on the impact of the unemployment reform.

5.1 Sequence analysis

5.1.1 Typologies of job trajectories

To construct a typologies of job trajectories in France, I use sequence analysis, and different dissimilarity measures and evaluate their performance. The analysis was conducted using R software and the TramineR and WeightedCluster packages.

The analytical sample consists of individual with at least one change of status during 2023. Because of the short amount of historical data and because people are only surveyed during 18 months I have a small proportion of individual knowing a change in their status. As a result, after accounting for this, I have a total of 11,680 individuals, giving a total of 11,680 sequences, each of length 12.

The first step consists in transforming the data into a sequence form. I have 2608 distinct sequences and a total of seven status including : employment, retirement, student, self-employment, unemployment, inactivity and disability and other inactivity.

The transition matrix can be calculated for individuals experiencing at least one change in their status in 2023. The transition matrix contains the probabilities of transitioning between different states.

Table 4: Transition matrix between the different states

	Emp	**Inac_dis**	**Oth_inac**	**Ret**	**Self_emp**	**Stu**	**Unemp**
Employment	0.867	0.009	0.019	0.012	0.020	0.010	0.060
Inactivity disability	0.034	0.800	0.008	0.006	0.002	0.001	0.200
Other inactivity	0.065	0.010	0.808	0.011	0.020	0.020	0.060
Retirement	0.003	0.000	0.002	0.994	0.0002	0.000	0.0007
Self employment	0.072	0.002	0.021	0.006	0.900	0.001	0.009
Student	0.082	0.001	0.029	0.000	0.002	0.900	0.030
Unemployment	0.092	0.030	0.025	0.003	0.006	0.008	0.800

Each line corresponds to the initial state of the individual, and each column the final state. The value in each cell represents the probability of transition from one state to another. For example, the probability that someone goes from employment to unemployment is 6%. Someone in inactivity and disability has a 20% chance to experience unemployment. The values on the diagonal represent the probability of staying in the same state. The probability that someone stays employed is 86.7%.

To ensure robustness I use several measures of dissimilarity, to see how the results vary. I use the euclidean distance, the ham distance, the length of the longest common subsequence (LCS) and optimal matching with costs based on transition rates. For all of these distance measures, I obtain dissimilarity matrices, to which I apply an agglomerative hierarchical clustering using Ward's method in order to build a typology of trajectories. Using cluster analysis with different dissimilarity measures should allow to see whether the trajectories are primarily structured by timing, duration or sequencing. To do so, it is possible to look at the average silhouette width(ASW). The best dissimilarity measure will be the one that has the highest ASW, resulting in the best quality of clustering.

I obtain the following table by calculating the ASW for the different clusters using different dissimilarity measures:

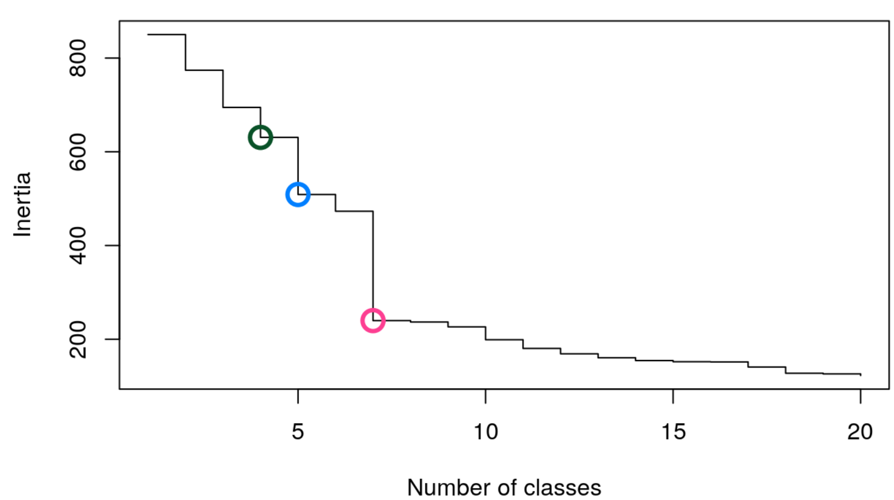
Table 5: Comparison of ASW for different dissimilarity measures

Distance measure	**Average silhouette width**	**Number of clusters**
OM	0.3	8
LCS	0.4	7
HAM	0.3	8
Euclidean	0.2	8

The LCS distance has the highest ASW among the dissimilarity measures. Therefore, I will focus exclusively on the steps and details of this measure in the rest of the report. The results obtained with the other dissimilarity measures will be included in the appendix section.

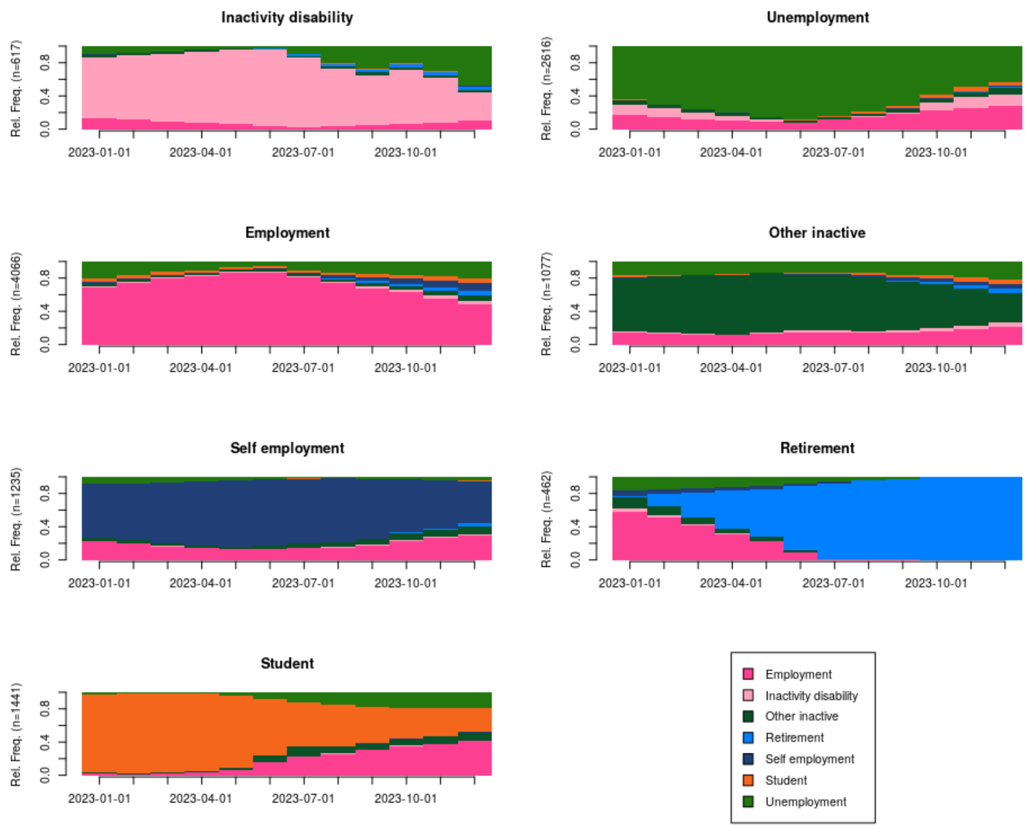
To determine the optimal number of clusters, the inertia jumps from the dendrogram can be represented according to the number of retained clusters. The primary purpose of this graph is to help determine the optimal number of clusters. The inertia represents the sum of the variances of the data points with respect to the overall data center.

Figure 5: Inertia plot for the LCS distance



It seems that seven is the optimal number of clusters, as it represents the biggest jump. After selecting the optimal number of clusters for this distance I use state distribution plots to show the typologies I have.

Figure 6: Typology of trajectories



These clusters reflect the diverse trajectories that individuals can experience in the labor market. Each class appears to be defined by a primary status. However, it is also possible to notice layers of other colors, indicating that these statuses may not be entirely stable over time.

Cluster 1: This cluster primarily includes individuals who experiences inactivity and disability. Over time, some of these individuals experience periods of unemployment.

Cluster 2: The second cluster is mainly characterized by unemployment. However, there are also instances of employment and periods of inactivity or disability.

Cluster 3: This cluster largely consists of individuals who are employed, but also includes those who experience unemployment during the observed period.

Cluster 4: This cluster mainly groups individuals categorized as "other inactive," with some also experiencing periods of employment.

Cluster 5: In this cluster, individuals are primarily engaged in self-employment and employment, indicating a focus on more independent or entrepreneurial work paths.

Cluster 6: The sixth cluster includes employed individuals, who then transition into retirement.

Cluster 7: The final cluster is mostly composed of students. After completing their studies, these individuals either find employment or face unemployment.

It is also interesting to know the distribution of the typology both in terms of counts and percentages. This table provides a breakdown of the number of individuals and their corresponding percentages within each job trajectory cluster.

Table 6: Typology distribution

Cluster	**n**	**Percentage (%)**
Inactivity disability	528	5.1
Unemployment	2731	26.5
Employment	3448	33.4
Other inactive	687	6.7
Self employment	1005	9.7
Retirement	585	5.7
Student	1333	12.9

Employment is the largest cluster, encompassing 33.4% of the subsample (3,448 individuals). It suggests that a significant portion of the population is primarily engaged in employment. Unemployment is the second-largest group and is characterized by unemployment, accounting for 26.5% of the subsample (2,731 individuals). The third biggest cluster is the student one, which represents 12.9% of the subsample (1,333 individuals), indicating that a notable segment of the population is either pursuing education or transitioning from education to the labor market. Other clusters, such as self-employment, retirement, other inactive, inactivity and disability, each account for less than 10% of individuals in the subsample.

Weights are applied to ensure that each cluster accurately represents the population in France. For this purpose, the *wpela_anr2* variable can be used, as it provides adjusted weights to reflect the 2023 population accurately.

Weighted typologies are shown in this table:

Table 7: Weighted typology distribution

Cluster	**Weighted amount**	**Percentage (%)**
1	356,185	4.64
2	2,037,779	26.54
3	2,638,172	34.36
4	520,189	6.78
5	759,561	9.89
6	363,077	4.73
7	1,002,340	13.06

The table clearly illustrates how French individuals are distributed among various clusters. Cluster 3 is notably the largest segment, whereas Clusters 1 and 6 represent the smallest shares. This distribution reveals diverse proportions across the clusters, offering useful insights for further analysis.

I then continue the class description by cross-referencing the typology with the gender variable :

Table 8: Typology distribution by gender

Cluster	**Men**	**Women**	**All**
Inactivity disability	4.6	6.0	5.4
Unemployment	22.3	23.1	22.7
Employment	36.6	34.2	35.3
Other inactive	6.2	12.1	9.4
Self employment	13.7	8.0	10.7
Retirement	3.9	4.1	4.0
Student	12.7	12.4	12.5
Total	100.0	100.0	100.0

The table offers a comprehensive breakdown of typology distribution across various categories, segmented by gender. It reveals several important insights into gender differences within these categories. Specifically, the distribution of students, retirees, unemployed individuals, and those employed is nearly equal between men and women, suggesting that these statuses are similarly prevalent across genders. However, in the "Other Inactive" category, women are represented twice as much as men. Additionally, men are more prominently represented in self-employment compared to women. These gender-based differences are essential for understanding the distribution of various statuses and for designing targeted policies.

5.1.2 Parametric bootstrap

An ASW of 0.4 is generally considered moderate, indicating that while there is some clustering structure present in the data, it is not particularly strong. This suggests that the clusters are somewhat well-defined, but there may be some overlap or uncertainty in the cluster assignments. To evaluate the quality of the clusters more rigorously, I employ parametric bootstrapping.

Parametric bootstrap represents a good validation technique for sequence analysis. It provides baseline values by clustering data that is similar but non clustered. To ensure consistency, the same clustering algorithm and distance measures used in my original analysis are applied. Specifically, I estimate the expected values of ASW by applying clustering to similar sequences that are not classified according to the combined randomization null model, using the LCS distance and Ward hierarchical clustering. The null model replicates the structure of sequence spells, including the frequencies and durations of each spell state. It is a generic null model reproducing the structuration of sequences in spells.

Due to computational constraints, I limited the number of bootstrap iterations to 100. This choice represents a trade-off between the number of iterations and execution time, as running more iterations proved too resource-intensive. Below are the results obtained from this analysis.

Figure 7: Chronogram of generated sequences with the combined randomization null model

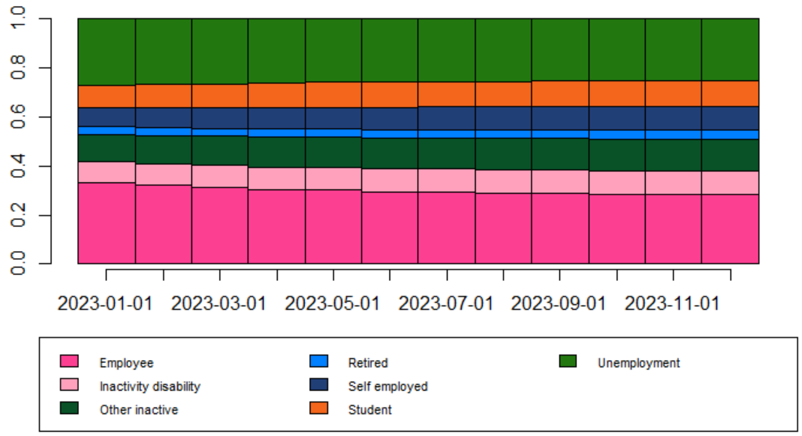


Figure 7 displays sequences produced by the null model for professional career data of the ERFS. The chronogram illustrates sequences generated across all bootstrap samples. This null model maintains the structure of spell sequences and preserves the frequencies and durations of each spell state, but it randomizes the order, duration, and timing of these spells. As anticipated, the null model does not replicate timing, duration, or sequencing; however, it does preserve the spell organization of the sequences. Although the state frequencies at each time point are not reproduced due to the lack of duration attachment to states, the employee state, shown in pink, remains the most prevalent.

Figure 8: Distribution of the standardized ASW null values for 2 to 7 clusters

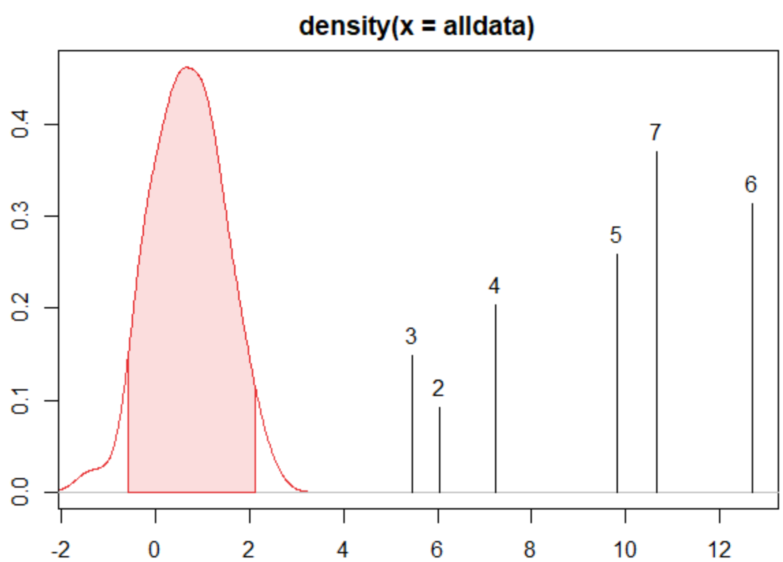


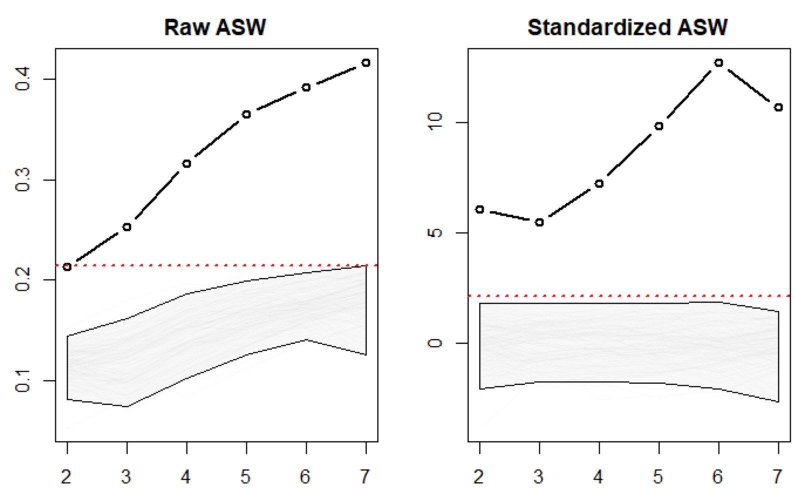
Figure 8 displays a density plot illustrating the null distribution for clustering results ranging from 2 to 7 clusters. The shaded area under the curve indicates the 95 percent confidence interval of these values. The ASW values derived from the original data are marked with a vertical black line, annotated with the corresponding number of clusters.

The quality of the typologies obtained with seven clusters is equal to an ASW of 0.4. It is quite similar to what I obtained with the null model, underlying that the clustering is weak. The results should be interpreted cautiously since they are based on only 100 bootstraps. To achieve more reliable results, at least 1,000 bootstraps would be necessary, but this was not feasible given the computational limitations of my computer.

By definition, any ASW value within the 95 percent confidence interval suggests a scenario where no significant clustering structure exists. In contrast, ASW values for clusters ranging from 2 to 7 groups fall well outside this interval, indicating a meaningful clustering structure in the data. I can reject the null model's assumption for the homogeneous sequencing and spell duration data. Therefore, considering 2 to 7 clusters, the typology suggests the presence of distinct timing, duration, or sequencing structures. The aim of standardization is to account for the changing behavior of the CQI measure and clustering algorithm for a different number of clusters. A high standardized value can then be interpreted as a high relative gain in the structure found for a given number of groups.

It is also possible to examine the ASW according to the number of clusters. The left hand side of figure 9 displays raw ASW of the typology for a varying number of groups using a solid black line. The 100 bootstrapped ASW values are represented with individual gray lines, and a 95 percent interval is computed and represented using a gray polygon. The horizontal red dotted line shows the upper bound of the null confidence interval. The right-hand side of the figure represents the standardized ASW values. According to this analysis, the 6 or 7 group solutions seem particularly well suited.

Figure 9: Observed and bootstrapped values of the ASW for different number of clusters



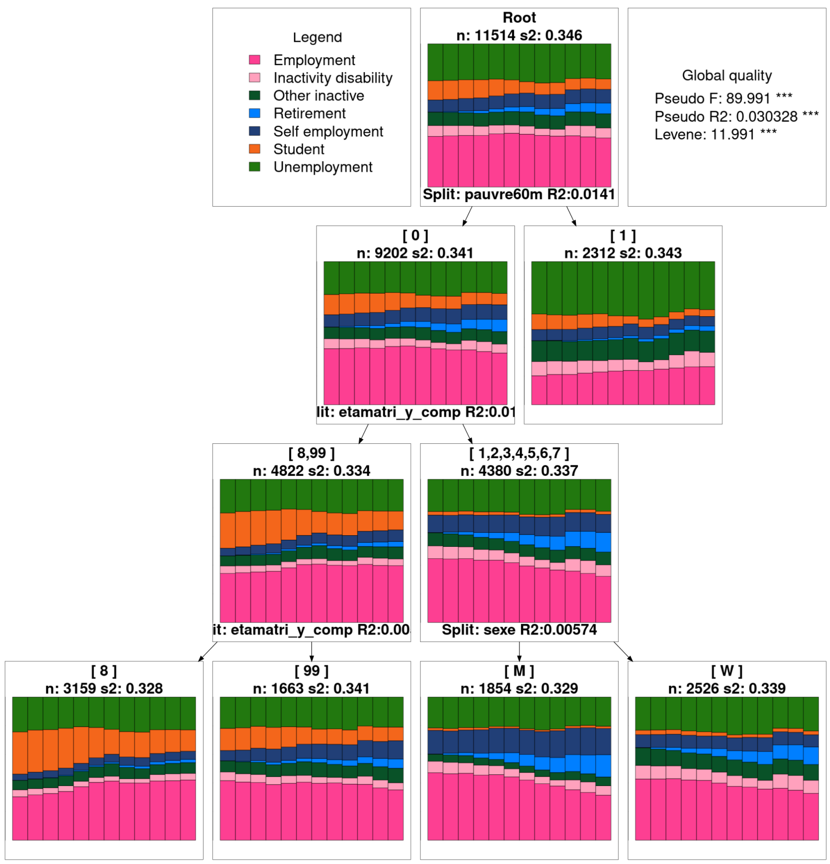
This figures also shows the evolution of the ASW with a varying number of classes when no clustering should be found in the data. In this case I find that the highest ASW is observed for 7 clusters, corresponding to the number of different states.

5.1.3 Tree-discrepancy analysis

Cluster analysis does not allow to measure the relationship between the sequences and the explanatory variables. That is why in this subpart I use tree-discrepancy analysis.

With several explanatory variables, I obtain the share of variance explained (or discrepancy) of dissimilarities explained by the variable (using a pseudo-R2) by all the variables and the decomposition of this share between the variables.

Figure 10: Sequence regression tree



In the initial node contains all the individuals, 11,514. Although the model demonstrates a high pseudo F statistic of 89.9, indicating that it significantly improves upon a baseline model, the pseudo R-squared value of 0.03 reveals that it explains only a small portion of the variance in the dependent variable. This suggests that while the model's predictors are statistically significant, its overall explanatory power is limited, and there may be other factors affecting the outcome that are not captured by the current model.

The first split is based on poverty status *pauvre60m*, indicating its importance in differentiating career trajectories. Subsequent splits involve marital status and gender, highlighting their roles in further differentiating groups.

5.2 Result of the analysis of the reform

The regression analysis was performed to evaluate the effect of a policy reform on disposable income by estimating the average difference in disposable income between two scenarios: with and without the reform. The results are displayed in Table 9.

Table 9: Regression Results

Residuals: **3Q**	**Min** **Max**	**1Q**	**Median**
92.4	-1124.5 828.2	92.4	92.4

Coefficients:	**Estimate**	**Std. Error**	**t value**	**Pr($\|t\|$)**
(Intercept)	-92.4	11.5	-8.01	0.0000000000000059

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 282 on 599 degrees of freedom

The intercept of the regression is -92.4. This value represents the average difference in disposable income between the two scenarios. In this case, the average disposable income is less under the reform scenario by about 92.4 euros per month.

The p-value of 0.75 is is much smaller than the conventional significance level of 0.05, indicating that the intercept is statistically significant. The regression model shows a statistically significant intercept with a very low p-value, suggesting a robust estimate. However, the variability in residuals suggests that the model's predictions have a substantial degree of error.

It is important to remember that these results apply specifically to the case-types analyzed and cannot be generalized to the entire French population. However, examining the reform through these case types provides insight into how changes in unemployment insurance calculations affect disposable income. The impact of the reform varies among different case types, with some being more significantly affected than others.

Overall, the unemployment benefit reform has increased the difficulty of accessing benefits by extending the required work period from 6 to 8 months. Additionally, the reference salary period for calculating benefits has been shortened, resulting in a reduction in the unemployment benefit amount. The maximum duration for receiving unemployment benefits has been reduced to 15 months. These changes collectively contribute to a decrease in individuals' disposable income. However, other social benefits, such as RSA and AL, are intended to help offset the impact of this decrease in disposable income, though they do not fully compensate for it.

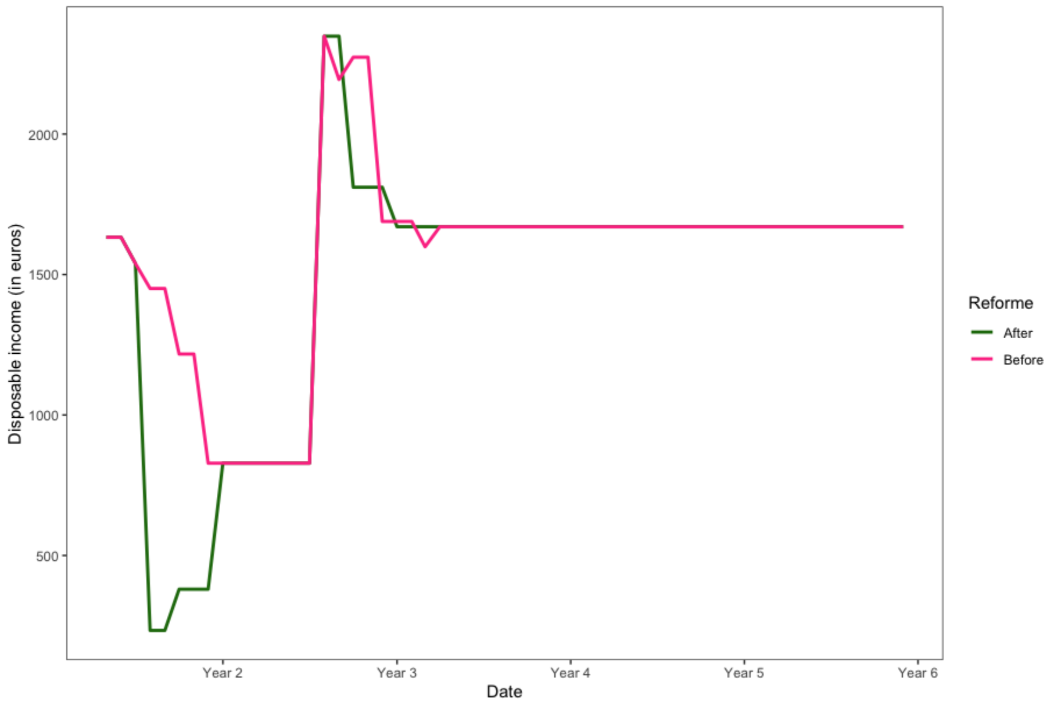
Figure 11 illustrates the disposable income for a 26-year-old student who works for 7 months with a gross monthly wage of 1,700 euros, followed by 1 year of unemployment [13]. After this period, the student secures a job with a gross monthly wage of 2,000 euros.

Before the second year, there is a significant difference in disposable income. The pink line, representing the current unemployment insurance system, is above the green line, which shows the situation

[13] Figures for the other case types are provided in the appendix.

after the reform. Under the existing system, the individual qualifies for unemployment benefits after meeting the 6-month work requirement, receiving benefits for 4 months. However, with the reform, the individual no longer qualifies for these benefits, leading to the initial gap in the graph. After this period, the disposable income levels before and after the reform become quite similar. This illustrates the impact of increasing the minimum required months of work to receive unemployment benefits, causing a sharp decrease in disposable income for a few months.

Figure 11: Student



6 Conclusion

6.1 Conclusion on the main results

This report delves into the complexity of career trajectories in France and evaluates the nuanced effects of potential unemployment policy reform on disposable income.

Through sequence analysis, the study mapped the diverse employment paths of individuals in the French labor market. By applying various dissimilarity measures, seven optimal clusters were identified, each representing distinct patterns of career progression. Understanding these career trajectories is essential for recognizing how individuals navigate the labor market and the factors influencing their employment status over time. Analyzing these paths provides valuable insights into the diversity and stability of different employment statuses. The research employed sequence analysis, using the LCS distance measure, to identify distinct career patterns, resulting in seven clusters that capture the variety of trajectories individuals experience. These clusters were examined to understand the predominant employment statuses and transitions over time, revealing that employment is the most prevalent status, with unemployment and student statuses also constituting significant portions of the sample. Gender-based differences emerged, with noticeable disparities in the distribution of certain statuses. The robustness of the clustering results was validated using parametric bootstrapping.

Additionally, examining the effects of policy reform, such as changes in unemployment benefits, offers an understanding of how such reforms impact individuals' financial well-being and overall economic stability. The analysis showed that the reform initially led to a significant reduction in disposable income; however, other social benefits, such as RSA and AL, helped to partially mitigate this impact, though they did not fully offset the loss.

The implications of these findings are significant for policymakers. A thorough understanding of diverse career trajectories and the effects of policy changes on disposable income is crucial for designing more targeted and effective labor market interventions. The study emphasizes the need to consider both immediate and long-term impacts when reforming unemployment insurance. Additionally, the observed gender disparities underscore the necessity for policies that address these inequalities and foster equitable employment opportunities.

In conclusion, this thesis enhances the broader understanding of career trajectories and the impact of unemployment reform in France. The insights from this analysis lay a foundation for future research. By continuing to explore the dynamics of employment and income trajectories, we can develop more inclusive and resilient social protection systems that better support individuals throughout their careers.

6.2 Improvement and limitations

While the study provides valuable insights, it is important to acknowledge certain limitations. The analysis was based on a relatively short observation period, which may not fully capture the long-term effects of career trajectories and policy changes. A primary issue was the limited length of historical data available for analyzing career trajectories. The microsimulation model had recently been transitioned from SAS to R, which meant that only the most recent survey from 2021 was accessible in R. Additionally, the available data covers only a short period for each individual, complicating the analysis of long-term career trajectories. However, future work will benefit from the DRM (Dispositif de Ressources Mensuelles), which is expected to provide comprehensive historical data on all French individuals and will significantly enhance the depth and breadth of trajectory analysis.

Additionally, the lack of multi-year data limited the application of more robust causal inference methods, such as difference-in-differences analysis. Future research should aim to extend the observation period to include multi-year data, enabling a more comprehensive analysis of the long-term effects of unemployment policies. Additionally, incorporating more diverse case types and exploring the impact of other socio-economic variables could provide a deeper understanding of how different groups are affected by policy reforms.

Using representative case types could provide a more accurate estimation of the effects of the unemployment reform. Initially, I considered leveraging the ERFS sample for this purpose; however, the limitation of having only one year of historical data made it challenging to extend the trajectories over time. I began by calculating the transition matrix for employment and unemployment, but I encountered an issue: the probabilities for remaining employed or unemployed were both 1, which hindered the extension of these trajectories. Due to these constraints, I decided to use theoretical case types to illustrate the potential impacts of the unemployment reform.

7 Bibliographic References

References

Abbott, A., Hrycak, A. (1990). Measuring resemblance in sequence data: An optimal matching analysis of musicians' careers. *American Journal of Sociology*, *96*(1), 144–185. doi: 10.1086/229495

Abbott, A., Tsay, A. (2000). Sequence analysis and optimal matching methods in sociology: Review and prospect. *Sociological Methods & Research*, *29*(1), 3–33. doi: 10.1177/0049124100029001001

Arthur, M. B., Rousseau, D. M. (1995). The boundaryless career: A new employment principle for a new organizational era. *Oxford University Press*.

Bessin, M., Legrand, J. (2010). Les classes préparatoires aux grandes Écoles: Un grand Écart entre la mixité sociale et les spécificités du milieu Étudiant. *Sociétés Contemporaines*, *80*(4), 49–71. doi: 10.3917/soco.080.0049

Bidwell, M., Briscoe, F. (2010). The dynamics of interorganizational job mobility and career outcomes. *Organization Science*, *21*(5), 1034–1053. doi: 10.1287/orsc.1090.0496

Brzinsky-Fay, C. (2007). Lost in transition? labour market entry sequences of school leavers in europe. *European Sociological Review*, *23*(4), 409–422. doi: 10.1093/esr/jcm011

Cahuc, P., Zylberberg, A. (2004). *Labor economics*. MIT Press.

Cousteaux, A. (2015). Optimal matching analysis of career paths and economic policies. *French Economic Journal*, *128*, 50–67.

Dupray, A., Moullet, S. (2004). Transitions from school to work in france: The changing influence of socio-economic background. *European Societies*, *6*(1), 53–82. doi: 10.1080/1461669032000176325

Duru-Bellat, M., Kieffer, A., Reimer, D. (2008). Patterns of social inequalities in access to higher education in france and germany. *International Journal of Comparative Sociology*, *49*(4-5), 347–368. doi: 10.1177/0020715208093081

Gautié, J., Schmitt, J. (2010). *Low-wage work in the wealthy world*. Russell Sage Foundation.

Giret, J.-F., Nohara, H., Van de Velde, C. (2006). Les jeunes face à l'emploi en france et au japon: Quels enseignements tirer des différences? *Revue Économique*, *57*(2), 245–274. doi: 10.3917/reco.572.0245

Hall, D. T. (2002). *Careers in and out of organizations*. Sage Publications.

Kahn, L. B. (2010). The long-term labor market consequences of graduating from college in a bad economy. *Labour Economics*, *17*(2), 303–316. doi: 10.1016/j.labeco.2009.09.002

McVicar, D., Anyadike-Danes, M. (2002). Predicting successful and unsuccessful transitions from school to work by using sequence methods. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *165*(2), 317–334. doi: 10.1111/1467-985X.00649

Moen, P., Sweet, S. (2002). Two careers, one employer: Couples working for the same corporation. *Journal of Vocational Behavior*, *61*(3), 466–483. doi: 10.1006/jvbe.2002.1881

Oreopoulos, P., von Wachter, T., Heisz, A. (2012). The short- and long-term career effects of graduating in a recession: Hysteresis and heterogeneity in the market for college graduates. *American Economic Journal: Applied Economics*, *4*(1), 1–29. doi: 10.1257/app.4.1.1

Piccarreta, R., Lior, O. (2010). Exploring sequences of categorical data: A comparison of nonparametric approaches. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *59*(2), 297–317. doi: 10.1111/j.1467-9876.2009.00700.x

Rapoport, B., Le Goff, J.-M. (2004). Career paths and labor market policies: A french perspective. *French Labor Review*, *47*, 121–144.

Stone, P. (2007). *Opting out?: Why women really quit careers and head home*. University of California Press.

Studer, M., Ritschard, G. (2010). A comparative review of sequence dissimilarity measures. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *173*(1), 479–511. doi: 10.1111/rssa.12125

Studer, M., Ritschard, G. (2016). What matters in differences between life trajectories: A comparative review of sequence dissimilarity measures. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *179*(2), 481–511. doi: 10.1111/rssa.12125

Wu, L. L. (2000). Some comments on "sequence analysis and optimal matching methods in sociology": Review and prospect. *Sociological Methods & Research*, *29*(1), 41–64. doi: 10.1177/0049124100029001003

8 Glossary

Glossary

Disposable income The income available to a household for consumption and saving. It includes net employment income after social contributions, unemployment benefits, pensions, income from assets (such as property and financial investments), and other social benefits received, all net of direct taxes..

Equivalence scale Ued to compare the living standards of households of various sizes and compositions, aim to account for the economies of scale that result from pooling resources and expenses within households..

France Travail The official public employment service (PES) and primarily aims to assist individuals in their job or training searches, collect job offers from companies, and connect these offers with job seekers.

Ines An acronym for "Insee-Drees," the two organizations that jointly develop the model. Since 2016, CNAF has also co-managed the model alongside Insee and Drees.

Standard of living A household's disposable income divided by the number of consumption units (CU). Consequently, the standard of living is the same for all individuals within the same household. The standard of living corresponds to what Eurostat refers to as 'equivalent disposable income'..

9 Appendix

The sum of squares SS can be expressed in terms of distances between pairs:

$$SS = \sum_{i=1}^{n}(y_i - \bar{y})^2 = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=i+1}^{n}(y_i - y_j)^2 = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=i+1}^{n} d_{ij}$$

Setting d_{ij} equal to LCS distance, I get a measure of dispersion.
Then, the Sum of Squares (SS) can be decomposed as follows:

$$SST = SSB + SSW$$

Using the formula shown earlier:

$$SST = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=i+1}^{n} d_{ij}$$

$$SSW = \sum_{g}\left(\frac{1}{n_g}\sum_{i=1}^{n_g}\sum_{j=i+1}^{n_g} d_{ij,g}\right)$$

$$SSB = SST - SSW$$

The Pseudo R^2 is defined as follows :

$$R^2 = \frac{SSB}{SST}$$

The Pseudo F statistic is given by:

$$F = \frac{\frac{SSB}{m-1}}{\frac{SSW}{n-m}}$$

where:

- SSB is the sum of squares between groups,
- SSW is the sum of squares within groups,
- m is the number of groups,
- n is the total number of observations.

Figure 12: Typology of trajectories with Euclidean distance

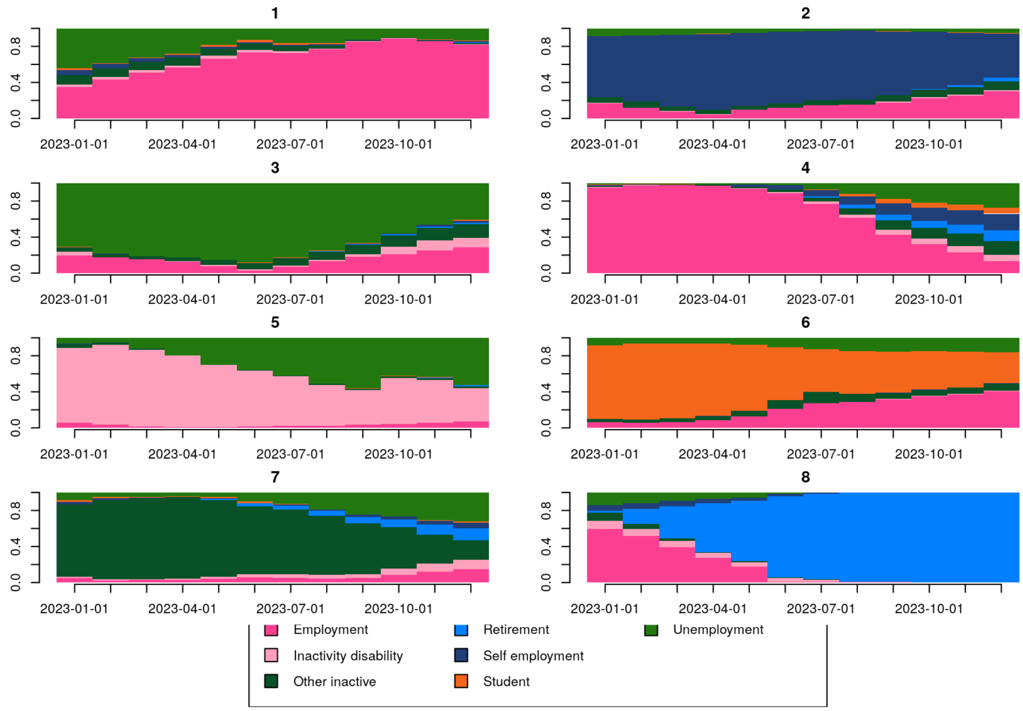


Figure 13: Typology of trajectories with ham distance

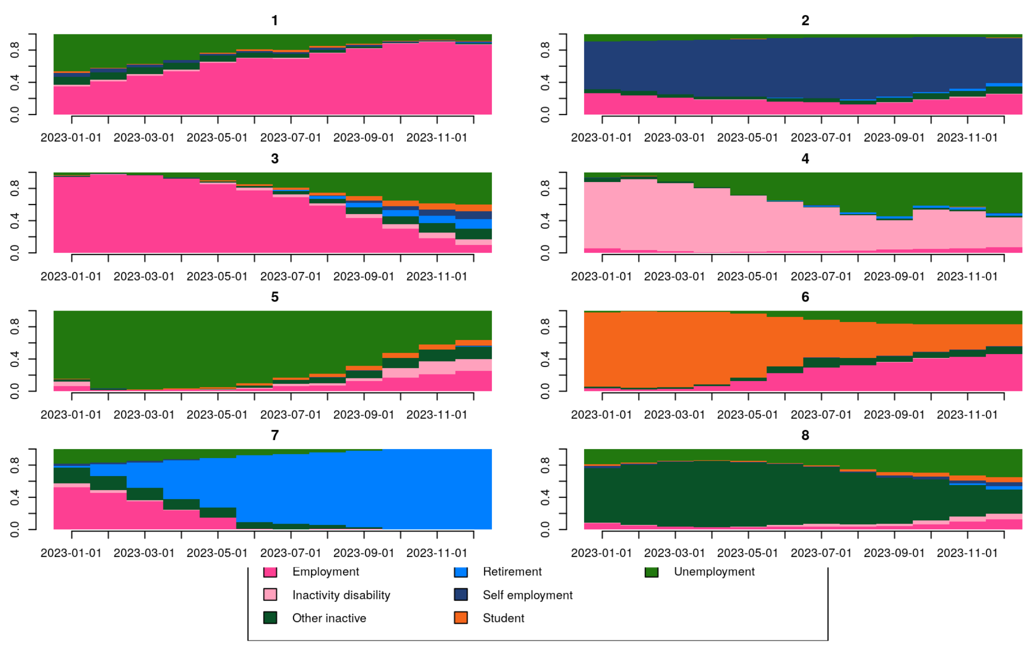


Figure 14: Typology of trajectories with OM distance

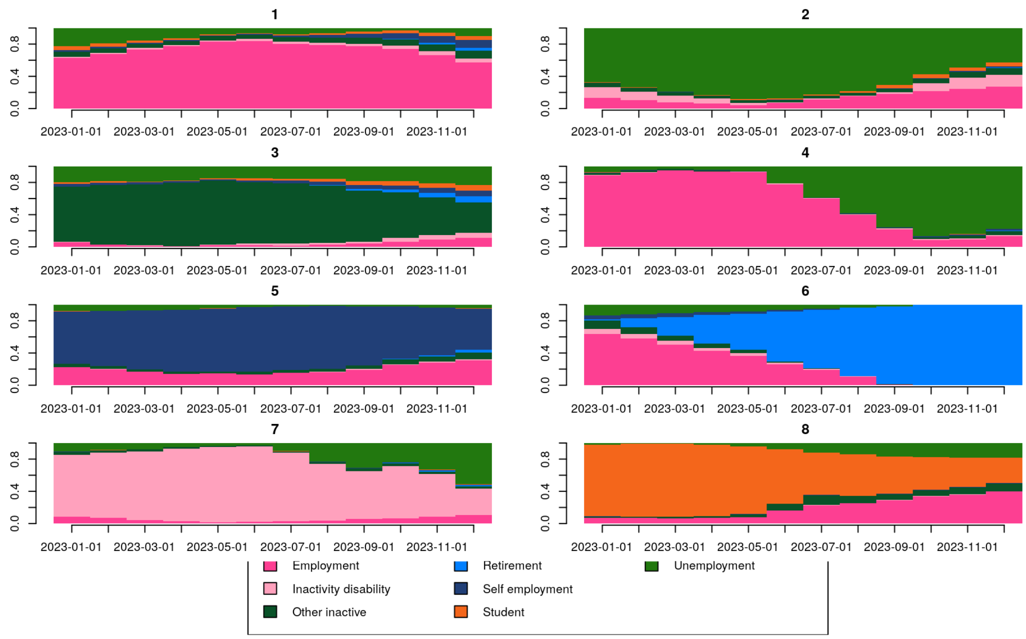


Figure 15: Single individual

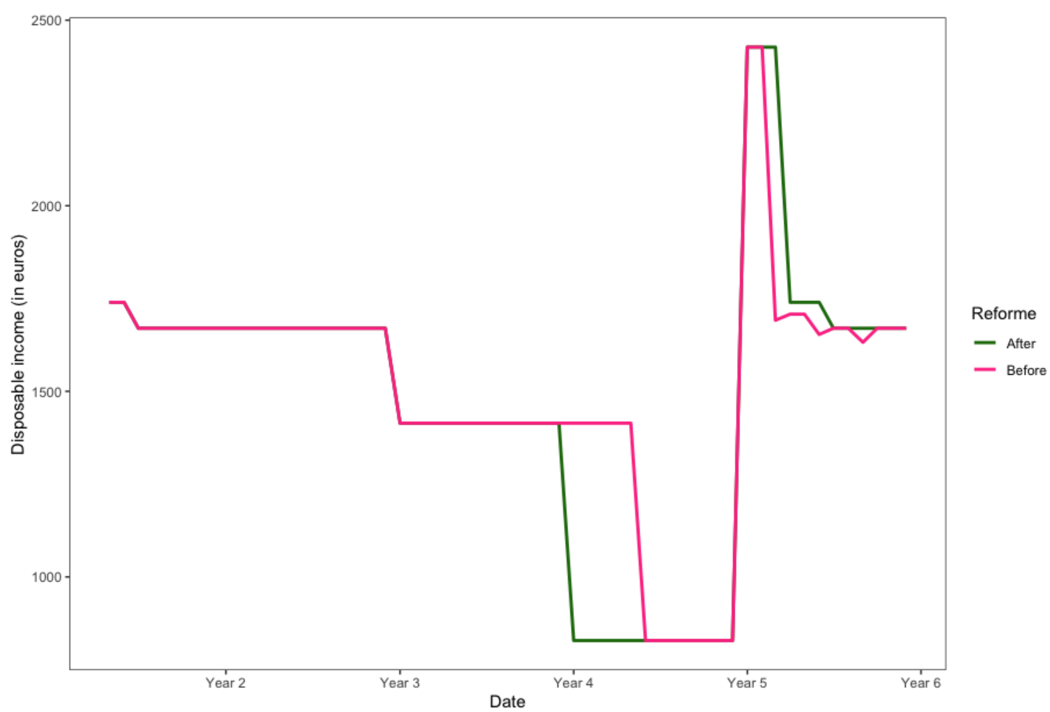


Figure 16: Couple with 2 children

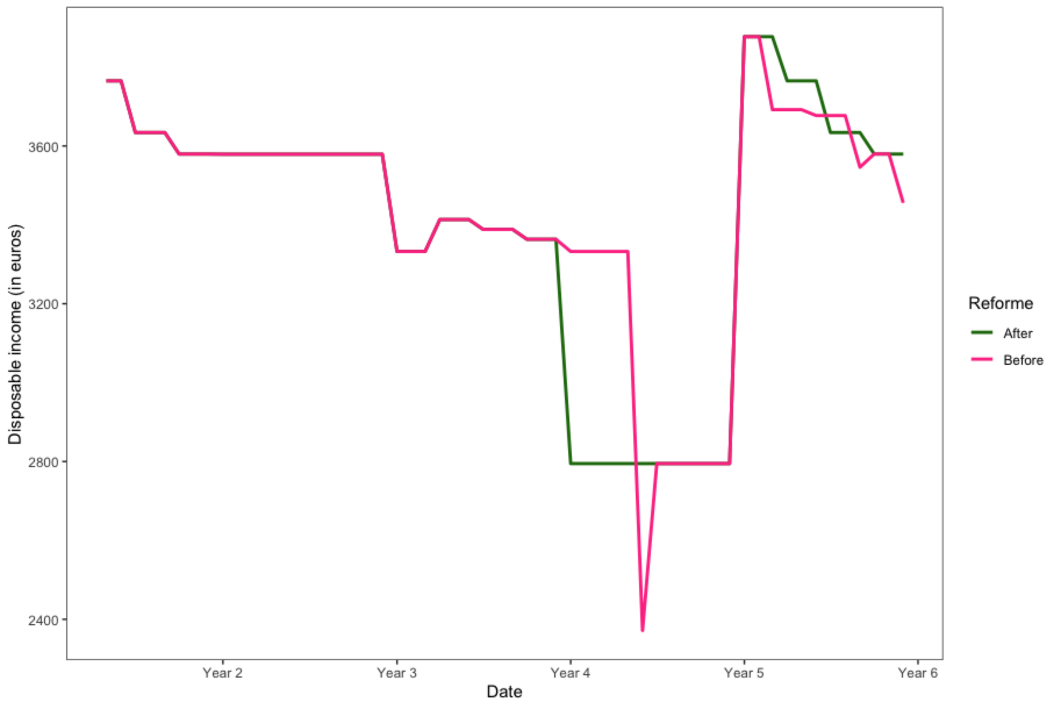


Figure 17: Seasonal couple without children

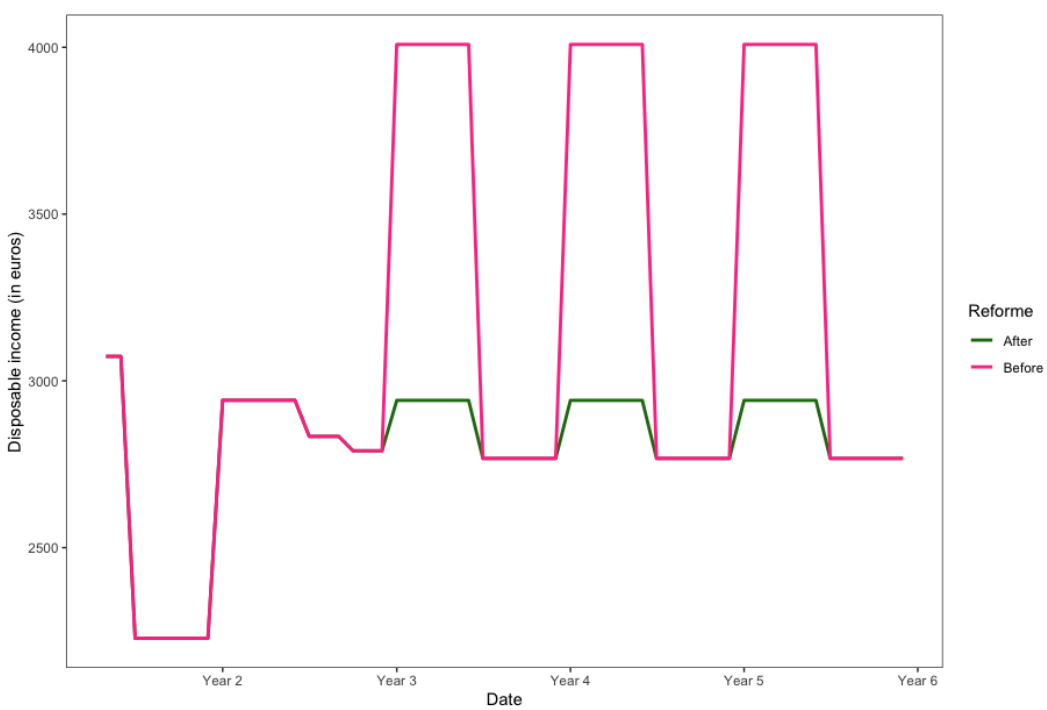


Figure 18: Single-parent family with 1 child

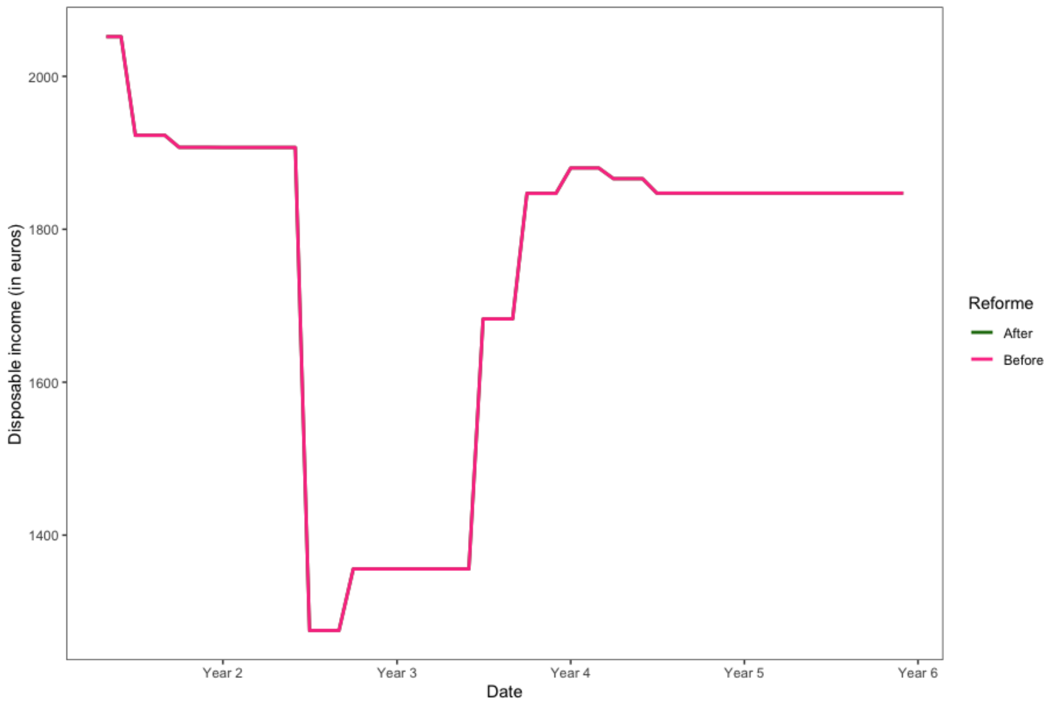


Figure 19: Single-parent family with 2 child

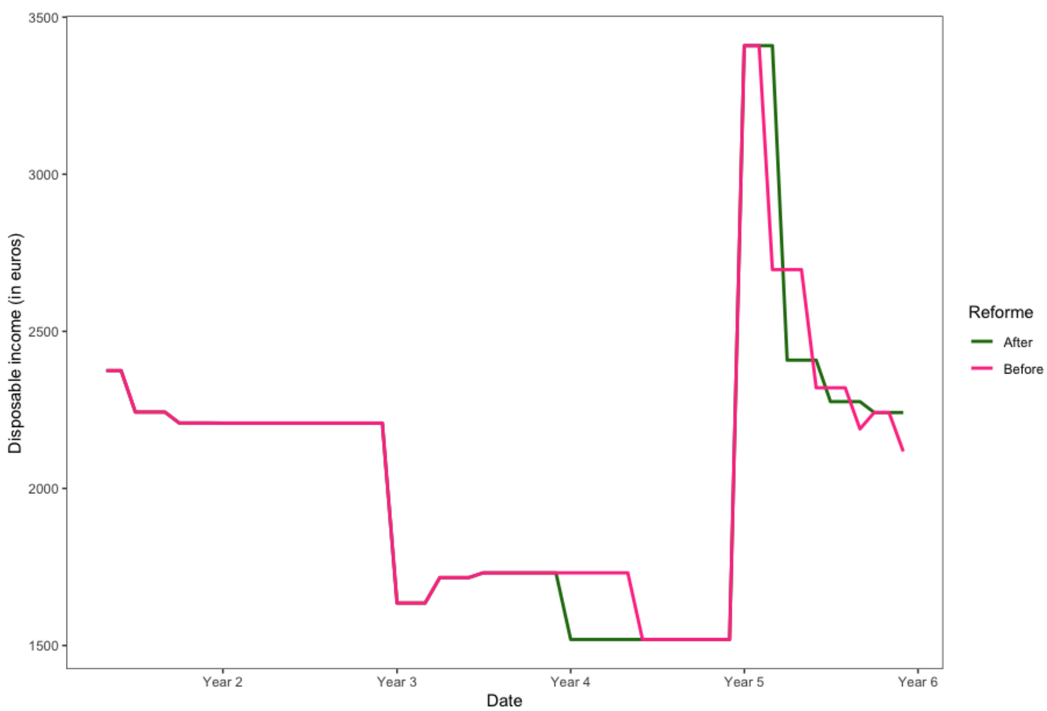


Figure 20: Single seasonal worker

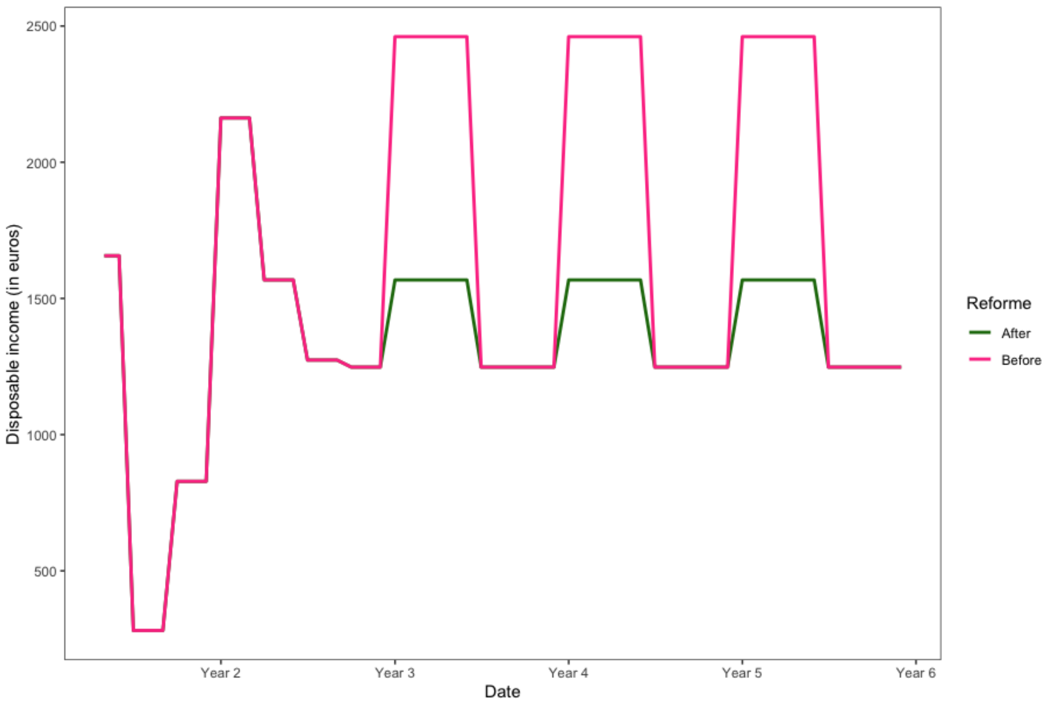


Figure 21: Single seasonal worker with 2 children