Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Federal Department of Home Affairs FDHA
Federal Statistical Office FSO
Data Science, AI and Statistical Methods / Statistical Methods
Health and Social Affairs / Social Welfare

Progression patterns in the Swiss social security system based on Machine Learning: methods for evaluating quality and model drift

Athanassia Chalimourda, Michael Leuenberger, Brandon Qorri Gonzalez, Luzius von Gunten

1 Introduction

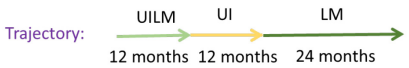
As part of the innovation project “Machine Learning – Social Security” (ML_SoSi) of the Swiss Federal Statistical Office (SFSO), we studied progressions of new unemployment insurance beneficiaries in the Swiss social security system. In the following, we will refer to *trajectories* instead of *progressions* and *recipients* instead of *beneficiaries,* since these terms correspond better to the official translation. These trajectories consist of the monthly states of an insured person over a period of four years. The four basic states and their more frequent combinations are shown in Table 1.

Table 1: The four basic states and possible combinations in the Swiss social security system

LM: Labour Market UI: Unemployment Insurance DI: Disability Incurance SA: Social Assistance	UI, UILM, DIUISA, DISA, DI, NENB UISA, DI, DIUISALM, DISALM, DILM UISALM, DIUI, UISALM, DILM, LM

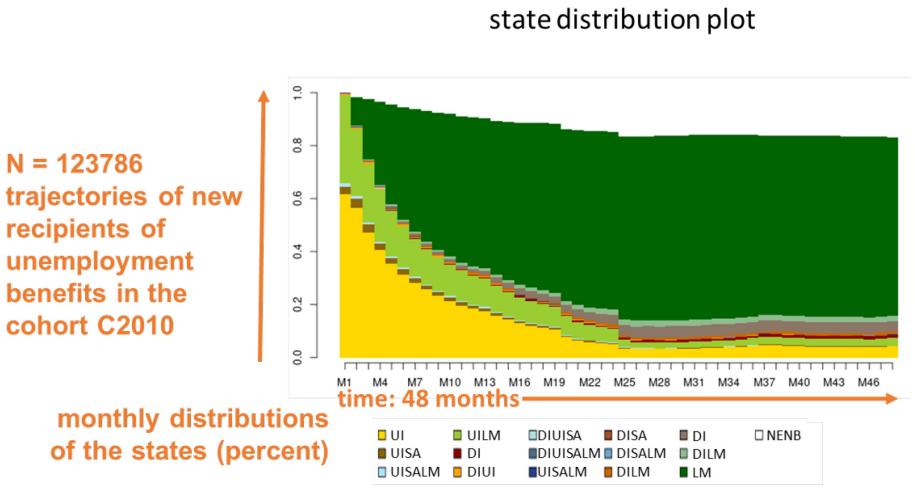
An insured person is included in a cohort of new recipients of unemployment insurance benefits for a certain year if their state contains "unemployment insurance" (UI) for the first time that year. From the month on that a person becomes a recipient of unemployment benefits, their monthly states are recorded over the observation period of 48 months. A recipient’s *trajectory* represents the sequence of 16 possible states over the four years observation time, see Figure 1 for an illustration.

Figure 1: An illustration of a recipient’s trajectory



An aggregated, summarized view of all individual trajectories of a certain year, a so-called *cohort*, is given by the *state distribution plot*. Figure 2 shows the state distribution plot for all recipients’ trajectories of the cohort of the year 2010. The x-axis extends over the 48 months of the observation time, while the y-axis shows the state percentages for each month. In the first month of the membership in the 2010 cohort over 60% of all beneficiaries have the state *unemployment insurance* (UI, in yellow), while around 35% of them have the composite state *unemployment insurance and labour market* (UILM, in light green). Minor states in the first month include *unemployment insurance and social assistance* (UISA, in brown) and *unemployment insurance, social assistance and labour market* (UISALM, in light blue). All monthly states sum up to 100%. The recipients who left the social security system have the state *no employment and no benefits* (NENB) which is shown in white, see Table 1.

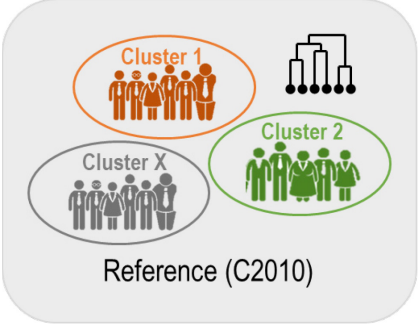
Figure 2: State distribution plot for the 2010 cohort of new recipients of the unemployment insurance.



Section 2 addresses the question, if it is possible to identify meaningful patterns in the 2010 cohort of new unemployment insurance recipients' trajectories. However, these patterns may change over time, due to individual factors, like changes in the recipients' population in terms of migration background or family composition, as well as to contextual factors, like changes in laws or economic shocks. Starting with the patterns found in the 2010 cohort of new unemployment benefits recipients as a reference, we investigated how this reference changes over time. To this end, we transferred the reference to future cohorts by means of a random forest model for prediction using the cohorts of 2011 and 2015 as examples, as described in Section 3. In Section 4 we outline methodological approaches that help quantify the evolution of the reference patterns in time in order to answer the principal question when the reference must be updated due to a possible model drift. In doing so, we rely on both internal measures for cluster validation, which characterize the quality of a clustering, and external measures, which compare the reference with alternative patterns in future cohorts of insured persons. Criteria regarding the internal and external measures that result in an update of the reference patterns must thus be defined. The transition to a new categorization poses additional challenges. For example, the correspondence between existing reference patterns and patterns of a new categorization is not always straightforward. We describe the challenges and give results on the internal and external measures in Section 5. The paper concludes with Section 6.

2 Data-driven identification of patterns in the 2010 cohort of new unemployment benefits recipients (reference)

Figure 3: Illustration of data-driven identification of trajectory patterns based on their similarity

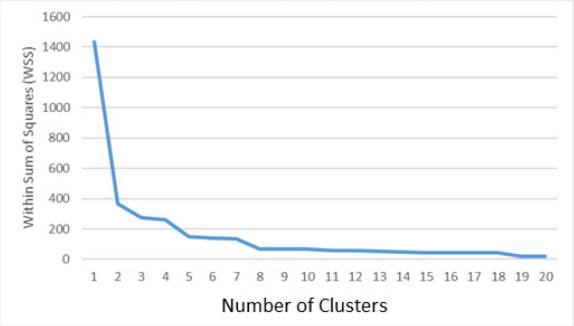


With the cohort of the new unemployment insurance recipients of the year 2010 as a reference, we posed the question if it is possible to identify patterns in the recipients' trajectories. We employed hierarchical clustering to form groups of recipients in a data-driven way. That means, the recipients' trajectories are grouped together based on their similarity as opposed to rule-driven building of trajectory groups. Figure 3 illustrates this idea.

In order to evaluate the similarity of the recipients' trajectories, we used the edit distance as a distance measure[1]. It is widely used to evaluate the similarity of sequences and strings (e.g., in translation algorithms). By interpreting the trajectories as

sequences, their similarity is calculated from the number of the operations insert, delete, substitute that are necessary to convert one trajectory into another[2].

Figure 4: The within clusters sum of squares as a function of the number of clusters for the beneficiaries' cohort of 2010 (C2010)



For the hierarchical clustering, we used the "Ward's minimum variance criterion", which forms the clusters by minimizing the within clusters variance. Based on this principle, plotting the within clusters sum of squares as a function of the number of clusters gives an indication on the optimal number of clusters. Figure 4 shows that the heterogeneity strongly decreases in cluster solutions with up to five clusters. A further decrease takes place between five and eight clusters. The number of clusters was finally fixed to ten following subject-specific considerations also due to their interpretability. In the formula below, S^2, is the within clusters sum of squares for a single cluster, $d(T_i, T_j)$ denotes the edit-distance between two trajectories, and n the number of trajectories in this cluster.

$$S^2 = \frac{2}{n(n-1)} \sum_{i,j=1}^{n} d(T_i, T_j)^2$$

We will refer to the resulting cluster solution with ten clusters based on the cohort of 2010 as *the reference*. Figure 5 shows the state distribution plots of the ten reference clusters with the percentages of the trajectories in each cluster with respect to all trajectories in the cohort of the year 2010.

The aggregated view of the state distribution plots reveals patterns in the monthly state distributions in each cluster which result from the similarity of the recipients' trajectories. The interpretation of the clusters and thus the patterns of the reference is carried out using the state distribution plots and indicators calculated on the trajectories. In summary, the clusters are roughly differentiated depending on the following:

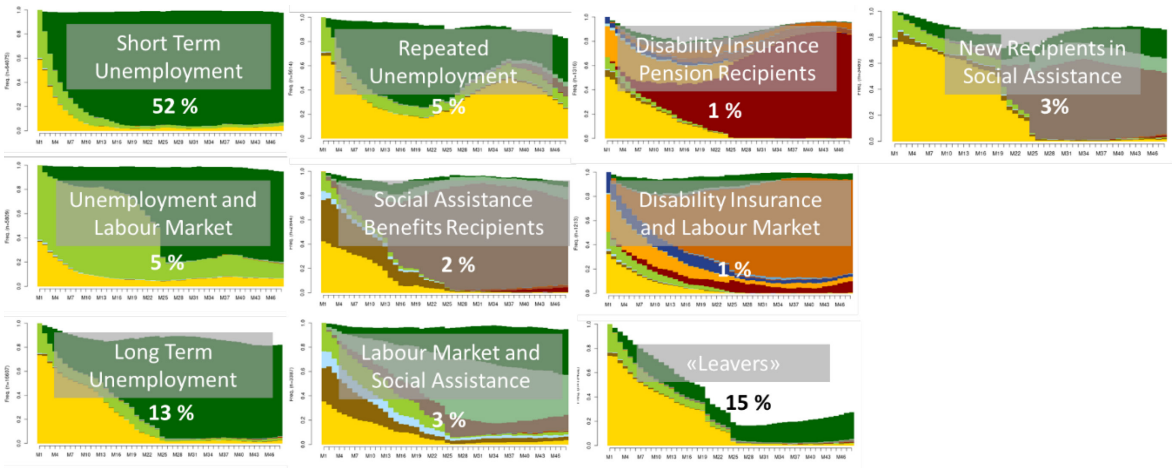
- Whether the recipients receive daily unemployment benefits alone or rather combined with other social benefits, like social assistance or disability insurance benefits
- The duration of receipt of daily unemployment benefits
- If there is a receipt of social benefits while the recipient is partly employed (see Cluster *Labour Market and Social Assistance* in Figure 5)

The biggest cluster of the reference includes 52% of the recipients' trajectories. It is labeled *Short-Term Unemployment* because it mainly contains trajectories of re-integration into the labour market after a rather short period of receiving daily unemployment benefits. On average, the first period of unemployment benefits lasts about four months, after which about 50% of the persons are fully reintegrated into the labour market. The average duration of gainful employment during the observation period is long.

For the cluster described as *Long-Term Unemployment*, the re-integration into the labour market takes place after a longer period of receiving daily unemployment benefits compared to *Short-Term Unemployment.* The first period of unemployment benefits lasts with 10 months longer than average. After about 14 months, around 50% of the recipients are fully integrated in the labour market again.

One of the smallest clusters of the reference with only 1% of the trajectories of the 2010 cohort is described as *Disability Insurance Pension*. It gathers trajectories that lead to a pension due to disability without gainful employment after an initial receipt of daily unemployment benefits for 10 months on average. Around 30% of the recipients in this group are dependent on bridging benefits from social assistance for an average of 5 months.

Figure 5: State distribution plots of each cluster of the initial cluster solution in the cohort 2010 (reference). The percentages of the trajectories in each cluster are also shown. The labels describe the patterns resulting from aggregating the individual trajectories in state distribution plots for each cluster.



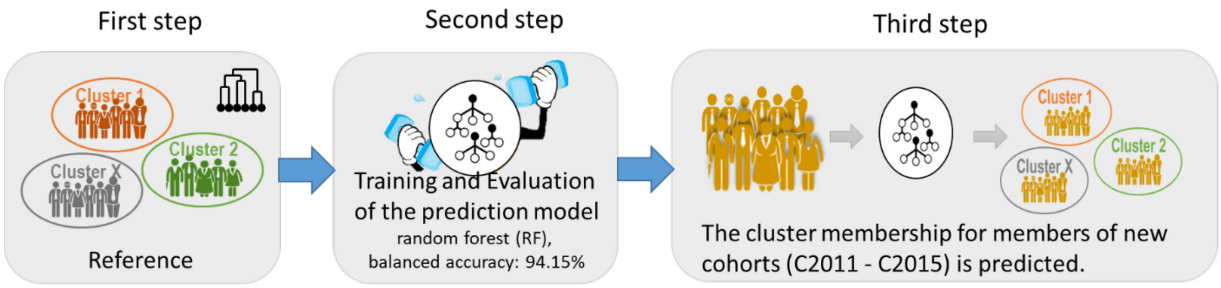
After giving detailed descriptions for three clusters in the cluster solution as examples, it can overall be said that around 70% of the persons in cohort 2010 find their way back into the labour market after initially receiving daily unemployment benefits without any major distortions in their employment biography. These are recipients from the clusters *Short Term Unemployment* (52%), *Long Term Unemployment* (13%) and *Unemployment and Labour Market* (5%). A further 16% of the cohort members withdraw from working life for various reasons or leave without needing further support from the unemployment insurance or without having or being able to appeal to the disability insurance or social assistance ("*leavers*"). In the case of a further 2%, a (partial) incapacity to work for health reasons is established during the initial daily allowance receipt, so that the recipients can claim financial benefits from the Disability Insurance (*Disability Insurance Pension* and *Disability Insurance and Labour Market*).

Since the reference clusters group similar recipients' trajectories, it is not surprising that they also exhibit similar properties. These are briefly expressed in the cluster labels of Figure 5 and emerge as patterns in the aggregated view of the state distribution plots. We therefore often refer in the following to a reference cluster as a *pattern*.

3 Transfer of the reference patterns in future cohorts of unemployment recipients

In this section we show how the patterns established in the reference clusters of the 2010 cohort can be transferred in future cohorts of new unemployment insurance recipients. To this end, a random forest is trained on the reference clusters thus providing a statistical prediction model which subsequently allows the cluster membership to be predicted for any trajectory of a cohort member. The random forest prediction model thus contains the "rules" for pattern allocation which are implicitly derived from the reference cluster solution. The prediction model is then applied to a new cohort assigning to each new recipient trajectory a predicted reference pattern. Since this assignment is associated with a certain probability, it is a probabilistic procedure in contrast to theory-based, explicit if-then rules in deterministic trajectory typologies. In a way, this procedure "projects" the ten reference patterns of the 2010 cohort in the future, thus detecting their representation in a future cohort. Differences between the reference and its prediction in future cohorts maybe due to the quality of the prediction model and differences in the data of the two cohorts. The transfer procedure is illustrated in the following figure.

Figure 6: Schematic illustration of the procedure for transferring the reference patterns of the C2010 in future cohorts, for example the cohorts of the years 2011 and 2015. In the first step the reference is established, in the second a prediction model is trained on the reference. Finally, in the third step, the model predicts the membership of a future recipient's trajectory to a reference pattern.



4 When should the reference patterns be updated?

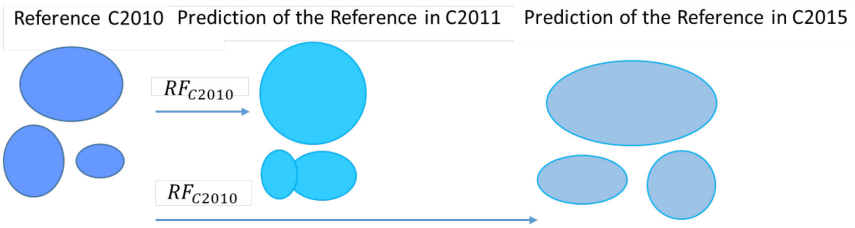
The predictions of the 2010 cohort by means of a random forest supervised machine learning model effectively transfer the reference patterns in future cohorts. However, trajectories in the social security system can change over time due to individual as well as contextual factors, as already mentioned. These changes may cause the initially established reference patterns to lose their relevance or validity. New cluster solutions representing new typical trajectory patterns may be more appropriate for the current cohort. Therefore, we need a procedure to decide when the validity of the reference patterns is no longer sufficient and these should be updated.

The potential relevance and validity loss of the reference over time, may manifest itself in two ways. First, the quality of the reference predictions may decline with time. Second, because of changes in the above-mentioned factors, new, more relevant patterns for the current cohort may emerge. These can be extracted by applying the same procedure of hierarchical clustering used to extract the reference.

4.1 Comparison of the reference with its predictions in future cohorts

We can evaluate the quality of the reference predictions by a set of measures which are characteristic for a cluster solution and thus are referred in the following as *internal measures of cluster validation*[3]. They are calculated and compared for both the reference and its predictions. A deterioration of the internal measures of the predictions with growing temporal distance from the reference might mean that its relevance diminishes with time, see Figure 7 for an illustration.

Figure 7: Schematic illustration of the transfer of the 2010 reference in future cohorts via prediction by a random forest model.



The internal measures of cluster validation in the present work consist of the number and percentage of recipients' trajectories in each cluster, average and maximum distances, proportions of the within clusters to overall variance and average silhouette coefficients for each cluster as well as for the whole cluster model. For example, an increase of the proportion of the within clusters' variance with respect to the overall variance may point to decreasing density and homogeneity in the reference patterns. An increase of the distances between trajectories in the patterns may also point to the same fact. On the other hand, a decrease of the average silhouette coefficient in the clusters may indicate a worsening of the separability of the patterns with time.

In order to define the silhouette coefficient[4] for each trajectory i, let a_i be its mean distance from all other trajectories in its cluster. Let b_i be its mean distance from all the trajectories of the nearest cluster, meaning the cluster with the smallest mean distance to i. The silhouette coefficient, s_i, of trajectory i is then defined as follows:

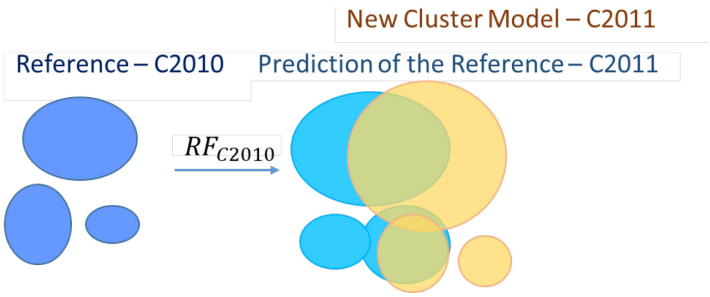
$$s_i = \frac{b_i - a_i}{\max(b_i, a_i)}.$$

Well classified trajectories have a small mean distance a_i to the trajectories of their own cluster and a large mean distance b_i to the trajectories of the nearest cluster, resulting to a s_i value close to 1. If s_i is close to zero, this trajectory could also be assigned to neighbouring clusters. A negative coefficient means that the trajectory is in the wrong cluster. The silhouette coefficient s_i takes values between -1 and 1. The average silhouette coefficient of a cluster is the average of the silhouette coefficients of its trajectories.

4.2 Comparison of reference predictions with new cluster models in future cohorts

A comparison of the reference via its prediction with a new cluster model which represents a potential new reference is schematically shown in Figure 8. If the prediction of the reference and the new cluster model differ significantly, then the reference patterns may have lost their relevance and validity and should be updated. Measures that compare two cluster solutions with each other, the so-called *external measures of cluster validation*[3], help to quantify this comparison. As indicated in Figure 8, this comparison poses additional challenges as not all clusters in a new cluster model have straightforward counterparts in reference patterns. Some clusters may exhibit an obvious correspondence with a reference pattern, while others are divided between different reference patterns or just do not have a correspondent among the reference patterns at all.

Figure 8: Illustration of the comparison of the prediction of the 2010 reference in the 2011 cohort and a new cluster model, a potential new reference, in the same cohort.



In the following section we present results on some measures based on the *confusion matrix*[5], the *Accuracy*, *Balanced Accuracy* and *Cohen's Kappa*. Each row of the confusion matrix represents the trajectories in a cluster of the reference prediction while each column represents the trajectories in a cluster of a new model for the cohort at hand. The *Accuracy* gives the proportion of the trajectories that coincide in all clusters of the prediction and the new cluster model with respect to all trajectories. The proportion of the common trajectories of the reference prediction and a cluster in the new model with respect to the prediction cluster shows how well the corresponding reference pattern is represented in the new cluster. The *Balanced Accuracy* is computed as the average of the above proportions for all clusters of the reference prediction. Since it is an average of proportions, small clusters like the *Disability Insurance Benefits Recipients* contribute equally to this measure as large ones like *Short Term Unemployment*, see Figure 5. *Cohen's Kappa*[6] measures the agreement between the prediction and the new cluster model by taking into account the agreement occurring by chance. Since *Short Term Unemployment* gathers more than half of the 2010 cohort's trajectories it makes sense to also use a measure which takes into account the random assignment of the trajectories to the reference patterns. Cohen's Kappa is defined as:

$$k = \frac{p_o - p_e}{1 - p_e},$$

where p_o is the relative observed agreement between the two cluster models and p_e the expected agreement in the case where the trajectories are randomly assigned to the reference patterns.

5 Results

5.1 Comparison of the reference with its predictions in future cohorts

The evolution of the reference in time is expressed by its predictions which project the reference patterns in future cohorts. A visual comparison of the reference with its predictions by means of the state distribution plots is the first step in the comparison between the reference and its predictions. It is followed by a comparison of the internal measures of cluster validation which monitor the quality of the predictions and quantify possible deterioration with time.

5.1.1 Comparison of the state distribution plots

Figure 9: Comparison of the state distribution plots of two patterns of the 2010 reference and their predictions in the cohorts 2011 and 2015.

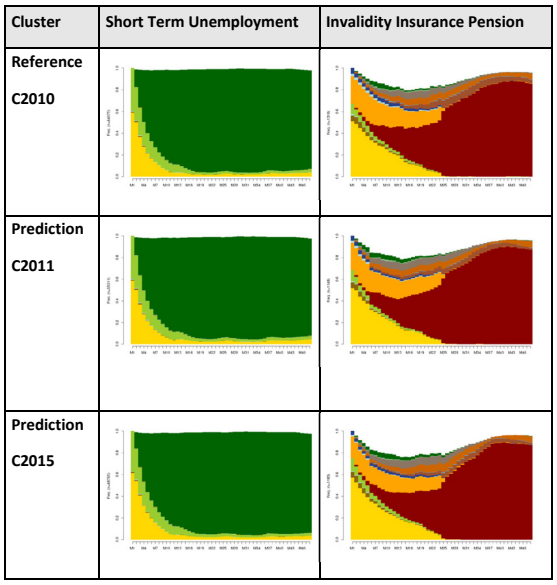


Figure 9 shows the state distribution plots of two clusters of the reference associated with the patterns of *Short-Term Unemployment* and *Disability Insurance Benefits Recipients* and their predictions in the cohorts 2011 and 2015. These two clusters correspond to the biggest (52%) and one of the smallest (1%) patterns of the reference. The similarity between the reference and the prediction patterns extends to all patterns of the reference, see Figure 5. It suggests that the random forest prediction model is able to group the trajectories in the cohorts C2011 and C2015 in a similar way to the respective reference patterns.

5.1.2 Comparison of the internal measures for the reference and its predictions

The comparison of the state distribution plots is followed by a quantitative comparison between the reference and its predictions. Table 2 shows an extract of some of the above-mentioned internal measures for three reference patterns and their predictions in the cohorts C2011 and C2015: the *Short-Term Unemployment*, the *Disability Insurance Benefits Recipients* and the *Long-Term Unemployment* which contain 52%, 1% and 13% of the recipients' trajectories respectively in the 2010 cohort.

We observe that the internal measures have similar values for the reference and its predictions. Thus, they do not seem to deteriorate in time because of the transfer of the reference to the future cohorts C2011 and C2015. Given the similarity of the state distribution plots illustrated in Figure 9, this is not surprising. The results are similar for the remaining eight clusters and the cohorts C2012 to C2014. Therefore, the internal measures do not give sufficient evidence for the need of the reference to be updated.

Table 2: Comparison of internal measures for the reference and its predictions in the cohorts 2011 and 2015. The percentage of the within-clusters to the overall variance quantifies the homogeneity of the patterns while the silhouette coefficient their separability within the cluster model.

	Cohort	Cluster Model	Short-Term Unemployment	Invalidity Insurance Pension	Long-Term Unemployment
Size (percentage)					
	2010	100	52	1	13
	2011	100	51	1	14
	2015	100	50	1	16
Homogeneity					
Variance (percentage)	2010	5.548	3.223	0.009	0.636
	2011	5.890	3.418	0.006	0.659
	2015	6.105	3.236	0.003	0.813
Separability					
Silhouette Coeffizient	2010	0.34	0.48	0.34	0.05
	2011	0.34	0.48	0.34	0.05
	2015	0.32	0.47	0.30	0.09

The average silhouette coefficient reaches its highest value for the biggest reference cluster that corresponds to the *Short-Term Unemployment* pattern. Its values for the remaining reference patterns range between near zero as for the pattern of *Long-Term Unemployment* and around 0.3, as for the pattern of *Invalidity Insurance Pension*, see Table 2. That means that some patterns are not well separated from the others. This effect does not seem due to the transfer of the reference or changes in factors of the social security system, it is already present in the reference patterns.

5.2 Comparison of the reference prediction with cluster models in future cohorts

A new reference of trajectories' patterns can be established in a future cohort following the same procedure of hierarchical clustering as for the 2010 cohort, described in section 3. If this, potentially new reference, is very different from the prediction of the 2010 reference in this cohort, we should consider its replacement by the new one. We refer to this replacement as the *actualization*, or simply *update* of the reference. The comparison will be done both visually and quantitatively as for the comparison between the reference and its predictions in Section 5.1.

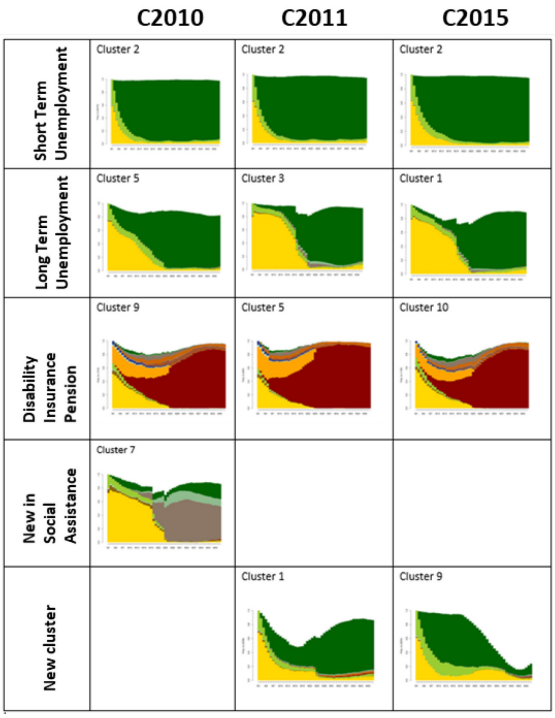
5.2.1 Comparison of the state distribution plots

The comparison of the state distribution plots of some reference patterns with the state distribution plots of new cluster models in the 2011 and 2015 cohorts is shown in Figure 10. There are reference patterns that correspond well to clusters of the new model, independently of their size, like *Short Term Unemployment* and *Disability Insurance Pension Recipients*. The largest reference pattern, *Short Term Unemployment,* consistently reappears in all cohorts between 2011 to 2015 retaining its relative size of about 50% of the respective cohort. Other patterns, however, like *New in Social Assistance* do not find a counterpart in the cohorts 2011 and 2015.

While most patterns find obvious counterparts in future cluster models, there are patterns, like *Long Term Unemployment,* for which this is not as straightforward. The 2011 cluster that shows a similar state distribution plot to the prediction of *Long-Term Unemployment* contains only 35% of its trajectories, while 34% are assigned to another reference pattern. The remaining 31% trajectories are scattered in the other clusters of the reference prediction. In ambiguous cases like these, we used a ratio based on the Jaccard index[7] in order to decide which cluster corresponds better to the prediction A, of a reference pattern, being *Long Term Unemployment* in our example. This is cluster X of the new cluster model which maximizes the ratio $A \cap X / A \cup X$ of the mutual to overall trajectories with respect to A. However, after this procedure, there may still be reference patterns, like *New in Social Assistance* which do not find counterparts in new models in future cohorts and vice versa, new clusters which cannot be associated with a reference pattern in the 2010 cohort.

Figure 10: State distribution plots of the 2010 reference and potential new references, i.e., new cluster models, in the 2011 and 2015 cohorts. The patterns shown are Short-Term Unemployment, Long-Term Unemployment,

Invalidity Insurance Pension and New (recipients) in Social Assistance, a cluster that includes 3% of the recipients' trajectories in 2010.



Ambiguity in the correspondence between reference patterns and clusters in future cohorts appears at least once in the comparison between reference prediction and new model in the cohorts 2011 to 2015. At least one pattern, like *New in Social Assistance* in Figure 10, does not have a counterpart in the new cluster model for the compared cohorts. Vice versa, clusters may appear in the new model which lack correspondence to a reference pattern.

5.2.2 Comparison of the external measures for cluster validation

Both the ambiguity in the correspondence as well as the lack of it between reference patterns and potentially new ones influence the external measures of cluster validation which compare whole cluster models with each other. In the case of ambiguity, the contribution of the associated clusters in the external measure is small while in the absence of correspondence it is zero. Table 3 shows the values of Accuracy, Balanced Accuracy and Cohen's Kappa for the comparison of the prediction of the 2010 reference to new cluster models in the 2011 and 2015 cohorts.

The higher value of Accuracy compared to Balanced Accuracy and Cohen's Kappa is due to the inherent imbalance in the cluster models. The largest reference cluster which corresponds to the *Short-Term Unemployment* pattern occupies around 50% in each cohort and is the most homogeneous and well separated cluster in all cluster models. It would still contain many of its trajectories even if these were assigned to the ten clusters purely by chance. All clusters participate equally to the value of Balanced Accuracy, independently of their size. Thus, the afore mentioned ambiguity and lack of correspondence which tentatively take place in smaller clusters affect Balanced Accuracy more than Accuracy. Cohen's Kappa also corrects the value of the Accuracy which stands for the relative observed agreement p_o, by the relative agreement as expected by chance, p_e.

Table 3: Accuracy, Balanced Accuracy and Cohen's Kappa for the comparison of the prediction of the reference to a new cluster model in the cohorts 2011 and 2015.

External Measures	Reference - CM 2011	Reference - CM 2015
Accuracy	0.77	0.77
Balanced Accuracy	0.66	0.67
Cohen's Kappa	0.67	0.66

Although the values of the three external measures are not particularly good, their stability in time is surprising. Since they do not show any deterioration in time, they do not indicate an update of the reference.

6 Conclusions

The present work shows that it is possible to find meaningful patterns in the trajectories of new recipients of unemployment benefits in a data-driven way based on their inherent similarities. Subsequent application of supervised machine learning transfers the patterns to future cohorts of new recipients by regrouping the latter according to the initial patterns' characteristics. Since new recipients' trajectories may change over time due to both individual and contextual factors, the question arises when the initial patterns, referred to as *the reference*, lose their relevance and validity over time and need to be updated. An example of such an update would be to replace the reference with a new one based on trajectories of a future cohort which are thought to better reflect the current situation in the social security system. To this purpose, we employed both internal and external measures for cluster validation. The internal measures showed that the predictions of the reference in future cohorts are very similar to the reference itself, thus indicating no need for its update. The comparison of the reference predictions with new potential references in future cohorts by means of external measures leads to the same conclusion. Neither the examined internal nor the external measures indicate a need for a reference update in five consecutive years. What became evident though, is that the reference itself should be improved since the initial separability of some of its patterns is weak and their correspondence to future patterns not straightforward. Possible improvements include, for example, adapting substitution costs in the distance measure to subject-specific expertise in the social security system as well as to project objectives. A solution with ten clusters was preferred from a subject-specific point of view but not necessarily indicated from a statistical point of view. The silhouette coefficient and additional measures on cluster stability could further support the decision on the optimal number of initial reference patterns. One should also keep in mind that the state distribution plots reveal patterns as an aggregated view of individual recipients' trajectories. The latter can combine characteristics of different patterns, especially if the trajectories cover a longer period of time. This might keep their assignment to a reference pattern challenging in spite of methodological improvements. The above-mentioned aspects and the resulting knowledge gain may improve the reference and make easier to identify when it needs to be updated.

7 References

1. Navarro, G (2001). "A guided tour to approximate string matching". *ACM Computing Surveys*. 33(1): 31–88.
2. Gabadinho A, Ritschard G, Müller N, Studer M (2011). "Analyzing and Visualizing State Sequences in R with TraMineR." *Journal of Statistical Software*, 40(4), 1-37.
3. Rendón E, Abundez I, Arizmendi A, Quiroz E M (2011). "Internal versus External Cluster Validation Indexes". *International Journal of Computers and Communications*. 5(1): 27-34.
4. Rousseeuw P J (1987). "Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis". *Computational and Applied Mathematics*. 20: 53–65.
5. Ting K M. (2017). Confusion Matrix. In: Sammut, C., Webb, G.I. (eds) *Encyclopedia of Machine Learning and Data Mining*. Springer, Boston, MA.
6. Cohen J (1960). *"A coefficient of agreement for nominal scales".* Educational and Psychological Measurement. 20 (1): 37–46.
7. Jaccard P (1912). "The Distribution of the Flora in the Alpine Zone.1". *New Phytologist*. 11 (2): 37–50.