FLEXIBLE EMPLOYMENT, A MACHINE LEARNING APPROACH

Marco Alfó [1], Dimitris Pavlopoulos[2] and Roberta Varriale[1]

[1] Department of Statistical Science, Sapienza University of Rome, (e-mail: marco.alfo@uniroma1.it, roberta.varriale@uniroma1.it)

[2] Department of Sociology, Vrije Universiteit Amsterdam, (e-mail: d.pavlopoulos@vu.nl)

ABSTRACT: Flexible employment is an important topic in scientific and political debate in Europe. The present work describes the use of machine learning techniques to predict the contract type, by using information coming from both survey and administrative data. Information on contract type come from linked data from the Labour Force Survey and the Employment Register of the Netherlands for the period 2007-2015.

KEYWORDS: flexible employment; machine learning; multi-source data

1 Introduction

Flexible employment is an important topic in scientific and political debate in Europe. In Eurozone countries, OECD statistics (https://stats.oecd.org) show that the incidence of temporary employment was 11.8 percent in 2021, while the probability of getting a job on a temporary contract increased by 36 percent between 2013 and 2019 (Latner, 2022). The Netherlands also saw a sharp increase in the incidence of temporary employment in 2021: from 13.7 percent in 2000 to 27.4 percent. In Italy, the increase was a bit lower, from 10.1 percent in 2000 to 16.6 percent in 2021. The role of temporary contracts can be assessed from a life course perspective: one of the main questions the research seeks to answer is whether temporary work is always a stepping stone to permanent employment or it should be rather considered as a trap of precarious jobs (Latner & Saks, 2022).

Data to study flexible employment dynamics may come from different sources, such as survey, administrative and statistical register data. Measurements from different sources may not agree for different reasons, including the presence of measurement error or misalignment in the definitions or between occasions of measurement. For example, there could be temporal misalignment of the sources, structural lack of administrative information on irregular work, misalignment of employment definition in the available sources.

Findings on mobility from temporary to permanent employment can be severely biased due to measurement error, usually present in the information used for analysis, coming from both survey and statistical register data (see, for example, Pavlopoulos & Vermunt, 2015; and Pankowska *et al.*, 2021). A possible approach to deal with measurement error when multiple data sources are available is based on the use of latent variable models. In particular, latent variable models can be used to predict the true target value (here, the current type of contract) given the observed measurements in the data sources when all these data sources contain information closely related to the target variable, but none can be assumed to be error free (Filipponi *et al.*, 2021). An alternative approach to deal with data coming from multiple possibly discordant sources is based on Machine Learning (ML) tools for supervised classification (Varriale & Alfo', 2023). ML tools may be used to predict the individual target variable, and to extract important information from the data to learn more about the phenomenon in the form of a selection of possibly important predictors of the response.

The present work describes the use of some ML techniques, including decision trees and random forests, to predict the individual contract type. We use linked data drawn from the Labour Force Survey (LFS) and the Employment Register of the Netherlands for the period 2007-2015. The aim of this paper is to show how ML techniques can be used with longitudinal data to extract important information for the purpose of estimating the probability of a temporary employment contract in the life course, and to learn more about the phenomenon.

2 The context

The data sources providing information on types of employment contracts are the Labour Force Survey (LFS) administered by Statistics Netherlands and the Employment Register of the Netherlands (ER). LFS represents the main source of information on the labour market for official statistics. It produces information on employment and the main aggregates of the job offer - profession, sector of economic activity, hours worked, type and duration of contracts, training. LFS is harmonized at the European level as established by the EU Regulation 2019/1700 of the European Parliament and the Council. In the Netherlands, the LFS has a rotating trimonthly scheme and it is representative for the Dutch population aged 15 or more. Since 1999, respondents are interviewed at 5 consecutive panel waves. The collected information refers to the moment of the interview, and the interviews are carried out during every week of the trimester. Table 1 show the LFS rotating scheme for two years.

Table 1. *LFS rotating scheme for two years.*

	$year_1$				$year_2$			
Sample	q_1	q_2	q_3	q_4	q_1	q_2	q_3	q_4
1	x	x	x	x	x	.	.	.
2	.	x	x	x	x	x	.	.
3	.	.	x	x	x	x	x	.
4	.	.	.	x	x	x	x	x
5	.	.	.	.	x	x	x	x
6	.	.	.	.	.	x	x	x
7	.	.	.	.	.	.	x	x
8	.	.	.	.	.	.	.	x

The ER is a register administered by the Institute for Employee Insurance (UWV), containing information on labour market and income for all insured workers in the Netherlands. The ER is constructed by collecting and matching information from various sources, i.e. the Tax Office, the Population Register and information drawn from temporary work agencies' registries (Bakker *et al.*, 2014). The submission of tax-reporting statements is compulsory for employers. However, while ER dataset contains monthly information, employers typically submit the relevant information only few times per year. This may, at least potentially, produce some errors, in particular for the information regarding the period between two consecutive submissions. Additional sources of measurement error in ER may result from administrative delays, wrong registration, and erroneous administrative procedures.

3 A machine learning approach

A ML approach for supervised classification is applied to predict individual contract type as a function of individual features and time. Categories of contract type can be classified as "permanent"/"non-permanent." The latter category can be divided into "fixed-term", "temporary or on-call", and "other". The response is contract type, and models considering both response at 2 and 4 categories as target variable are considered and estimated.

Let y_{ijt} denotes the binary indicator for contract type j at occasion t for individual i ($i = 1,\ldots,n$, t=1,. . . ,T, $j = 1,\ldots,m$). In this work $T = 32$, each time t corresponding to a specific quarter of the year. We use multiple strategies of analysis. The first strategy involves using y_{ijt} as the target variable and all the information available in previous times as covariates. Therefore, we want to model the conditional expectation $E(y_{ijt}|x_{it},x_{it-1},\ldots,x_{i1})$. The second strategy uses as covariate also the information on the target variable y at time $t-1$, in order to take into account the longitudinal structure of the data by defining a formal for the conditional expectation $E(y_{ijt}|y_{ijt-1},x_{it},x_{it-1},\ldots,x_{i1})$. In particular, we are assuming that the evolution of the contract type is governed by a first order Markov chain with transition matrix that do not depend on time. Last, we will consider a first order non homogeneous Markov chain where transition probabilities may depend on individual features and or time. ML techniques are applied using the R software.

The aim is to show how ML can be used with longitudinal data, both for prediction and to extract the relevant information.

References

BAKKER, B.F., VAN ROOIJEN, J., & VAN TOOR, L. 2014. The system of social statistical datasets of statistics netherlands: An integral approach to the production of register-based social statistics. *Statistical Journal of the IAOS*, **30**(4), 411–424.

FILIPPONI, D., GUARNERA, U., & R., VARRIALE. 2021. Latent Mixed Markov Models for the Production of Population Census Data on Employment. *Book of short papers SIS 2021*, 112–117.

LATNER, J.P. 2022. Temporary employment in Europe: stagnating rates and rising risks. *European Societies*, **24**(4), 383–408.

LATNER, J.P., & SAKS, N. 2022. The wage and career consequences of temporary employment in Europe: Analysing the theories and synthesizing the evidence. *Journal of European Social Policy*, **32**(5), 514–530.

PANKOWSKA, P., BAKKER, B., OBERSKI, D., & PAVLOPOULOS, D. 2021. Dependent interviewing: a remedy or a curse for measurement error in surveys? *Survey Research Methods*, **15**(2), 135–146.

PAVLOPOULOS, D., & VERMUNT, J.K. 2015. Measuring Temporary Employment. Do Survey or Register Data Tell the Truth. *Survey Methodology*, **41**(1), 197–214.

VARRIALE, R., & ALFO', M. 2023. Multi-source statistics on employment status in Italy, a machine learning approach. *METRON*.