A finite mixture latent trajectory model for hirings and separations in the labor market

Bacci, Silvia and Bartolucci, Francesco and Pigini, Claudia and Signorelli, Marcello

Department of economics, University of Perugia

November 2014

A finite mixture latent trajectory model for hirings and separations in the labor market

Silvia Bacci, Francesco Bartolucci, Claudia Pigini and Marcello Signorelli

Abstract We propose a finite mixture latent trajectory model to study the behavior of firms in terms of open-ended employment contracts that are activated and terminated during a certain period. The model is based on the assumption that the population of firms is composed by unobservable clusters (or latent classes) with a homogeneous time trend in the number of hirings and separations. Our proposal also accounts for the presence of informative drop-out due to the exit of a firm from the market. Parameter estimation is based on the maximum likelihood method, which is efficiently performed through an EM algorithm. The model is applied to data coming from the Compulsory Communication dataset of the local labor office of the province of Perugia (Italy) for the period 2009-2012. The application reveals the presence of six latent classes of firms.

1 Introduction

Recent reforms of the Italian labor market [4] have shaped a prevailing dual system where, on the one side, workers with an open-ended contract benefit from a high degree of job security (especially in firms with more than 15 employees) and, on the other, temporary workers are exposed to a low degree of employment protection.

Silvia Bacci
Department of Economics, University of Perugia, Via A. Pascoli 20, 06123 Italy, e-mail: silvia.bacci@unipg.it

Francesco Bartolucci
Department of Economics, University of Perugia, Via A. Pascoli 20, 06123 Italy, e-mail: francesco.bartolucci@unipg.it

Claudia Pigini
Department of Economics, University of Perugia, Via A. Pascoli 20, 06123 Italy, e-mail: pigini@stat.unipg.it

Marcello Signorelli
Department of Economics, University of Perugia, Via A. Pascoli 20, 06123 Italy, e-mail: marcello.signorelli@unipg.it

Several policy interventions have been carried out with the purpose of improving the labor market performance and productivity outcomes. The effects of employment protection legislation in Italy have been investigated mainly with respect to firms' growth and to the incidence of small firms [9]. The empirical evidence points towards a mild effect of these policies on firms' growth: Schivardi and Torrini [9] state that firms avoid the costs of highly protected employment by substituting permanent employees with temporary workers; Hijzen, Mondauto and Scarpetta [4] find that employment protection has a sizable impact on the incidence of temporary employment. In this context, the analysis of open-ended employment turnover may shed some light on whether the use of highly protected contracts has declined especially in relation to the recent economic crisis.

In order to analyze the problem at issue we use the data from the Compulsory Communication (CC) dataset of the labor office of the province of Perugia (Italy) in the period 2009-2012 and we introduce a latent trajectory model, which is based on a finite mixture of logit and log-linear regression models. A logit regression model is specified to account for the informative drop-out due to the exit of a firm from the market in a certain time window, mainly due to bankruptcy, closure of the activity, or termination. Besides, conditionally to the presence of a firm in the market, two log-linear regression models are defined for the number of open-ended hirings and separations observed at every time window. Finally, we assume that firms are clustered in a given number of latent classes that are homogeneous with respect to the behavior of firms in terms of open-ended hirings and separations, other than in terms of probability of exit from the market.

The paper is organized as follows. In Section 2 we describe the data. In Section 3 we first illustrate the model assumptions and, then, we describe the main aspects related to the model estimation and to the selection of the number of latent classes. In Section 4 we apply the proposed model to the data from the Compulsory Communication dataset of local labor office of Perugia. Finally, we conclude the work with some remarks.

2 Data

The CC dataset is an Italian administrative longitudinal dataset collected by the Ministry of labor, health, and social policies from local labor offices. With the ministerial decrees n. 181 and n. 296, since 2008 Italian firms and Public Administrations (PAs) are required to transmit a telematic communication for each hiring, prolongation, transformation, or separation (i.e., firing, dismissal, retirement) to the qualified local labor office. The local labor office of Perugia provided all the communications from January 2009 to December 2012 sent by firms and PAs operating in the province of Perugia. The dataset contains information on the single contracts as well as the workers concerned by each communication and the firms/PAs sending the record to the local labor office.

The single CC represents the unit of observation for a total of 937,123 records. In order to avoid a possible distortion due to new-born firms in the period 2009-2012, we consider only firms/PAs that sent at least one communication in the first quarter of 2009 and those communicating separations of contracts that started before 2009. Once these firms have been selected, we end up with 34,357 firms/PAs in our dataset. Note that if firms/PAs do not send any record between 2009 and 2012 they do not appear in the dataset. The number of firms and PAs entering the dataset in each quarter is reported in the first column of Table 1. In addition, firms exiting the market must be accounted for: relying on the information on the reason for the communicated separations, if the firm communicates a separation for closing in a given quarter and no communications are recorded for the following quarters, we consider the firm closed from the quarter of its latest communication onward. The number of firms closing is 1,132.

In our analysis, we only consider open-ended contracts: for each firm we retrieve the number of open-ended contracts activated and terminated in each quarter. The total number of hirings and separations are reported in Table 1 for each quarter. The other available information at the firm level in the CC dataset concern the sector of the economic activity and the municipality in the province of Perugia where the firm/PA is operating. Sectors are identified by the ATECO (ATtività ECOnomiche) classification used by the Italian Institute of Statistic since 2008 (Table 2). The number of firms/PAs in each municipality is displayed in the second column of Table 2.

Table 1 Data description, by quarter

Quarter	Number of firms	Hirings	Separations	Quarter	Number of firms	Hirings	Separations
2009:q1	5,487	2,403	3,740	2011:q1	962	1,280	1,910
2009:q2	2,947	1,450	2,616	2011:q2	673	1,055	1,551
2009:q3	2,086	1,018	2,397	2011:q3	522	773	1,369
2009:q4	2,659	1,215	3,220	2011:q4	658	1,059	1,641
2010:q1	1,664	1,345	2,342	2012:q1	6,936	11,749	17,405
2010:q2	1,116	1,149	1,971	2012:q2	2,753	9,001	15,257
2010:q3	875	953	1,823	2012:q3	2,049	9,956	17,526
2010:q4	1,065	986	2,147	2012:q4	1,905	7,150	13,131

3 The latent trajectory model

The application concerning the behavior of firms/PAs (in the following we use the term firm to indicate both firms and PAs) in terms of open-ended hirings and separations during the period 2009-2012 relies on a finite mixture latent trajectory model, the assumptions of which are described in the following. Then, we give some details on the parameter estimation based on the maximization of the log-likelihood, and, finally, we deal with the issue of model selection.

Table 2 Sectors of economic activity and municipalities

Sector	Number of firms	Municipality	Number of firms
Accommodation and food	2,770	Assisi	1,152
Activities of extraterritorial organizations	10	Bastia Umbra	944
Activities of households as employers	6,793	Castiglione del Lago	546
Administrative and support activities	1,057	Città di Castello	1,780
Agriculture, forestry and fishing	1,690	Corciano	819
Arts, sports, entertainment and recreation	705	Foligno	2,221
Constructions	4,144	Gualdo Tadino	552
Education	568	Gubbio	1,295
Electricity, gas, air-conditioning supply	47	Magione	515
Financial and insurance activities	425	Marsciano	655
Health and social work activities	607	Perugia	7,795
Information and Communication	958	Spoleto	1,763
Manufacturing products	4,723	Todi	781
Mining and quarrying products	46	Umbertide	708
Other personal service activities	1,829	Other	12,831
Professional, scientific, technical activities	1,388		
Public administration and defense	247		
Real estate activities	202		
Transport and storage	1,377		
Waste-management	124		
Wholesale and retail trade	4,647		

3.1 Model assumptions

We denote by i a generic firm, $i = 1,\dots,n$, and by t a generic time window, $t = 1,\dots,T$; in our application, we have $n =$34,357 and $T = 16$. Moreover, let S_{it} be a binary random variable for the status of firm i at time t, with $S_{it} = 0$ for a status in which the firm is operating and $S_{it} = 1$ in case of firm's activity cessation in that quarter. For a firm performing well we expect to observe all values of S_{it} equal to 0. Finally, we introduce the pair of random variables, Y_{1it} and Y_{2it}, denoting the number of open-ended employment contracts that firm i activated and terminated at time t, respectively. The observed number of hirings and separations is denoted by y_{1it} and y_{2it} and it is available for $i = 1,\dots,n$ and $t = 1,\dots,T$ when $S_{it} = 0$, whereas when $S_{it} = 1$ no value is observed because the firm left the labor market.

To account for different behaviors in terms of open-ended hirings and separations during the time period since the first trimester 2009 to the last trimester 2012, we adopt a latent trajectory model ([2], [7], [8]) where firms are assumed to be clustered in a finite number of unobservable groups or latent classes, which are homogeneous with respect to their behavior and their status [6]. Let U_i be a latent variable that indicates the cluster of firm i. This variable has k support points, from 1 to k, and corresponding weights $\pi_u = p(U_i = u)$, $u = 1,\dots,k$. Then, the proposed model is based on two main assumptions that are illustrated in the following.

First, assume the following log-linear models for the number of hirings and separations, respectively:

$$Y_{hit}|U_i = u \sim \text{Poisson}(\lambda_{htu}), \;\; \lambda_{htu} = \exp(\mathbf{x}_t'\boldsymbol{\beta}_{hu}), \quad h = 1,2, \tag{1}$$

with $\boldsymbol{\beta}_{1u}$ and $\boldsymbol{\beta}_{2u}$ vectors of regression coefficients driving the time trend of hirings and separations for each latent class u and $\mathbf{x}_t$ denoting a column vector containing the terms of an orthogonal polynomial of order r, which in our application is equal to 3.

Second, we account for the informative drop-out through a logit regression model, which is specified for the status of firm i at time t as follows:

$$\operatorname{logit} p(S_{it} = 1|S_{i,t-1} = 0, U_i = u) = \mathbf{x}_t'\boldsymbol{\gamma}_u, \tag{2}$$

where the vector of regression parameters $\boldsymbol{\gamma}_u$ is specific for each latent class u.

3.2 Estimation

Parameters of the latent trajectory model described in the previous section are estimated by maximizing the log-likelihood function, which is expressed as

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n} \log f(\mathbf{s}_i, \mathbf{y}_{1i,obs}, \mathbf{y}_{2i,obs}),$$

where $\boldsymbol{\theta}$ denotes the vector of model parameters, that is, $\boldsymbol{\gamma}_u, \boldsymbol{\beta}_{1u}, \boldsymbol{\beta}_{2u}, \pi_u$ for $u = 1,\dots,k$, $\mathbf{s}_i = (s_{i1},\dots,s_{iT})'$ is a column vector describing the sequence of status observed for firm i along the time, and $\mathbf{y}_{hi,obs}$ ($h = 1,2$) is obtained for vector $\mathbf{y}_{hi} = (y_{hi1},\dots,y_{hiT})'$ omitting the missing values. Therefore, if $\mathbf{s}_i = \mathbf{0}$, then $\mathbf{y}_{hi,obs} \equiv \mathbf{y}_{hi}$, otherwise elements of $\mathbf{y}_{hi,obs}$ represent a sub-set of those of $\mathbf{y}_{hi}$.

The manifest distribution of the proposed model is obtained as

$$f(\mathbf{s}_i, \mathbf{y}_{1i,obs}, \mathbf{y}_{2i,obs}) = \sum_{u=1}^{k} \pi_u f(\mathbf{s}_i, \mathbf{y}_{1i,obs}, \mathbf{y}_{2i,obs}|U_i = u),$$

with the conditional distribution given the latent variable U_i defined as follows:

$$f(\mathbf{s}_i, \mathbf{y}_{1i,obs}, \mathbf{y}_{2i,obs}|U_i = u) = \prod_{t=1}^{T} p(s_{it}|U_i = u) \prod_{t=1:\, s_{it=0}}^{T} p(y_{1it}|U_i = u) p(y_{2it}|U_i = u),$$

for $u = 1,\dots,k$, where $p(s_{it}|U_i = u)$ is defined in (2) and $p(y_{1it}|U_i = u)$ and $p(y_{2it}|U_i = u)$ are defined according to (1).

The maximization of function $\ell(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$ may be efficiently performed through the Expectation-Maximization (EM) algorithm [3], along the usual lines based on alternating two steps until convergence:

E-step: it consists in computing the expected value, given the observed data and the current values of parameters, of the complete data log-likelihood:

$$\ell^*(\boldsymbol{\theta}) = \sum_{i=1}^{n} \sum_{u=1}^{k} z_{iu} \log \left[\pi_u f(\mathbf{s}_i, \mathbf{y}_{1i,obs}, \mathbf{y}_{2i,obs} | U_i = u) \right],$$

where z_{iu} is an indicator variable equal to 1 if firm i belongs to latent class u.

M-step: it consists in maximizing such an expected value with respect to $\boldsymbol{\theta}$ and the result is used to update the estimates at the E-step.

Finally, we remind that the EM algorithm needs to be initialized in a suitable way. Several strategies may be adopted, based on deterministic or random values. We suggest to use both, so to effectively face the well-known problem of multimodality of the log-likelihood function that characterizes the finite mixture models [6]. For instance, in our application we choose the starting values for π_u as $1/k$ for $u = 1, \dots, k$, under the deterministic rule, and as random drawings from a uniform distribution between 0 and 1, under the random rule.

3.3 Model selection

A crucial issue is the choice of the number k of latent classes. The prevailing approaches in the literature are based on information criteria, obtained through a penalization of the maximum log-likelihood, so to balance model fit and parsimony. Among these criteria, the most common are the Akaike Information Criterion (AIC; [1]) and the Bayesian Information Criterion (BIC; [10]), although several alternatives have been developed in the literature (for a review, see [6], Ch. 8). In particular, we suggest to use BIC, which is based on the minimization of the index $BIC = -2\hat{\ell} + \log(n)\#\text{par}$, where $\hat{\ell}$ is the maximum log-likelihood estimate obtained as illustrated in Section 3.2 and #par is the number of free parameters. With respect to AIC, BIC tends to be more parsimonious. Moreover, it is asymptotically consistent for finite mixture models under certain regularity conditions [5].

On the basis of this criterion, the selected number of latent classes is the one corresponding to the minimum value of BIC. In practice, in our application we fit the model for increasing values of k until the index begins to increase or, in presence of decreasing indices, until the change in two consecutive indices is less than 1%, and we take the previous value of k as the optimal one.

4 Results

In order to choose the number of latent classes we proceed as described above and fit the latent trajectory model for values of k from 1 to 9. The results of this preliminary fit are reported in Table 3. On the basis of these results, we choose $k = 6$ latent classes, as for values of k greater than 6 the reduction of BIC index is less than 1%.

Table 3 Model selection: number of mixture components (k), log-likelihood, number of free parameters (#par), BIC index, and difference between consecutive BIC indices (delta)

k	log-likelihood	#par	*BIC*	delta
1	-476783.18	8	953649.90	–
2	-383685.95	17	767549.44	-0.1951
3	-361696.26	26	723664.05	-0.0572
4	-356020.51	35	712406.53	-0.0156
5	-348313.32	44	697086.15	-0.0215
6	-344502.01	53	689557.51	-0.0108
7	-341997.83	62	684643.13	-0.0071
8	-341091.09	71	682923.64	-0.0025
9	-339680.21	80	680195.87	-0.0040

As shown in Table 4, that describes the average number of hirings and separations for each latent class and the corresponding weight, most firms come from class 1 ($\hat{\pi}_1 = 0.524$), followed by class 3 ($\hat{\pi}_1 = 0.220$) and class 2 ($\hat{\pi}_1 = 0.198$), and do not exhibit relevant movements either in incoming or in outgoing. Indeed, the estimates of the average number of hirings and separations, obtained as $\bar{\lambda}_{hu} = \frac{1}{T}\sum_{t=1}^{T}\lambda_{1tu}$, $h = 1,2$, are strongly less than 1. On the contrary, classes 5 and 6, that gather just the 1.4% of total firms, show a different situation. Firms in class 5 hire 1.5 open-ended employees per quarter, whereas 2.4 employees per quarter stop their open-ended relation with the firm. As concerns firms in class 6, the average number of hirings and separations equal 6.95 and 9.89 per quarter, respectively. Besides, we observe that the separations tend to be higher than the hirings for all the classes.

Table 4 Estimated average number of hirings ($\hat{\bar{\lambda}}_{1u}$) and separations ($\hat{\bar{\lambda}}_{2u}$) and weights ($\hat{\pi}_u$) by latent class

	$u=1$	$u=2$	$u=3$	$u=4$	$u=5$	$u=6$
$\hat{\bar{\lambda}}_{1u}$	0.019	0.032	0.147	0.504	1.501	6.950
$\hat{\bar{\lambda}}_{2u}$	0.057	0.055	0.228	0.792	2.429	9.894
$\hat{\pi}_u$	0.524	0.198	0.220	0.044	0.013	0.001

As concerns the time trend of dropping out from the market, plot in Figure 1 (top) shows that the probability of drop-out is increasing during year 2009, then it reduces and it again increases since the beginning of 2012. However, the estimated probabilities are always very small, being not higher than 2.5%. Classes 5 and 6 are characterized by the highest probabilities of drop-out during the first two years, although firms in class 6 show the smallest probabilities of drop-out in the last year. On the contrary, class 3 shows an increase of these probabilities during year 2012, so that it has the highest probability of drop-out during the last observed quarter. Finally, firms in class 1 constantly preserve very low values.

As concerns the time trend of hirings and separations (Figure 1 middle and bottom, respectively), both of them tend to increase along the time, although this phenomenon is evident only for classes 5 and 6. More in detail, the maxima values of hirings and separations for firms from class 6 are achieved in the last quarter of 2012 and equal 23.9 and 36.5, respectively.

In order to further characterize the latent classes, we analyze the distribution of firms by economic sector (Table 5). Class 1 distinguishes for a greater presence of extraterritorial organizations and of firms operating in the following sectors: agriculture, forestry, and fishing; arts, entertainment and recreation; electricity, gas, steam, and air-conditioning supply; financial and insurance activities; health and social work activities; information and communications; professional, scientific, and technical activities; and real estate activities. In class 2 there is a prevalence of activities characterized by households as employers, whereas in class 3 there is a greater presence of activities related to accommodation and food, construction, manufacturing products, mining and quarrying products, and waste-management. Finally, both classes 5 and 6 show a prevalence of public administration and defense activities, other than education in case of class 5 and arts, entertainment, and recreation in case of class 6. Finally, we outline that no special difference comes out between municipalities (output here omitted).

Table 5 Distribution of firms by economic sector and latent class (row frequencies)

	$u=1$	$u=2$	$u=3$	$u=4$	$u=5$	$u=6$
Accommodation and food	0.422	0.148	0.306	0.099	0.024	0.001
Extraterritorial organizations	0.700	0.200	0.100	0.000	0.000	0.000
Activities of households as employers	0.495	0.306	0.194	0.004	0.001	0.000
Administrative and support activities	0.553	0.141	0.192	0.085	0.025	0.006
Agriculture, forestry and fishing	0.785	0.106	0.094	0.012	0.004	0.000
Arts, sports, entertainment and recreation	0.697	0.123	0.119	0.038	0.013	0.010
Constructions	0.486	0.158	0.279	0.064	0.013	0.001
Education	0.585	0.040	0.134	0.090	0.148	0.004
Electricity, gas, air-conditioning supply	0.617	0.149	0.128	0.064	0.043	0.000
Financial and insurance activities	0.642	0.172	0.142	0.028	0.017	0.000
Health and social work activities	0.604	0.191	0.147	0.033	0.017	0.008
Information and Communication	0.652	0.172	0.152	0.020	0.003	0.001
Manufacturing products	0.483	0.155	0.278	0.070	0.014	0.001
Mining and quarrying products	0.435	0.152	0.304	0.087	0.022	0.000
Other personal service activities	0.627	0.215	0.140	0.011	0.005	0.002
Professional, scientific, technical activities	0.667	0.201	0.111	0.019	0.002	0.000
Public administration and defense	0.494	0.101	0.202	0.097	0.089	0.016
Real estate activities	0.704	0.166	0.121	0.000	0.010	0.000
Transport and storage	0.491	0.194	0.231	0.062	0.017	0.005
Waste-management	0.517	0.125	0.275	0.050	0.033	0.000
Wholesale and retail trade	0.560	0.185	0.215	0.033	0.006	0.000

5 Conclusions

The different trends of open-ended hirings and separations of a set of Italian firms in every quarter of the time period 2009-2012 has been analyzed through a finite mixture latent trajectory model. Six latent classes of firms were detected, which have specific trends for the probability of drop-out from the market and of hirings and separations. In all clusters, the probability of drop-out increases during the first year, then it decreases and, during the last year, it increases again. Moreover, the number of hirings and separations is usually very close to zero, although it tends to increase during the time. The two smallest classes, which account for just the 1.4% of the total set of firms, distinguish from the others for a higher probability of drop-out during the first two years and for a greater number of hirings and separations during all the period 2009-2012.

Acknowledgements Authors acknowledge the financial support from the grant *Finite mixture and latent variable models for causal inference and analysis of socio-economic data* (FIRB - *Futuro in ricerca* 2012) funded by the Italian Government (grant RBFR12SHVV). They also thank the Province of Perugia (Direction for "Work, Training, School and European Policies") for permitting to extract a specific database from the "Compulsory Communication dataset of the Employment Service Centers".

References

1. Akaike, H.: Information theory and an extension of the maximum likelihood principle. In: Petrov, B. N., Caski, F. (eds.) Proceeding of the Second International Symposium on Information Theory, pp. 267-281. Akademiai Kiado, Budapest (1973)
2. Bollen, K.A., Curran, P.J.: Latent curve models: A structural equation perspective. Wiley, Hoboken, NJ (2006)
3. Dempster, A. P., Laird, N. M., Rubin, D. B.: Maximum likelihood from incomplete data via the EM algorithm (with discussion). Journal of the Royal Statistical Society, Series B, **39**, 1–38 (1977)
4. Hijzen, A., Mondauto, L., Scarpetta, S.: The Perverse Effects of Job-Security Provisions on Job Security in Italy: Results from a Regression Discontinuity Design. IZA Discussion Paper number 7594 (2013) Available via http://ftp.iza.org/dp7594.pdf. Cited 4 Nov 2014
5. Keribin, C.: Consistent estimation of the order of mixture models. Sankhya: The Indian Journal of Statistics, Series A **62**, 49–66 (2000)
6. McLachlan, G., Peel, D.: Finite mixture models. Wiley, Hoboken, NJ (2000)
7. Muthén, B.: Latent variable analysis: Growth mixture modelling and related techniques for longitudinal data. In: Kaplan, D. (eds.) Handbook of Quantitative methodology for the social sciences, pp. 345-368. Sage, Newbury Park, CA (2004)
8. Muthén, B., Shedden, K.: Finite mixture modelling with mixture outcomes using the EM algorithm. Biometrics **55**, 463–469 (1999)
9. Schivardi, F., Torrini, R.: Identifying the effects of firing restrictions through size-contingent differences in regulation. Labour Economics **15**, 482–511 (2008)
10. Schwarz, G.: Estimating the dimension of a model. The Annals of Statistics **6**, 482–511 (1978)

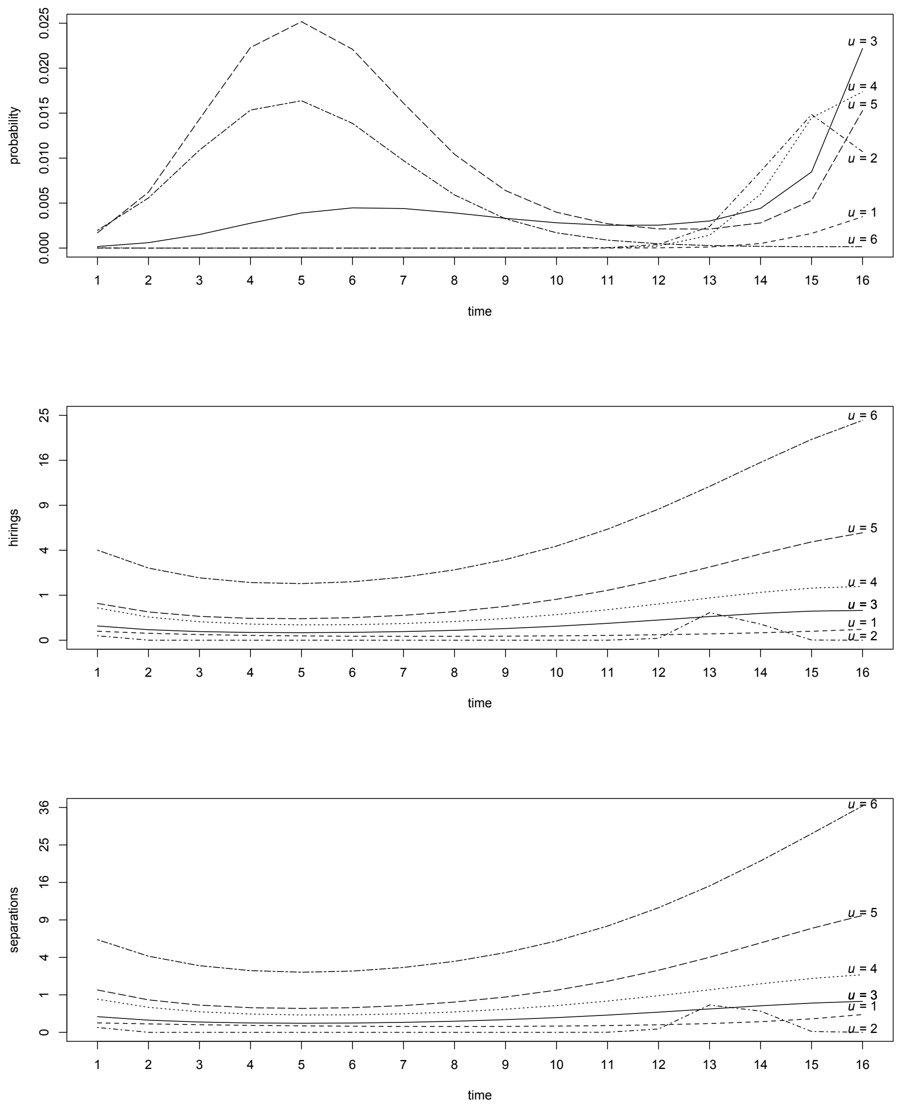

Fig. 1 Trend of the probability of leaving the market (top) and trends of the number of open-ended hirings (middle) and separations (bottom), by latent class