Appendices

Appendix A. Estimation of unemployment duration

Appendix A-1. Methodology

Specification

In order to estimate the different unemployment-to-work hazard rates, we use Weibull model. This choice results from the necessity to have a model enough flexible, while being applicable to data of million observations. The essential parameter of Weibull model is the hazard function that gives the instantaneous unemployment-to-work hazard rate. It is defined as follow:

$$h_i(t) = h_0(t)\exp(X_i b),$$

where $h_0(t)$ is the baseline hazard, which depends only on time and all the explanatory variables for individual *i*. In the case of the Weibull model, the hazard function takes the following specific form:

$$h_0(t) = \alpha\, t^{\alpha-1},\ \ \alpha > 0$$

According to the value of α, the baseline hazard may be increasing or decreasing. The particular case $\alpha = 1$ is the exponential model, characterized by the lack of relationship between the unemployment duration and the unemployment-to-work hazard rate. In the case where α is below 1, it means that the instantaneous unemployment-to-work hazard rate decreases with unemployment duration.

For our estimations, we use two others quantities. The first is the survival function that gives the probability that unemployment duration is below a given threshold:

$$S_i(t) = \exp\left(-\exp(X_i b)t^{\alpha}\right),$$

This function is used to calculate raw and net rates, which give the probability of leaving unemployment before *t* months for individual’s characteristics X_i. We can also prefer to express the rapidity of unemployment-to-work transition by using number of months spent in unemployment from the registration date, which is equal to:

$$E(T) = \exp(X_i b / \alpha)\Gamma(1 + 1/\alpha)$$

Estimation

In order to estimate this model, we use the maximum likelihood method. As all unemployment durations are not observed until their end, we have to deal with censored observations. There are two main sources of censorship in our study: in one hand, people are still unemployed at the final date of the ANPE’s file; on the other hand, people leave the file for another reason that the return to employment.

The full duration is noted t_i, the duration beyond which the data are censored is noted $\bar{t}_i$, and the duration that we observe is equal to:

$$y_i = \min\left(t_i, \bar{t}_i\right)$$

The density of observation of a complete period y_i is given by:

$$f(y_i) = h(y_i)S(y_i),$$

and the probability of observing a censored duration is given by $S(y_i)$. By noting a dummy variable equal to 1 if the data is censored and 0 otherwise, the observation density of any duration is defined by:

$$L_i = f(y_i)^{1-c_i} S(y_i)^{c_i},$$

The log likelihood of the sample is then equal to the following quantity in the general case:

$$\ell = \sum_{i=1}^{N} \ln L_i = \sum_{i=1}^{N} (1-c_i)\ln h(y_i) + \ln S(y_i).$$

For the Weibull model, we obtain:

$$\ell = \sum_{i=1}^{N} (1-c_i)(X_i b + \ln\alpha + (\alpha-1)\ln y_i) - \exp(X_i b) y_i^{\alpha}$$

The parameters (α, b) are obtained by maximizing this function.

Raw and net rates

Raw and net rates are calculated taking into account problems of censored data. In order to do this, we estimate a model that contains only dummies for municipalities. Let be d_j the dummies for municipalities ($j = 1,...,J$), we estimate the model defined by the survival function:

$$S_i(t) = \exp\left\{-\exp\left(\sum_{j=1}^{J} \gamma_{0,j} d_{j,i}\right) t^{\alpha_0}\right\}.$$

This model is equivalent to apply a specific exit rate at each locality j. The raw rates are then obtained by the formula:

$$\hat{S}_j^B(t) = \exp\left\{-\exp\left(\hat{\gamma}_{0,j} d_{j,i}\right) t^{\hat{\alpha}_0}\right\}, \quad j = 1,..., J$$

In order to calculate net rates, we estimate the model with individuals' explanatory variables $(X_{1i},..., X_{Ki})$, dummies for municipalities $(d_{1i},..., d_{Ji})$, then we set individual variables at the level of regional average $(\overline{X}_1,...,\overline{X}_K)$ so as to keep differences that come from municipalities. The estimated model is then:

$$S_i(t) = \exp\left\{-\exp\left(\sum_{j=1}^{J} \gamma_{1,j} d_{j,i} + \sum_{k=1}^{K} X_{ki}\beta_k\right) t^{\alpha_1}\right\},$$

We note that coefficients for the dummy variables are different in a model with explanatory variables. We then calculate net rate at the municipality level as:

$$\hat{S}_j^N(t) = \exp\left\{-\exp\left(\hat{\gamma}_{1,j} d_{j,i} + \sum_{k=1}^{K} \overline{X}_k \hat{\beta}_k\right) t^{\hat{\alpha}_1}\right\}.$$

Appendix A-2. Individual determinants of unemployment-to-work transitions

	Removal from list		Return to work	
	Coefficient	**T-stat**	**Coefficient**	**T-stat**
α	0,917	2252,53	0,843	1148,88
Age (years)	-0,018	236,17	-0,036	234,27
Permanent contract	**réf**		**réf**	
Limited-term contract	-0,382	125,96	-0,491	87,52
Seasonal	-0,104	37,21	-0,168	31,29
Degree level VI	**réf**		**réf**	
Level I et II	-0,001	0,40	0,364	59,17
Level III	0,032	11,30	0,361	66,17
Level IV	-0,030	13,02	0,186	40,06
Level V	-0,051	30,29	0,074	19,93
Without children	**réf**		**réf**	
One child	-0,077	41,31	0,017	4,50
Two children	-0,079	37,41	0,224	56,22
Three or more children	-0,055	22,75	0,235	47,71
Man	**réf**		**réf**	
Woman	-0,062	40,20	-0,223	77,02
Non-disabled	**réf**		**réf**	
Disabled	-0,274	98,01	-0,621	94,96
Single, widowed	**réf**		**réf**	
Divorced, separated	0,031	12,44	-0,009	1,83
Married, de facto married	-0,003	1,51	-0,011	3,21
ROME : Serv persons and community	**réf**		**réf**	
Administrative and sales	0,024	10,00	0,039	8,01
Hotels restaurants	0,313	105,82	0,499	84,00
Sales and distribution	0,124	52,34	0,151	30,27
Arts and entertainment	-0,523	102,18	-1,013	86,48
Initial and continuing education	-0,073	13,71	-0,072	7,56
Social work devt local employment	0,042	11,06	0,022	2,93
Paramedical	0,205	37,32	0,315	31,95
Medical	0,025	2,16	0,144	7,26
Managers admin/ communic. information	-0,060	15,70	-0,090	12,47
Managers sales	-0,028	6,21	-0,004	0,50
Agriculture and fisheries	0,102	24,17	0,229	27,35
Public works and extraction	0,190	55,82	0,323	45,34
Transport and logistics	0,010	3,66	0,096	16,82
Mechanical electrical electronic	0,049	14,74	0,094	14,20
Processing	-0,088	20,16	-0,010	1,20
Other manufacturing	0,005	0,97	0,113	9,89
Personal artisanal	0,206	45,12	0,309	34,14
Industrial management	0,117	8,61	-1,873	153,72
Industrial technician	0,037	8,31	0,002	0,20
Management technical industries	0,069	12,28	0,080	8,25
Technical managers outside manufacturing	0,146	27,45	0,195	20,66
Lay-offs for financial reasons	**réf**		**réf**	
Other lay-offs	0,053	18,65	-0,042	8,27
Resignations	0,507	153,49	0,389	63,94
End of contracts	0,292	110,40	0,421	89,42
End of temp work	0,275	86,04	0,236	39,60
First entry	0,568	166,56	0,363	53,66
Return to work of more than 6 months	0,489	115,46	0,309	35,25
Other cases	0,367	137,21	0,153	30,34
Manual workers	**réf**		**réf**	
Skilled workers	0,027	11,12	0,185	36,97
Unskilled employees	-0,008	3,34	-0,051	9,25

Skilled employees	-0,025	10,17	0,144	27,55
Technician, supervisor	-0,003	0,96	0,204	30,85
Manager	-0,030	6,99	0,155	18,80
Non RMI	**réf**		**réf**	
RMI	-0,212	105,27	-0,587	114,12
Full-time	**réf**		**réf**	
Part-time	-0,226	120,70	-0,555	132,22
Nationality French	**réf**		**réf**	
EU 15	0,066	14,39	0,094	10,35
Rest of world	-0,002	0,79	-0,197	35,26

Sources: Historical statistical file of Pôle Emploi (2001 cohort)

Reading: Results of estimations of Weibull model by Maximum-Likelihood. The coefficients apply to rates of exit from unemployment (i.e. hazard function) in relation to the modality of reference indicated in the table.

Appendix B. Employment area in the Paris region

Appendix C. Construction of explanatory variables

Appendix C-1. Measurement of local dynamism

The volume of gross creation of employment C_{ct} in the municipality *i* between the dates *t-1* and *t* is:

$$C_{ct} = \sum_{e \in C^+} \Delta N_{ect}$$

where C^+ is the sub-total of firms *e* of the municipalities *i* for which the number of jobs at the end of the period is more than the number of jobs at the beginning of the period of observation, and Δ operates the difference between *t-1* et *t*. Similarly, the volume of gross destruction of jobs D_{ct} is:

$$D_{ct} = \sum_{e \in C^-} |\Delta N_{ect}|$$

where C^- is the sub-total of firms *e* of the municipalities *i* that experience a negative variation in employment during the year.

Appendix C-2. Measurement of Skills Mismatch

Occupational status and diploma

Occupational status	under-qualification	Median qualification	over-qualification
Liberal professions	4	6	6
Public service executives	4	5	6
Executives	3	4	5
Public service intermediate professions	4	5	6
Intermediate and commercial professions	3	4	5
Technicians	3	4	5
Supervisors and foremen	2	3	4
Public service employees	2	3	4
Administrative employees	2	3	4
Commercial employees	2	3	4
Services and sales workers	1	2	3
Skilled workers	2	3	4
Unskilled workers	1	2	3
Farm workers	1	2	3
Observations		139 111	

Source: 1999 Population census (INSEE).

Lecture: The value for under-qualification means that 25% of the workers have less than this diploma level for a given occupational status. For example, 25% of executives have the French CAP or BEP or an inferior diploma level. The value for over-education means that 25% of the workers have more than this diploma level for a given occupational status.

Notes: 1=Primary school certificate (CEP); 2=First cycle certificate (BEPC); 3=Certificate of professional aptitude (CAP), Diploma of Occupational Studies (BEP); 4=General of Technological baccalaureate; 5=First stage of university studies; 6=Second and third stage of university studies.

Percentages of occupational statuses and under-/over-educated workers

Occupational status	Percentage in total employment (France)	Percentage in total employment (Paris region)	Percentage mismatched (Paris region)
Liberal professions	0,04	0,06	10,34
Public service executives	0,82	1,18	18
Executives	0,14	0,31	45,92
Public service intermediate professions	10,90	11,47	35,41
Intermediate and commercial professions	9,32	14,52	30,08
Technicians	5,24	6,38	20,11
Supervisors and foremen	3,18	3,13	41,26
Public service employees	14,21	14,08	35,45
Administrative employees	10,30	14,1	44,12
Commercial employees	5,11	4,93	38,19
Services and sales workers	6,80	7,09	48,18
Skilled workers	20,17	14,73	35,1
Unskilled workers	12,44	7,84	10,69
Farm workers	1,32	0,19	14,39
Observations	759 026	139 111	

Source: 1999 Population census (INSEE).

Lecture: The first and second columns present the percentage of each occupational status in our samples. The third column presents the percentage of under- and over-educated workers of each occupational status.

Appendix D. Spatial auto-correlation issue

In order to measure the problem of spatial auto-correlation, we calculate the Moran's coefficient *I* for the unemployment durations, which could be interpreted as the relation of the covariance between contiguous observations and the total variance observed in the sample (Jayet, 1993). It is given by:

$$I = \frac{n}{\sum_i \sum_j w_{ij}} \frac{\sum_i \sum_j w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_j (x_i - \bar{x})^2}$$

Where n is the number of municipalities and w_{ij} is a weight which permits us to take into account the geographical proximity of spatial units *i* and *j*.

When $I > E[I] = (n-1)^{-1}I$ (respectively $I < E[I]$), the values taken by durations are not randomly located but are close (respectively distant) for two neighbouring spatial units. The geographically close spatial units are also statistically close (respectively distant) and we conclude to a presence of a positive spatial auto-correlation (respectively negative). When *I* is close to $E[I]$, we conclude to an absence of spatial auto-correlation. In this case, we cannot establish a link between the statistical proximity and the geographical proximity of spatial units.

In fact, the calculation of the Moran index *I* is sensitive to the definition of the matrix of spatial weights $W(w_{ij})$. There are effectively several criteria in order to determine the spatial units that will be considered as neighbours: contiguity, nearest neighbours, distance. The following table presents the auto-correlation coefficients (Moran *I*) of the net unemployment durations obtained for different spatial weight matrices.

Global spatial auto-correlation of unemployment duration

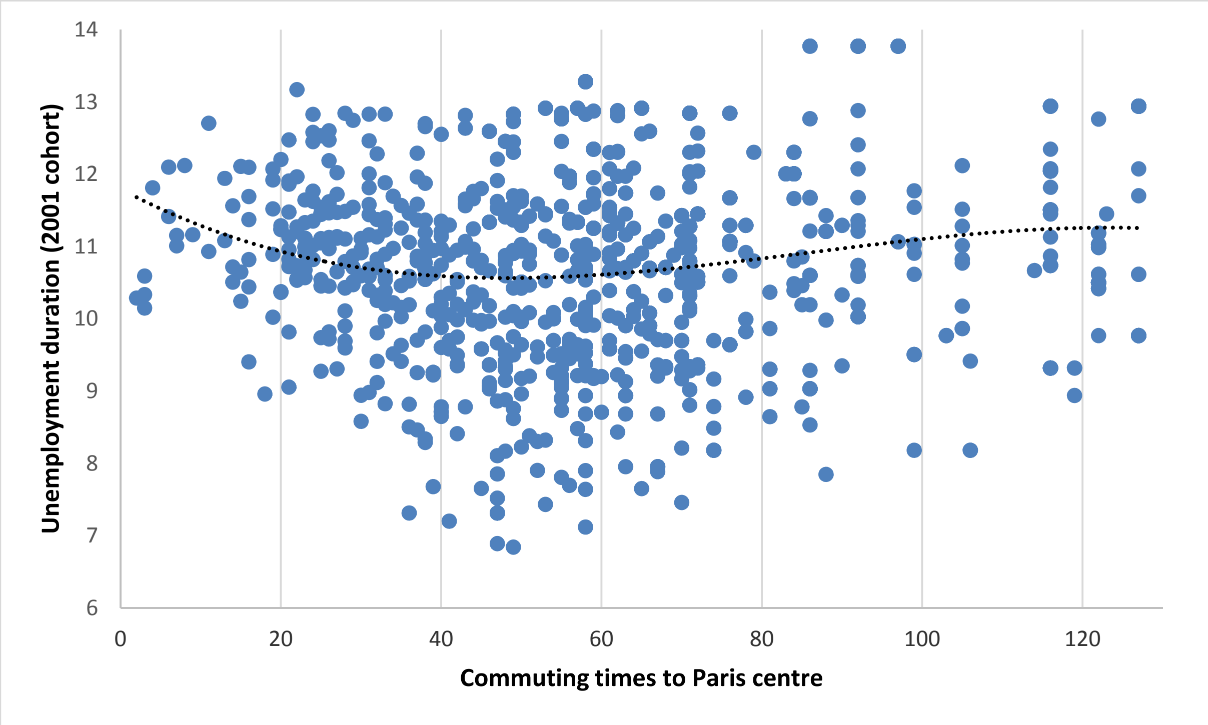
Matrix W	**Moran I**	**Standard-error**	**p-value**
Queen (order 1)	0,653	0,0208	0,001
Queen (order 2)	0,492	0,0148	0,001
Nearest neighbours (6)	0,623	0,0174	0,001
Distance (10 km)	0,481	0,0089	0,001

Source: Historical statistical records of Pôle Emploi.
Notes: E[I] = -0,0010.

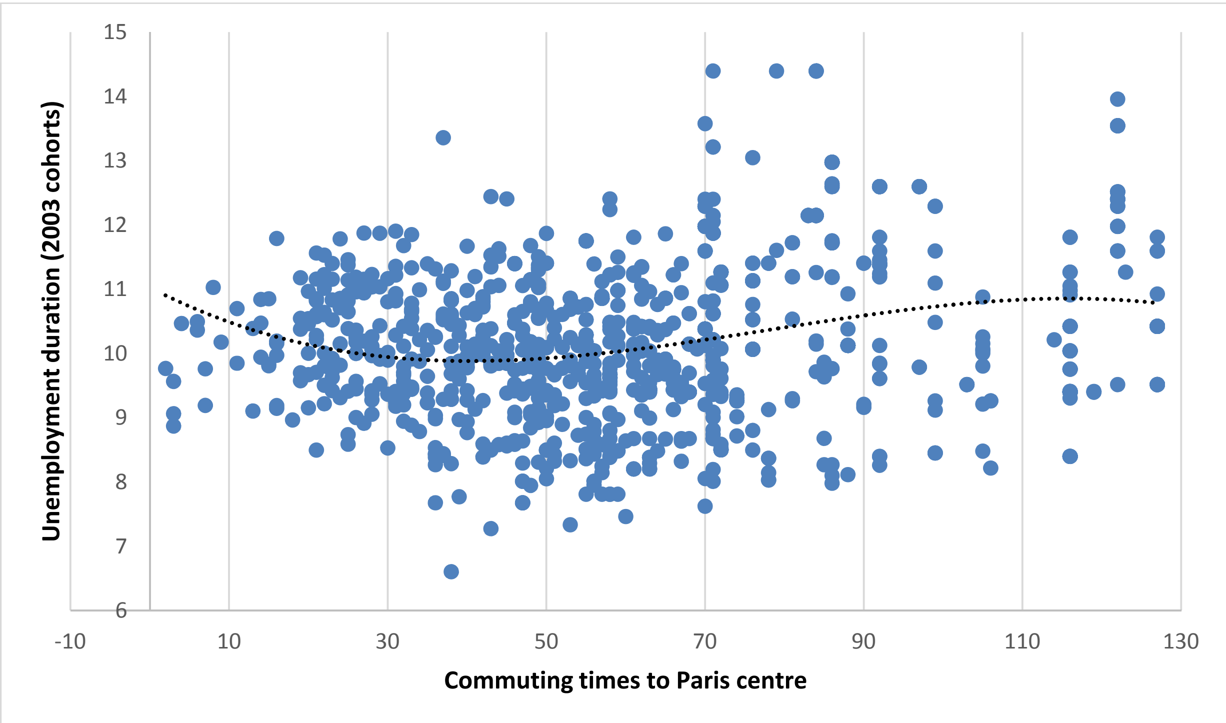
Whatever the matrix used, we see that the unemployment durations present a significant and relatively high positive spatial auto-correlation value. Then, the geographically neighbouring municipalities are also neighbours in terms of unemployment duration. To consider the spatial auto-correlation problem in our data, we must use an appropriate model. Without surprise, we see that when the number of neighbours is increasing, the spatial auto-correlation measured with the index is decreasing. We retain the queen matrix at order 2. We believe that spillover effects may concern close municipalities and not exclusively municipalities that are immediate neighbours.

Appendix E. Unemployment duration and commuting times to Paris centre

E-1. 2001 cohort

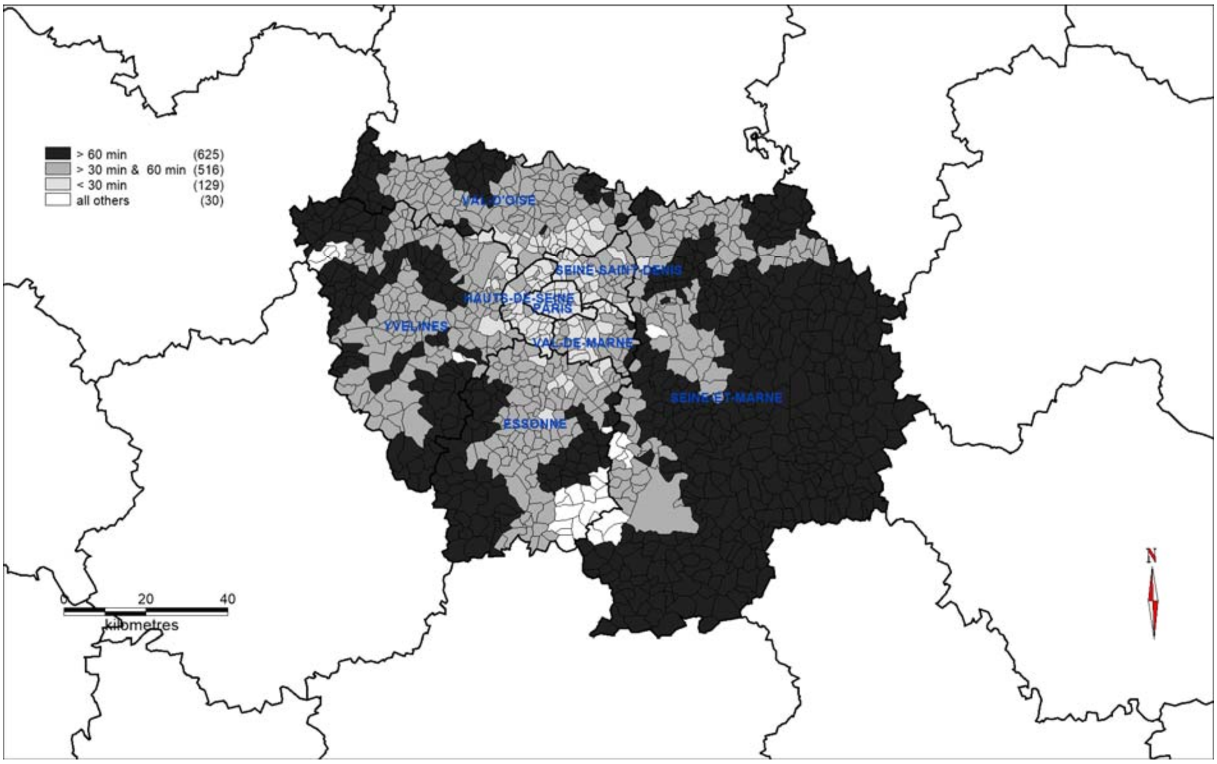


E-2. 2003 cohort



Sources: Historical statistical file of Pôle Emploi (2001 and 2003 cohorts) and DREIF's time travel matrices (2003). *Reading:* Unemployment durations are expressed in months and represent the estimated duration for each municipality. Commuting times are expressed in minutes and represent the time necessary to go Paris centre by public transport, for each municipality.

Appendix F. The three regimes



Source: DREIF's time travel matrices (2003).
Notes: Commuting times are expressed in minutes and represent the time necessary to go Paris centre by public transport, for each municipality.

Appendix G. Estimations using 2003 cohort

G-1. Explaining net unemployment duration disparities

Variables	**OLS-White**	**SEM model**	**GS2SLS**
Constant	-0,043	-2,822**	-1,749
	1,078	*1,214*	*1,261*
Socio-economic characteristics			
Percentage of young people	-3,967**	-1,062	-2,446
	1,635	*1,521*	*1,587*
Percentage of low-skilled people	5,89***	3,386***	4,519***
	1,039	*0,979*	*1,040*
Percentage of households executives	5,227***	3,736**	3,481**
	1,284	*1,181*	*1,252*
Percentage of households with a car	-0,682	2,222**	0,917
	0,776	*0,951*	*0,974*
Percentage of public housing	-1,759***	-1,169**	-1,709***
	0,556	*0,535*	*0,560*
Local dynamism			
Creation rate	-0,019	-0,016	-0,019
	0,051	0,044	0,046
Destruction rate	-0,008	-0,014	-0,012
	0,042	0,037	0,037
Spatial Mismatch			
Percentage of accessible jobs	-1,293***	-1,021**	-1,124**
(in 60 min)	*0,290*	*0,336*	*0,376*
Skill Mismatch			
Percentage of over- and under-qualified	31,655***	36,226***	35,412***
	1,392	*1,291*	*1,563*
λ		0,756***	0,651***
		0,049	*0,034*
R^2	0,419	0,404	0,415
Log likelihood		-2 049,73	
AIC		4 119,47	
Breusch-Pagan	207,517***		
Observations	**1 075**	**1 075**	**1 075**

Sources: Pôle Emploi's Historical statistics files (2001 and 2003 cohorts), DADS 2002-2005, DREIF's time travel matrices (2003), population census 1999 (INSEE).
Notes: ***, ** and * denote significance at the 1%, 5% and 10% levels respectively. Standard errors are presented in italics.

G-2. Explaining net unemployment durations – Analysis by spatial regime

	SEM Model			GS2SLS		
Variables						
Constant	12,702***	-2,662	15,564***	13,266***	3,259	22,896***
	2,781	*2,029*	*2,92*	*3,199*	*2,071*	*3,751*
Socio-economic characteristics						
Percentage of young people	-11,975***	-1,521	-0,497	-12,270***	-1,603	1,149
	3,335	*2,494*	*2,249*	*3,348*	*2,553*	*2,44*
Percentage of low-skilled people	3,338	3,046***	1,586	3,294	3,292**	0,815
	2,394	*0,083*	*1,449*	*2,429*	*1,708*	*1,557*
Percentage of households executives	4,751	2,901	2,633	4,59	3,564*	3,276*
	3,565	*1,931*	*1,639*	*3,613*	*2,003*	*1,748*
Percentage of households with a car	-2,708***	1,445	-1,794	-2,899***	2,049	-0,532
	0,841	*1,748*	*2,399*	*0,943*	*1,776*	*2,592*
Percentage of public housing	-1,166**	-0,836	-4,507***	-1,258**	-0,775	-4,318***
	0,594	*0,795*	*1,193*	*0,607*	*0,809*	*1,269*
Local dynamism						
Creation rate	-0,07	-0,026	0,002	-0,071	-0,026	-0,012
	0,097	*0,079*	*0,069*	*0,097*	*0,079*	*0,074*
Destruction rate	0,036	-0,013	0,009	0,036	-0,013	0,025
	0,064	*0,069*	*0,055*	*0,064*	*0,069*	*0,059*
Spatial Mismatch						
Percentage of accessible jobs	-2,347***	-0,787*	0,262	-2,261**	-0,421	1,526
(in 60 min)	*0,738*	*0,452*	*0,918*	*0,766*	*0,505*	*1,233*
Skills Mismatch						
Percentage of over- and under-qualified	8,411**	37,577***	-13,272***	7,256	36,452***	45,001***
	4,019	*1,554*	*4,196*	*5,141*	*1,701*	*4,171*
	0,104	0,389***	0,162*	0,091	0,384***	0,181***
	0,203	*0,083*	*0,086*	*0,173*	*0,066*	*0,048*
Log likelihood	-130,917	-832,964	-986,49	-146,904	-720,73	-818,927
AIC	281,84	1 685,93	1 992,98	313,81	1 461,46	1 657,85
Chow test		234,691***			241,842***	
Observations	127	448	500	127	448	500

Sources: Pôle Emploi's Historical statistics files (2001 and 2003 cohorts), DADS 2002-2005, DREIF's time travel matrices (2003), population census 1999 (INSEE).
Notes: ***, ** and * denote significance at the 1%, 5% and 10% levels respectively. Standard errors are presented in italics.