A New Light into Regional Unemployment Disparities in Belgium: Longitudinal Analysis of Grouped Duration Data.*

Muriel Dejemeppe† and Yves Saks‡
Institut de Recherches Economiques et Sociales,
Université catholique de Louvain

July 29, 2002

Abstract

In this paper, we investigate whether the diverging evolution of unemployment in the two main regions of Belgium, Flanders and Wallonia, is driven by different evolutions in their average unemployment duration and/or their unemployment incidence. To that purpose, we proceed in two stages. In the first stage, we estimate a mixed proportional hazard model by region, and decompose variations in the aggregate outflow rate over calendar time between a general effect, i.e. changes in the exit rate of all currently unemployed, and a composition effect, i.e. fluctuations in the average quality of entrants. We also specify a non-proportional model to check whether the general effect is the same for unemployed workers with different schooling levels and sub-region of living, in each region. In the second stage, we decompose variations of the unemployment stock in Flanders and Wallonia into an incidence effect and the duration effect estimated in the first stage. We base our analysis on yearly exit probabilities of male workers aged 25-44 and flowing into unemployment in June, each year from 1972 to 1992. The use of aggregate data covering 21 years and stratified by a set of relevant characteristics allows us to control for changes in the inflow composition without relying (completely) on functional form assumptions.

JEL classification: C41, J64

Keywords: Unemployment duration, unemployment incidence, business cycle, composition effect.

---

*This paper has benefited from the comments of Bart Cockx, Marc Gurgand, Thierry Kamionka, Henri Sneessens, Gerard van den Berg, Bas Van der Klaauw and Bruno Van der Linden. Financial support from the 'Fonds National de la Recherche Scientifique' is gratefully acknowledged. This research is also part of a programme supported by the Belgian government (Poles d'Attraction inter-universitaires PAI P4/01). We thank the Office National de l'Emploi for providing the data.

†Fonds National de la Recherche Scientifique and Institut de Recherches Economiques et Sociales, Université catholique de Louvain, Place Montesquieu, 3, 1348 Louvain-la-Neuve, Belgium. Tel : 32.10.47.41.00, Fax : 32.10.47.39.45, E-mail : dejemeppe@ires.ucl.ac.be.

‡Centrum voor Economische Studiën, Katholieke Universiteit Leuven, Naamsestraat 69, 3000 Leuven, Belgium. Tel : 32.16.32.68.18, Fax : 32.16.32.67.96, E-mail : yves.saks@econ.kuleuven.ac.be.

1 Introduction

The unemployment rate[1] increased steadily in Belgium after the first and the second oil price shocks: From 3% in 1973 to 19% in 1983. This evolution is illustrated in Figure 1. With the economic recovery of the mid eighties, the unemployment rate decreased to a level of 12%, which still lied far above the level reached in 1973. From 1990, the economic downturn pushed the unemployment rate up again. This evolution masks an important disparity between the two main regions of Belgium: Flanders and Wallonia. Up to 1983, the two regions had roughly the same performance in terms of unemployment, worse than in our bordering countries (Germany, France and the Netherlands). From 1983, a diverging evolution seems to show up: The unemployment rate decreased markedly in Flanders, while unemployment remained at a high level in Wallonia. In 2000, the unemployment rate was 20% in Wallonia and 6% in Flanders.

INSERT FIGURE 1 APPROXIMATELY HERE

In this study, we investigate whether the diverging evolution of unemployment in Flanders and Wallonia is driven by different evolutions in their average unemployment duration and/or their unemployment incidence. To that purpose, we proceed in two stages (see Figure 2). In the first stage, the average unemployment duration is modelled through the specification of the hazard rate. We estimate a Mixed Proportional Hazard (MPH) model by region, and decompose variations in the aggregate outflow rate over time between a general effect, i.e. changes in the exit rate of all currently unemployed, and a composition effect, i.e. fluctuations in the average quality of entrants into unemployment, leaving the individual exit rates constant. We also specify a non-proportional model to check whether the general effect is the same for unemployed workers with different schooling levels and sub-region of living, in each region. In the second stage, we decompose variations of the unemployment stock in Flanders and Wallonia into an incidence effect and the duration effect estimated in the first stage.

INSERT FIGURE 2 APPROXIMATELY HERE

We base our analysis on aggregate administrative data. We have Belgian census data relative to all unemployed male workers aged 25-44 and flowing into unemployment in June, each year from 1972 to 1992. These data also contain a number of personal characteristics of the unemployed people. It is however impossible to track each individual spell through different years. We therefore group the data into homogeneous cohorts with respect to the year of inflow and a set of relevant characteristics. The latter include the age, the schooling level, the province of living and the sector occupied before becoming unemployed. For each cohort, the flows from unemployment are counted on a yearly basis. Each observation in our data set therefore corresponds to a non-parametric estimation of the annual outflow probability for a given cohort at a given duration (and by construction, calendar time).

[1] The unemployment rate is the ratio between the full-time registered unemployed people, entitled to unemployment benefits and actively looking for a job, and the population insured against unemployment.

In this paper, we propose a formal decomposition of the variations in the aggregate hazard over time, allowing for observed and unobserved changes in the quality of the inflow. In our MPH model, the general effect is the estimated calendar time dependence of the hazard rate controlling for composition effects. With our micro data covering a long timespan (1972-1993), we capture part of the fluctuation in the composition of the inflow by conditioning the hazard on observed characteristics. The remaining variation, the unobserved composition effect, is accounted for by making the mean of the distribution of unobserved heterogeneity dependent on a business cycle indicator at entry.

Given the recent interest of European researchers in this question, there are few studies which decompose the time variations in unemployment duration for European countries (see, for France, Abbring *et al.* 2001a, Van den Berg and Van der Klaauw 2001; for the UK, Kalwij 2001; for Denmark, Rosholm 2001; for Belgium, Cockx and Dejemeppe 2002). This question has been investigated on US data since the mid 1980s already (see, e.g., Darby *et al.* 1985; Dynarski and Sheffrin 1990; Baker 1992; Imbens and Lynch 1993; Abbring *et al.* 2001b). Studies based on European data conclude that compositional effects are of minor importance. Only Kalwij (2001) finds that changes in the quality of entrants account for a significant part of the cyclicality in the aggregate exit rate for the UK. Among studies exploiting US data, results are rather mixed. Only Darby *et al.* (1985) find that the cyclicality in average unemployment duration is predominantly explained by compositional effects. Using the same data, but covering a more recent period, Abbring *et al.* (1999) highlight that cyclical variability in the quality of entrants accounts for a significant share of the fluctuations in the US unemployment duration. On the other side, Imbens and Lynch (1993) conclude that the variation in the average duration of young people is general and cannot be driven by compositional effects.

Our analysis breaks new ground in the empirical literature on the dynamics of unemployment duration. Below, we contrast our approach with the existing studies by focusing on those which decompose variations in the outflow probability in a formal way.

The first contribution of our paper is related to the richness of the data. The use of aggregate data covering 21 years and stratified by a set of relevant characteristics allows us to control for changes in the inflow composition without relying (completely) on functional form assumptions. Most of the studies exploit aggregate time-series data with a limited number of characteristics on the unemployed people such as sex, race and age group (see, e.g., Abbring *et al.* 2001a,b; Kalwij 2001). Variations in the composition of inflow are then accounted for by specifying a MPH model where the heterogeneity distribution depends on calendar time at entry. Even if both the distribution of unobserved characteristics and its calendar time dependence are specified non-parametrically (see e.g. Abbring *et al.* 2001a), parametric restrictions are required for identification purposes. Studies based on macro data are forced to assume the absence of a trend in the composition of inflow. As a consequence, the structural part of the general effect is not necessarily rid of variations in the quality of entrants into unemployment.

One exception is the study of Van den Berg and Van der Klaauw (2001) which combines aggregate time-series data with a micro data set covering a short time period. They model jointly the distribution of the observed characteristics at entry into unemployment and the hazard rate. As in the previous studies, the macro data are used to identify the way the inflow distribution moves over time. But contrary to typical macro analysis, there is no need to impose restrictions on its time dependence: The use of another data set allows to relax the identification conditions[2]. However, their estimated general effect can still reflect changes in the unobserved quality of entrants since the shape of the unobserved heterogeneity distribution is not allowed to vary over time.

Some empirical studies also use longitudinal survey data to decompose calendar time variations in the aggregate outflow probability (see e.g. Imbens and Lynch 1993). Even if they do not cover all the population of unemployed workers, their sample size is sufficiently large to approximate the composition of inflow. Yet, none of them allow for an unobserved composition effect, so that their estimated general effect can still have two interpretations. One exception is Rosholm (2001) who exploits micro Danish data covering the period 1981-1991. Contrary to our model, his unobserved compositional effect is not built into the model though: The mean of the mixing distribution is not explicitly allowed to move with calendar time at the inflow. It is constructed in a somewhat ad-hoc way based on the parameter estimates of his model.

The unobserved composition effect can be identified at the cost of some hypothesis. In this study, we choose to parametrise the time dependence of the distribution of unobserved characteristics by the number of entrants into unemployment. If this indicator does not vary only according to an exponential trend, identification is assured. We could have specified the variability of the mixing distribution in a more flexible way with time dummies forced to be orthogonal to a linear trend (see Abbring *et al.* 2001a). If changes in the composition of inflow closely follow variations in incidence, our specification turns out to be a better alternative. For, it allows the unobserved composition effect to follow a trend (i.e. the trend in incidence), separately from the trend in the general effect.

The second contribution of our study comes from the estimation of a non-proportional hazard model in which the covariate effects may vary over time. This model allows us to check whether the general effect of the calendar time is the same for unemployed workers with different personal characteristics. Only Rosholm (2001) also interacts the covariate effects with the time dependence of the hazard rate. However, only cyclical interactions are allowed for, ruling out the possibility that structural evolutions can affect differently the various groups of unemployed people. Some authors (see e.g. Imbens and Lynch 1993, Van den Berg and Van der Klaauw 2001) argue that if variations in unemployment duration are mainly driven by a general effect, targeted policies are unlikely to be useful. Our findings suggest that even if composition effects are unimportant, there is a scope for labour market policies aimed at specific groups of unemployed workers.

The plan of the paper is as follows. The next section describes the data and gives some descriptive statistics on duration and incidence. The third section is devoted to the derivation of the MPH model, provides the formal decomposition of the aggregate hazard over calendar time and discusses the estimation results. In the fourth section,

[2] Since the duration dependence is identified from the micro dataset, the macro data identify the calendar time dependence of both the hazard rate and the inflow distribution without any functional form restrictions.

we extend our MPH model by allowing for non-proportional calendar time effects. The fifth section contains the decomposition of the unemployment stock into a duration and an incidence effect. The concluding part summarises our findings.

2 The data

2.1 Unemployment insurance scheme in Belgium and data collection

Each year since 1972, the Belgian Employment Agency (ONEm-RVA) takes a census of all unemployed workers on June 30. These data also contain a number of individual characteristics of the unemployed people, such as sex, age, sub-region of living, education level, sector of previous employment, and the starting month of their unemployment spell. The unemployed workers are registered in different categories according to their administrative status. The largest group are full-time unemployed persons actively looking for a job and entitled to unemployment benefits. On average, 81% of entrants belongs to this category.

In Belgium, benefits entitlement is conditional on a sufficiently long employment record[3]. The period of entitlement to benefits is then in principle indefinite. One exception to this rule concerns the cohabitants (living with a working partner or with their parents). They can loose entitlements on grounds of 'excessive' unemployment duration (i.e. when elapsed duration becomes larger than 1.5 times the regional average duration). The level of the benefit is based on the level of previous labour earnings and the household type of the worker, but is bounded by a ceiling and a floor, depending on the household situation. Unemployment benefits for those who are not head of a family drop after 1 year and once more after 18 months if one is cohabitant. The data do not include any information about previous wage or unemployment benefits paid to the individual. Finally, it should be noted that the rules governing the unemployment benefit scheme and the functioning of the labour market are the same in all regions of Belgium.

The data at our disposal are the unemployment stock for the years 1972 up to 1993. They were made anonymous such that is impossible to track individual spells through different years. As explained below, this requires a grouping of the data.

2.2 Data selection

In order to avoid the initial condition problem, we construct an inflow data set of unemployment spells. We select all entrants in the month of June, each year from 1972 to 1992. As the yearly census of the unemployed workers occurs on June 30, the number of entrants excludes those who left unemployment before the end of June. Since we aim to analyse the two main regions of Belgium (Flanders and Wallonia) separately, two data sets are created.

We only follow male workers aged between 25 and 44 years old at the begin of their unemployment spell. The main reason for this choice is that we cannot distinguish between

[3] Graduated students who never worked can also be eligible after some waiting period.

spells ending in employment, participation in labour market programmes or withdrawal from the labour force. This is not likely to bias our results given that employment is the most frequent state of exit for the male unemployed workers: About 77% of them leave for employment, against 47% for the women (see FOREM 1995). In our sample, this fraction must be larger since the reported figure includes workers older than 44 years. These workers often enter a kind of early-retirement scheme, a policy heavily used after the mid eighties in Belgium for older unemployed people.

Employment is also the most frequent destination for the young unemployed workers aged less than 25 years old. However, the share of prime-aged men becoming employed in the flow from unemployment is larger than the share of young people since the latter participate more to education and training programmes. Furthermore, the legislation concerning young people changed substantially more than for the prime-aged group during the 1972-1993 period (change in the age limit with respect to mandatory schooling, eligibility to benefits based on schooling etc.). These changes could have affected significantly both the composition and the number of young entrants into unemployment.

A second reason to focus on the prime-aged male unemployed is that most of them are head of household or live alone. They are therefore less likely to be sanctioned on ground of excessive unemployment duration[4].

To reach more homogeneity in the data, only Belgian unemployed persons are retained in our analysis. Finally, we restrict ourselves to people officially registered as unemployed persons and who satisfy the OECD definition of unemployment, i.e. without employment, immediately available for employment and actively looking for a job. Our selection covers more than 94% of all entrants into unemployment.

2.3 Grouping of the data and induced biases

Since it is impossible to track individual spells of unemployment through time, we group the data into homogeneous cells with respect to some relevant characteristics. So we construct synthetic cohorts of unemployed workers, a cohort being defined by its year of inflow l and a set of individual characteristics x. For each cohort (l, x), the flows from unemployment are counted on a yearly basis, from the time of inflow until the end of June 1993, date at which all spells are right censored. We have that $l \in \{0, 1, ..., 20\}$. The variable k denotes unemployment duration, measured in yearly intervals from the month of inflow l. By construction of the data, $k \in \{1, 2, ..., 21\}$.

The individual characteristics x are the level of education, as a proxy for the skill level, the sub-region where the unemployed person lives, the sector of previous employment[5] and the age at entry. Each characteristic is categorised into classes. We distinguish four levels of education: elementary school, junior high school, completed high school and higher education. The subregion of living is described by the province. The provinces

[4] In 1992, less than 25% of the prime-aged men were cohabitants against more than 65% for the female unemployed workers (see ONEm 1992). These are estimates since the distribution of the household types by age is not available.

[5] The sub-region of living and the sector of previous employment are not necessarily those in which the worker finds a new job.

are, for Flanders: Oost-Vlaanderen, West-Vlaanderen, Antwerpen, Limburg, Vlaams-Brabant and, for Wallonia: Hainaut, Liège, Luxembourg, Namur and Brabant wallon. It is more difficult to group the sectors of previous employment into meaningful categories. We group them into 13 broad classes. These 13 categories are (1) Agriculture, (2) Iron-steel industry and mining, (3) Chemical industry, (4) Metal manufacture, (5) Manufacture of non-metallic mineral products, glass and ceramics, (6) Manufacture of consumption and investment goods[6], (7) Other manufacturing industries, (8) Energy and water production, (9) Building, (10) Private services like wholesale dealing, retail distribution, distribution trade, banking, insurance, transport, (11) Public services including education, (12) Others and (13) No previous employment. Age at entry is categorised in months. The grouping is thus very fine. As a result, 93% of the cohorts are made up of only one individual at the moment of inflow into unemployment.

Let u_{klx} be the number of individuals in cohort (l, x) who are still unemployed at the start of the k^{th} duration interval and at risk of leaving unemployment within the $(l+k)^{th}$ quarter. Let f_{klx} be the number of these individuals flowing from unemployment within the k^{th} duration interval. We can then estimate the annual outflow probability, P_{klx}, by the Kaplan-Meier estimator of the hazard adapted to grouped data:

$$\widehat{P}_{klx} = \frac{f_{klx}}{u_{klx}} \tag{1}$$

However, the non-parametric measure of the outflow probability $\widehat{P}_{klx}$ is subject to biases. For, the non-observability of the individual spells of unemployment, and the consecutive grouping of the data, induces two sources of spurious outflows: (1) temporary exits out of unemployment (of less than three months) and (2) changes in the value of an individual characteristics x during the unemployment spell. Below, we give some light on these problems and the following correction of the data. In Appendix 1, we discuss in details the correction procedure applied to $\widehat{P}_{klx}$.

The first source of spurious outflows is due to temporary exits. In Belgium, workers who exit unemployment for a period of less than three months keep the same starting date of their unemployment spell at their return. Given that the census of the unemployed workers occurs on June 30, it happens that some unemployed are not counted at a given year because they are in a training scheme or have a temporary job that day. They are still unemployed though and counted as such in a following census. Temporary exits introduce biases in our measure of the number of individuals in cohort (l, x) that leave unemployment during the k duration interval, f_{klx}. If an individual counted in u_{klx} exits unemployment temporarily around the start of the $k+1$ duration interval and returns within the next three months, then the number of outflows f_{klx} increases spuriously of one unit. Since the majority of cells in our data set counts at most one individual, the greater part of the temporary exits can be spotted and corrected for by setting the identified spurious risk sets to 1.

[6] Including food, drink, tabacco, textile, leather, footwear and clothing, wooden furniture, paper and related industries, manufacture of motor vehicules and related activities.

Change in the value of an individual characteristics x during the unemployment spell is the second source of spurious outflows. We illustrate the induced bias with the following example. Imagine that an unemployed worker i moves from province A to province B (A and B being in the same region), k years after his entry into unemployment, without flowing out of unemployment. This event is recorded as follows in our data set. On one hand, the number of outflows in the cohort of individual i increases spuriously of one unit at an elapsed duration of k years. On the other hand, the number of individuals at risk of leaving unemployment in the cohort which shares all the characteristics of the initial cohort but the province (province B), increases spuriously by one unit. Actually, these changes can result from administrative problems and do not necessarily reflect true modifications in the situation of the unemployed workers. In particular, we know that they cannot be attributed mostly to one variable, as for example moves between provinces or changes in education level during the unemployment spell.

Since a majority of cohorts are made of one individual, most of the individuals with a time-varying characteristic are identified by picking out cohorts where the risk-set is zero at the date of entry into unemployment and strictly positive in at least one of the following duration intervals. More formally, the number of spurious outflows in cohort l at duration interval k, $f^S_{kl\cdot} = \sum_x f^S_{klx}$, is approximately the number of observations with $u_{klx} < u_{k+1lx}$ (see Appendix 1 for details). Figure 3 gives the number of spurious outflows aggregated over duration and in proportion of the inflow over our observation period. The size of the problem is slightly higher in Wallonia - on average 10% of entrants - than in Flanders - on average 7%. This difference is due to the years 1985 and 1986 where this proportion reaches more than 30% in Wallonia[7].

INSERT FIGURE 3 APPROXIMATELY HERE

It should be emphasised that the importance of the bias in $\widehat{P}_{klx}$ due to time-varying characteristics is inversely related to the level of aggregation. In particular, if the cohorts were defined only in terms of duration and year of inflow, the Kaplan-Meyer estimate of the aggregate outflow probability would be unbiased. On the other hand, the fine grouping of the data - mainly induced by measuring age on a monthly basis - allows us to spot individuals with time-varying characteristics.

In sum, the correction procedure applied to our data works as follows. We first correct the data for temporary exits. Second we calculate the number of spurious outflows in cohort l at duration interval k, $f^S_{kl\cdot}$. Third, we suppress all observations with $u_{klx} < u_{k+1lx}$ since they violate the condition: $0 \leq \widehat{P}_{klx} \leq 1$. Finally, we apply the following correction to the non-parametric outflow probability (see Appendix 1):

$$\hat{P}^*_{klx} = (1 - \frac{f^S_{kl\cdot}}{f_{kl\cdot}})\widehat{P}_{klx} \tag{2}$$

[7] The strong increase in the proportion of entrants with changing variables in 1985 and 1986 is partly explained by an important drop in the number of entrants these years (see Figure 3.5). We have no additional explanation of this finding at this point.

where $f_{kl\cdot} = \sum_x f_{klx}$. This correction lies on the assumption that the (unobserved) number of spurious exits in cohort (l, x) during the k^{th} interval, f^S_{klx}, is proportional to $f^S_{kl\cdot}$ with a factor of proportionality $(\frac{f_{klx}}{f_{kl\cdot}})$. In the sequel, we always use the data set with the corrected $\widehat{P}^*_{klx}$. In order to lighten the notation, we note $\widehat{P}_{klx}$. Notice that we apply this correction in each region separately, i.e. we made the implicit assumption that movements of unemployed people between regions are negligible with respect to movements within regions[8]. The low degree of labour mobility between Flanders and Wallonia is supported by the study of Candelon *et al.* (2000).

The obtained 63793 empirical probabilities $(\widehat{P}_{klx})$ for Flanders and 45987 for Wallonia summarize the labour market histories of 40981 and 25334 unemployed persons[9] respectively.

2.4 Descriptive statistics

In this sub-section, we present some descriptive statistics about unemployment, incidence and duration which are directly relevant to our formal analysis.

Figure 4 shows the total stock of unemployment for men aged 25-44 in our data set, in Flanders and Wallonia. As already stressed in the introduction, unemployment increased steadily in both region between 1973 and 1983. From 1983, a diverging evolution seems to show up: Unemployment decreased markedly in Flanders, while it remained at a high level in Wallonia. However, when we de-trend the two series, the picture is different: The divergence between the two regions seems to be driven by a diverging trend, while the cyclical variability in unemployment is similar across regions.

INSERT FIGURE 4 APPROXIMATELY HERE

The number of people flowing into unemployment during the month of June l and still present at the end of the month, $u_{1l} = \sum_x u_{1lx}$, is our measure of the size of inflow into unemployment. Figure 5 shows the evolution of the inflow into unemployment in Flanders and Wallonia. The number of entrants is strongly increasing over time, slightly more markedly in Wallonia: at rates equal to 3.9% and 4.6% yearly in Flanders and Wallonia respectively[10]. So, incidence cannot be the main cause of divergence in the evolution of unemployment between the two main regions of Belgium. Incidence also show a similar countercyclical behaviour in the two regions. In the second half of the 1970s there is large increase in the inflow, particularly in Flanders, with a peak in 1981 (after the second oil shock). From 1982, the inflow into unemployment decreases in both regions, more importantly in Flanders, and rises again in the late eighties.

INSERT FIGURE 5 APPROXIMATELY HERE

[8] Applying other corrections (and assumptions) confirms that the chosen correction performs better.

[9] What corresponds to 37674 cohorts for Flanders and 23459 for Wallonia.

[10] These rates are obtained by regressing the logarithm of the number of entrants in both regions on a constant term and a trend.

In order to see the evolution of unemployment duration in both regions, we calculate the Kaplan-Meyer estimate of the aggregate outflow probability, in the first duration interval, $\widehat{P}_{1l}$. Given that unemployment duration is measured in years, its evolution gives a fair view of the pattern of the exit probability for the population of interest. In the empirical analysis, we specify a MPH model. For purposes of comparison with the model estimates, we transform the empirical outflow probability into its continuous equivalent[11]. Figure 6 shows the monthly empirical hazard[12] in the first year of unemployment. Both regions exhibit a downward trend in the outflow rate over the period, at rates equal to -3.0% and -1.6%, in Flanders and Wallonia respectively[13]. So duration seems to drive the different structural evolutions in unemployment across regions. The cyclical variability in the exit rate is roughly similar in Flanders and Wallonia. In the next section, we try to determine the part of this structural and cyclical variation which is due to changes in the outflow rate of all currently unemployed workers and the part due to fluctuations in the average quality of entrants into unemployment.

INSERT FIGURE 6 APPROXIMATELY HERE

The distribution in the June inflow of the personal characteristics x, i.e. $\frac{u_{1lx}}{u_{1l}}$, is used in our formal analysis to evaluate the impact of changes in the observed quality of entrants on the aggregate outflow rate. How does this distribution evolve during the observation period? The composition of the inflow is depicted in Appendix 3 for Flanders and Wallonia: Figures 14, 15 (level of education), Figures 16, 17 (province of residence) and Figures 18, 19 (sector of previous employment).

The distribution in the inflow of the *schooling level* changes significantly over the whole period. There is a decrease of entrants with at most an elementary degree, from 60% on average over the first half of the period (1972-1983) to 35% over the second half of the period in Flanders, and from 57% to 40% in Wallonia. All other schooling levels similarly increase their shares in the inflow. In both regions, the inflow distribution of the *province of residence* is in line with their shares in the active population. In Wallonia, the large majority of entrants live in Hainaut and Liège while in Flanders, the number of entrants by province of living is more evenly distributed. The distribution of the province of living remains fairly constant over the period in Wallonia. In Flanders, there is a rise of entrants living in Vlaams-Brabant and Limburg which is compensated by a decline in the number of entrants living in West-Vlaanderen. There is some changes in the distribution of entrants by *sector of previous employment* over the period. In both regions, the share of entrants working previously in the building industry is large and declines over the whole period, from 30% on average over the first half of the period to 15% over the second half[14]. The large share of entrants coming from the public service

[11] The transformation is: $\widehat{h}_{1l} = -\ln\left(1 - \widehat{P}_{1l}\right)$.

[12] The monthly aggregate hazard is simply the yearly aggregate hazard divided by twelve.

[13] These rates are obtained by regressing the logarithm of the hazard rate in both regions on a constant term and a trend.

[14] This negative trend could be due to the increased severity of the entitlement rules which reduces the incentive to register for temporary laid-off workers.

is partly a characteristic of the June inflow: At the end of this month, teachers who do not hold a fixed position have to register as unemployed workers. The proportion of workers previously employed in the public services rises slightly over the period in both regions. This can reflect the importance of subsidised jobs created by the government to fight unemployment. The share of entrants from the steel-iron industry is slightly higher in Wallonia than in Flanders (on average 16% and 14%), but remains stable over time. The proportion of workers previously employed in the private sector is similar in both regions (on average 18%) and increases weakly. Finally, the mean *age* of entrants is 33 years old in Flanders and Wallonia and does not change significantly over the period.

In sum, variations in the share of unemployed workers by schooling level, and to a less extent by sector of previous employment, could induce significant changes in the aggregate outflow rate in both regions. On the other hand, changes in the distribution of age and province of living are unimportant to account for fluctuations in the aggregate exit rate. Variations in the composition of inflow is, however, only relevant in respect of personal characteristics that affect the exit probability. Our duration analysis allows us to evaluate whether the schooling level and the sector of previous employment have a significative impact on the outflow rate.

3 The mixed proportional hazard model

3.1 Notations and assumptions

Let the variables l and t denote respectively the calendar time at the moment of inflow and unemployment duration. These variables are measured in years. The way the inflow is constructed in our data requires us to model calendar time at the moment of inflow, l, as a discrete time process. We have that $l \in \{0, 1, ..., 20\}$ and by definition, that $t \equiv 0$ at the moment of inflow l. Although we analyse discrete data, we treat unemployment duration, t, as a continuous time process. So, $t \in R^+$. Otherwise, the specification would be sensitive to the time unit of the data (see Flinn and Heckman 1982). Finally, when conditioning on the unemployment history such as required in mixture models, each individual is treated 'as if' s/he did not change characteristics.

We denote the hazard of an individual unemployed for t years and at risk of leaving unemployment at calendar time $l + t$, conditional on his set of observed characteristics x and on his unobserved characteristics v, by:

$$h(t\,|l + t, x, v) \tag{3}$$

and the survivor function at unemployment duration t by:

$$S(t\,|x, v) = \exp\left[-\int_0^t h(s\,|l + s, x, v)\,ds\right] \tag{4}$$

Here and in the sequel, the calendar time dependence is implicit in the notation of the survivor function.

We then specify a mixed proportional hazard (MPH) model in which we impose a continuous baseline hazard and constant calendar time effects within yearly intervals:

$$\forall t \in [k-1, k) : h(t \,|l + t, x, v) = \exp\left[\varphi_1(k) + \varphi_2(l+k) + x\beta\right] v \equiv h_{klx} v \tag{5}$$

where $k \in \{1, ..., 21\}$, $\varphi_1(k) = \int\limits_{k-1}^{k} \varphi_1(t)\, dt$ and the functions $\varphi_1(.)$ and $\varphi_2(.)$ represent respectively the duration and calendar time dependence of the conditional hazard.

We assume that calendar time at the moment of inflow affects the shape of the mixing distribution (as in Abbring *et al.* 2001a). By doing so, we allow for variations in the unobserved characteristics of entrants into unemployment. Using (4) and (5), the mixture survivor function at the end of year k can then be found to be:

$$S^m(k \,|l, x) = \int\limits_0^\infty \exp\left[-v \sum_{j=1}^{k} h_{jlx}\right] dF_l(v) \tag{6}$$

where $F_l(.)$ is the heterogeneity distribution. Below, we will specify this distribution as well as how it depends on calendar time at the inflow.

We parametrise $\varphi_1(k)$ as a piecewise-constant function:

$$\varphi_1(k) = c + \sum_{j=2}^{21} \left(\gamma_j - c\right) \delta_{jk}, \tag{7}$$

where δ_{jk} is the Kronecker delta and where we impose that $\forall j \geqslant 4 : \gamma_j = \gamma_4$. The latter restriction avoids erratic variation of the parameters induced by a too small number of observations at the tail of the duration distribution.

In the line of Abbring *et al.* (2001b), we represent the calendar time dependence of the hazard, $\varphi_2(l+k)$, by the following flexible parametric specification:

$$\varphi_2(l+k) = \sum_{i=1}^{12} \alpha_i p_i(l+k) \tag{8}$$

The effect of calendar time at exit on the hazard is modelled by a flexible twelfth degree polynomial in calendar time, $p_i(l+k)$, capturing both trend and business cycle effects. To avoid multi-collinearity problems, we follow Abbring *et al.* (2001b, pp.439-440) and use the orthogonal Chebyshev polynomial instead of an ordinary polynomial.

3.2 The statistical model

We choose to estimate our model by Minimum Chi-Square (see Berkson 1944, Amemiya 1981, Cockx 1997). In a nutshell, this method consists in regressing, for each cohort, the empirical exit probability to its theoretical counterpart. The theoretical probability at the k^{th} duration interval for a worker of type x and at risk of leaving unemployment within the $(l+k)^{th}$ year is:

$$P_{klx} \equiv \Pr(k-1 \leq T < k \,|T \geq k-1, l+k, x) \tag{9}$$

Using (6), we then link this probability to the conditional hazard rate:

$$P_{klx} = \frac{S^m(k-1\,|l,x) - S^m(k\,|l,x)}{S^m(k-1\,|l,x)} = 1 - \frac{\int_0^\infty \exp\left[-v\sum_{j=1}^{k} h_{jlx}\right] dF_l(v)}{\int_0^\infty \exp\left[-v\sum_{j=1}^{k-1} h_{jlx}\right] dF_l(v)} \tag{10}$$

We assume that V (random variable) is a Gamma variate[15] of mean, $\mu_l = \frac{\lambda}{\delta_l}$, and variance, $\sigma_l^2 = \frac{\lambda}{\delta_l^2}$. As such, we allow the mean and the variance of V to vary proportionally over calendar time at the moment of inflow, l (as in Cockx and Dejemeppe 2002). By normalisation, the mean is equal to one for $l = 0$. We therefore impose $\lambda = \delta_0$. Since the moment generating function of this Gamma variate is (see Lancaster 1990, p.328):

$$M_v \equiv E\left(e^{sV}\right) = \left(1 - \frac{s}{\delta_l}\right)^{-\lambda} \tag{11}$$

we are able to simplify greatly the last expression in (10).

Before presenting the implied model, we specify how the variables δ_l depends on calendar time at the inflow. We first introduce the following notation: $\delta_l = \exp(-\varphi_3(l))$. Unlike for the effect of calendar time on the hazard at the outflow, we do not specify a polynomial to capture the effect of calendar time at entry. We choose to characterise the effect of calendar time at the moment of inflow by a business cycle indicator:

$$\varphi_3(l) = \eta_0 + \eta_1 b(l) \tag{12}$$

The business cycle indicator, $b(l)$, is the logarithm of the number of flows into unemployment every month of June[16]. We normalise $b(0) = 0$, so that $\lambda = \exp[-\eta_0]$. If this indicator does *not* vary *only* according to an exponential trend, identification is assured (see Abbring *et al.* 2001a). The plausibility of our parametric assumption can be verified by checking whether variations in the observed composition of entrants coincide with variations in incidence.

Combining (10), (11) and (12), the outflow probability then becomes:

[15] The choice of a Gamma distribution for $F(v)$ can be justified by the result that, for a continuous v and under mild regularity conditions, the unobserved heterogeneity distribution among survivors at t converges to a Gamma distribution if $t \to \infty$ (see Abbring and Van den Berg 2001).

[16] We also estimated a model with a second-order polynomial in $b(l)$. The coefficients of this polynomial are however very imprecisely estimated.

$$P_{klx} = 1 - \left[1 - \frac{\sigma_0^2 \mu_l \exp\left[\varphi_1(k) + \varphi_2(l+k) + x\beta\right]}{1 + \sigma_0^2 \mu_l \sum\limits_{j=1}^{k} \exp\left[\varphi_1(j) + \varphi_2(l+j) + x\beta\right]} \right]^{\frac{1}{\sigma_0^2}} \tag{13}$$

where $\sigma_0^2 = \exp[\eta_0]$ and $\mu_l = \exp[\eta_1 b(l)]$. By construction, $\sigma_l^2 = \sigma_0^2 \exp[2\eta_1 b(l)]$.

Finally, we relate the theoretical outflow probability to the empirical one. If we replace the probability of leaving unemployment, P_{klx}, by its estimate, $\hat{P}_{klx}$, given in (1), the equality of equation (13) does no longer holds exactly. For the observed aggregate exit probability is calculated on the basis of a finite sample. So we follow Cockx (1997) and obtain the following non-linear heteroskedastic regression model:

$$\hat{P}_{klx} = P_{klx} + \omega_{klx} + \varepsilon_{klx} \tag{14}$$

where ω_{klx} is the approximation error. The error ε_{klx} is a specification error which allows for random deviation from the true specification of the hazard rate. We assume that the latter has a distribution such that $E(\varepsilon_{klx}) = 0$ and $E\left(\varepsilon_{klx}^2\right) = s_\varepsilon^2$, and that they are not autocorrelated. Cockx (1997) shows that $E(\omega_{klx}) = 0$ and that a consistent estimate of the variance of ω_{klx} is:

$$\hat{s}_{klx}^2 = \left[\left(\hat{P}_{klx}\right)\left(1 - \hat{P}_{klx}\right)\right] \frac{1}{u_{klx}} \tag{15}$$

The estimation procedure consists in two-steps (see Amemiya and Nold 1975). In a first step we estimate the model (14) by Ordinary Least Squares (OLS). On the basis of the OLS residuals $(\hat{v}_{klx})$ and of the estimated variance of the approximation error $(\hat{s}_{klx}^2)$, we can calculate a consistent estimate of s_ε^2:

$$\hat{s}_\varepsilon^2 = \frac{1}{N} \left\{ \sum_x \sum_{l=0}^{20} \sum_{k=1}^{21-l} \left[(\hat{v}_{klx})^2 - \hat{s}_{klx}^2 \right] \right\} \tag{16}$$

where N denotes the number of cells (i.e. observations) in our data set. In a second step, we estimate the statistical model (14) by Generalized Non-Linear Least Squares[17]. The obtained estimator is a modified version of the Minimum Chi-square estimator, first applied to duration data by Cockx (1997).

[17] Since the majority of cells counts only one individual in our data set, the empirical exit probability will mostly take two values: $\hat{P}_{klx} = 1$ or $\hat{P}_{klx} = 0$. As a consequence, the weight based on the variance of the approximation errors, $u_{klx}\left[\left(\hat{P}_{klx}\right)\left(1 - \hat{P}_{klx}\right)\right]^{-1}$, is no longer defined for these cases. We deal with this problem by replacing $\hat{P}_{klx}$ by $P_{klx}\left(\hat{\theta}\right)$ which is simply P_{klx} evaluated at the previous round estimated values of the set of unknown parameters $(\hat{\theta})$. The advantage of this method is that $\hat{P}_{klx}\left(\hat{\theta}\right)$ will be never exactly equal to one or zero so that the weights are always defined (see Cockx 1997, p.395).

Because we have grouped data, we could use a χ^2-goodness-of-fit test to evaluate the model specification. The size of our cells is however too small for applying large sample theory. Recall that 93% of the cells are made up of only one individual at entry. The normal approximation of the binomial distribution on which the χ^2 test relies is good as long as $u_{klx}P_{klx} \simeq f_{klx} > 5$ (see Hoel 1971, p.82). Testing therefore requires a re-grouping of the data[18]. However, whatever the considered grouping this condition remains violated for most cells with an unemployment duration larger than one year. This is because only very few long-term unemployed workers leave unemployment, implying $f_{klx} < 5$ for $k > 1$. We therefore choose to average $\hat{P}_{klx}$ over age and sector of previous employment and calculate a new weighted sum of squared residuals for observations with $k = 1$ only. The fraction of observations still violating $f_{klx} > 5$ is equal to 1% and 6% in Flanders and Wallonia respectively[19].

The adjusted weighted sum of squared residuals ($WSSR^*$) is distributed $\chi^2_{N^*-p^*}$ if the model is correctly specified, where N^* and p^* is the number of observations and parameters after re-grouping the data. The latter number is obtained by subtracting from the total number of parameters those concerning the variables used in the re-grouping and the duration parameters. Moreover, we can test the acceptability of restrictions on the parameters, for instance the absence of unobserved composition effects (i.e. $\eta_1 = 0$). If the constrained model is well specified, then the difference between the adjusted weighted sum of squared residuals of the constrained and the unconstrained model ($WRSS_0^* - WRSS^*$) is distributed $\chi^2_{(p^*-p_0^*)}$.

3.3 Decomposition of the aggregate hazard over calendar time

In this section, we show the theoretical decomposition of the aggregate hazard over calendar time. As stated in the introduction, the calendar time dependence of the aggregate hazard combines both a general effect, i.e. changes in the individual (or conditional) hazard, and a composition effect, i.e. fluctuations in the average quality of entrants leaving the individual hazard constant. The latter effect is made of variations in the observed and unobserved characteristics of those flowing into unemployment. In our modelling framework, the aggregate hazard is the conditional hazard of an individual with average observed and unobserved characteristics, both averages being allowed to vary over calendar time at the moment of inflow. Focusing on unemployed workers in their first duration class (i.e. $k = 1$)[20] and denoting by $\Pr(X = x \mid k = 1, l)$ the empirical density of X at $k = 1$ and l, the aggregate hazard at calendar time $l + 1$ is given by:

[18] We are grateful to Bas van der Klaauw for pointing this out to us.

[19] A grouping according to the variable 'age' only still leaves around 80% of cells with $f_{klx} < 5$ among observations with $k = 1$.

[20] We do not calculate the aggregate hazard averaged over a larger number of duration intervals for the following reason. Since we have an inflow data set, not all duration classes are represented at the begin of the observation period. Therefore, considering more duration intervals decreases the timespan on which the aggregate hazard can be calculated. We have checked that the results of the decomposition are not affected by calculating the aggregate hazard on the basis of three duration intervals.

$$\bar{h}_{1l} = \sum_x \left[\int_0^1 h(t\,|l+t,x)\, E(v\,|l,T>t)\, dt \right] P(X=x\,|k=1,l) \tag{17}$$

where the empirical density of X at l is estimated by:

$$\Pr\left(X=x \mid k=1,l\right) = \frac{u_{1lx}}{u_{1l}} \tag{18}$$

Since the expression in square brackets is the mixture hazard evaluated at $k=1$, we can write:

$$\bar{h}_{1l} = \sum_x \left[-\ln\left(S^m(1\,|l,x)\right)\right] \frac{u_{1lx}}{u_{1l}} = \sum_x \left[-\ln\left(1-P_{1lx}\right)\right] \frac{u_{1lx}}{u_{1l}} \tag{19}$$

Substituting the expression of the outflow probability given in (13) into (19) and evaluating it at $k=1$, the aggregate hazard at calendar time $l+1$ is therefore:

$$\bar{h}_{1l} = \sum_x \left[\frac{1}{\sigma_0^2} \ln \left(1 + \sigma_0^2 \exp \left[\underbrace{c + \sum_{i=1}^{12} \alpha_i p_i\,(l+1)}_{\text{G}} + \underbrace{\eta_1 b\,(l)}_{\text{UC}} + \underbrace{x\beta}_{\text{OC}} \right] \right) \right] \underbrace{\frac{u_{1lx}}{u_{1l}}}_{\text{OC}} \tag{20}$$

The *general (G) effect* is the instantaneous effect of business conditions at exit on the hazard controlling for composition effects. In our model, we capture part of the variation in the composition of the inflow by conditioning the hazard on observed characteristics. The remaining variation, the unobserved composition effect, is accounted for by making the mixing distribution dependent on a business cycle indicator at entry. The polynomial function $\sum_{i=1}^{12} \alpha_i p_i\,(l+1)$ therefore measures the general effect of the calendar time on the aggregate hazard. The *observed composition (OC) effect* is the variation in the aggregate hazard due solely to changes in observed characteristics of those becoming unemployed through time. Variations of the empirical density of X in the inflow, $\frac{u_{1lx}}{u_{1l}}$, captures the observed composition effect of the calendar time on the aggregate hazard. Note that changes in the inflow density affect the aggregate hazard *only if* these changes concern characteristics that have a significant impact on the conditional hazard. The *unobserved composition (UC) effect* is the variation in the aggregate hazard which is only explained by changes in unobserved characteristics of those entering unemployment through time. The linear function $\eta_1 b\,(l)$ measures the unobserved composition effect of the calendar time on the aggregate hazard, which simply reflects the calendar time dependence of the mean (and the variance) of the Gamma distribution.

We compute synthetic series of the aggregate hazard rate in order to determine the relative role of each of its time-varying component. Each series is obtained by fixing the other components to their value at the first calendar time interval, i.e. $l = 0$. This yields to the following time-series of the aggregate hazard (where superscript C refers to the sum of the observed and unobserved components):

$$\bar{h}_{1l}^{G} = \sum_{x} \left[\frac{1}{\sigma_0^2} \ln \left(1 + \sigma_0^2 \exp \left[c + \sum_{i=1}^{12} \alpha_i p_i \left(l+1 \right) + x\beta \right] \right) \right] \frac{u_{10x}}{u_{10}} \tag{21}$$

$$\bar{h}_{1l}^{OC} = \sum_{x} \left[\frac{1}{\sigma_0^2} \ln \left(1 + \sigma_0^2 \exp \left[c + \sum_{i=1}^{12} \alpha_i p_i \left(1 \right) + x\beta \right] \right) \right] \frac{u_{1lx}}{u_{1l}} \tag{22}$$

$$\bar{h}_{1l}^{UC} = \sum_{x} \left[\frac{1}{\sigma_0^2} \ln \left(1 + \sigma_0^2 \exp \left[c + \sum_{i=1}^{12} \alpha_i p_i \left(1 \right) + \eta_1 b \left(l \right) + x\beta \right] \right) \right] \frac{u_{10x}}{u_{10}} \tag{23}$$

$$\bar{h}_{1l}^{C} = \sum_{x} \left[\frac{1}{\sigma_0^2} \ln \left(1 + \sigma_0^2 \exp \left[c + \sum_{i=1}^{12} \alpha_i p_i \left(1 \right) + \eta_1 b \left(l \right) + x\beta \right] \right) \right] \frac{u_{1lx}}{u_{1l}} \tag{24}$$

We can then quantify the relative importance of the general and the composition effects in the total variation of the aggregate hazard over the whole period by an analysis of variance. Since our observation period spans 20 years, structural and cyclical evolutions are separately investigated. We calculate the exponential trend in the three synthetic series ($\bar{h}_{1l}^{G}$, $\bar{h}_{1l}^{OC}$, $\bar{h}_{1l}^{UC}$) and we assume that the sum of these trends is equal to the trend in the aggregate hazard ($\bar{h}_{1l}$). This assumption is an approximation since the trend in the aggregate series does not sum up exactly the trend in the general and composition effects due to the non-linearity of each series. We then regress the de-trended logarithm of the aggregate hazard on the de-trended logarithm of the three synthetic series. On the basis of this model estimates, we calculate the part of the total variation in the aggregate hazard that is accounted for by a structural and a cyclical variation. Each type of variation is decomposed into the proportion explained by a general effect, an observed composition effect and an unobserved composition effect. The decomposition procedure is detailed in Appendix 2.

3.4 Empirical results

The estimation results for the mixture regression model specified in (14) are reported in Table 7 in Appendix 4, for Flanders and Wallonia. The constant term allows to calculate the hazard rate of a reference individual with an elementary school degree, living in Brabant wallon (resp. Vlaams-Brabant for Flanders), previously employed in the chemical industry and with the average unobserved characteristic (i.e. $v = 1$).

Consider first the performance of the models on the basis of the goodness-of-fit statistic. At a significance level of 5%, the model is rejected against the saturated model for both regions. This result was expected given the large number of observations used in the estimation. There is no evidence of specification errors as their estimated standard deviation is zero in both regions. Finally, we reject that the mean of the unobserved characteristics is constant over time ($\eta_1 = 0$) at a signifiance level of 5% in both regions. So changes in the unobserved quality of entrants can partly explain variations in the aggregate hazard in Flanders and in Wallonia.

3.4.1 Unobserved heterogeneity, duration dependence and covariate effects

Before we discuss the decomposition of the calendar time variation in the aggregate hazard, we first examine the shape of the baseline hazard and the estimated effect of the individual characteristics on the conditional hazard.

The result for the estimated variance of the heterogeneity distribution indicates a mild disparity between the unobserved characteristics of those flowing into unemployment. Since the variance of the Gamma distribution is allowed to fluctuate over time, we report its average estimated value over our observation period. In Flanders, the estimated variance is small but significantly positive: $\hat{\sigma}^2 = 0.20$ with a 95% confidence interval equals to $[0.15, 0.26]$. The variance is somewhat larger, but not significantly, in Wallonia: $\hat{\sigma}^2 = 0.29$ with a 95% confidence interval equals to $[0.21, 0.40]$.

The conditional hazard is significantly decreasing over unemployment duration, more markedly in Flanders. There is a 40% decrease in the individual hazard after two years of unemployment in Flanders while this declines is only equal to 25% in Wallonia. With respect to Wallonia and prime-aged men, Cockx and Dejemeppe (2002) obtain similar results on the basis of quarterly aggregate exit probabilities for the 1989-1994 period. After four years of unemployment the conditional hazard has dropped by more than 60% in both regions with respect to its initial level. There is an important spurious negative duration dependence in the unconditional hazard: If we did not control for heterogeneity in the characteristics of entrants into unemployment, there would be a 50% and 40% decrease after two years of unemployment in Flanders and Wallonia respectively, and 80% decrease after four years[21].

The estimates of the covariate effects indicate a more important disparity between the different groups of unemployed workers in Wallonia as compared to Flanders. In the MPH model, the covariate effects give the average effect of the observed characteristics on the exit rate over the 1972-1993 period. In the next stage, we aim to examine these disparities more carefully by analysing the way they have moved over time.

We first consider the effect of schooling on the hazard. The exit rate rises significantly with the level of educational attainment in both regions. However, the difference between the re-employment probabilities of workers by schooling level is much larger in Wallonia. For instance, having a higher education degree increases the re-employment probability by 25% in Wallonia while this increase amounts to 7% only in Flanders.

In Wallonia, the outflow rate is much lower for the workers who live in Hainaut, and in a less extent in Liège and Namur, than for those who live in Brabant wallon and Luxembourg. Living in Hainaut decreases the outflow rate by 25% as compared to living in the latter provinces. The difference between the exit rates according to the province

[21] The estimation is available upon request.

of residence is less pronounced in Flanders, with the worst performance for Limburg.

The dispersion of the sectorial effects is also higher in Wallonia. For instance, there is a significant disparity in Wallonia between the outflow rates of those coming from the steel and mining industry and of those previously employed in the private services (the latter are 40% more likely to leave unemployment than the former), while this difference is not significative in Flanders. It should be stressed that the sector occupied before becoming unemployed is not necessarily the sector of re-employment. Therefore, differences between the re-employment probabilities in the sectorial dimension can be explained either by differential sectorial business conditions and/or by differential mobility of the unemployed workers between sectors.

Finally, the outflow rate decreases significantly with age, more importantly in Flanders. The exit rate of 40 years old workers is 33% and 26% lower than the one of the 25 years old workers in Flanders and Wallonia respectively[22]. So even if we restrict our analysis to the 25-44 age group, age within this group still influences markedly the outflow probability.

3.4.2 Decomposition of the aggregate hazard over calendar time

In this section, we look at the relative importance of macroeconomic conditions and compositional changes in the inflow to explain the variations in the aggregate outflow rate over calendar time. Figures 7 and 8 show the estimated monthly aggregate hazard[23] in the first year of unemployment without composition effects ($\bar{h}_{1l}^{G}$), with only composition effects ($\bar{h}_{1l}^{C}$) and with the combination of both ($\bar{h}_{1l}$), for Flanders and Wallonia respectively. The decomposition between the observed ($\bar{h}_{1l}^{OC}$) and unobserved composition effects ($\bar{h}_{1l}^{UC}$) is depicted in Figures 9 and 10.

INSERT FIGURES 7, 8, 9 and 10 APPROXIMATELY HERE

Variations in the aggregate hazard seem to be driven by a negative trend in Wallonia while cyclical changes are relatively more important in Flanders. This observation is confirmed by the analysis of variance: 52% of the variation in the aggregate hazard ($\bar{h}_{1l}$) is accounted for by a structural change while this share is only equal to 13% in Flanders.

The graphical decomposition also indicates that variations in the aggregate hazard are driven, in both regions, by fluctuations in the individual exit rate of all currently unemployed rather than by changes in the quality of entrants into unemployment. Still, the composition effect accounts for a significant part of the structural variations in the aggregate hazard. According to the analysis of variance, changes in the composition of entrants explain 27% and 24% of the structural evolution in the aggregate hazard, in Flanders and Wallonia respectively, and only 13% and 14% of its cyclical variability[24].

[22] We also estimated a model with a second-order polynomial in age. The coefficients of this polynomial are however very imprecisely estimated.

[23] The monthly aggregate hazard is simply the estimated yearly aggregate hazard divided by twelve.

[24] In the cyclical decomposition, there is a negative covariance between the general effect and both, the observed and unobserved composition effects. This negative covariance indicates that the cyclicality of the general and the composition effects works in the opposite direction, thereby reducing the total

As suggested by Figures 9 and 10, the unobserved component accounts for a larger share of the structural variation in the average quality of entrants than the observed one: 60% in Flanders and 70% in Wallonia. Changes in the unobserved quality also drive most of the cyclical variability in the composition effect: Only 5% and 30% of its variability comes from changes in the observed quality of entrants[25].

It should be noticed that the results of the graphical decomposition and the variance analysis are hardly sensitive to the reference date chosen (i.e. $l = 0$)[26].

Below, we discuss the structural evolution in the components of the aggregate hazard by analysing their exponential trend, and their cyclicality by comparing their estimated cycle to a business cycle indicator.

Flanders and Wallonia exhibit both a decreasing trend in the aggregate outflow rate over 1973-1993 period at rates equal to -1.6% and -3.7% yearly, respectively[27] (see Table 1). This downward trend only results from a structural decrease in the individual exit rate (i.e. in the *general effect*), at rates equal to -2.8% and -5.9% yearly in Flanders and Wallonia respectively. The strong decline in the individual outflow rate over this period was mitigated at the aggregate level by an increase in the quality of those flowing into unemployment in both regions. The rising share of entrants with good re-employment prospects since 1972 has induced an annual increase in the aggregate outflow rate of 1% in Flanders and 1.8% in Wallonia. In the absence of this positive composition effect, the aggregate exit rate would have declined by 43% and 69% with respect to its initial level in Flanders and Wallonia respectively.

INSERT TABLES 1 and 2 APPROXIMATELY HERE

Table 2 shows that the upward trend in the composition of inflow is due to an improvement in both observed and unobserved quality of entrants[28], the trend in the latter component being much larger in Wallonia. The structural rise in the observed quality of entrants is mainly attributed to an increase in the share of high educated workers in the unemployed population, who have on average a larger exit rate. The variance analysis indeed reveals that over the 1972-1993 period, changes in the distribution of schooling among entrants account for 80% and 90% of the increase in the observed quality of entrants in Flanders and Wallonia respectively.

In both regions, the cyclical variability in the aggregate hazard is driven by variations in the individual exit rate (i.e. in the *general effect*) rather than by composition effects.

variability in the aggregate hazard.

[25] Since we parametrise the unobserved composition effect by the number of entrants, the cyclical variation in the unobserved quality of entrants follows the cycle of incidence. The amplitude of the cycle in the unobserved composition effect is given by the estimated value of the parameter η_1.

[26] By taking $l = 10$ (i.e. June 1982) as reference date in the decomposition analysis, the relative role of the composition effect in the structural variation of the aggregate hazard is 31% and 27% in Flanders and Wallonia respectively, and is unchanged in the cyclical decomposition.

[27] As compared to the Figure 6 reported in Section 2.4, there is a slight overestimation of the downward trend in the aggregate hazard in Wallonia as compared to the trend in the empirical aggregate hazard.

[28] The variance of the unobserved characteristics of entrants also increases over time, since it is simply twice the evolution of the mean.

Given their respective role, we first examine the cyclicality in the general effect. Figure 11 shows the de-trended general effect in the two main regions of Belgium. In order to evaluate its cyclical pattern, we compare it to a Belgian business cycle indicator[29], the Kredietbank indicator. This indicator starts in 1980, i.e. after the oil price shocks which coincide with an important rupture in the long-run growth of the Belgian economy[30].

INSERT FIGURE 11 APPROXIMATELY HERE

The individual exit rate appears to be procyclical, its pattern being very similar in Flanders and in Wallonia over the whole period. The exit rate is sharply decreasing between 1974 and 1983: It declines by more than 60% in 1983 compared to its level in 1974. This strong decline of the outflow rate in Belgium coincides with the first and second oil price shocks in 1973 and in 1979. Over this period, employment decreases strongly in Belgium, the employment losses being concentrated in the industry. During the period of recovery, which starts with the devaluation of the Belgian Franc in 1982 and continues up to 1990, Wallonia lags Flanders continuously. Yet, the business revival on the outflow rate is observable in the two regions. From 1990, the economic downturn pushed the exit rate down again in both regions.

In Appendix 3, Figures 20 and 21 show the estimated cyclical variation in the observed quality of entrants for Flanders and Wallonia respectively. Even if its impact on the aggregate hazard is marginal, we can state that the observed composition of the inflow is countercyclical in both regions: The average quality of entrants into unemployment increases during recessions and falls in upturns. The positive sign of $\hat{\eta}_1$ (see Table 7 in Appendix 4) also indicates that the unobserved quality of entrants improves as labour market conditions deteriorate (i.e. when incidence increases). These results suggest that employers fire the most able workers only if they have no alternative, i.e. in a recession. Unable workers are made redundant all the time.

Our results concerning the unobserved composition effect could be sensitive to our parametric assumption: We assume that changes in the unobserved quality of entrants are proportional to changes in incidence. The pattern of the observed composition effect confirms the plausibility of this assumption. For, variations in the observed quality of entrants coincide with variations in incidence, in both the cyclical evolution and the structural evolution. As a result, our specification turns out to be a better alternative than a more flexible specification since it allows the unobserved composition effect to follow a trend, separately from the trend in the general effect.

We finish the discussion of the estimation results by comparing them to those obtained in other studies. A more detailed interpretation of the results is provided in the next section. Our results are in line with most of the studies based on European data. With the exception of Kalwij (2001) for the UK, they all conclude that composition effects are of minor importance to explain *cyclical* fluctuations in the aggregate hazard (see the introduction). So the English-speaking countries seem to remain the exception,

[29] There is no regional business cycle indicator in Belgium.

[30] Over the 1960-1975 period, the long-run growth of the Belgian economy is 4.5% while it is only 1.9% in the 1976-1996 period (see Bodart 2000).

with variations in the quality of entrants driving a significant part of the variations in unemployment duration. It should, however, be stressed that none of the reviewed studies decompose formally the structural variation in the aggregate outflow rate into a general and a composition effect. As is shown in this study, the role of the composition effect is, in this respect, relatively important.

Although they analyse a shorter timespan, Abbring *et al.* (2001a) also found that the individual exit rate trends downward significantly in France between 1982 and 1994, for the male unemployed workers. However, Abbring *et al.* estimate the composition effect non-parametrically, so that they were forced to assume the absence of a trend in the composition of inflow. If the quality of entrants had also improved over their observation period, the estimated trend of the individual exit rate in France would be upward biased. Using the same data, but combining them with a micro data set, Van den Berg and Van der Klaauw (2001) actually found a slightly increasing trend in the observed quality of entrants over the 1982-1994 period in France.

The cyclical fluctuations in the individual hazard in Belgium seems similar to the one estimated in France (see Abbring *et al.* 2001a) and in the UK (see Kalwij 2001). In both countries, the individual exit rate is procyclical and drives the cyclical variation in the aggregate hazard. As regards to the cyclicality of the composition effect, our results are close to the one obtained in Cockx and Dejemeppe (2002), Van den Berg and Van der Klaauw (2001) and Kalwij (2001). They also find that the quality of entrants is countercyclical. It should be noted that Abbring *et al.* (2001a) found a slightly procyclical composition effect on the basis of the same data as in Van den Berg and Van der Klaauw (2001). However, the estimated composition effect of the former authors combines both an observed and unobserved component since they cannot condition the hazard rate on observed characteristics. Darby *et al.* (1985) and Rosholm (2001) also obtain a procyclical variation in the quality of those becoming unemployed. They attribute their findings to changing proportions of voluntary quitters and laid off workers over the cycle. Note that we do not capture voluntary quits in our data, since these workers are not eligible for unemployment benefits.

4 Decomposition of the general effect

In this section, we question whether the general effect of the calendar time on the aggregate hazard affects all unemployed in the same proportion in both regions of Belgium. For that purpose, we relax the proportionality assumption and allow the calendar time dependence of the conditional hazard to be specific to the personal characteristics x. This amounts to replace $\varphi_2(l+k)$ and $x\beta$ in the conditional hazard (5) by $\varphi_2(x, l+k)$, which is specified as follows:

$$\varphi_2(x, l+k) = x\beta_0 + \sum_{i=1}^{12} x\beta_i p_i(l+k) \tag{25}$$

Allowing for a separate polynomial in calendar time for each value of the personal characteristics would lead to imprecise and intractable parameter estimates. In a benchmark model, we therefore specify a separate twelfth degree polynomial by schooling level and allow for different trends only for the other explanatory variables[31]. Moreover, we assume that the age effect is constant over our observation period.

4.1 Empirical results

The estimation results for the mixture regression model (14) in which $\varphi_2(x, l+k)$ replaces $\varphi_2(l+k)$ are reported in the first column of Table 8 and Table 9 in Appendix 4, respectively for Flanders and for Wallonia. The estimated trends by sub-region of living and sector of previous employment are expressed in deviation from the reference trend (one by schooling level).

Upon comparison with the estimates of Table 7, we conclude that the introduction of time-varying covariate effects does not influence the other parameter estimates. In particular, the estimated value of the parameter capturing unobserved composition effects, η_1, is similar to the one estimated in the MPH model. Also notice that the performance of the model improves on the basis of the goodness-of-fit statistic. Nevertheless, the non-proportional model is still rejected against the saturated model at a significance level of 5%, for both regions. Before entering into a more detailed interpretation of the estimation results, we first question whether we can impose some restrictions on the time dependence of the covariate effects based on the χ^2 test[32].

With respect to the *schooling level*, results differ between regions. In Wallonia, the exit rate of the highest educated workers evolves differently from the three other schooling levels: At a significance level of 5%, we cannot reject that both the cyclical and structural patterns of the exit rate are the same for the three lowest education levels while they differ significantly for those who completed higher education. However, the difference between these two groups is mild as regards to the cyclical variability in their exit rates. In Flanders, we cannot reject that the structural evolution of the exit rate is the same for the two lowest schooling levels while it is significantly different for those who have at least an upper secondary degree. On the other hand, the cyclical evolution of the outflow rate is similar across schooling groups in Flanders. With respect to the *province of living*, only unemployed people living in Antwerp and Limburg in Flanders, Hainaut and Namur in Wallonia, have a different trend in their exit rate in comparison with the reference province. Finally, as regards to the *sector of previous employment*, we cannot reject that the effects of all sectors on the exit rate evolve similarly over our observation period at a signifiance level of 5% in Wallonia. In Flanders, we cannot reject that the exit rate of workers coming from the iron and steel industry has a different trend than the exit rate of those coming from the other sectors.

The second column of Tables 8 and 9 in Appendix 4, respectively for Flanders and

[31] We assume the same trend for the sectors 'public employment' and 'energy' since this latter concerns a too small number of observations.

[32] In order to test for the signifiance level of the sectoral trends, we average $\hat{P}_{klx}$ over age, province of living and schooling level and calculate a new weighted sum of squared residuals on observations with $k = 1$.

Wallonia, presents the estimation results of our extended regression model in which we set the restrictions not rejected according to the χ^2 test. For purposes of comparison between the two regions, we also impose the same cyclical evolution in the exit rates of all schooling groups in Wallonia and the absence of specific trends in the sectorial dimension in Flanders. The interpretation of the results is based on this restricted model.

The interest of the non-proportional hazard model is to decompose the structural evolution in the MPH general effect into group specific trends. To that purpose, we compute the general effect of the aggregate hazard rate, $\bar{h}^G_{1l}$ in (21), by group of workers with a diverging evolution in their outflow rate, for instance the low educated workers in the province of Hainaut. We then calculate the exponential trend in each series. The trend estimated in the MPH general effect is then, approximately, a weighted sum of the trends in the group specific general effects, where the weights are the average share of each group in the inflow into unemployment.

Table 3 shows that the estimated trend of the MPH general effect masks a different trend for the low and high educated workers in both regions. In Wallonia, the outflow rate of less educated job seekers (below higher education) deteriorated more importantly than the exit rate of the highest educated workers on the whole period: at rates equal to -6.1% and -4.1% yearly[33], respectively. Since the low educated workers represent, on average, 85% of entrants, the trend in the MPH general effect closely follows the pattern of their outflow rate. In Flanders, the outflow rate of less educated job seekers (below upper secondary) decreased at a rate equals to -3.1% yearly while it declined more mildly for the higher educated unemployed workers, at a rate of -1.9% yearly.

INSERT TABLE 3 APPROXIMATELY HERE

Table 4 compares the marginal impact of each schooling degree on the exit rate in 1973 and in 1993, in both regions. We also report the average effect, such as estimated in the MPH model, in order to see the importance of allowing for time-varying covariate effects. In Wallonia, the disparity between exit rates by schooling level was already important in 1973, and reinforced markedly between the highest and the lower education levels over the whole period. For, having a higher education degree did not raise the re-employment rate in 1973, while it increased this rate by 42% in 1993. Table 4 also indicates that the disparity between the outflow rate by schooling level increased slightly in Flanders, between the lower and the upper secondary level. Having a upper secondary degree reduced the re-employment probability by 5% in 1973, while it increased the outflow rate by 19% in 1993. Nevertheless, the disparity by educational attainment remained at a relatively low level in Flanders as compared to Wallonia.

INSERT TABLE 4 APPROXIMATELY HERE

The difference between the exit rates by province of residence also rose in Wallonia. Table 5 gives the effect on the outflow rate of living in other provinces than Brabant

[33] These rates are means over provinces.

wallon (resp. Vlaams-Brabant in Flanders)[34], in 1973 and 1993. The re-employment prospects of those living in Hainaut were the worst among the provinces of Wallonia in 1973 and continued to deteriorate over the period. The outflow rate of workers living in Namur was similar to the one estimated in Brabant wallon in 1973, but got closer to the one estimated in Liège in 1993. On the contrary, the divergence between the re-employment prospects by province of residence reduced in Flanders. The re-employment prospects of those living in Antwerpen were the most favourable in 1973, but this ranking was reverse at the end of the period in favour of workers living in Limburg.

INSERT TABLE 5 APPROXIMATELY HERE

4.2 Tracks of interpretation for the regional divergences

As shown in the previous section, the larger decrease in the aggregate outflow rate in Wallonia over the 1973-1993 period is only due to a stronger trend in the individual re-employment prospects, which is twice as large the one estimated in Flanders. This section also reveals that the deterioration of the re-employment prospects reflects, in both regions, different evolutions across skill groups and provinces, mainly in Wallonia. In this sub-section, we review some possible causes of the regional disparities in Belgium and investigate whether our results can reinforce the existing evidence[35].

Note first that the stronger negative trend in the individual exit rate in Wallonia cannot be attributed to a migration (with a change in the place of residence) of the workers with good re-employment prospects to Flanders, which would have affected the composition of the unemployed population in the two regions. For, the mobility of workers between Flanders and Wallonia is very weak (see Candelon *et al.* 2000). Linguistic and cultural barriers play an important role in that respect.

The stronger deterioration of the re-employment prospects in Wallonia is reflected in the employment growth across regions of Belgium. Table 6 shows the evolution of salaried employment, by sector and sub-periods, in Flanders and Wallonia. Over the 1973-1995 period, employment declined by 1.3% in Wallonia while it increased by 20% in Flanders. Table 6 also suggests that the poor employment performance in Wallonia over this period reflects both a more pronounced and persistent shock in the industry sector and an insufficient employment growth in the private services.

INSERT TABLE 6 APPROXIMATELY HERE

Based on a shift-share analysis, some authors argue that the sectorial structure cannot account for the diverging employment growth in Flanders and Wallonia (see Binon *et al.* 1998, Mignolet and Vieslet 2000). This diverging evolution only results from a poor

[34] Luxembourg (resp. West-Vlaanderen in Flanders) is not reported in these tables since both the level and the evolution of the outflow rate in this province is not significantly different than the one estimated in Brabant wallon (resp. Vlaams-Brabant).

[35] As emphasised in the data section, the rules governing the unemployment benefit scheme and the functioning of the labour market are the same in all regions of Belgium. These factors can therefore not be put forward to explain the regional divergences in Belgium.

employment performance within sectors. The effect of the sectorial structure appears to be weak or even favourable to Wallonia. This structure is against this region for declining sectors in which the employment rate was high as compared to Flanders, like the iron-steel industry and mining. But at the same time, there are other declining sectors in which Flanders was more specialised. The shift-share analysis does not however provide the reasons of the poor employment performance in Wallonia.

According to Sneessens *et al.* (1999), structural imbalances between skill groups and sub-regions play a crucial role to explain the weakness of job creation in Wallonia[36]. Their measure of mismatch is based on unemployment rates by level of education and province. The disparity between the demand and the supply of skills is, on average, lower in Flanders than in Wallonia, and increased more markedly in the latter region over the 1973-1993 period. With respect to the regional mismatch, the disparity between provinces became more pronounced in the two regions since the late seventies, but they disappeared progressively in Flanders since the mid eighties. As it is shown above, we obtain similar results based on re-employment probabilities by skill level and province.

The stronger rise of imbalances in the Walloon labour market is partly due to the nature of the shocks that hit this region in the seventies and early eighties (see Binon *et al.* 1998 and Sneessens *et al.* 1999). Unlike Flanders, employment losses were concentrated in some important sectors and provinces in Wallonia, particularly the steel and mining industry in the province of Hainaut. The extent of the de-industrialisation, and its concentration in large area, have certainly hampered the development of productive activities in Wallonia and the re-orientation of dismissed workers to growing sectors.

Our results can give some lights on the latter fact. Both regions of Belgium lost employment in the steel and mining industry over the 1973-1993 period while they gained jobs in the private services (see Table 6). However, the re-employment prospects of workers previously employed in the steel and mining industry in Flanders are not significantly different from those of workers coming from the private services, contrary to Wallonia (see Section 3.4). Although this result partly reflects larger job destructions in the Walloon industry, it can also indicate a better ability of Flanders to provide dismissed workers from declining sectors with skills suited to the new developing activities. An insufficient effort put into education and training in Wallonia must have played an important role in that respect (see Binon *et al.* 1998).

Concerning the schooling system, it appears that the vocational education in the upper secondary school in Wallonia attracts a larger share of students with poor learning ability as compared to Flanders. This composition effect can prevent the acquisition of skills that are directly valuable on the labour market. Wallonia is also characterised by a lower number of young people who achieve the upper secondary school, particularly with vocational and technical degrees. In a context of skill biased technological change, this feature of the educational system may discourage the emergence of new activities.

[36] Other explanations are the lack of private investment in Wallonia in the eighties and the large share of public funds invested to declining sectors in this region (see Binon *et al.* 1998). There is no evidence that different evolutions in competitivity (in terms of wage costs) can account for the diverging employment growth in Flanders and Wallonia.

These factors can explain why the re-employment prospects of workers with an upper secondary degree evolve similarly to those of the lowest educated ones in Wallonia while they evolve similarly to those of workers with higher education in Flanders. Another cause of the deficit of skilled labour in Wallonia is that the share of unemployed workers completing a training scheme is much lower in this region than in Flanders[37].

Finally, note that the strong disparity between the Walloon provinces also indicates that some of them perform relatively well. With a shift-share analysis, Toulemonde (2000) shows that over the 1973-1992 period, two Walloon provinces (Luxembourg and Brabant wallon) registered one of the best employment performance in Belgium while one Flemish province (Antwerp) was performing badly. According to our results, the aggregate exit rate estimated in Antwerp was actually similar to the one estimated in Luxembourg and Brabant wallon, in 1993. However, the relative size of these provinces in the Walloon labour market is relatively small.

5 Simulated stock of unemployment over calendar time

In this last section, we simulate the unemployment stock in June based on the previously estimated MPH model. The objective of this exercise is to determine the relative importance of incidence and duration in the dynamics of unemployment in Flanders and Wallonia. To that purpose, we decompose the variations in the simulated unemployment stock into an incidence and a duration effect by the analysis of variance. Structural and cyclical evolutions are separately investigated. Abbring *et al.* (2001b) follows a similar approach to decompose the cyclical variations in the U.S. unemployment rate into various incidence and duration components.

The unemployment stock is calculated from our inflow data set, i.e. it only accounts for entrants in the month of June. The simulated stock of unemployed workers at calendar time l, u_l, is the sum of the workers entered into unemployment at l and of those entered at $l-k$, $\forall k \in (1, 2, ..., K)$, and surviving from $l-k$ up to l:

$$u_l = u_{1l} + \sum_{k=1}^{K} u_{1l-k} \sum_x \exp\left[-\sum_{i=1}^{k} \int_{i-1}^{i} h(t\,|l-k+t, x)\, E(v\,|l-k, T>t)\, dt \right] \frac{u_{1lx}}{u_{1l}} \quad (26)$$

where K is the number of durations accounted for in the simulated stock. The expression which comes after u_{1l-k} is simply the aggregate (mixture) survival function at the end of the duration interval k.

In order to determine the relative role of incidence and duration, we simulate two synthetic series of unemployment. Each series is obtained by fixing the other component to its value at the first calendar time interval, i.e. at $l = 0$. The simulated stock under the assumption that only incidence varies, u_l^{IN}, writes as follows:

[37] Training policies are mostly within the competence of the regions in Belgium.

$$u_l^{IN} = u_{1l} + \sum_{k=1}^{K} u_{1l-k} \sum_x \exp\left[-\sum_{i=1}^{k} \int_{i-1}^{i} h(t\,|1,x)\, E(v\,|0,T>t)\, dt\right] \frac{u_{10x}}{u_{10}} \qquad (27)$$

where both the general and the composition effects of the calendar time on the aggregate hazard are fixed at their value at $l = 0$.

The simulated stock under the assumption that only duration varies, u_l^{DU}, is:

$$u_l^{DU} = u_{10}.\left(1 + \sum_{k=1}^{K} \sum_x \exp\left[-\sum_{i=1}^{k} \int_{i-1}^{i} h(t\,|l-k+t,x)\, E(v\,|l-k,T>t)\, dt\right] \frac{u_{1lx}}{u_{1l}}\right) \qquad (28)$$

where the expression between brackets can be interpreted as the average unemployment duration (over K duration intervals) at calendar time l. Its variation captures both the general and the composition effects of the calendar time.

Since we have an inflow data set, simulating the stock taking into account K duration intervals decreases our series by K simulated values. This could bias the relative share of structural and cyclical evolutions in the total variation of unemployment. For instance, the number of entrants is increasing at rates equal to 3.9% and 4.6% yearly over the 1972-1992 period, in Flanders and Wallonia respectively, while these rates are only equal to 0.8% and 2.3% over the 1975-1992 period[38]. This would justify to simulate the unemployment stock for short term unemployed workers only ($K = 1$), which allows for a decomposition over the 1973-1992 period. However, the role of duration will be underestimated by considering only one duration interval in the unemployment stock. For, variations in incidence lead to variations in the average unemployment duration when the baseline hazard is decreasing. By focusing on the first duration interval only, we ignore this relationship and, as a result, we decrease the variability of duration.

In order to limit the potential biases, we decide to examine the relative importance of structural and cyclical evolutions based on the stock of short term unemployment. The relative role of duration and incidence in the variability of unemployment is analysed for a more representative unemployed population. In the benchmark simulation, we choose $K = 3$, which allows for a simulation of unemployment over the 1975-1992 period, i.e. $\forall l \in \{3, 4, ...20\}$. We then investigate the sensitivity of our results by including more duration intervals in the unemployment stock.

In order to quantify the relative importance of each component of the unemployment stock dynamics, we use the same analysis of variance than the one developed in Appendix 2. We calculate the exponential trend in the two synthetic series (u_l^{IN}, u_l^{DU}) and we assume that the sum of these trends is equal to the trend in the total stock (u_l). We then regress the de-trended logarithm of the total stock on the de-trended logarithm of the two synthetic series. On the basis of these model estimates, we calculate the part of the total variation in the unemployment stock that is accounted for by a structural and a cyclical variation. Each type of variation is decomposed into the proportion explained by an incidence and a duration effect.

[38] These rates are the same whether they are calculated on u_l^{IN} or u_{1l}.

Changes in unemployment are dominated by a positive trend in Wallonia while cyclical changes appear to be relatively more important in Flanders: 54% of the variation in unemployment is accounted for by a structural variation while this share is only equal to 27% in Flanders. This difference is explained by a stronger positive trend in duration and incidence in Wallonia. As regards to the relative role of duration and incidence, we simulate the unemployment stock up to $K = 7$ to obtain stable results. In Wallonia, 57% of changes in unemployment results from changes in duration[39], while 43% results from variations in incidence. The decomposition is similar in Flanders: 55% of the total variability in unemployment is due to changes in duration[40] and 45% in incidence.

So our results indicate that *both* duration and incidence matter in the variation of unemployment in Belgium. The common belief is that changes in incidence is largely irrelevant in Europe (see, e.g., Layard *et al.* 1991; Mortensen and Pissarides 1999), such as claimed by Burgess and Thuron (2000, p.2) in their critique on the referred authors. Our estimated figures are actually close to those obtained by Abbring *et al.* (2001b) for U.S. Adopting a similar decomposition approach, they find that 57% of the cyclicality in the U.S. unemployment rate is due to changes in duration and 43% to incidence.

The role of duration dependence on the simulated stock is illustrated in Figures 12 and 13, for Flanders and Wallonia respectively. These Figures compare the total stock of unemployment in June such as reported by the official statistics and the simulated stock of unemployment for $K = 1$, $K = 3$ and $K = 7$.

INSERT FIGURES 12 and 13 APPROXIMATELY HERE

When considering more duration intervals, the divergence between the simulated and total stock decreases due to the smaller bias in the unemployment duration. Even if it only accounts for workers entered into unemployment in the month of June, the simulated stock fits the total stock of unemployment reasonably well. There is one exception, in 1986, where the simulated stock is decreasing while the total stock is mildly increasing. This divergence comes from a different evolution of the June inflow as compared to the evolution of the monthly average inflow, such as reported by the public employment agency. The inflow in June 1986 was substantially lower than the monthly average.

6 Conclusion

In this paper, we investigated whether the diverging evolution in the unemployment stock in Flanders and Wallonia can be explained by different evolutions in their average unemployment duration and/or their unemployment incidence. To that purpose, we proceeded in two stages. In the first stage, we estimated a MPH model by region, and decompose variations in the aggregate outflow rate over calendar time between a general effect and a composition effect. We also specified a non-proportional model to check

[39] The relative contribution of duration is 60% and 56% in the structural and cyclical dimension respectively.

[40] The relative contribution of duration is 51% and 69% in the structural and cyclical dimension respectively.

whether the general effect of the calendar time was the same for unemployed workers with different schooling levels and sub-region of living, in each region. In the second stage, we decomposed variations of the unemployment stock in Flanders and Wallonia into an incidence effect and the duration effect estimated in the first stage.

Over the 1973-1993 period, unemployment increased more markedly in Wallonia due to a stronger trend in average unemployment duration and, to a less extent, a more positive trend in incidence. As a consequence, changes in unemployment are dominated by an upward trend in Wallonia while cyclical changes are relatively more important in Flanders. Our results indicate that the composition of the unemployed population cannot explain the larger trend in the average unemployment duration in Wallonia. On the contrary, the structural rise in duration would have been higher in both regions if the average quality of entrants had not improved over the period. The larger increase in the average duration in Wallonia is only due to a stronger trend in individual durations, which is twice as large as the one estimated in Flanders.

On the basis of our results, we also conclude that disemployability effects, even if they play a role, cannot explain a stronger trend in unemployment duration in Wallonia. For, genuine negative duration dependence is less important in this region than in Flanders. In the latter region, the relatively large degree of true negative duration dependence could partly account for the slight positive trend in its average unemployment duration: The existence of disemployability could have somewhat tempered the employment growth in Flanders over the eighties.

According to our results, the deterioration of the hiring probability in both regions reflects different evolutions across skill groups and provinces. In Wallonia, having a higher education degree did not raise the re-employment rate in 1973, while it increased this rate by 42% in 1993. The disparity between the outflow rate by schooling level also widened in Flanders, but remained at a low level as compared to Wallonia. Moreover, the difference between the exit rates according to the province of residence rose in Wallonia while it reduced in Flanders. The hiring prospects of unemployed workers living in Hainaut were the worst among the provinces of Wallonia in 1973 and deteriorated steadily over the period. Our findings therefore suggest that even if composition effects cannot account for the structural drop in the outflow rate, there is a scope for labour market policies aimed at specific groups of unemployed workers.

The cyclical nature of unemployment is very similar in the two regions of Belgium and is characterised by a strong countercyclicality, in which variations in unemployment duration are slightly more important. In both regions, the cyclical variability in duration is driven by cyclical variations in the individual durations rather than by composition effects. Our results are thus in line with most of the studies based on European data, which conclude that composition effects are of minor importance to account for cyclical fluctuations in the aggregate hazard. Individual durations are countercyclical, their pattern being very similar in Flanders and in Wallonia over the whole period. Even if its impact on the aggregate hazard is marginal, we can state that the composition of the inflow is countercyclical in both regions: The average quality of entrants into unemployment increases during recessions and falls in upturns.

One important conclusion of our results is also that *both* duration and incidence matter in the (cyclical and structural) variation of unemployment in Belgium while the common belief is that changes in incidence is largely irrelevant in Europe.

7 References

Abbring, J.H. and G.J. van den Berg (2001), "The unobserved heterogeneity distribution in duration analysis", Working Paper, Free University Amsterdam, Amsterdam.

Abbring, J.H., G.J. van den Berg, and J.C. van Ours (2001a), "The anatomy of unemployment dynamics", *European Economic Review*, forthcoming.

Abbring, J.H., G.J. van den Berg, and J.C. van Ours (1999), "Business cycles and compositional variation in U.S. unemployment", Discussion Paper n°9965, CentER for Economic Research, Tilburg.

Abbring, J.H., G.J. van den Berg, and J.C. van Ours (2001b), "Business cycles and compositional variation in U.S. unemployment", *Journal of Business and Economic Statistics*, 19 (4), 436-448.

Abramowitz, M. and I.A. Stegun, eds (1970), *Handbook of mathematical functions; with formulas, graphs and mathematical tables*, Dover Publications, New York.

Amemiya, T. (1981), "Qualitative response models: A survey", *Journal of Economic Literature*, 19, 1483-1536.

Amemiya, T. and F. Nold (1975), "A modified logit model", *Review of Economics and Statistics*, 57, 255-57.

Baker, M. (1992), "Unemployment duration: Compositional effects and cyclical variability", *American Economic Review*, 82, 313-321.

Berkson, J. (1944), "Application of the logistic function to bio-assay", *Journal of the American Statistical Association*, 39, 357-365.

Binon, F., J.-B. Charles, R. Deschamps, S. Delaigle, J.-C. Jacquemin, O. Lohest, M. Mignolet, H.R. Sneessens, and B. Van der Linden (1998), "Recherche relative aux facteurs explicatifs de l'évolution de l'emploi en Wallonie et de la divergence interrégionale pour la période 1974-1995", Rapport au Ministre wallon du Budget et Finances, Emploi et Formation J.-C. Van Cauwenberghe, IRES (UCL) et CREW (FUNDP).

Bodart, V. (2000), "Le comportement cyclique de l'économie belge: 1960-1996", Etude spéciale du Service d'Analyse économique IRES-UCL, Mars.

Burgess S. and H. Turon (2000), "Unemployment dynamics, duration and equilibrium: Evidence from Britain", CEPR Discussion Paper n°2490, Centre for Economic Policy Research, London.

Candelon, B., A. Hecq, and O. Lohest (2000), "Labor mobility in Belgium: An Empirical analysis of the relationship between provincial employment dynamics and migration", IRES Discussion Paper n°$2000-29$, Institut de Recherches Economiques et Sociales, Université Catholique de Louvain.

Cockx, B. (1997), "Analysis of transition data by the Minimum Chi-Square method: An application to welfare spells in Belgium", *The Review of Economics and Statistics*, 79 (3), 392-405.

Cockx, B. and M. Dejemeppe (2002), "Duration and calendar time dependence of the exit rate out of unemployment in Belgium. Is it true or sprurious?", *mimeo*, Institut de Recherches Ecomomiques et Sociales, Université catholique de Louvain.

Darby, M.R., J. Haltiwanger, and M. Plant (1985), "Unemployment rate dynamics and persistent unemployment under rational expectations", *American Economic Review*, 75, 614-637.

Dynarski, M. and S.M. Sheffrin (1990), "The behavior of unemployment durations over the cycle", *Review of Economics and Statistics*, 72, 350-356.

Flinn, C.J. and J.J. Heckman (1982), "Models for the analysis of labor force dynamics", in Basmann, R. and G. Rhodes (eds.), *Advances in Econometrics*, 1, 35-95.

FOREM (1995), Enquête sur les sorties d'inscription des demandeurs d'emploi, Service Etudes et Statistique, FOREM.

Gourieroux, C. and A. Monfort (1995), *Statistics and Econometric Models*, Cambridge University Press, Cambridge.

Hoel, P.G. (1971), *Introduction to mathematical statistics*, Fourth edition. New York: Wiley.

Imbens, G.W. and L.M. Lynch (1993), "Re-employment probabilities over the business cycle", NBER Working Paper, 4585.

Kalwij, A. S. (2001), "Individuals' Unemployment Durations over the Business Cycle", IZA Discussion Paper n°369, Institute for the Study of Labor, Bonn, Germany.

Lancaster, T. (1990), *The econometric analysis of transition data*, Cambridge University Press, Cambridge.

Layard, R., S.J. Nickell, and R. Jackman (1991), *Unemployment*, Oxford University Press, Oxford.

Merz, M. (1999), "Time series evidence of unemployment flows: The sample period matters", *Journal of Business and Economic Statistics*, 17 (3), 324-334.

Mignolet, M. et L. Vieslet (2000), "Les disparités interrégionales", dans *Que nous est-il arrivé? Un demi-siècle d'évolution de l'économie belge*, *Reflets et Perspectives de la vie économique*, De Boeck Université.

Mortensen, D.T. and C.A. Pissarides (1999), *Job reallocation, employment fluctuations and unemployment*, Handbook of Macroeconomics, J. Taylor and M. Woodford (eds.), Amsterdam: North-Holland.

ONEm (1992), Rapport annuel.

Rosholm, M. (2001), "Cyclical variations in unemployment duration", *Journal of Population Economics*, 14, 173-191.

Sider, H. (1985), "Unemployment duration and incidence: 1968-82", *American Economic Review*, 75, 461-472.

Sneessens, H.R., R. Nicolini, et F. Shadman-Metha (1999), "Innovation et chômage en région wallonne : aspects économiques", dans *Des idées et des hommes : pour construire l'avenir de la Wallonie et de Bruxelles*, Academia-Bruylant.

van den Berg, G.J. and B. van der Klaauw (2001), "Combining micro and macro unemployment duration data", *Journal of Econometrics*, 102, 271-309.

8 Figures and tables

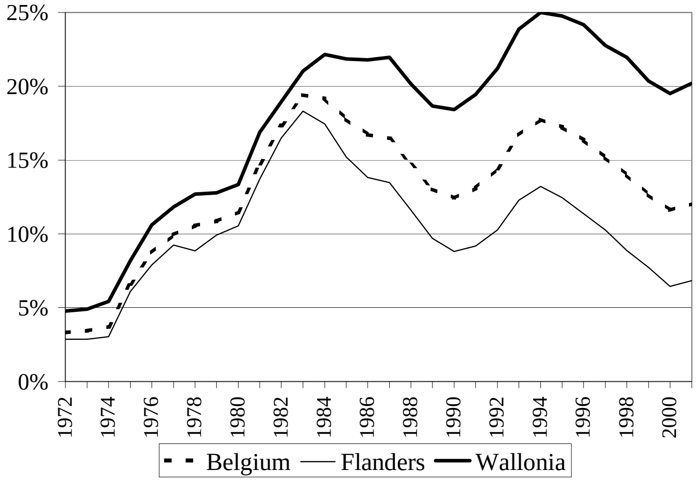

Figure 1: Unemployment rate in Belgium, Flanders and Wallonia

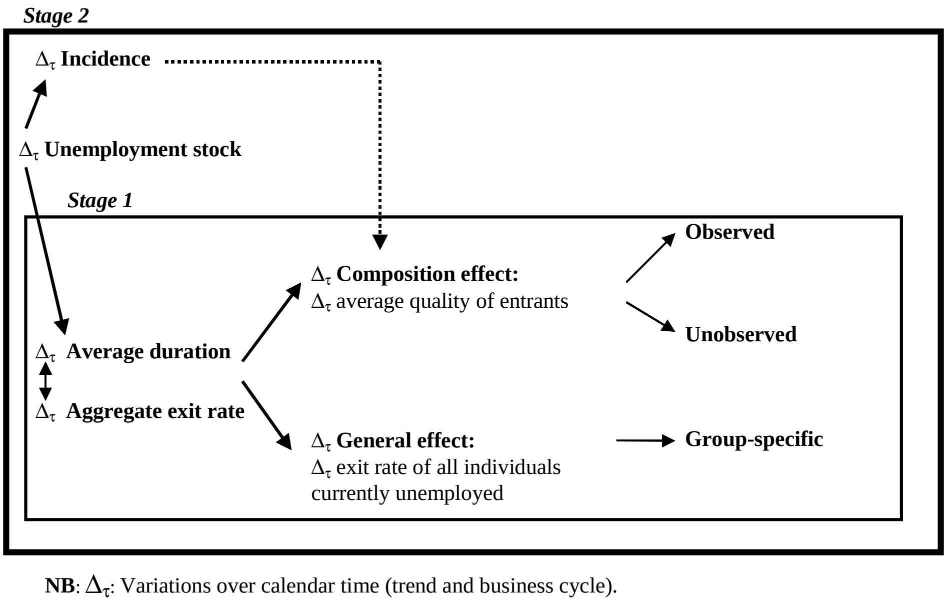

Figure 2: Full decomposition of the unemployment stock

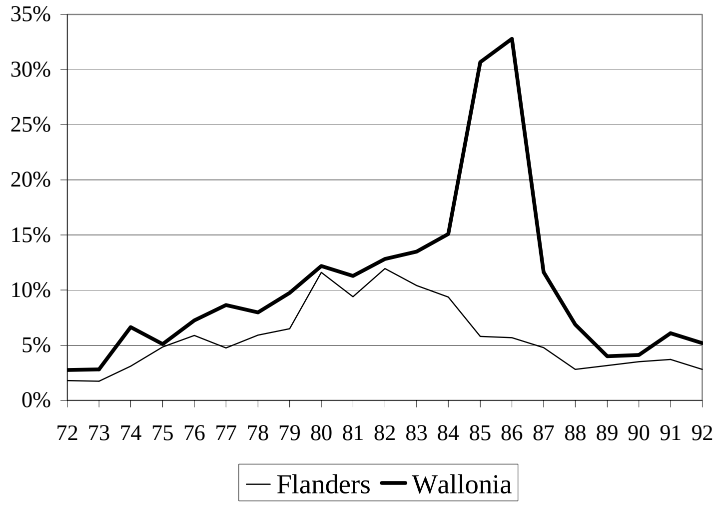

Figure 3: Proportion of entrants with changing value of their personal characteristics

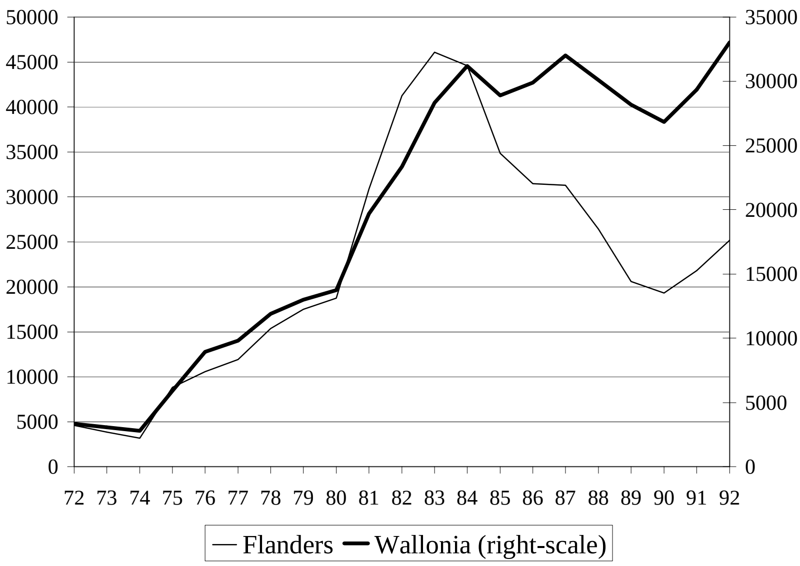

Figure 4: Total stock of unemployment in June (male aged 25-44)

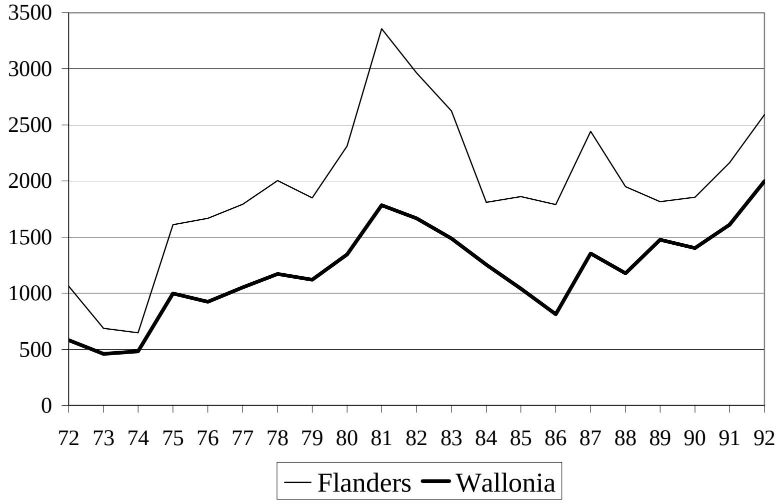

Figure 5: Inflow into unemployment in June

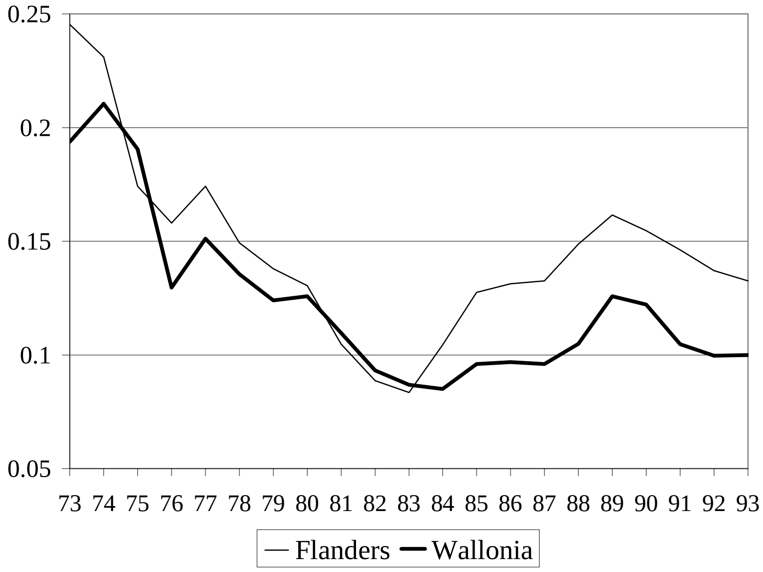

Figure 6: Empirical monthly hazard in the first duration interval

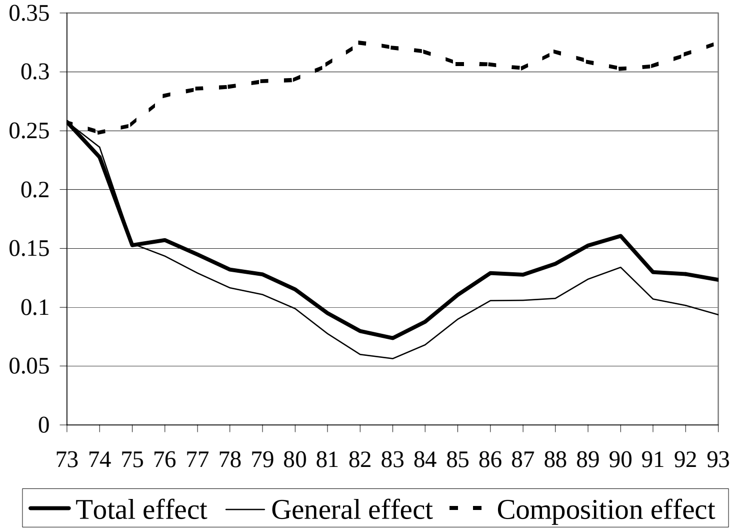

Figure 7: Decomposition of the monthly aggregate hazard (K=1) - Flanders

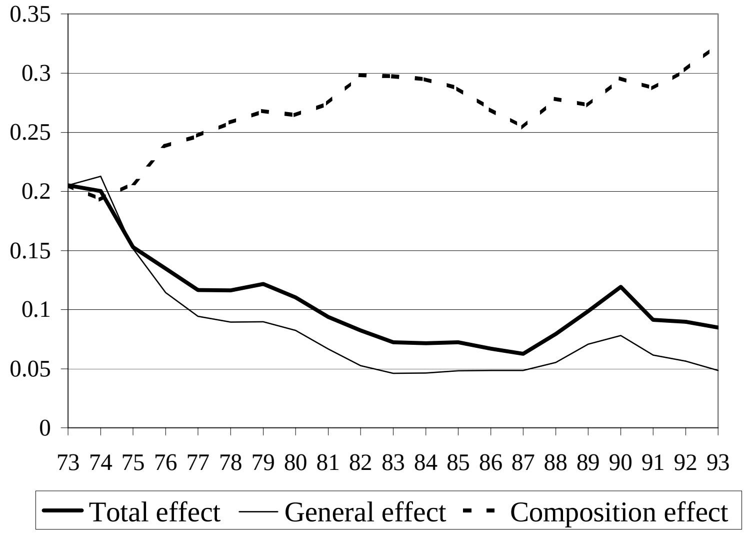

Figure 8: Decomposition of the monthly aggregate hazard (K=1) - Wallonia

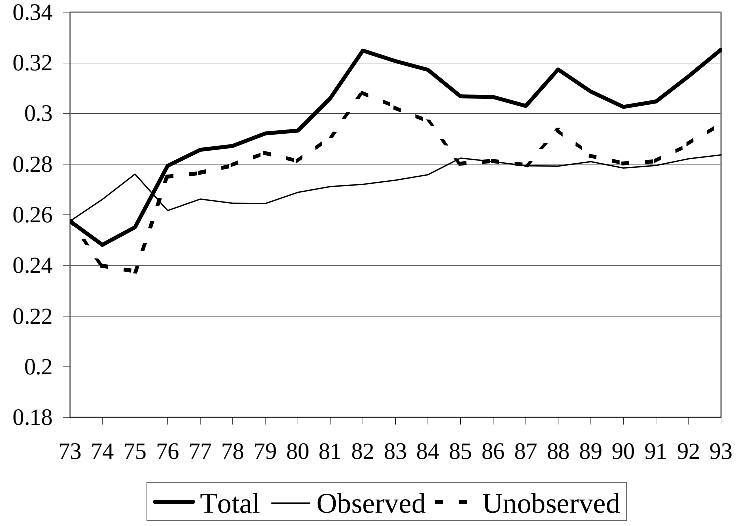

Figure 9: Composition effect - Flanders

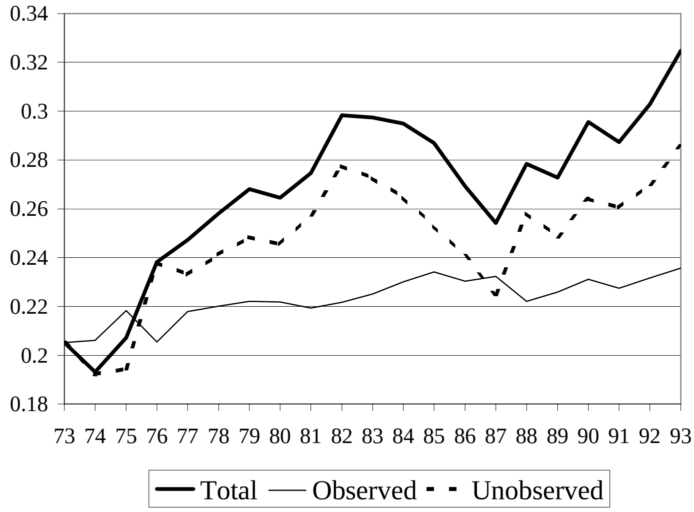

Figure 10: Composition effect - Wallonia

Table 1: Structural evolution of the monthly aggregate hazard over 1973-1993

	Total effect		General effect		Composition effect	
	Wallonia	Flanders	Wallonia	Flanders	Wallonia	Flanders
Total variation	-53%	-28%	-69%	-43%	43%	22%
Yearly variation	-3.7%	-1.6%	-5.9%	-2.8%	1.8%	1.0%

Table 2: Structural evolution of the composition effect over 1973-1993

	Total effect		Observed		Unobserved	
	Wallonia	Flanders	Wallonia	Flanders	Wallonia	Flanders
Total variation	43%	22%	12%	8%	28%	13%
Yearly variation	1.8%	1.0%	0.6%	0.4%	1.3%	0.6%

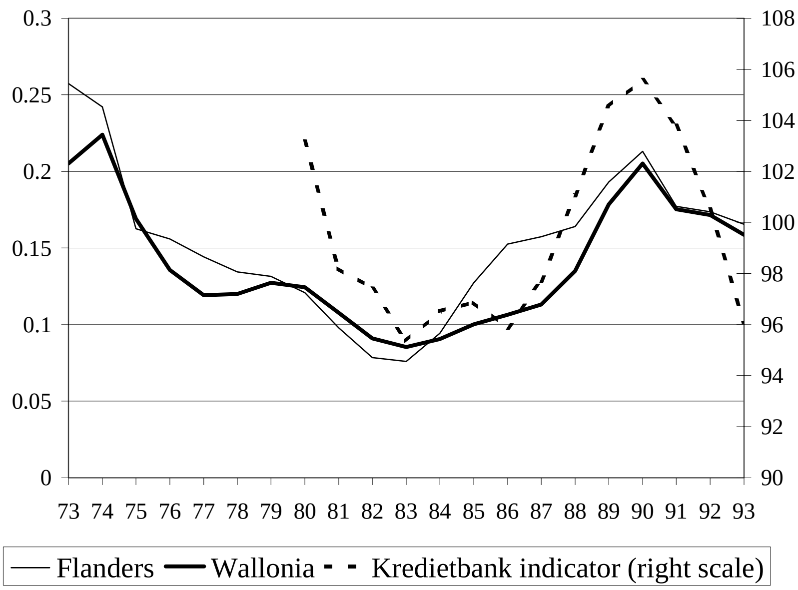

Figure 11: General effect over the cycle

Table 3: Structural evolution of the general effect over 1973-1993

		General effect	
		Wallonia	Flanders
MPH	Total variation	-69%	-42%
	Yearly variation	-5.8%	-2.7%
Low-educated*	Total variation	-70%	-46%
	Yearly variation	-6.1%	-3.0%
High-educated	Total variation	-56%	-32%
	Yearly variation	-4.1%	-1.9%

* Low-educated in Wallonia (resp. Flanders): Below higher education (resp. upper secondary).

Table 4: Marginal effect of each schooling degree on the hazard

	Wallonia			Flanders		
	1973	1993	MPH	1973	1993	MPH
Higher education	-1%	42%	25%	25%	25%	25%
Upper secondary	18%	18%	18%	-5%	19%	9%
Lower secondary	31%	31%	31%	7%	7%	7%

Table 5: Effect of living in other provinces than Brabant wallon in Wallonia (resp. Vlaams-Brabant in Flanders) on the hazard

Wallonia	1973	1993	MPH	Flanders	1973	1993	MPH
Hainaut	-18%	-29%	-25%	Antwerpen	5%	-20%	-11%
Liège	-17%	-17%	-17%	Limburg	-37%	-1%	-17%
Namur	1%	-18%	-11%	Oost-Vlanderen	-11%	-11%	-11%

Table 6: Evolution of salaried employment, by sector and sub-periods, in % of the initial regional employment

	Industry		Private services		Public services		Total	
	Wallonia	Flanders	Wallonia	Flanders	Wallonia	Flanders	Wallonia	Flanders
1984/73	-19.1%	-12.5%	3.1%	7.3%	7.2%	5.8%	-8.8%	0.6%
1995/84	-6.6%	-1.0%	10.5%	17.0%	3.7%	3.4%	7.6%	19.4%
1995/73	-25.7%	-13.5%	13.5%	24.3%	10.9%	9.2%	-1.3%	20.0%

Source: Sneessens and Shadman-Mehta (2000, p.117)

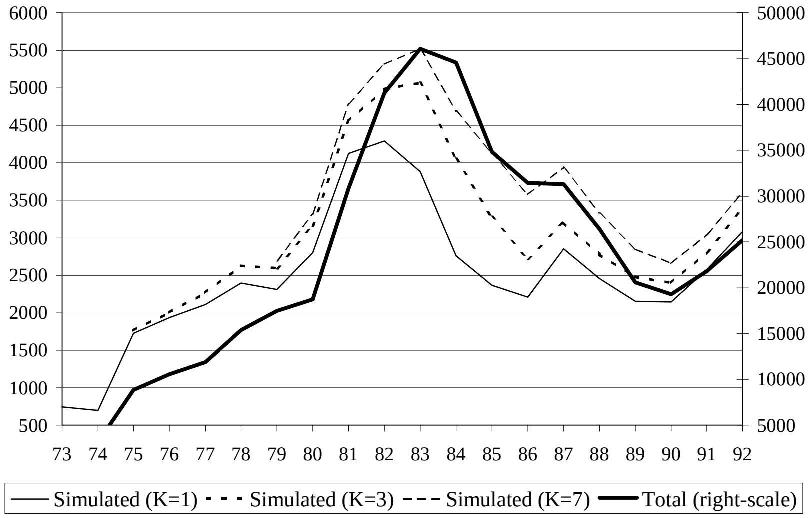

Figure 12: Stock of unemployment in June - Flanders

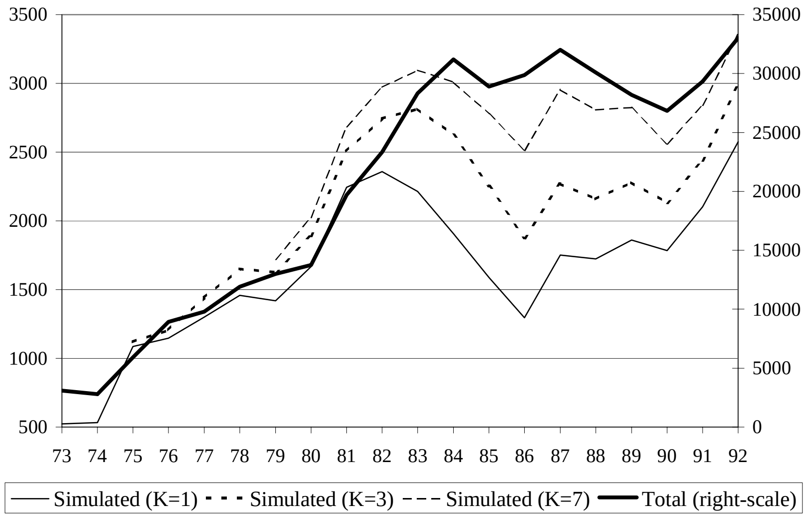

Figure 13: Stock of unemployment in June - Wallonia

9 Appendix

9.1 Appendix 1: Data correction

In this appendix, we describe the correction procedure applied to the non-parametric outflow probabilities, $\widehat{P}_{klx}$. The reason for correcting the data relates to their particular feature. The non-observability of the individual spells of unemployment, and the consecutive grouping of the data, induces two sources of spurious outflows: (1) Changes in time-varying characteristics during the unemployment spell and (2) temporary exits out of unemployment (of less than three months). In the sequel of this appendix, we first assume the absence of temporary exits and derive the appropriate correction to $\widehat{P}_{klx}$. We then analyse the consequences of temporary exits. An identification problem leads us to consider two alternative corrections of the data.

9.1.1 No temporary exits

Denote u_{kl}^{yx} the number of individuals with entry date l and characteristics x at the beginning of duration interval k, who had characteristics y at the beginning of duration $k-1$, where x and y are two vectors of characteristics which can be identical. Also denote $u_{kl}^{\cdot x} = \sum\limits_{y} u_{kl}^{yx}$, $u_{kl}^{y\cdot} = \sum\limits_{x} u_{kl}^{yx}$, and $u_{kl}^{\cdot\cdot} = \sum\limits_{x}\sum\limits_{y} u_{kl}^{yx} = \sum\limits_{x}\sum\limits_{y} u_{kl}^{xy}$.

The true number of individuals with entry date l and characteristics x at the start of duration interval k that leave unemployment during this interval is given by:

$$\widetilde{f}_{klx} = u_{kl}^{\cdot x} - u_{k+1l}^{x\cdot} \tag{a}$$

In our data set, we can only condition on the characteristics of the current state since we cannot track individual spells through time. In other words we do not observe $u_{k+1l}^{x\cdot}$ but only $u_{k+1l}^{\cdot x}$. The number of individuals with entry date l and characteristics x at the start of duration interval k that leave unemployment during this interval is therefore approximated by:

$$f_{klx} = u_{kl}^{\cdot x} - u_{k+1l}^{\cdot x} \tag{b}$$

This approximation is subject to two sources of bias which can possibly cancel out. First, an individual counted in $u_{kl}^{\cdot x}$ can move from x to y at the k^{th} duration interval without flowing out of unemployment. In that case, the number of outflows f_{klx} increases spuriously of one unit ($u_{k+1l}^{\cdot x}$ decreases of one unit). Second, an individual counted in $u_{kl}^{\cdot y}$ can move from y to x at the k^{th} duration interval without flowing out of unemployment. Then, the number of outflows f_{klx} decreases spuriously of one unit ($u_{k+1l}^{\cdot x}$ increases of one unit). The number of spurious outflows in the k^{th} duration interval conditional on entry date l and characteristics x at the start of duration interval k is given by:

$$
\begin{aligned}
f^S_{klx} &= f_{klx} - \widetilde{f}_{klx} = -u^{\cdot x}_{k+1l} + u^{x\cdot}_{k+1l} \\
&= -u^{xx}_{k+1l} - \sum_{y\neq x} u^{yx}_{k+1l} + u^{xx}_{k+1l} + \sum_{y\neq x} u^{xy}_{k+1l} \\
&= \sum_{y\neq x}(-u^{yx}_{k+1l} + u^{xy}_{k+1l}) \qquad \text{(c)}
\end{aligned}
$$

where u^{yx}_{k+1l} (resp. u^{xy}_{k+1l}) are the number of individuals that become of type x (resp. y) in the k^{th} duration interval conditional on being of another type at the beginning of this interval. The term u^{yx}_{k+1l} (resp. u^{xy}_{k+1l}) leads to an underestimation (resp. overestimation) of the non-parametric outflow probability, $\hat{P}_{klx}$.

Notice that at the aggregate level, there is no spurious outflows (by use of (c)):

$$
f^S_{kl\cdot} = \sum_x f^S_{klx} = \sum_x(-u^{\cdot x}_{k+1l} + u^{x\cdot}_{k+1l}) = -u^{\cdot\cdot}_{k+1l} + u^{\cdot\cdot}_{k+1l} = 0 \qquad \text{(d)}
$$

This amounts to saying that over- and underestimations cancel out at the aggregate level:

$$
\sum_{y\neq x}(- u^{y\cdot}_{k+1l} + u^{\cdot y}_{k+1l}) = 0 \Leftrightarrow \sum_{y\neq x} u^{y\cdot}_{k+1l} = \sum_{y\neq x} u^{\cdot y}_{k+1l} \qquad \text{(e)}
$$

The aggregate non-parametric outflow probability, $\hat{P}_{kl\cdot}$, is therefore unbiased:

$$
\hat{P}_{kl\cdot} \equiv \frac{\sum_x \widetilde{f}_{klx}}{\sum_x u^{\cdot x}_{kl}} = \frac{\sum_x f_{klx}}{\sum_x u^{\cdot x}_{kl}} \qquad \text{(f)}
$$

Finally note that we can only use observations with f_{klx} such that $0 \leq f_{klx} \leq u^{\cdot x}_{kl}$ to ensure that $0 \leq \widehat{P}_{klx} \equiv \frac{f_{klx}}{u^{\cdot x}_{kl}} \leq 1$. This implies that before calculating $\widehat{P}_{klx}$, we suppress all observations which do not satisfy the condition:

$$
u^{\cdot x}_{k+1l} \leq u^{\cdot x}_{kl} \qquad \text{(g)}
$$

Now suppose that cohorts are defined such that they are unique, i.e. $0 \leq u^{\cdot x}_{kl} \leq 1$. In that case, if $u^{\cdot x}_{k+1l} = u^{yx}_{k+1l} = 1$ $(y \neq x)$, then $u^{\cdot x}_{kl} = 0$. Since the vector x is unique and the individual had characteristics y in interval k $(u^{\cdot y}_{kl} = 1)$, there must be nobody with characteristics x in interval k. If we insert this in condition (g), we immediately see that this condition is always violated:

$$
u^{\cdot x}_{k+1l} = 1 > u^{\cdot x}_{kl} = 0 \qquad \text{(h)}
$$

The term $u^{\cdot x}_{k+1l}$ should lead to an underestimation of the non-parametric outflow probability $\hat{P}_{klx}$ (see (c)). However, this type of observation will never be considered in the construction of $\widehat{P}_{klx}$ since it violates condition (g).

On the other hand, if $u^{x\cdot}_{k+1l} = u^{xy}_{k+1l} = 1$ $(y \neq x)$, then $u^{\cdot x}_{kl} = 1$ and condition (g) can never be violated:

$$u^{\cdot x}_{k+1l} = 0 \leq u^{\cdot x}_{kl} = 1 \tag{i}$$

In that case, $u^{\cdot x}_{k+1l} = 0$ since the vector x is unique.

Consequently, when observations violating (g) are suppressed and cohorts are unique, spurious outflows always leads to an overestimation of the outflow (see (c)):

$$f^S_{klx} = \sum_{y \neq x} u^{xy}_{k+1l} \geq 0 \tag{j}$$

Under the assumption of uniqueness, the number of spurious outflows in cohort l at duration interval k is therefore given by:

$$f^S_{kl.} = \sum_{y \neq x} u^{\cdot y}_{k+1l} = \sum_{y \neq x} u^{y\cdot}_{k+1l} \tag{k}$$

where the last equality follows from (e). The number of spurious outflows is simply the number of transitions in cohort l at interval k from some y to x $(y \neq x)$. This figure is known by our assumption: It is the number of individuals who violate condition (g):

$$f^S_{kl.} = \sum_{x} u^{\cdot x}_{k+1l} \text{ for all observations s.t. } u^{\cdot x}_{kl} < u^{\cdot x}_{k+1l} \tag{l}$$

We now assume that the number of spurious exits in cohort (l, x) during the k^{th} interval is proportional to this number with a factor of proportionality $(\frac{f_{klx}}{f_{kl\cdot}})$[41]:

$$\widehat{f}^S_{klx} = \left(\frac{f_{klx}}{f_{kl\cdot}}\right) f^S_{kl\cdot} \tag{m}$$

This assumption of proportionality is arbitrary. Imagine for example that a majority of changes occur almost systematically between two levels of education. In that case, the proportional form (m) would clearly lead to large biases in the estimate of the true f^S_{klx}. We have however no reason to think that changes follow systematic patterns. In particular, we know that they cannot be attributed mostly to one variable. The proportionality assumption appears as the best alternative given the lack of better information.

The corrected non-parametric outflow probability, $\hat{P}^*_{klx}$, writes therefore as follows:

[41] The factor of proportionality depends on the fraction of observed transitions $\frac{f_{klx}}{f_{kl.}}$ rather than on $\frac{u_{klx}}{u_{kl.}}$ since we need only correct when there are effectively transitions in x.

$$\hat{P}^*_{klx} = \frac{f_{klx} - \widehat{f}^S_{klx}}{u^{\cdot x}_{kl}} = (1 - \frac{f^S_{kl\cdot}}{f_{kl\cdot}})\frac{f_{klx}}{u^{\cdot x}_{kl}} = (1 - \frac{f^S_{kl\cdot}}{f_{kl\cdot}})\widehat{P}_{klx} \tag{n}$$

If some cohorts are not unique, spurious outflows may lead to an underestimation of the outflows ($u^{yx}_{k+1l} > 0$ with condition (g) not violated) and $f^S_{kl.}$ in (l) will not give (as a consequence) all the transitions in cohort l at interval k from some y to x ($y \neq x$). However, we will assume that if $u^{yx}_{k+1l} > 0$, condition (g) is always violated. We believe that this assumption is justified since a large majority of the cohorts are unique in our data set (93% at the entry into unemployment).

Departing from the uniqueness assumption requires however to adapt expression (l) in the following way:

$$f^S_{kl.} = \sum_x \left(u^{\cdot x}_{k+1l} - u^{\cdot x}_{kl}\right) \text{ for all observations s.t. } u^{\cdot x}_{kl} < u^{\cdot x}_{k+1l} \tag{o}$$

since if $u^{yx}_{k+1l} > 0$, $u^{\cdot x}_{kl}$ is no more necessarily equals to 0 (although $u^{\cdot x}_{kl} < u^{yx}_{k+1l}$ by assumption).

9.1.2 Temporary exits

In Belgium, unemployed people who exit unemployment for a period of less than three months keep the same starting date of their unemployment spell at their return. The purpose of this rule is to maintain for the unemployed person the right to benefit from measures which are conditional to a given elapsed duration into unemployment[42]. Given that the census of the unemployed workers occurs on June 30, it happens that some unemployed are not counted at a given year because they are in a training scheme or have a temporary job that day. However they are still unemployed and counted as such in a following census.

Temporary exits introduce an additional bias in our measure of the number of individuals in cohort (l, x) that leave unemployment during the k duration interval, f_{klx} given in (b). If an individual counted in $u^{\cdot x}_{kl}$ exits unemployment temporarily around the start of the $k+1$ duration interval and returns within the next *three months*, then the number of outflows f_{klx} increases spuriously with one unit (since $u^{\cdot x}_{k+1l}$ spuriously decreases of one unit). On the other hand, if an individual counted in $u^{\cdot x}_{k+1l}$ was temporarily out of unemployment around the start of the k duration interval and returned within the next three months, the number of outflows f_{klx} increases spuriously with one unit (since $u^{\cdot x}_{kl}$ spuriously increased with one unit).

Suppose again that cohorts are defined such that they are unique, i.e. $0 \leq u^{\cdot x}_{kl} \leq 1$. In that case, a temporary exit around the $k+1$ duration interval always induces an overestimation of the outflows at the k^{th} interval: $u^{\cdot x}_{kl} = 1$ and $u^{\cdot x}_{k+1l} = 0$. In other respects, a temporary exit around the k duration interval followed by a *sufficiently long*

[42] Like the right to participate to active job-search workshop, vocational training and in-depth personal counseling.

unemployment spell always leads to a violation of condition (g): $u_{kl}^{\cdot x} = 0$ and $u_{k+1l}^{\cdot x} = 1$. We can therefore identify which $u_{kl}^{\cdot x} = 0$ corresponds to a temporary exit and correct appropriately for it. Of course, if an individual really flows out of unemployment during the year that follows his temporary exit, we will not be able to spot this temporary exit since condition (g) is not violated.

Formally, if condition (g) is violated at the k duration interval and the following condition holds simultaneously:

$$\exists j \in \{1, ..., k-1\} \mid u_{jl}^{\cdot x} = 1 \tag{p}$$

then $u_{kl}^{\cdot x} = 0$ is a temporary exit. If $j < k-1$, the individual was temporary out of unemployment on June 30 several years consecutively, and all $u_{il}^{\cdot x} = 0$ with $i \in \{j+1, ..., k-1\}$ are temporary exits. Notice that under the uniqueness assumption, the problem of temporary exits differs from the one of changing variables by condition (p). For, this condition is never satisfied when condition (g) is violated in the latter case. We simply correct the identified spurious risk sets by setting them to 1. This correction is done in the first place. We then calculate the number of spurious outflows due to changing characteristics (see (l)) and apply the correction (n) to the non-parametric outflow probability, $\widehat{P}_{klx}$, by suppressing beforehand observations which violate (g).

When some cohorts are not unique, temporary exits are observationally equivalent to a problem of changing characteristics. Notice first then that departing from the uniqueness assumption requires to adapt condition (p) to the case of several individuals per cohort:

$$\exists j \in \{1, ..., k-1\} \mid u_{jl}^{\cdot x} > u_{kl}^{\cdot x} \tag{q}$$

A violation of condition (g) at the k duration interval together with condition (q) satisfied are now necessary to have a case of temporary exit(s). If we neglect the problem of changing variables, the correction procedure is the following. First select the largest j such that condition (q) is satisfied. The number $u_{jl}^{\cdot x} - u_{kl}^{\cdot x}$ is then the number of individuals who exit temporarily. If $j = k-1$, we correct the risk sets $u_{kl}^{\cdot x}$ as follows: $(u_{kl}^{\cdot x})^* = u_{kl}^{\cdot x} + Min\left(u_{jl}^{\cdot x} - u_{kl}^{\cdot x}, u_{k+1l}^{\cdot x} - u_{kl}^{\cdot x}\right)$. If $j < k-1$, the individual(s) was (were) temporary out of unemployment on June 30 several years consecutively, and we correct all risk sets $u_{il}^{\cdot x}$ with $i \in \{j+1, ..., k\}$ as follows: $(u_{il}^{\cdot x})^* = u_{il}^{\cdot x} + Min\left(u_{jl}^{\cdot x} - u_{il}^{\cdot x}, u_{k+1l}^{\cdot x} - u_{il}^{\cdot x}\right)$. This procedure has to be repeated until it is no more possible to find a j such that condition (q) is satisfied. All the temporary exits on June 30 from the consecutive unemployment censuses are then corrected.

However, a violation of condition (g) alongside with condition (q) fulfilled can be now be interpreted as follows: An individual entered into unemployment at l moves from y to x in the k^{th} duration interval while there was another individual in cohort (l, x) at some duration $j < k$ who flows out of unemployment since then. However, we will assume that if condition (g) is violated and condition (q) satisfied, we have a case of temporary exit. We believe that this assumption is justified for the following reason. Since a large majority of the cohorts are unique in our data set (93% at entry), most of the spurious outflows due to changing characteristics will be picked out in cohorts where both conditions (g) and (p) are violated. In our data set, this concerns 7% of the total number of cohorts (and the same share of the total number of entrants). And from preliminary analysis on the data base, we approximate the number of individuals with changing variables to around 8% of the total number of entrants. In other respects, around 3% of the total number of cohorts satisfies the conditions required to pick out temporary exits. We think that this number is too high to be, in a large majority, attributed to a problem of time-varying characteristics.

We therefore correct the spurious risk sets which can be attributed to temporary exits in the way explained supra. This correction is done in the first place. We then calculate the number of spurious outflows due to changing characteristics (see (o)) and apply the correction (n) to the non-parametric outflow probability, $\widehat{P}_{klx}$, by suppressing beforehand observations which violate (g). As a sensitivity analysis, we construct another data set in which we assume the absence of temporary exits. This means that we consider the 3% of the total number of cohorts that violates conditions (g) and satisfies (q) as cases of time-varying characteristics. So we only follow the correction procedure of the first section of this appendix. We estimate the mixed proportional hazard model (14) based on these two alternative data sets. The 95%-confidence intervals of the estimated parameters in the two models overlap. For the reason outlined above, we choose to use the first data set.

9.2 Appendix 2: Variance decomposition

In this appendix, we derive the part of the total variation in the aggregate hazard, $\bar{h}_{1l}$ in (20), that is accounted for by a structural and a cyclical variation. Each type of variation is decomposed into the part explained by a general effect, an observed and unobserved composition effect. To that purpose, we first introduce the following notations: $y_l = \ln\left(\bar{h}_{1l}\right)$, $x_{1l} = \ln\left(\bar{h}_{1l}^{G}\right)$, $x_{2l} = \ln\left(\bar{h}_{1l}^{OC}\right)$ and $x_{3l} = \ln\left(\bar{h}_{1l}^{UC}\right)$. We then regress the de-trended logarithm of the aggregate hazard on the de-trended logarithm of the three synthetic series:

$$y_l - \hat{b}\left(l - \bar{l}\right) = a + c_1 e_{1l} + c_2 e_{2l} + c_3 e_{3l} + \varepsilon_l$$
$$with \;\; e_{1l} = x_{1l} - \hat{\beta}_1 l;\; e_{2l} = x_{2l} - \hat{\beta}_2 l;\; e_{3l} = x_{3l} - \hat{\beta}_3 l;\; \hat{b} = \hat{\beta}_1 + \hat{\beta}_2 + \hat{\beta}_3$$

where $\bar{l}$ is the average value of l. In this regression, we assume that the sum of the exponential trends in the synthetic series is equal to the exponential trend in the aggregate hazard. The total variation in the aggregate hazard then decomposes as follows:

$$\sum_l (\hat{y}_l - \hat{a})^2 =$$

$$\left\{\begin{array}{c} \underbrace{\hat{\beta}_1 \hat{b} \sum_l \left(l - \bar{l}\right)^2}_{G} + \underbrace{\hat{\beta}_2 \hat{b} \sum_l \left(l - \bar{l}\right)^2}_{OC} + \underbrace{\hat{\beta}_3 \hat{b} \sum_l \left(l - \bar{l}\right)^2}_{UC} \\ + \hat{c}_1 \hat{b} \sum_l \left(l - \bar{l}\right) e_{1l} + \hat{c}_2 \hat{b} \sum_l \left(l - \bar{l}\right) e_{2l} + \hat{c}_3 \hat{b} \sum_l \left(l - \bar{l}\right) e_{3l} \end{array}\right\} Trend$$

$$\left\{\begin{array}{c} \underbrace{+\hat{c}_1^2 \sum_l (e_{1l})^2 + \hat{c}_1 \hat{c}_2 \sum_l e_{1l} e_{2l} + \hat{c}_1 \hat{c}_3 \sum_l e_{1l} e_{3l}}_{G} \\ \underbrace{+\hat{c}_2^2 \sum_l (e_{2l})^2 + \hat{c}_1 \hat{c}_2 \sum_l e_{1l} e_{2l} + \hat{c}_2 \hat{c}_3 \sum_l e_{2l} e_{3l}}_{OC} \\ \underbrace{+\hat{c}_3^2 \sum_l (e_{3l})^2 + \hat{c}_1 \hat{c}_3 \sum_l e_{1l} e_{3l} + \hat{c}_2 \hat{c}_3 \sum_l e_{2l} e_{3l}}_{UC} \\ + \hat{c}_1 \hat{b} \sum_l \left(l - \bar{l}\right) e_{1l} + \hat{c}_2 \hat{b} \sum_l \left(l - \bar{l}\right) e_{2l} + \hat{c}_3 \hat{b} \sum_l \left(l - \bar{l}\right) e_{3l} \end{array}\right\} Cycle$$

where $Trend$ and $Cycle$ refer to the part of the total variation in y_l that is accounted for a structural and a cyclical evolution respectively, and where G, OC and UC refer to the part explained by the general, observed composition and unobserved composition effects in each type of evolution. The total variation of y_l in the structural and cyclical dimensions is simply the sum of G, OC and UC in each of these dimensions. The ratio between each component and its corresponding total variation measures the proportion of the variance in the aggregate hazard accounted for each component.

Some components of the variance of y_l could, however, be negative, reflecting either a negative estimated parameter ($\hat{\beta}_i$) in the trend decomposition or a negative covariance between two effects in the trend or cyclical decomposition. Intuitively, a negative (resp. positive) covariance between two components reduces (resp. amplifies) the variability in the aggregate hazard since both components moves in the opposite (resp. same) direction. If one part of the variance of y_l is negative, we assume that the total variability in y_l is a weighted sum of each of its components, in which the weights sum to one. These weights measure the proportion of the variance of y_l accounted for by each component.

9.3 Appendix 3: Figures

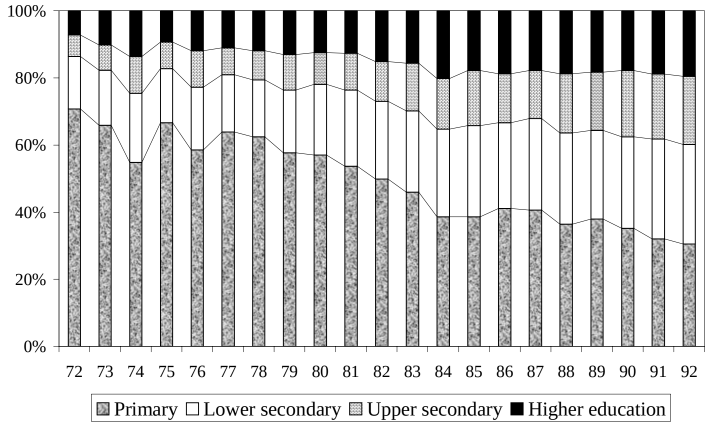

Figure 14: Distribution of the level of education at entry - Flanders

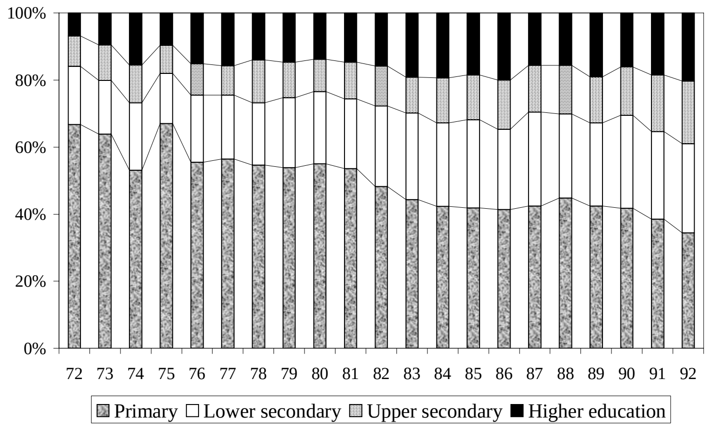

Figure 15: Distribution of the level of education at entry - Wallonia

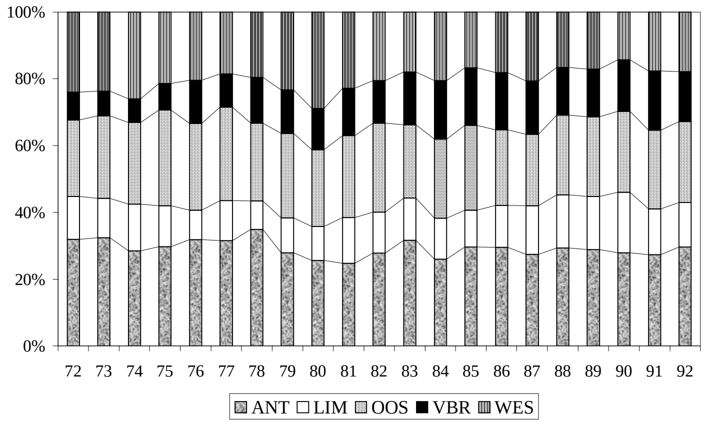

Figure 16: Distribution of the province of residence at entry - Flanders

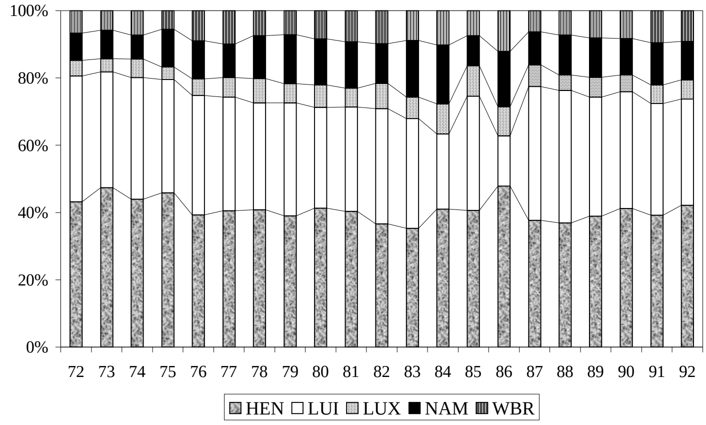

Figure 17: Distribution of the province of residence at entry - Wallonia

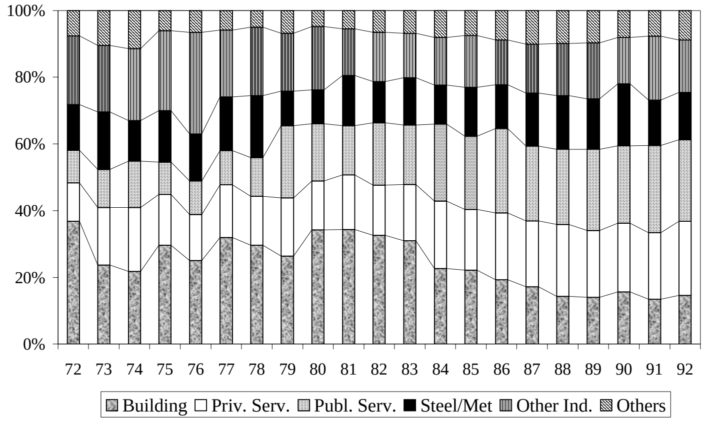

Figure 18: Distribution of the sector at entry - Flanders

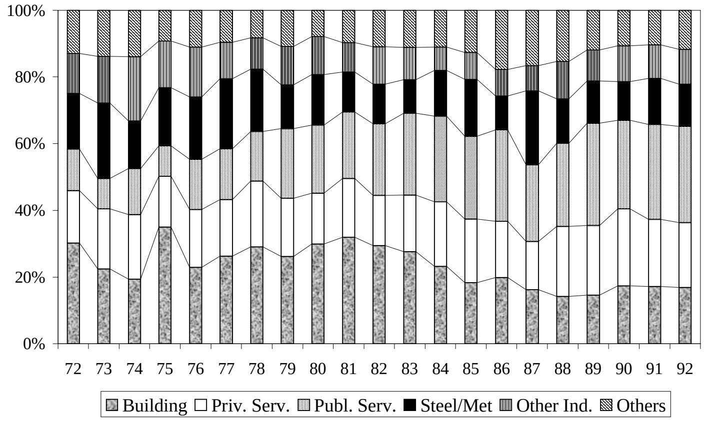

Figure 19: Distribution of the sector at entry - Wallonia

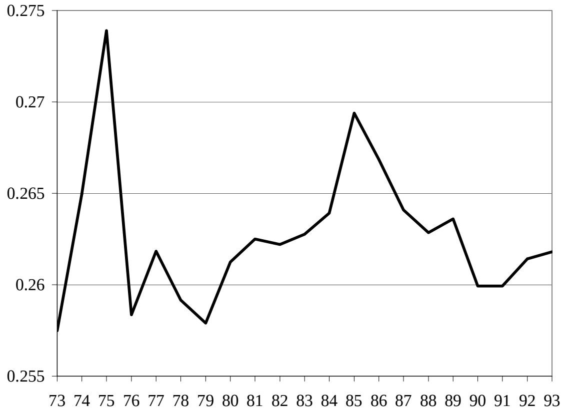

Figure 20: Observed composition effect over the cycle - Flanders

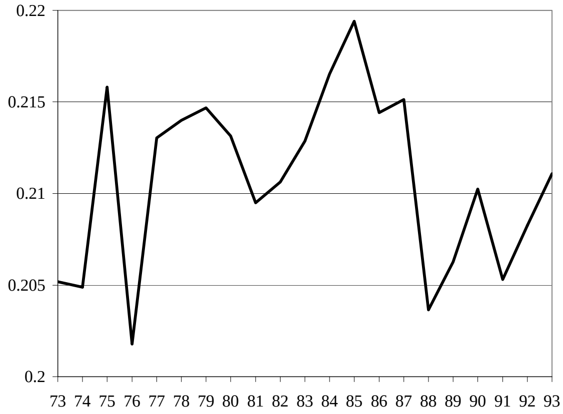

Figure 21: Observed composition effect over the cycle - Wallonia

9.4 Appendix 4: Estimation results

Table 7: Hazard model estimates - Mixed Proportional Hazard Model

Variables	Flanders	SD	Wallonia	SD
Constant term*	**3.42**	0.16	**2.25**	0.18
1. Covariate effects				
Age (in months - log)	**-0.87**	0.04	**-0.64**	0.05
Education				
Junior high school	**0.22**	0.02	**0.27**	0.02
Completed high school	**0.31**	0.02	**0.43**	0.03
Higher education	**0.37**	0.02	**0.65**	0.03
Sub-region of living				
Antwerpen	**-0.08**	0.02		
Limburg	**-0.19**	0.02		
Oost-Vlaanderen	**-0.12**	0.02		
West-Vlaanderen	0.04	0.02		
Liège			**-0.19**	0.03
Luxembourg			0.02	0.04
Namur			**-0.11**	0.03
Hainaut			**-0.29**	0.03
Sector of previous employment				
Agriculture	**0.41**	0.06	**0.18**	0.08
Iron-steel industry and mining	**0.13**	0.06	**-0.14**	0.06
Metal manufacture	**0.08**	0.04	0.07	0.05
Manufacture of non-metallic products	0.01	0.07	**-0.16**	0.08
Manufacture of consumption and investment goods	0.07	0.04	0.01	0.06
Other industries	**0.18**	0.06	0.06	0.09
Energy and water production	-0.12	0.12	-0.11	0.13
Building	**0.12**	0.04	**0.14**	0.05
Private services	**0.19**	0.04	**0.24**	0.05
Public services	0.02	0.04	0.04	0.05
Others	**0.20**	0.06	**0.52**	0.07
No previous employment	**0.14**	0.05	0.05	0.07

*Reference individual in Flanders (resp. Wallonia): Elementary school, Vlaams-Brabant (resp. Brabant wallon), Chemical industry, $v = 1$.

Note: In **bold**, estimates not significantly different from zero at 5%.

(Table continued on next page)

Table 7: (Continued)

2. Calendar time at outflow				
α_1	**-0.42**	0.02	**-0.70**	0.03
α_2	**0.49**	0.02	**0.46**	0.02
α_3	**-0.23**	0.01	**-0.12**	0.02
α_4	**-0.08**	0.01	**-0.10**	0.02
α_5	-0.01	0.01	**-0.07**	0.01
α_6	**0.12**	0.01	0.00	0.01
α_7	**0.03**	0.01	**0.11**	0.01
α_8	**-0.11**	0.01	**-0.05**	0.01
α_9	**0.06**	0.01	0.02	0.01
α_{10}	**-0.04**	0.02	-0.03	0.02
α_{11}	-0.02	0.01	**-0.06**	0.01
α_{12}	**-0.13**	0.01	**-0.08**	0.01
3. Duration (in years)				
2 $(\gamma_2 - c)$	**-0.50**	0.03	**-0.30**	0.03
3 $(\gamma_3 - c)$	**-0.82**	0.04	**-0.67**	0.05
4 $(\gamma_4 - c)^{**}$	**-1.18**	0.05	**-1.05**	0.06
4. Variance of the Gamma at *l* =0				
η_c	**-1.84**	0.14	**-1.73**	0.16
$\sigma^2_0=\exp(\eta_c)$	**0.16**	0.02	**0.18**	0.03
5. Calendar time at inflow				
η_1	**0.20**	0.04	**0.35**	0.05
Number of estimated parameters	38		38	
Number of observations	63793		45987	
Weighted sum of squared residuals	52089.6		35944.5	
Adjusted number of parameters (after regrouping)	22		22	
Adjusted number of observations (after regrouping)	420		420	
Adjusted weighted sum of squared residuals (after regrouping)	541.66		544.76	
P-value of the goodness-of-fit test	0.00		0.00	

** $\gamma_{21} = ... = \gamma_4$

Table 8: Hazard model estimates - Mixed Non-Proportional Hazard Model - Flanders

Variables	Benchmark	SD	Restrictions	SD
Constant term*	**3.46**	0.16	**3.46**	0.16
1. Covariate effects (time dependent)				
Age (in months - log)	**-0.88**	0.04	**-0.88**	0.04
Education				
Elementary school				
β1	**-0.36**	0.08	**-0.44**	0.03
β2	**0.52**	0.03	**0.49**	0.02
β3	**-0.23**	0.02	**-0.23**	0.01
β4	**-0.08**	0.02	**-0.08**	0.01
β5	-0.02	0.01	-0.01	0.01
β6	**0.13**	0.01	**0.12**	0.01
β7	**0.05**	0.01	**0.03**	0.01
β8	**-0.10**	0.02	**-0.11**	0.01
β9	**0.08**	0.02	**0.06**	0.01
β10	**-0.05**	0.02	**-0.04**	0.02
β11	-0.02	0.02	-0.02	0.01
β12	**-0.15**	0.02	**-0.13**	0.01
Junior high school				
β0	**0.19**	0.03	**0.23**	0.02
β1	**-0.30**	0.09	**-0.44**	0.03
β2	**0.45**	0.05	**0.49**	0.02
β3	**-0.19**	0.03	**-0.23**	0.01
β4	**-0.08**	0.03	**-0.08**	0.01
β5	0.00	0.03	-0.01	0.01
β6	**0.10**	0.03	**0.12**	0.01
β7	0.02	0.03	**0.03**	0.01
β8	**-0.12**	0.03	**-0.11**	0.01
β9	0.02	0.03	**0.06**	0.01
β10	-0.02	0.04	**-0.04**	0.02
β11	-0.05	0.03	-0.02	0.01
β12	**-0.10**	0.03	**-0.13**	0.01

*Reference individual: Elementary school, Vlaams-Brabant, Chemical industry, $v = 1$.

Note: In **bold**, estimates not significantly different from zero at 5%. In grey, restricted parameters.

(Table continued on next page)

Table 8: (Continued)

Completed high school				
β0	**0.27**	0.05	**0.29**	0.02
β1	**-0.29**	0.11	**-0.33**	0.03
β2	**0.45**	0.07	**0.49**	0.02
β3	**-0.24**	0.05	**-0.23**	0.01
β4	-0.07	0.05	**-0.08**	0.01
β5	0.01	0.04	-0.01	0.01
β6	0.03	0.04	**0.12**	0.01
β7	**0.10**	0.04	**0.03**	0.01
β8	**-0.20**	0.05	**-0.11**	0.01
β9	**0.10**	0.05	**0.06**	0.01
β10	-0.03	0.05	**-0.04**	0.02
β11	-0.01	0.04	-0.02	0.01
β12	**-0.09**	0.04	**-0.13**	0.01
Higher education				
β0	**0.30**	0.04	**0.35**	0.02
β1	**-0.18**	0.10	**-0.33**	0.03
β2	**0.44**	0.06	**0.49**	0.02
β3	**-0.24**	0.05	**-0.23**	0.01
β4	-0.08	0.05	**-0.08**	0.01
β5	-0.01	0.04	-0.01	0.01
β6	**0.14**	0.04	**0.12**	0.01
β7	-0.02	0.04	**0.03**	0.01
β8	-0.04	0.04	**-0.11**	0.01
β9	0.03	0.04	**0.06**	0.01
β10	-0.02	0.04	**-0.04**	0.02
β11	-0.01	0.03	-0.02	0.01
β12	**-0.11**	0.04	**-0.13**	0.01
Sub-region of living				
Antwerpen (β0)	**-0.05**	0.02	**-0.06**	0.02
(β'1–β1)	**-0.15**	0.04	**-0.12**	0.03
Limburg (β0)	**-0.22**	0.02	**-0.23**	0.02
(β'1–β1)	**0.15**	0.04	**0.23**	0.03
Oost-Vlaanderen (β0)	**-0.11**	0.02	**-0.12**	0.02
(β'1–β1)	-0.07	0.04	0.00	
West-Vlaanderen (β0)	**0.05**	0.02	0.04	0.02
(β'1–β1)	-0.04	0.04	0.00	

(Table continued on next page)

Table 8: (Continued)

Sector of previous employment				
Agriculture ($\beta 0$)	**0.42**	0.06	**0.41**	0.06
($\beta'1-\beta 1$)	-0.05	0.10	0.00	
Iron-steel industry and mining ($\beta 0$)	0.01	0.06	0.10	0.06
($\beta'1-\beta 1$)	**0.29**	0.11	0.00	
Metal manufacture ($\beta 0$)	**0.09**	0.04	0.08	0.04
($\beta'1-\beta 1$)	-0.04	0.07	0.00	
Manufacture of non-metallic products ($\beta 0$)	0.01	0.07	0.00	0.07
($\beta'1-\beta 1$)	-0.20	0.13	0.00	
Manufacture of consumption and investment goods ($\beta 0$)	0.08	0.04	0.07	0.04
($\beta'1-\beta 1$)	-0.05	0.07	0.00	
Other industries ($\beta 0$)	**0.15**	0.06	**0.15**	0.06
($\beta'1-\beta 1$)	**-0.32**	0.11	0.00	
Energy and water production ($\beta 0$)	-0.12	0.12	-0.13	0.12
($\beta'1-\beta 1$)	-0.04	0.07	0.00	
Building ($\beta 0$)	**0.13**	0.04	**0.12**	0.04
($\beta'1-\beta 1$)	-0.07	0.07	0.00	
Private services ($\beta 0$)	**0.20**	0.04	**0.19**	0.04
($\beta'1-\beta 1$)	-0.06	0.07	0.00	
Public services ($\beta 0$)	0.02	0.04	0.02	0.04
($\beta'1-\beta 1$)	-0.04	0.07	0.00	
Others ($\beta 0$)	**0.20**	0.06	**0.21**	0.06
($\beta'1-\beta 1$)	0.01	0.11	0.00	
No previous employment ($\beta 0$)	**0.14**	0.06	**0.13**	0.05
($\beta'1-\beta 1$)	-0.06	0.10	0.00	
2. Duration (in years)				
2 (γ_2-c)	**-0.49**	0.03	**-0.49**	0.03
3 (γ_3-c)	**-0.81**	0.04	**-0.81**	0.04
4 (γ_4-c)**	**-1.15**	0.06	**-1.16**	0.05
3. Variance of the Gamma at *l* =0				
η_c	**-1.81**	0.14	**-1.83**	0.14
$\sigma^2_0=\exp(\eta_c)$	**0.16**	0.02	**0.16**	0.02
4. Calendar time at inflow				
η_1	**0.20**	0.04	**0.20**	0.04
Number of estimated parameters	89		41	
Number of observations	63793		45987	
Weighted sum of squared residuals	51821.0		51988.8	
Adjusted number of parameters (after regrouping)	62		25	
Adjusted number of observations (after regrouping)	420		420	
Adjusted weighted sum of squared residuals (after regrouping)	425.67		472.39	
P-value of the goodness-of-fit test	0.008		0.004	
P-value of the restrictions test			0.13	

** $\gamma_{21} = ... = \gamma_4$

Table 9: Hazard model estimates - Mixed Non-Proportional Hazard Model - Wallonia

Variables	Benchmark	SD	Restrictions	SD
Constant term*	**2.25**	0.18	**2.27**	0.18
1. Covariate effects (time dependent)				
Age (in months - log)	**-0.64**	0.05	**-0.65**	0.05
Education				
Elementary school				
β1	**-0.57**	0.10	**-0.68**	0.04
β2	**0.47**	0.03	**0.45**	0.02
β3	**-0.11**	0.02	**-0.12**	0.02
β4	**-0.07**	0.02	**-0.10**	0.02
β5	**-0.07**	0.02	**-0.07**	0.01
β6	0.03	0.02	0.00	0.01
β7	**0.11**	0.02	**0.11**	0.01
β8	**-0.08**	0.02	**-0.05**	0.01
β9	0.01	0.02	0.02	0.01
β10	-0.02	0.02	-0.03	0.02
β11	**-0.07**	0.02	**-0.06**	0.01
β12	**-0.08**	0.02	**-0.07**	0.01
Junior high school				
β0	**0.26**	0.04	**0.27**	0.02
β1	**-0.57**	0.11	**-0.68**	0.04
β2	**0.42**	0.05	**0.45**	0.02
β3	**-0.10**	0.04	**-0.12**	0.02
β4	**-0.12**	0.04	**-0.10**	0.02
β5	-0.05	0.03	**-0.07**	0.01
β6	-0.02	0.03	0.00	0.01
β7	**0.13**	0.03	**0.11**	0.01
β8	-0.06	0.03	**-0.05**	0.01
β9	-0.01	0.03	0.02	0.01
β10	-0.04	0.04	-0.03	0.02
β11	**-0.09**	0.03	**-0.06**	0.01
β12	**-0.07**	0.03	**-0.07**	0.01

*Reference individual: Elementary school, Brabant wallon, Chemical industry, $v = 1$.

Note: In **bold**, estimates not significantly different from zero at 5%. In grey, restricted parameters.

(Table continued on next page)

Table 9: (Continued)

Completed high school					
β0		**0.41**	0.05	**0.44**	0.03
β1		**-0.60**	0.12	**-0.68**	0.04
β2		**0.41**	0.07	**0.45**	0.02
β3		**-0.15**	0.06	**-0.12**	0.02
β4		-0.09	0.05	**-0.10**	0.02
β5		-0.08	0.04	**-0.07**	0.01
β6		-0.05	0.04	0.00	0.01
β7		**0.12**	0.04	**0.11**	0.01
β8		0.00	0.05	**-0.05**	0.01
β9		0.02	0.05	0.02	0.01
β10		0.00	0.06	-0.03	0.02
β11		-0.03	0.04	**-0.06**	0.01
β12		-0.05	0.04	**-0.07**	0.01
Higher education					
β0		**0.62**	0.05	**0.61**	0.03
β1		**-0.42**	0.13	**-0.50**	0.04
β2		**0.46**	0.08	**0.45**	0.02
β3		-0.08	0.06	**-0.12**	0.02
β4		**-0.19**	0.05	**-0.10**	0.02
β5		-0.01	0.04	**-0.07**	0.01
β6		-0.05	0.04	0.00	0.01
β7		**0.11**	0.04	**0.11**	0.01
β8		0.01	0.05	**-0.05**	0.01
β9		0.09	0.05	0.02	0.01
β10		-0.08	0.06	-0.03	0.02
β11		-0.04	0.04	**-0.06**	0.01
β12		**-0.12**	0.04	**-0.07**	0.01
Sub-region of living					
Liège (β0)		**-0.17**	0.03	**-0.19**	0.03
(β'1–β1)		-0.06	0.05	0.00	
Luxembourg (β0)		0.00	0.04	0.02	0.04
(β'1–β1)		0.08	0.08	0.00	
Namur (β0)		**-0.09**	0.04	**-0.09**	0.03
(β'1–β1)		**-0.13**	0.06	**-0.10**	0.04
Hainaut (β0)		**-0.26**	0.03	**-0.27**	0.03
(β'1–β1)		**-0.11**	0.05	**-0.07**	0.03

(Table continued on next page)

Table 9: (Continued)

Sector of previous employment				
Agriculture ($\beta0$)	**0.21**	0.09	**0.18**	0.08
($\beta'1-\beta1$)	-0.13	0.14	0.00	
Iron-steel industry and mining ($\beta0$)	-0.12	0.06	**-0.14**	0.06
($\beta'1-\beta1$)	-0.09	0.11	0.00	
Metal manufacture ($\beta0$)	0.10	0.06	0.07	0.05
($\beta'1-\beta1$)	-0.13	0.09	0.00	
Manufacture of non-metallic products ($\beta0$)	-0.15	0.08	**-0.16**	0.08
($\beta'1-\beta1$)	0.01	0.13	0.00	
Manufacture of consumption and investment goods ($\beta0$)	0.03	0.06	0.01	0.06
($\beta'1-\beta1$)	-0.08	0.10	0.00	
Other industries ($\beta0$)	0.09	0.09	0.06	0.09
($\beta'1-\beta1$)	-0.18	0.15	0.00	
Energy and water production ($\beta0$)	-0.11	0.13	-0.12	0.13
($\beta'1-\beta1$)	-0.02	0.09	0.00	
Building ($\beta0$)	**0.15**	0.05	**0.14**	0.05
($\beta'1-\beta1$)	-0.06	0.09	0.00	
Private services ($\beta0$)	**0.24**	0.05	**0.24**	0.05
($\beta'1-\beta1$)	-0.02	0.09	0.00	
Public services ($\beta0$)	0.04	0.05	0.04	0.05
($\beta'1-\beta1$)	-0.02	0.09	0.00	
Others ($\beta0$)	**0.47**	0.07	**0.52**	0.07
($\beta'1-\beta1$)	0.17	0.12	0.00	
No previous employment ($\beta0$)	**0.17**	0.08	**0.03**	0.07
($\beta'1-\beta1$)	**-0.31**	0.12	0.00	
2. Duration (in years)				
2 (γ_2-c)	**-0.30**	0.03	**-0.30**	0.03
3 (γ_3-c)	**-0.66**	0.05	**-0.66**	0.05
4 (γ_4-c)**	**-1.04**	0.06	**-1.04**	0.06
3. Variance of the Gamma at *l* =0				
η_c	**-1.74**	0.16	**-1.73**	0.16
$\sigma^2_0=\exp(\eta_c)$	**0.18**	0.03	**0.18**	0.03
4. Calendar time at inflow				
η_1	**0.34**	0.05	**0.34**	0.05
Number of estimated parameters	89		41	
Number of observations	45987		45987	
Weighted sum of squared residuals	35808.1		35875.2	
Adjusted number of parameters (after regrouping)	62		25	
Adjusted number of observations (after regrouping)	420		420	
Adjusted weighted sum of squared residuals (after regrouping)	446.55		502.62	
P-value of the goodness-of-fit test	0.001		0.000	
P-value of the restrictions test			0.02	

** $\gamma_{21} = ... = \gamma_4$