ELSEVIER

Journal of Econometrics 102 (2001) 271–309

JOURNAL OF Econometrics

www.elsevier.com/locate/econbase

Combining micro and macro unemployment duration data

Gerard J. van den Berg[a,b,c,d,e], Bas van der Klaauw[a,f,*]

[a] *Department of Economics, Free University Amsterdam, De Boelelaan 1105, NL-1081 HV, Amsterdam, Netherlands*

[b] *Tinbergen Institute, Amsterdam, Netherlands*

[c] *CEPR, London, UK*

[d] *IFAU, Uppsala, Sweden*

[e] *INSEE-CREST, Paris, France*

[f] *Scholar, Amsterdam, Netherlands*

Received 3 September 1998; revised 19 November 2000; accepted 23 January 2001

Abstract

We combine micro and macro unemployment duration data to study the effects of the business cycle on the outflow from unemployment. We allow the cycle to affect individual exit probabilities of unemployed workers as well as the composition of the total inflow into unemployment. We estimate the model using (micro) survey data and (macro) administrative data from France. The distribution of the inflow composition is estimated along with the other parameters. The estimation method deals with differences between the micro and macro unemployment definitions. The results also show to what extent the unemployment duration distributions corresponding to the two definitions can be described by the same model. © 2001 Elsevier Science S.A. All rights reserved.

JEL classification: J64; C41; E24; C51

Keywords: Unemployment composition; Heterogeneity; Duration dependence; Business cycle; Seasons; Unemployment definition

* Corresponding author.

E-mail address: klaauw@tinbergen.nl (B. van der Klaauw).

0304-4076/01/$ - see front matter
PII: S0304-4076(01)00055-0

Nomenclature

t, T	unemployment duration
τ	calendar time
x, X	vector of personal characteristics
$\theta(.)$	monthly individual exit probability out of unemployment
$\psi_1(.)$	effect of t on individual exit probability
$\psi_2(.)$	effect of τ on individual exit probability
β	parameter vector for effect of x on individual exit probability
v	unobserved heterogeneity term
x_i, X_i	elements of vectors x and X
$\bar{x}_i$	largest possible value of x_i
$h(.)$	Hermite density
$\alpha_{i_1 \cdots i_n}$	parameters of h
$c_\tau^i(.)$	threshold values in limdep specification of distribution of $X_i \mid \tau$
$U(.)$	number of unemployed
N_τ	size of inflow into unemployment
$\Theta(.)$	quarterly aggregate exit probability out of unemployment
$\tilde{U}(.), \tilde{\Theta}(.)$	observations of U and Θ in macro data
ε, e	errors in $\tilde{U}$ and $\tilde{\Theta}$
$\psi_{1,i}$	parameters of ψ_1
$\psi_{2,s}(.)$	seasonal effect in ψ_2
$\psi_{2,b}(.)$	business cycle effect in ψ_2
ω_s	parameters of $\psi_{2,s}$
η_i	parameters of $\psi_{2,b}$
$f_i(.)$	orthogonal polynomials
$d_s^i(.)$	seasonal effect parameters in c_τ^i
$d_b^i(.)$	business cycle effect parameters in c_τ^i
v_i, p_i	parameters of distribution of v
δ	difference in level of micro and macro exit probabilities
$d_{<'87}$	dummy in $\Theta(0 \mid \tau)$ for pre-1987
σ	standard deviation of $\log \varepsilon$

1. Introduction

The recently expanding macro-economic literature on aggregate flows between labor market states stresses that the distribution of unemployment durations changes markedly over the business cycle, and it acknowledges the importance of heterogeneity in both stocks and flows of unemployed workers.

Empirically, the average duration is typically found to be countercyclical (see for example Layard et al., 1991). This may be because in a recession the exit probability out of unemployment decreases for all workers, or because in a recession the composition of the (heterogeneous) inflow shifts towards individuals who have low exit probabilities. Darby et al. (1985) argue that the latter is the major cause of the observed exit probabilities being low in recessions.

Typical macro time-series data are not sufficiently informative to study this, because they do not contain information on the composition of the heterogeneous inflow into unemployment. Typical longitudinal micro data are neither sufficiently informative to study this issue, for the reason that they do not cover a sufficiently long time span.[1] In micro-economic analyses of individual variation in unemployment duration, it is typically assumed that the parameters are independent of macro-economic conditions, and these conditions are at most included by way of an additional regressor (see Devine and Kiefer (1991) for a survey).

In this paper we combine micro and macro unemployment duration data in order to study the effects of the business cycle on the outflow from unemployment. We allow the business cycle to affect the individual exit probabilities of all unemployed workers, and we simultaneously allow it to affect the composition of the total inflow into unemployment. Both may lead to different aggregate exit probabilities. We allow the individual exit probabilities out of unemployment to depend on (i) the elapsed unemployment duration, (ii) calendar time, and (iii) personal characteristics. The dependence on calendar time is modeled by way of a product of a flexible high-order polynomial in calendar time (capturing business cycle effects) and dummy variables capturing seasonal effects.

We also model the joint distribution in the inflow into unemployment of the personal characteristics that affect the exit probabilities, including the way in which this distribution varies over time. In duration analysis it is standard practice to condition on explanatory variables such as personal characteristics. Here however their distribution is of interest. We allow for business cycle effects as well as seasonal effects on this distribution. Note that what really matters is not simply whether the inflow distribution of particular personal characteristics changes over time, but rather whether it changes for those characteristics that affect the exit probabilities. The composition of the inflow is only relevant in respect of personal characteristics that affect the exit probabilities. It is thus insufficient to investigate whether the composition changes

[1] In addition, the sample sizes may not be sufficiently large to observe the composition of the inflow in, say, a given quarter, and the data may be subject to endogenous attrition. Admittedly, the problems with the time span and sample sizes of micro data may be more serious for European countries than for the US.

by way of graphical checks on the proportion of certain types of individuals in the inflow. Instead, it is necessary to estimate a joint model for the composition of the inflow and the duration until outflow.

On a macro level, personal characteristics are unobserved. Observed explanatory characteristics at the micro level constitute unobserved heterogeneity at the aggregate level. Thus, the distribution of personal characteristics enters the expression for the probability distribution of the observed macro unemployment durations. For the distribution of personal characteristics we use a specification based on Hermite polynomials. Such a specification is sufficiently flexible while being computationally feasible as well. In an extended version of our model we also allow for heterogeneity that is unobserved in the micro data. To enhance the empirical analysis we exploit the fact that multiple unemployment spells are observed for some individuals in the micro data.

Ideally, the macro data provide the exact aggregate unemployment duration distributions in the population. Thus, ideally, these data are deterministically equal to the corresponding model expressions, and all parameters may be deduced from such equations. Unfortunately, the actual situation is more complicated than this. In most OECD countries, the official unemployment statistics follow an unemployment definition that differs from the definition in micro labor force surveys. In particular, as a rule, official national statistics count registrations at public employment agencies, whereas alternative statistics are based on self-reported unemployment in labor force surveys of sampled individuals.[2] In this paper, we have to face this problem, as the micro data we use are from the French longitudinal labor force panel survey whereas the macro data concern French registered unemployment. The macro unemployment concept deviates from the micro concept in a number of respects.

Indeed, the second motivation of this paper concerns the nature of the differences between the measures of unemployment based on the micro and macro definition, respectively (note that this motivation logically precedes the economic motivation described earlier in this section). The behavior over time of the difference in the *levels* of these two measures has been well documented (European Commission, 1994; CSERC, 1996). In this paper we analyze any differences on a deeper level. The full model contains a number of overidentifying restrictions, and by estimating the determinants of the

[2] The simultaneous use of different measures has led to demands for more clarity, in public opinion as well as in the scientific literature. For example, according to the European Commission (1994), the differences 'are a source of confusion and misunderstandings'. Labor market researchers have repeatedly advocated more clarity in the publication of unemployment statistics (CSERC, 1996; Le Monde, 1997). In his survey on European unemployment, Bean (1994) concludes that 'there needs to be a more deliberate attempt to identify the extent to which apparent differences in fit are due to different variable definitions'.

duration distributions associated with both measures, we are able to describe and explain to what extent they are dissimilar.

Some of the differences between both unemployment measures relate to features of the individual search behavior, some to decisions by the employment agency, and some to practical measurement issues. It would be very difficult to model these on an individual level, and it would therefore be even more difficult to derive macro duration distributions from individual duration distributions for the unemployment population corresponding to the macro definition. We therefore take a different approach. Basically, we take the observed macro exit probabilities to be equal to a perturbed version of the probabilities that would prevail if the macro definition would be the same as the micro definition, and we allow for correlated measurement errors in the macro data.[3]

To date, a number of empirical studies using micro survey data have been published that focus on one or more of the issues we deal with in the present paper. It should be noted from the outset that all of this empirical literature is based on US data, except for Lollivier (1994a). The studies by Dynarski and Sheffrin (1990), Imbens and Lynch (1992) and Lollivier (1994a) use micro data to estimate the effect of business-cycle indicators like the unemployment rate on the unemployment duration distribution. By conditioning on personal characteristics, the effect of the business cycle (or calendar time in general) on the individual exit probability can in principle be singled out. In Dynarski and Sheffrin (1990) and Lollivier (1994a), the time span covered by the data is relatively short. Imbens and Lynch (1992) use longitudinal US data (the NLS Youth Cohort) covering 11 years to study the effect of calendar time and individual duration determinants on the duration of *non*-employment (i.e. unemployment plus non-participation) among youths. Their estimation results enable an assessment of the extent to which the quality of the inflow into non-employment among youths changes over the business cycle. From a graphical check they conclude that this change is not substantial, apart from seasonal variation.

Darby et al. (1985) examine US micro data from the CPS surveys, which cover a long time span. Using a somewhat informal approach, they estimate an equation for the exit probability as a function of a proxy of the average 'quality' of the inflow (this varies over the cycle) as well as other business cycle indicators.[4] They conclude that changes in the composition of the

[3] Imbens and Lancaster (1994) develop a methodology for the joint empirical analysis of micro and macro data that is more suitable if the macro data provide (features of) exact aggregate distributions in the population and if one is not interested in the (determinants of the) distribution of the explanatory variables. See Laisney and Lechner (1996) for an application.

[4] Specifically, they use the lagged fraction of short-term unemployed as an indicator of the average quality of the inflow.

inflow are a primary determinant of cyclical variations in the exit probability. Baker (1992) and Davis et al. (1996) examine CPS data as well, and both studies conclude that the composition varies across the cycle in terms of the reason of inflow, age, and gender. In particular, a relatively large part of the inflow in recessions consists of permanently laid-off workers and prime-aged men. Laid-off persons have lower exit probabilities out of unemployment, and from this Davis et al. (1996) conclude that changes in the composition are an important cause of the countercyclicality of aggregate unemployment durations (they also find strong seasonal effects on the composition of the inflow). Baker (1992) provides a more formal analysis of the determinants of the cyclicality of durations. Specifically, the estimated variation in durations is decomposed into a part due to a changing composition and a part due to cyclical effects on the exit probability. Different individual-specific characteristics are analyzed in separate decompositions. He concludes that cyclical variation in unemployment durations is mainly driven by the effect of the cycle on individual exit probabilities (rather than by the effect on the composition). Note that this literature does not adopt a formal multivariate framework to test whether a personal characteristic x has an inflow distribution that varies over the cycle while at the same time x itself affects the individual exit probability. Moreover, even if both of these would be significant, it remains to see whether x is actually quantitatively important as a determinant of the variation in unemployment durations over the cycle.

The outline of the paper is as follows. In Section 2 we examine the definition of unemployment in both data sets in detail, and we discuss the observation of unemployment durations in the data sets. The model specification is presented in Section 3. Section 4 describes the data used to estimate the model. Section 5 contains the estimation results, and Section 6 concludes.

2. Definition and measurement of unemployment in the micro and macro data

2.1. The micro data

The French Labor Force Survey (*Enquête sur l'emploi*) is a longitudinal panel survey on labor supply behavior over time, collected by INSEE (National institute of statistics and economic studies). In its present form, this panel survey runs since 1991. In March every year, members of around 60,000 French households are interviewed. One third of the household sample is renewed each year, such that a given individual is interviewed in three consecutive years. We use the data of those who entered the survey in 1991.

An effort is made to collect extensive information on the labor market behavior of individual respondents in the year preceding the moment of the

interview. In particular, the respondents are asked to report the main labor market state (*situation principale*) they were in, for each month in that year, including the month of the interview. The respondent can choose between eight states, one of which is unemployment. Four of the other states concern employment (including self-employment and employment in regular jobs and paid training jobs), whereas the remaining three states concern non-participation (including retirement, unpaid training, and being housewife or student).[5] The respondent must choose a single state for each month. It is thus likely that a respondent who has worked less than 50% of the time in a given month and who has been unemployed for the remainder of the time will classify himself as unemployed for that month. It is important to note that a respondent may assign himself to unemployment when he is not registered as such at the public employment agency. The answers on the monthly labor-market state questions are generally consistent with the preceding questions on past and current labor market behavior (see Lollivier, 1994b).

By comparing individual labor market states of consecutive months in the period from March 1990 to March 1993, individual unemployment durations can be constructed; these always consist of an integer number of calendar months. Personal characteristics of the respondent are recorded at the first interview. Unit nonresponse in the labor market survey has been rather low (on average 6%).

2.2. The macro data

The macro data concern quarterly unemployment data over the period 1982.IV–1993.I, collected by the French public employment agencies (ANPE), and subsequently reported by the Ministry of Social Affairs and Employment (see ILO, 1989, for an extensive description). The data are collected at the final date of each quarter. For each gender, they provide the total number of individuals in the population at that moment who have completed a given number of quarters of unemployment duration in their current spell. So, for example, they provide the number of men who are unemployed for more than 3 and less than 4 quarters, on December 31, 1990. These data obviously allow for the reconstruction of individual unemployment durations, although the inflow and outflow dates can only be traced back to lie in three-month intervals.

[5] See the working paper version Van den Berg and van der Klaauw (1998) for details. The (implicit) definition of unemployment used here is similar to the definition used by the International Labor Organization (ILO). The latter requires the individual to classify himself as (1) without employment, (2) seeking employment, and (3) currently available for employment (see ILO, 1982). Unemployment statistics based on the ILO definition are reported in the media and used by international organizations like the European Union.

We now turn to the definition of unemployment in the macro data. When individuals (voluntarily) register at a public employment agency, they state that they are seeking work. At the moment of registration, the individual is classified by the agency into one of five categories, according to his current situation and his desired type of employment. One of these categories is 'Without employment, and immediately available for employment, and actively searching employment, and seeking permanent full-time employment'.[6] Our macro unemployment data cover only this category. Indeed, the number of individuals in it defines the official ('registered') unemployment statistic.

Of the other four categories, two concern employment and non-participation, whereas the remaining two capture individuals seeking part-time employment and individuals seeking temporary or seasonal employment (while being without employment, immediately available for employment, and actively searching employment). We do not have detailed duration information for the latter two categories, so we cannot include these in the empirical analysis. Of course, the corresponding individuals may well classify themselves as being unemployed, so the micro unemployment data may include unemployed individuals seeking part-time or temporary employment. However, according to available data on the over-all outflow of individuals from the three unemployment categories in the macro data, the two categories of individuals seeking part-time or temporary employment are quantitatively unimportant, in particular for men. For example, in 1994.IV, the male outflow out of the three categories consisted for 94% of individuals 'seeking permanent full-time employment'. For women in 1994.IV, this figure is 87%. Because of this, we restrict attention to men in this study.

Reasons for removing an individual out of our unemployment data category include (in addition to finding suitable work or movement to another category) failure to comply with register continuation requirements. Registration with ANPE is a necessary condition for the receipt of any unemployment benefits (with the exception of individuals over 55 years of age).

Since 1982, the registration process and the operationalization of the definitions of the categories have been changed a number of times. Notably, the procedures concerning the continuation of individual registration have become stricter. At the same time, changes in the unemployment benefits system have affected the incentive to register. There is evidence that all this has affected the composition and number of individuals in the register (see ILO (1989) and Van den Berg and van der Klaauw (1998) for details). In October 1986,

[6] Here, 'immediately' means 'within 15 days', and 'full-time' means 'more than 30 hours per week'. If the individual is employed, but employment is known to terminate within 15 days, then the individual is registered to be without employment. Individuals who have worked more than half of the time during the month can therefore be registered as being unemployed, at least, in our sample period.

procedures regarding the timing of the data collection and the statistical processing of the raw data changed substantially (see ILO (1989) for details). It turns out that the data display a discontinuity around this date, and we return to this below.

Because the classification into the five categories was introduced in 1982.IV, and because there were major changes before 1982.IV in the relation between registration and receipt of unemployment benefits, we do not use data before 1982.IV (the time series display a discontinuity between 1982.III and 1982.IV).

3. The model

3.1. Modelling individual exit probabilities

Throughout the paper we use two measures of time, each with a different origin. The variable t denotes unemployment duration as measured from the moment of inflow into unemployment. The variable τ denotes calendar time, which has its origin somewhere in the past.

In the micro data as well as in the macro data, unemployment duration and calendar time are both measured in discrete units. For a given unemployment spell in the data we only know the months or quarters within which they started and/or ended. Both t and τ are therefore taken to be discrete variables, and we define the month to be the unit of time and duration. We define $t{:=}0$ in the first month of unemployment. So, in general, $t \in \{0, 1, 2, \ldots\}$.

It is unattractive to have a model that is not invariant to changes in the time unit. We therefore specify our discrete-time model as a continuous-time model in which time and duration are aggregated over monthly intervals. However, we do not interpret the data as being realized by some underlying continuous-time process that is imperfectly observed. This is because otherwise we would have to take account of the fact that spells may cover only part of a month, and that there are spells starting and ending within the same month. This would greatly complicate the analysis (see below).

The basic elements in the model are the exit probabilities at the individual level. It is assumed that all variation in the individual exit probabilities out of unemployment can be explained by the prevailing unemployment duration t and calendar time τ, and by heterogeneity across individuals. We denote the monthly probability that an individual leaves unemployment right at t months of unemployment, given that he is unemployed for t months at calendar time τ, and conditional on his observed characteristics x, by $\theta(t \mid \tau, x)$. For convenience, we only allow for characteristics x that are time-invariant at the individual level (although of course their distribution in the inflow may vary

over time).[7] Suppose for the moment that all heterogeneity across individuals is observable. We assume that $\theta(t \mid \tau, x)$ can be written as

$$\theta(t \mid \tau, x) = 1 - \exp(-\psi_1(t)\psi_2(\tau)\exp(x'\beta)) \tag{1}$$

with ψ_1 and ψ_2 positive. This specification can be derived from a continuous-time Proportional Hazard (PH) model. Consider a continuous-time PH model with, in obvious notation, the individual exit rate $\psi_1(t)\psi_2(\tau)\exp(x'\beta)$.[8] Now consider an individual with characteristics x who is unemployed for t months at τ. The conditional probability of leaving unemployment between τ and $\tau + 1$ then equals

$$1 - \exp\left(-\int_t^{t+1} \psi_1(u)\psi_2(\tau + u)\exp(x'\beta)\,\mathrm{d}u\right) \tag{2}$$

(Lancaster, 1990). If ψ_1 and ψ_2 are constant within monthly intervals, this probability equals the expression for $\theta(t \mid \tau, x)$ of (1). In fact, we will assume that both ψ_1 and ψ_2 are constant within quarterly intervals.

Expressions for the individual unemployment duration distribution follow from (1). Let T denote the random duration of a completed spell. For an individual with characteristics x, the probability that the duration T equals t months if the individual has entered unemployment at τ equals

$$\Pr(T = t \mid \text{inflow at } \tau; x) = \theta(t \mid \tau + t, x)\prod_{u=0}^{t-1}(1 - \theta(u \mid \tau + u, x)). \tag{3}$$

The model is readily extended to allow for unobserved heterogeneity on the micro level; that is, for the presence of personal characteristics v that affect unemployment duration like x but that are not recorded in the micro data. Assume that both the individual v and the distribution of v are time-invariant, and that v is independent of x. By analogy with the paragraphs above it is obvious that the specification

$$\theta(t \mid \tau, x, v) = 1 - \exp(-\psi_1(t)\psi_2(\tau)\exp(x'\beta)\exp(v)) \tag{4}$$

can be derived from a continuous-time Mixed Proportional Hazard specification. The micro data can be interpreted as aggregates over v. The exit probability $\theta(t \mid \tau, x)$ at duration $T = t$, given $T \geqslant t$ and x, and given inflow

[7] The micro data do not show how personal characteristics vary over time. The model framework can be extended to the case where there are observed time-varying explanatory variables, provided that suitable exogeneity and identifiability conditions are satisfied (Van den Berg, 2001). However, computing the aggregated exit probabilities requires modelling the moment at which explanatory variables change values. This complicates the model enormously.

[8] This multiplicative specification in t, τ and x has been used before by Imbens (1994).

at calendar time $\tau - t$, equals

$$\theta(t \mid \tau, x) \equiv \Pr(T = t \mid T \geqslant t;\ \text{inflow at } \tau - t; x)$$
$$= \frac{\mathrm{E}_v[\Pr(T = t \mid \text{inflow at } \tau - t; x; v)]}{\mathrm{E}_v[\Pr(T \geqslant t \mid \text{inflow at } \tau - t; x; v)]} \tag{5}$$

in which the expectations E_v are taken with respect to the distribution of v in the inflow. The probabilities on the right-hand side are easily expressed in terms of $\theta(\cdot \mid \cdot, x, v)$. For example, $\Pr(T = t \mid \text{inflow at } \tau; x; v)$ is given by (3), provided we replace x by x, v.

3.2. Modelling the composition of the inflow

In this subsection we model the joint distribution in the inflow of the personal characteristics x affecting the exit probabilities, including the way it changes over time. We assume that these personal characteristics are described by a set of discrete variables $x_1, \ldots, x_n$. This is not restrictive, because the micro data do not contain continuous explanatory variables.[9] We normalize the model by imposing that the set of possible values of x (i.e., the locations of the mass points of the n-dimensional multivariate discrete distribution of x) does not shift over time (for example, a dummy is always either zero or one, and not zero or one in the beginning and one or two later on). The calendar time effect is modelled as affecting the probabilities of the different values of x.

On the one hand, it is clear that the number of unknown parameters in the model becomes too large if no restrictions are imposed on the multivariate discrete distribution of x and its variation between cohorts. On the other hand, it is important to allow for sufficient flexibility. It would be too restrictive to assume independence of the x or to suppose that a recession affects all n marginal distributions of the elements of x in the same way. To proceed, we adopt a specification based on Hermite series. This specification is related to a specification for distribution functions that is used in the popular semi-nonparametric estimation method of Gallant and Nychka (1987).

We denote the random variable associated with x_i by X_i and its possible values by X_i by $\{0, 1, 2, \ldots, \bar{x}_i\}$. We assume that the joint distribution of $X_1, \ldots, X_n$ in the inflow at cohort date τ can be written as

$$\Pr(X_1 = x_1, \ldots, X_n = x_n \mid \tau)$$
$$= \int_{c_\tau^1(x_1)}^{c_\tau^1(x_1+1)} \cdots \int_{c_\tau^n(x_n)}^{c_\tau^n(x_n+1)} h(u_1, \ldots, u_n)\, \mathrm{d}u_1 \cdots \mathrm{d}u_n. \tag{6}$$

[9] The framework developed in this subsection can be extended to the case where there are both continuous and discrete explanatory variables.

There are two types of determinants of the right-hand side: the 'core density' h on the one hand, and the 'threshold values' $c_\tau^i(x_i)$ on the other. For the threshold values, the super-index refers to the explanatory variable at hand, whereas the argument refers to the realized value of this explanatory variable. The threshold values are such that $c_\tau^i(0)=-\infty$, $c_\tau^i(x_i) < c_\tau^i(x_i+1)$, and $c_\tau^i(\bar{x}_i+1)=\infty$. Intuitively, the threshold values are closely linked to the shapes of the marginal distributions of $X_1,\ldots,X_n$ whereas the density h is closely linked to the way in which the elements of $X_1,\ldots,X_n$ are interrelated. Obviously, for a given $h(u_1,\ldots,u_n)$, the threshold values are identified from the marginal distributions of $X_1,\ldots,X_n$ (all for a given τ). As a special case, if $n=1$ then the distribution of X_1 is as in an ordered probit model, which becomes clear in the remainder of this subsection: h is standard normal and does not have unknown parameters, and the threshold values divide the support of h into intervals such that probabilities of the intervals correspond to probabilities of realizations of X_1. Note that, in general, if h factorizes in terms of $u_1,\ldots,u_n$ then $X_1,\ldots,X_n$ are jointly independent.

The threshold values specify how the joint distribution changes over calendar time τ. To illustrate this, consider a binary characteristic x_i, and suppose that $c_\tau^i(1)$ increases over calendar time. Then the proportion of the newly unemployed individuals who have $x_i=0$ increases over the calendar time. In Section 3.4 we examine in detail how we model the dependence of the threshold values on calendar time. Somewhat loosely one may state that, by making the threshold values rather than h dependent on τ, we impose that the business cycle affects the distribution of X mostly by shifting the marginal distributions, whereas the interrelations between $X_1,\ldots,X_n$ are less affected.

The density $h(u)\equiv h(u_1,\ldots,u_n)$ is modeled by way of a Hermite series. Specifically, for some set $V\in\mathbb{N}^n$,

$$h(u)=\frac{1}{S}\left(\sum_{(i_1\cdots i_n)\in V}\alpha_{i_i\cdots i_n}u_1^{i_1}\cdots u_n^{i_n}\right)^2\exp\left(-\frac{u_1^2}{\delta_1^2}-\cdots-\frac{u_n^2}{\delta_n^2}\right), \tag{7}$$

where S is a normalizing constant ensuring that h integrates to one. We can then normalize further by fixing $\alpha_{0\cdots 0}=1$. Moreover, the unidentified scale of h can also be normalized, and we set $\delta_1=\cdots=\delta_n=\sqrt{2}$. Now the shape of the density only depends on the values of $\alpha_{i_1\cdots i_n}$ and thus on the set V. A large number of elements in V gives more flexibility. We take

$$V=\{(i_1\cdots i_n)\,|\,i_1,\ldots,i_n\in\{0,1\}\cup i_1+\cdots+i_n\leqslant 2\}. \tag{8}$$

It is now easy to show that

$$S=(2\pi)^{n/2}\sum_{(i_1\cdots i_n)\in V}\alpha_{i_i\cdots i_n}^2. \tag{9}$$

We can now also normalize the unidentified location of the density function h to zero. This is achieved by imposing that $\alpha_{i_1\cdots i_n}=0$ for every combination

for $i_1, \ldots, i_n$ with $i_1 + \cdots + i_n = 1$. We are subsequently left with only $n(n-1)/2$ unknown parameters in h: $\alpha_{i_1 \cdots i_n}$ with $i_1 + \cdots + i_n = 2$. These parameters can be interpreted as indicators of the signs of the interrelations between the elements of X (although they also affect other moments of the joint distribution).

Note that if $n = 1$ then h equals a standard normal density function. As another example, consider $n = 2$. Then $h(u)$ has only one unknown parameter: α_{11}. Specifically,

$$h(u) = \frac{1}{2\pi(1 + \alpha_{11}^2)}(1 + 2\alpha_{11}u_1u_2 + \alpha_{11}^2 u_1^2 u_2^2)\mathrm{e}^{-(1/2)u_1^2 - (1/2)u_2^2}.$$

The correlation between u_1 and u_2 equals $\alpha_{11}/(1 + 3\alpha_{11}^2)$. If X_1 and X_2 are dummy variables then

$$\begin{aligned}
&\Pr(X_1 = 0, X_2 = 0 \mid \tau) \\
&\quad = \Phi(c_1)\Phi(c_2) + \frac{\alpha_{11}}{2\pi(1 + \alpha_{11}^2)}[(2 + \alpha_{11}c_1c_2)\mathrm{e}^{-(1/2)c_1^2 - (1/2)c_2^2} \\
&\qquad - \alpha_{11}\sqrt{2\pi}(c_1\mathrm{e}^{-(1/2)c_1^2}\Phi(c_2) + c_2\mathrm{e}^{-(1/2)c_2^2}\Phi(c_1))],
\end{aligned}$$

where Φ denotes the standard normal c.d.f., and c_i is shorthand notation for $c_\tau^i(1)$.

A major advantage of the specification proposed above is its computational convenience. Note that all integrals in (6) can be expressed analytically. Moreover, the specification for the distribution of X does not automatically impose that time has the same effect on the marginal distributions of the elements of X, and it does not restrict the signs of the correlations between elements of X. However, the specification has the disadvantage that there is no simple relation between the parameters and moments of X. In particular, because every parameter influences every element of the variance–covariance matrix of X, testing for specific correlation structures is not straightforward.

3.3. Modelling measurement and specification errors in the macro data

We take the unemployment definition used in the micro data as the most relevant definition (recall that this definition resembles the ILO definition), and we assume the model of Sections 3.1 and 3.2 to describe these micro data. As a consequence, the parameters and functions of interest are β, the functions $\psi_1(t)$, $\psi_2(\tau)$, the α-parameters and the $c_\tau^i(x_i)$ as functions of τ.

It is useful to start this subsection with a derivation of the model expressions for the observables in the macro data *as if* the macro data concern the population from which the micro data are sampled. Recall that the macro data measure durations in quarters at quarterly time intervals. We thus have to aggregate the exit probabilities over time as well as over individuals. It is useful to introduce some notation. We denote the number of unemployed

with a duration of $t, t+1$ or $t+2$ months, at calendar time τ, by $U(t\,|\,\tau)$ (for $t \in \{0,3,6,\ldots\}$ and for τ equal to the third month of a quarter). These numbers constitute the macro data. Let N_τ denote the size of the inflow into unemployment at month τ.

$$U(t\,|\,\tau) = \sum_{i=0}^{2} N_{\tau-t-i} \Pr(T \geqslant t+i \,|\, \text{inflow at } \tau-t-i). \tag{10}$$

From the values of $U(t\,|\,\tau)$ one can calculate the proportion of individuals who are unemployed for $t,\ldots,t+2$ months at calendar time τ who leave unemployment before the end of the next quarter. This fraction equals the quarterly exit probability out of unemployment among the workers who are unemployed for $t,\ldots,t+2$ months at calendar time τ. We denote this probability by $\Theta(t\,|\,\tau)$,

$$\Theta(t\,|\,\tau) = \frac{U(t\,|\,\tau) - U(t+3\,|\,\tau+3)}{U(t\,|\,\tau)}. \tag{11}$$

Assume that the size of the inflow into unemployment is constant within a quarter, so $N_{\tau-2} = N_{\tau-1} = N_\tau$, for any τ equal to the third month of a quarter. Then, using Eq. (10), $\Theta(t\,|\,\tau)$ can be rewritten as

$$\Theta(t\,|\,\tau) = \frac{\sum_{i=0}^{2} \Pr\,(T \in [t+i, t+i+2] \,|\, \text{inflow at } \tau-t-i)}{\sum_{i=0}^{2} \Pr\,(T \geqslant t+i | \text{inflow at } \tau-t-i)}.$$

This can be rewritten in order to highlight the fact that the macro data concern aggregates of different individuals (so we integrate over x). Obviously, there is a strong analogy to the introduction of unobserved heterogeneity in Section 3.1. Let us ignore such heterogeneity v for the moment.

$$\Theta(t\,|\,\tau) = \frac{\sum_{i=0}^{2} \mathrm{E}_{x|\tau-t-i}[\Pr\,(T \in [t+i, t+i+2] | \text{inflow at } \tau-t-i; x)]}{\sum_{i=0}^{2} \mathrm{E}_{x|\tau-t-i}[\Pr\,(T \geqslant t+i | \text{inflow at } \tau-t-i; x)]} \tag{12}$$

in which the expectations $\mathrm{E}_{x|\tau-t-i}$ are over the distribution of x (or, equivalently, the distribution of $\exp(x'\beta)$) in the inflow at $\tau-t-i$. The probabilities on the right-hand side of this equation are easily expressed in terms of $\theta(\cdot|\cdot, x)$, using the fact that $\Pr(T = t | \text{inflow at } \tau; x)$ is given by (3). As an example, the denominator of the right-hand side of (12) for $t = 0$ equals,

$$\begin{aligned} &1 + \mathrm{E}_{x|\tau-1}[1 - \theta(0|\tau-1, x)] \\ &\quad + \mathrm{E}_{x|\tau-2}[(1 - \theta(0|\tau-2, x))(1 - \theta(1|\tau-1, x))]. \end{aligned}$$

Suppose we observe $U(t\,|\,\tau)$ for n duration classes $0, 3, \ldots, 3n-3$. Then (12) can be thought to represent $n-1$ different equations, namely for $\Theta(0|\tau)$ until and including $\Theta(3n-6|\tau)$. The loss of information when going from

n duration classes for U to $n-1$ equations for Θ (which is a first difference of U) concerns the *level* of unemployment, say at $t=0$. There is a one-to-one correspondence between $U(0|\tau)$ and the size N_τ of the monthly inflow during the quarter. We are not interested in the latter. For our purposes it can therefore be stated that the macro data consist of the observed values of $\Theta(t\,|\,\tau)$.

One may argue that the macro data provide exact population quantities, and that therefore the observed values of $\Theta(t\,|\,\tau)$ are deterministically equal to the corresponding model expressions.[10] The unknown parameters (to the extent that they are identified) can then be deduced from this nonlinear system of equations.

However, the situation is more complicated than this. First, recall from Section 2 that the macro definition deviates from the micro definition in a number of respects, and, consequently, that it describes a different set of individuals. A number of types of individuals satisfy the micro definition but not the macro definition, whereas other types satisfy the macro definition but not the micro definition.[11] In such cases, an individual permanently satisfies one definition and not the other. However, it is also possible that an individual changes his behavior at a certain point of time in such a way that a transition into or out of unemployment occurs according to one definition but not according to the other. In addition to this, the macro definition itself is not time-invariant.

In sum, there is a large number of fundamental differences between both unemployment measures. Some of these relate to features of the individual search behavior, some to decisions by the employment agency, and some to measurement procedures. Clearly, it is impossible to model all this on an individual level. It is therefore also impossible to derive macro duration distributions from individual duration distributions for the unemployment population corresponding to the macro definition. We therefore take a different approach. First of all, we establish the relation between the model and the macro data by taking the observed macro exit probabilities to be equal to a perturbed version of the probabilities $\Theta(t\,|\,\tau)$ that would prevail if the macro definition would be the same as the micro definition. Since $\Theta(t\,|\,\tau)$ is derived from $U(t\,|\,\tau)$, we achieve this by allowing for errors in the latter. From now on we place a $\sim$ on top of observed values of macro variables, in contrast to the corresponding 'true' values. We assume that

$$\tilde{U}(t\,|\,\tau) = U(t\,|\,\tau)\varepsilon_{t,\tau} \tag{13}$$

[10] Alternatively, the macro data are a sample from a hypothetical population of possible worlds.

[11] See the working paper version Van den Berg and van der Klaauw (1998) for a long list of examples.

with

$$\log \varepsilon_{t,\tau} \sim \mathrm{N}(0, \sigma^2).$$

Here, $\varepsilon_{t,\tau}$ captures measurement errors in $\tilde{U}(t \mid \tau)$ as well as effects of the differences between the unemployment definitions and the changes in the macro definition over time (below we introduce additional parameters for these effects). We assume normality for convenience. As we shall see, the estimate of σ is informative on the fit of the model to the macro data.

The observed exit probability out of unemployment $\tilde{\Theta}(t \mid \tau)$ equals the right-hand side of Eq. (11) with U replaced by $\tilde{U}$. By substituting Eq. (13) into this, we obtain

$$\log(1 - \tilde{\Theta}(t \mid \tau)) = \log(1 - \Theta(t \mid \tau)) + e_{t,\tau}, \tag{14}$$

where $e_{t,\tau} := \log \varepsilon_{t+3,\tau+3} - \log \varepsilon_{t,\tau}$. Eq. (14) links the observed macro exit probabilities to the model. Note that $e_{t,\tau}$ is normally distributed with mean zero. The errors in Eq. (14) are correlated. In particular, $\mathrm{Corr}(e_{t,\tau}, e_{t+3,\tau+3}) = -\frac{1}{2}$ (all other correlations are zero).

In the empirical analysis we also allow for differences between the 'micro' and 'macro' parts of the model by allowing certain parameters to have different values in both parts. This is feasible because some parameters are well identified from either data (for example, the level of the exit probability at low durations). Such an approach is informative on systematic differences in the determinants of the duration distributions associated with both unemployment concepts, in contrast to the 'perturbation' approach above.

3.4. Parameterization

The baseline duration dependence function $\psi_1(t)$ is parameterized as a piecewise constant function that is constant on three-monthly intervals,

$$\psi_1(t) = \sum_{i=1,2,\ldots}^{11} \psi_{1,i} \mathrm{I}(3i - 3 \leqslant t < 3i),$$

$\mathrm{I}(\cdot)$ being the indicator function. This is a very flexible specification with a duration dependence parameter for each quarterly duration interval. The duration dependence is assumed to be constant after 30 months. Note that the maximum possible observed completed duration in the micro data equals 35 months. In the empirical analyses, we use macro data on the first 12 quarterly duration classes, to obtain observations on $\tilde{\Theta}(0|\tau), \tilde{\Theta}(3|\tau), \ldots, \tilde{\Theta}(30|\tau)$. The maximum monthly duration in the macro data is thus 35 as well.

The calendar time effect $\psi_2(\tau)$ on the individual exit probability is modeled as the product of a seasonal effect and a business cycle effect,

$$\psi_2(\tau) = \psi_{2,s}(\tau)\psi_{2,b}(\tau).$$

The seasonal effect is written as

$$\psi_{2,s}(\tau) = \exp\left\{\sum_{s=1}^{4} \omega_s \mathrm{I}_s(\tau)\right\}$$

where the ω_s are unknown parameters and $\mathrm{I}_s(\tau)$ is an indicator function for season s. Business cycle effects (or cyclical and trend effects) are represented by a flexible polynomial of degree, say, 5. We could specify this polynomial in the standard way as a sum of terms $\eta_i \tau^i$, $i = 0, \ldots, 5$. However, as the terms τ^i are not mutually orthogonal, estimation of the parameters η_i suffers from multicollinearity. To avoid this, we use Chebyshev polynomials of the first kind. Thus, we specify the polynomial as the sum of terms $\eta_i f_i(\tau)$, $i = 1, \ldots, 5$, where $f_0(\tau), f_1(\tau), \ldots, f_5(\tau)$ are mutually orthogonal fully specified polynomials of indexed degree.[12] The business cycle effect $\psi_{2,b}(\tau)$ at month τ is then specified as the value attained by

$$\psi_{2,b}(\tau) = \exp\left\{\sum_{i=0}^{5} \eta_i f_i(\tau)\right\}$$

at the beginning of the quarter within which τ lies. As a result, $\psi_{2,b}(\tau)$ is a piecewise constant function with a shape determined by the polynomial expression above. We choose to take the value of the expression at the beginning of the quarter instead of the value at the beginning of the month (or the average value within the month) for computational reasons.

Note that one could model the dependence of the individual exit probability on the business cycle by way of an observable business cycle indicator like the capital utilization ratio. However, the present approach is flexible, as it does not impose this dependence to be a simple parametric function of such an indicator. A polynomial of sufficiently high degree is able to mimic the behavior of such indicators.

Calendar time affects the composition of the inflow by way of the threshold values $c_\tau^i(x_i)$ (see Eq. (6)). We allow the composition of the inflow to vary over seasons and over the cycle, so we specify $c_\tau^i(x_i)$ as the sum of a seasonal and a cyclical component. In particular,

$$c_\tau^i(x_i) = \sum_{s=1}^{4} d_s^i(x_i)\mathrm{I}_s(\tau) + d_b^i(x_i)\psi_{2,b}(\tau), \tag{15}$$

[12] More specifically, we first linearly transform the calendar time domain to the domain of orthogonality of the Chebyshev polynomial, $[-1, 1]$, by means of $\hat{\tau}(\tau) = 2(\tau - \tau_0)/(n_\tau - 1) - 1$, where n_τ is the number of calendar time periods considered. The series of orthogonal polynomials is then generated by (see Abramowitz and Stegun, 1970, Table 22.3) $f_0(\tau) = 1$, and $f_k(\hat{\tau}) = k/2 \sum_{i=0}^{[k/2]} (-1)^i (k-i-1)!/(i!(k-2)!)(2\hat{\tau})^{k-2i}$ for $k = 1, 2, \ldots, 5$.

where the parameter sets $d_s^i(x_i)$ and $d_b^i(x_i)$ denote the effect of the season and the business cycle, respectively, on the distribution of X_i in the inflow into unemployment at calendar time τ. The $d_s^i(x_i)$ parameters include the constant term for $c_\tau^i(x_i)$ as a function of τ.

Note that the function $\psi_{2,b}$ is thus assumed to affect the business cycle dependence of the composition of the inflow into unemployment. However, we do not impose that this effect is in any sense equal or proportional to the direct effect of $\psi_{2,b}$ on the individual exit probabilities. The parameters $d_b^i(x_i)$ are unknown and are to be estimated. Moreover, we allow for a different business cycle effect for each covariate in the inflow (in the application this amounts to 9 parameters). The reason for not introducing a separate polynomial specification for the dependence of the composition of the inflow on the business cycle is purely practical: such a separate polynomial would increase the number of parameters even more.

Now consider the distribution of unobserved heterogeneity on a micro level. We take this to be discrete with unrestricted mass point locations (or points of support). The latter are denoted by v_j and the associated probabilities by $\Pr(v=v_j)=p_j$, where $0 \leqslant p_j \leqslant 1$ and $\sum p_j=1$. Discrete mixture distributions are flexible and attractive from a computational point of view.[13]

3.5. Some remarks on identification

We start by examining the case in which there is no unobserved heterogeneity at the micro level. We normalize the components of the individual exit probabilities by imposing $\psi_{1,1}=1$, $\omega_1=0$, and $\eta_0=\eta_2-\eta_4$. The latter ensures that $\psi_{2,b}=1$ in the calendar-time mean in the sample.

It is obvious that if the time span of the micro data is sufficiently long then the micro data identify the full model. In general, the micro duration data conditional on x identify $\psi_1, \psi_{2,s}$ and β.[14] The micro inflow data identify $d_s^i(x_i)$ (which includes the constant term for $c_\tau^i(x_i)$ as a function of τ) and the $\alpha_{\ldots}$ parameters of the joint distribution of the covariates.

The micro duration data conditional on x also contain information on the function $\psi_{2,b}$ on the time interval covered by the micro-data sample. Similarly, the micro inflow data contain information on the parameters $d_b^i(x_i)$, from a comparison of the inflow distribution of $X|\tau$ and $\psi_{2,b}(\tau)$ on the time

[13] We do not allow the distribution of v in the inflow to depend on the moment of inflow, because in the absence of a much longer micro sample identification of this dependence would strongly rely on functional form assumptions. If we would observe micro data for a much longer time span, we could easily allow the distribution of v in the inflow to vary over the cycle, in particular since for many individuals we observe multiple unemployment spells.

[14] Indeed, the model is overidentified to the extent that interactions between e.g. duration dependence and covariate effects are identified as well.

interval covered by the sample. However, recall that the latter interval is rather short. In particular, it is shorter than a full business cycle. This means that from the micro data it is difficult to obtain estimates of the shape of $\psi_{2,b}$ and the values of $d_b^i(x_i)$ that are not strongly dependent on functional form assumptions. To advance on this, consider the macro data. The quarterly exit probabilities $\Theta(t\,|\,\tau)$ can be thought of as being complicated functions of the elapsed duration t, the current calendar time τ, and the moment of inflow $\tau - t$. Obviously, one cannot identify the separate effects of t, τ and $\tau - t$ on an observable without any functional form restrictions. However, there is no need to impose such restrictions, since the (duration dependence) effect of the elapsed duration t has already been identified from the micro data. Thus, the macro data allow identification of the effects of τ and $\tau - t$, which translates into identification of both business cycle effects over the whole macro-data time interval. In particular, the effect of τ on $\Theta(t\,|\,\tau)$ identifies the shape of $\psi_{2,b}$ over the whole macro-data time interval, while the effect of $\tau - t$ on $\Theta(t\,|\,\tau)$ identifies the compositional effect of the distribution of $X|(\tau - t)$ on the whole macro-data time interval.

Of course, the effect of the distribution of $X|(\tau - t)$ is only captured to the extent to which it is revealed in the distribution of $\exp^{X'\beta}|(\tau - t)$. Identification of the effect of the business cycle on the distribution of $\exp(X'\beta)$ does not entail identification of all effects of the business cycle on the full distribution of X in the inflow. The estimates of the $d_b^i(x_i)$ parameters (which capture the business cycle effect on the full distribution of X in the inflow) may therefore be sensitive to the choice of time interval for the micro sample. Together, however, these parameters capture the effect of the business cycle on the distribution of $\exp(X'\beta)$, and this effect is well-identified. In our discussion of the results we will therefore not focus on the estimates of the separate $d_b^i(x_i)$ parameters, but rather on the implied behavior of the distribution of $\exp(X'\beta)$ over the cycle.

Finally, consider the presence of unobserved heterogeneity at the micro level. Here we exploit the fact that the micro data provide multiple unemployment spells for some respondents. Honoré (1993) shows that multiple spells enable identification of Mixed Proportional Hazard models under weak assumptions if the individual heterogeneity term is fixed across spells.

Note that some parameters, like those describing seasonal effects, are identified from either the micro and the macro data. These overidentifying restrictions are used for specification tests.

4. Data description

In this section we describe the micro and macro samples. The original micro database contains 27,962 individuals. We select men who reported inflow

into unemployment at least once during the observation period from April 1990 to March 1993. We create a so-called inflow sample of unemployment durations: we only include spells starting within this period. This avoids initial conditions problems (see e.g. Lancaster, 1990). The resulting sample consists of 1536 men, who experienced 2192 spells of unemployment. For 457 individuals more than one spell of unemployment is observed. The maximum number of unemployment spells experienced by a single individual is 7.

As has been mentioned above, at each interview the respondents describe their labor market history of the past 12 months. Consequently, two answers are available on the labor market state in March 1991 (and also March 1992): the answer given at the March 1991 (1992) interview and the retrospective answer given at the March 1992 (1993) interview. In approximately 10% of all cases the two answers differ. It is clear that individuals who often change between labor market states are more likely to make such recall or memory errors. Such individuals are also more likely to experience at least one spell of unemployment. Our sample contains 490 unemployment spells with at least one recall error, which is approximately 22% of the total number of spells in the sample.

Most of the studies that use the recent French Labor Force Survey data refer to the existence of the recall errors. Lollivier (1994a) excludes spells containing recall errors from the sample, whereas D'Addio (1997) and Magnac (2000) neglect the recall errors in the analysis. However, neglecting recall errors leads to large outflow in March while excluding the spells is selective in a sense that presumably many spells that end in the period shortly after March are excluded. Magnac and Visser (1999) focus on recall errors more in general. They assume an underlying Markov chain describing the true transition process between the labor market states and assume that the data are observed with a measurement error of which the variance depends on the time to the next interview. Note that our true transition process may not be a Markov chain because of duration dependence. For simplicity, we here apply a solution which is in line with Van den Berg (1990). Like Magnac and Visser (1999), we assume that if the two answers on the labor market state in March differ, then the retrospective answer is incorrect and the other answer is correct. By assumption we rule out that transitions between labor market states can be forgotten, so we assume that in case of inconsistency the transition occurs in the period shortly after March. Now, we distinguish between recall errors at the end of an unemployment spell and recall errors at the beginning of an unemployment spell. If a recall error is observed at the end of a spell we assume that with a probability of 0.35 the transition out of unemployment occurs in March, with a probability of 0.2 in April, with a probability of 0.2 in May, with a probability of 0.15 in June and with a probability of 0.10 in July. This probability distribution is chosen arbitrarily, but we found that our results are insensitive to modest changes in it. We follow a

similar procedure for recall errors at the beginning of a spell, taking account of the fact that the spell may be observed to end shortly after an interview date. After correcting for the recall errors we have verified the consistency of the data, i.e. all spells have a positive duration and a new unemployment spell does not start before the previous spell finishes.

From the first interview in March 1991 we select a number of personal characteristics that are assumed to be time-constant over the period April 1990 –March 1993. The set of characteristics contains indicator functions for living in the agglomeration Paris, having a non-French nationality, being married, and having children. Furthermore, age at March 1991, level of education, and profession are divided into three categories each, for which we include dummy variables. Some of the previous studies mentioned in the introduction find that the distribution of the individual-specific reason of inflow into unemployment changes substantially over the cycle, and Davis et al. (1996) even argue that the latter is an important determinant of the cyclical variation in durations and the unemployment rate. Our micro data do not contain a variable with exactly the same definition as used in those studies (that definition distinguishes between layoffs, quits, job losers, new entrants and re-entrants). We do however observe the labor market state before entering unemployment, and we include this in x. Note that this state is a spell-specific characteristic. We distinguish between 4 categories: (i) inflow after permanent employment, (ii) after temporary employment, (iii) after being a student or in military service, and (iv) after any other non-participation state. Table 1 provides a brief summary of the sample.

Now let us turn to the macro data. Fig. 1 shows the over-all quarterly exit probability of leaving unemployment over the macro sample period.[15] Clearly, this exit probability varies over calendar time. Between 1987 and 1990 the exit probability is higher than in the period before that, and it again decreases after 1990. This follows the conventional macro-economic business cycle indicators for France, like for example real GDP growth per year or capacity utilization rate. Note from the figure that the seasonal effects dominate the cyclical effects.

We may compare the raw duration distributions in both data sets for the individuals flowing in at a particular quarter. Specifically, from both the micro and the macro data we select the cohort of individuals who were unemployed in June 1990 for less than 3 months. For both of these we compute the Kaplan–Meier estimate of the survivor function after June 1990. These are plotted in Fig. 2. The survivor function of the micro data is slightly higher than the survivor function of the macro data. This suggests higher exit probabilities for the macro data.

[15] Here we include individuals in duration classes corresponding to more than 12 quarters, whereas in the estimation we only use the first 12 quarters.

Table 1
Summary statistics on the personal characteristics in the micro data

Inhabitant	
Paris	16%
Other	84%
Nationality	
French	89%
Non-French	11%
Marital status	
Married	41%
Not married	59%
Age	
15–30	50%
31–45	33%
46–65	17%
Education	
High	8%
Intermediate	8%
Low	84%
Children	
Children	47%
No children	53%
Profession	
Civil servant and high skill	28%
Intermediate skill	45%
Low skill and farmers	27%
Labor market state before inflow	
Temporary employment	39%
Permanent employment	43%
Student/Military service	12%
Other	6%
# Individuals	1536
# Spells	2192

For a more formal description of the differences between the micro data and the macro data, we aggregate the micro data by computing the numbers of individuals who are unemployed for a certain number of quarters t at the end of a certain quarter τ. These are the counterparts of the $\tilde{U}(t \mid \tau)$ values that are observed in the macro data, and they can be used to calculate the counterparts

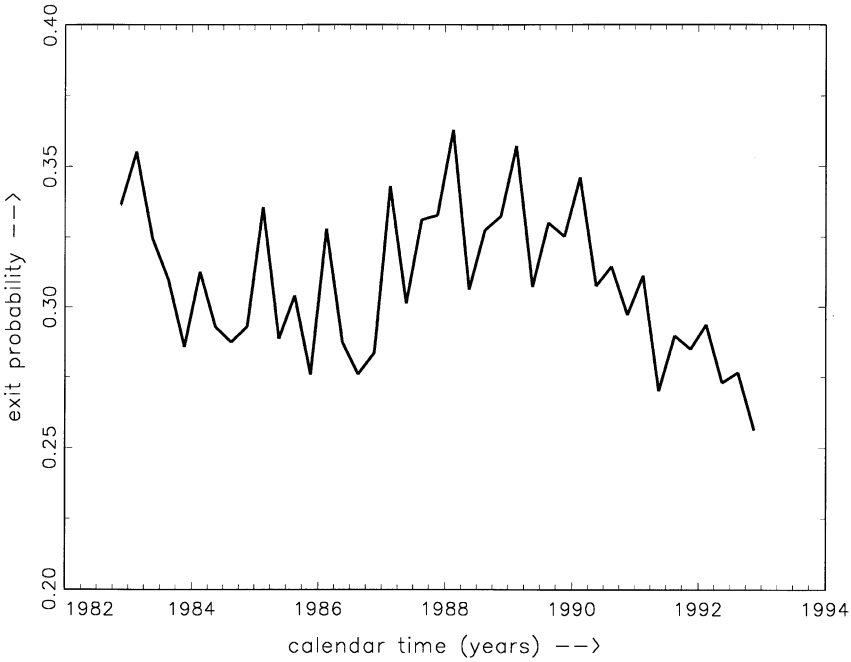

Fig. 1. The quarterly over-all exit probability in the macro data.

of the exit probabilities $\tilde{\Theta}(t \mid \tau)$. We regress the difference between the macro exit probability and the micro exit probability on an intercept and the elapsed quarterly duration t, and we include dummies for the season at τ and the season at $\tau - t$ (i.e., at the moment of inflow into unemployment). The results are in Table 2. The parameter estimates are jointly insignificant and relatively small. 'Leaving unemployment during the third season' is the only significant variable, although its effect is small.

5. Estimation of the full model

5.1. Preliminary issues

For computational reasons, we omit from x those personal characteristics that turned out to be insignificant in a duration analysis of the micro data.[16]

[16] This concerns 'living in Paris', 'having children' and profession dummies. In this analysis with the micro data we do not allow for cyclical effects, but the model is otherwise the same as the full model. See Van den Berg and van der Klaauw (1998) for the estimates.

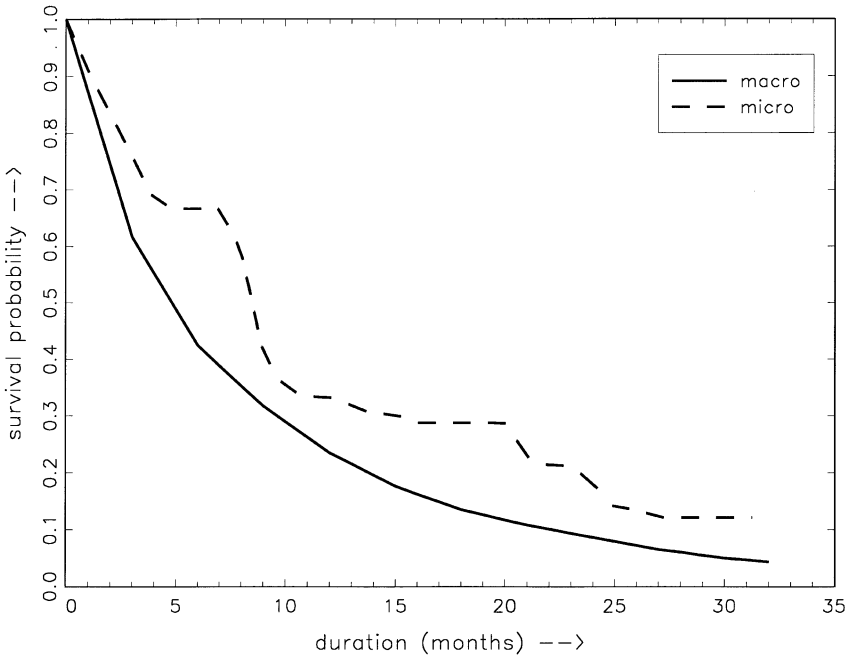

Fig. 2. Kaplan–Meier estimate of the outflow in the micro and the macro data of individuals who were unemployed in June 1990 for less than 3 months.

As a result, x consists of indicators of nationality, age, being married, education, and the state before inflow into unemployment.[17]

Recall from Section 4 that during the period for which we observe both micro and macro data, the over-all macro exit probabilities are higher than the over-all aggregated micro exit probabilities. To investigate whether there is a systematic difference in the levels of the corresponding individual exit probabilities, we allow the $\theta(t \mid \tau, x, v)$ appearing in the macro expressions to

[17] As explained below, the full model has 88 unknown parameters. Each additional personal characteristic would give rise to at least 11 additional parameters. If one would want to include a large number of variables in x then it is advisable to impose some restrictions on the way in which the inflow composition varies over seasons, e.g. by merging seasons or imposing that these seasonal effects are the same for certain variables in x. This could be guided by a separate analysis of the micro data. An interesting topic for further research would be to investigate whether Bayesian data augmentation methods could be helpful as an alternative way to overcome the computational obstacles associated with a large number of variables in x (Tanner and Wong, 1987).

Table 2
Estimation results for the OLS regression of the difference between the quarterly exit probability in the macro data and the corresponding quarterly exit probability in the aggregated micro data[a]

Intercept	0.045	(0.040)
Quarterly duration	−0.0019	(0.0058)
Contemporaneous season (*i.e.*, *at* τ)		
Second season	0.0031	(0.032)
Third season	0.098	(0.039)
Fourth season	−0.018	(0.036)
Season at the moment of inflow (*i.e.*, *at* $\tau - t$)		
Second season	−0.021	(0.032)
Third season	−0.013	(0.035)
Fourth season	−0.037	(0.037)

[a] *Note*: Standard errors are in parentheses.

differ from those in the micro expressions, as follows:

$$\theta_{\text{micro}}(t \mid \tau, x, v) = 1 - \exp(-\psi_1(t)\psi_2(\tau)\exp(x'\beta)\exp(v)),$$

$$\theta_{\text{macro}}(t \mid \tau, x, v) = 1 - \exp(-\psi_1(t)\psi_2(\tau)\exp(x'\beta)\exp(v)\exp(\delta)).$$

The unknown parameter δ gives the relative difference in the exit rates of the underlying continuous-time models. Note that θ_{micro} above is specified as in Section 3.1.[18]

As noted in Section 2.2, the procedure of collecting the data changed in late 1986, and as a result, the time series on $\tilde{U}(t \mid \tau)$ exhibit ruptures at 1986.IV. This turns out to be particularly important for the series on $\tilde{U}(0|\tau)$. We therefore add to the model a dummy variable $d(\tau)$ which is one if and only if τ is before 1987. Specifically, we multiply the expressions for $\Theta(0|\tau)$ in the corresponding model equations by $(d_{<'87})^{d(\tau)}$, in which $d_{<'87}$ is a parameter to be estimated. The results turn out to be insensitive with respect to small changes of the calendar time point defining the areas in which the dummy variable equals zero and one, respectively.

The results below are conditional on v having two mass points v_1 and v_2.[19] The unknown parameters in the model are $\psi_{1,i}$ $(i = 2, \ldots, 11)$, η_i $(i = 1, \ldots, 5)$, ω_s $(s = 2, 3, 4)$, β, $\alpha_{i_1 \cdots i_5}$ $((i_1, \ldots, i_5) \in V)$, the parameter sets $d_s^i(x_i)$

[18] We also estimated an alternative specification in which δ is a multiplicative factor in the individual monthly exit probabilities; $\theta_{\text{macro}}(t \mid \tau, x, v) = \exp(\delta)\theta_{\text{micro}}(t \mid \tau, x, v)$. This gave similar conclusions.

[19] We allowed for additional mass points, but these invariably converged to one of the others during the iterations of the estimation procedure. This also occurred when we adopted procedures similar to Baker and Melino (2000).

and $d_b^i(x_i)$, v_1, v_2, p_1, σ, $d_{<'87}$ and δ. We estimate the full model by maximum likelihood (ML), where the likelihood function is the product of the likelihood functions of the two data sets (see the appendix for a more detailed description of the likelihood function). Note that, as a result of the latter, the likelihood contributions concern drawings from fundamentally different distributions. On the one hand, each individual in the micro data provides a drawing from the joint distribution of personal characteristics and the duration of unemployment (possibly censored, possibly with multiple spells). On the other hand, each calendar time period in the macro data provides drawings from the distribution of measurement and specification errors (we even allow for correlated drawings here). Both types of drawings are informative on the same set of parameters.[20], [21]

5.2. Estimation results

The parameter estimates are in Table 3. The parameter δ, which indicates the level difference between the macro and the micro exit probabilities, is significantly larger than zero. This implies significantly larger exit probabilities in the macro data, which is consistent with the results found in Section 4. The individual exit probability is about 1.3 times larger in the macro data than in the micro data. As noted above, this may be because of errors in the measurement of transitions in either data set, or because of systematic differences in the underlying populations. We return to this below.

The estimated duration dependence ($\psi_1(t)$) is such that during the first 9 months the individual exit probability decreases. Between 9 and 24 months it slowly increases, and after 24 months it increases up to a level that is above the initial level. However, for the higher durations the standard errors are quite large.

[20] The usual asymptotic results for ML estimators hold in many cases where the separate contributions are not identically distributed. It is important that asymptotically the separate contribution of a single observation becomes ignorable. See Gnedenko and Kolmogorov (1954) for details. From another point of view, if both samples are drawn simultaneously then one may define a single drawing as a joint observation of one calendar time period in the macro data and say 1000 respondents in the micro survey data. Such an interpretation of the data-collection process is in fact not unreasonable if both data sets are collected for the single purpose of studying unemployment in all of its facets.

[21] If data from fundamentally different sources are used to study the same set of parameters then the Bayesian approach to statistical inference can be fruitfully applied. In Van den Berg and van der Klaauw (1998) we show that the ML approach for estimation of the full model is equivalent to a Bayesian estimation method. In the Bayesian approach we start with a noninformative prior distribution, and this is subsequently updated with the likelihoods of the macro and micro data sets.

Table 3
Estimation results for the full model[a]

Duration dependence $\psi_1(t)$		
$\psi_{1,1}$	1	
$\psi_{1,2}$	0.86	(0.039)
$\psi_{1,3}$	0.75	(0.035)
$\psi_{1,4}$	0.77	(0.047)
$\psi_{1,5}$	0.89	(0.065)
$\psi_{1,6}$	0.90	(0.084)
$\psi_{1,7}$	0.90	(0.10)
$\psi_{1,8}$	0.92	(0.12)
$\psi_{1,9}$	1.02	(0.14)
$\psi_{1,10}$	1.13	(0.16)
$\psi_{1,11}$	1.19	(0.16)
Contemporaneous cyclical effect $\psi_{2,b}(\tau)$		
η_1	−0.29	(0.043)
η_2	0.061	(0.016)
η_3	−0.088	(0.012)
η_4	0.020	(0.0094)
η_5	0.036	(0.0053)
Contemporaneous seasonal effect $\psi_{2,s}(\tau)$		
ω_1	0	
ω_2	0.15	(0.024)
ω_3	0.050	(0.021)
ω_4	0.016	(0.025)
Observed personal characteristics β		
Non-French	−0.36	(0.11)
Married	0.23	(0.078)
Age 31–45	−0.29	(0.075)
Age 46–65	−0.74	(0.098)
High education	−0.0092	(0.11)
Intermediate education	0.22	(0.10)
Labor market state before inflow:		
Temporary employment	0.54	(0.12)
Permanent employment	0.18	(0.11)
Student/military service	0.22	(0.14)
Unobserved heterogeneity		
v_1	−3.86	(0.36)
v_2	−2.10	(0.13)
p_1	0.055	(0.030)
p_2	0.94	(0.51)
δ	0.27	(0.032)
$d_{<'87}$	0.80	(0.030)
σ	0.035	(0.0013)

Table 3 (*Continued*).

Joint distribution of the observed heterogeneity $X\|\tau$		
Joint-dependence parameters		
α_{11000}	0.24	(0.049)
α_{10100}	0.094	(0.040)
α_{10010}	−0.081	(0.052)
α_{10001}	−0.020	(0.028)
α_{01100}	0.69	(0.053)
α_{01010}	0.018	(0.041)
α_{01001}	−0.017	(0.026)
α_{00110}	0.017	(0.034)
α_{00101}	−0.10	(0.024)
α_{00011}	0.071	(0.026)
Seasonal effect $d_s^i(x_i)$ on threshold values		
Non-French (season 1)	1.90	(0.44)
Non-French (season 2)	2.04	(0.45)
Non-French (season 3)	2.16	(0.45)
Non-French (season 4)	2.03	(0.45)
Married (season 1)	−0.32	(0.65)
Married (season 2)	−0.43	(0.68)
Married (season 3)	−0.044	(0.67)
Married (season 4)	−0.32	(0.66)
Age 31–45 (season 1)	−0.0095	(0.51)
Age 31–45 (season 2)	−0.090	(0.53)
Age 31–45 (season 3)	0.32	(0.52)
Age 31–45 (season 4)	0.23	(0.52)
Age 46–65 (season 1)	1.67	(0.39)
Age 46–65 (season 2)	1.55	(0.40)
Age 46–65 (season 3)	2.12	(0.39)
Age 46–65 (season 4)	1.92	(0.40)
High education (season 1)	0.28	(0.49)
High education (season 2)	0.18	(0.50)
High education (season 3)	0.30	(0.49)
High education (season 4)	0.31	(0.48)
Intermediate education (season 1)	1.12	(0.54)
Intermediate education (season 2)	1.18	(0.56)
Intermediate education (season 3)	1.13	(0.55)
Intermediate education (season 4)	1.20	(0.54)
Temporary employment (season 1)	−0.27	(0.80)
Temporary employment (season 2)	−0.089	(0.81)
Temporary employment (season 3)	−0.15	(0.83)
Temporary employment (season 4)	−0.63	(0.82)
Permanent employment (season 1)	0.59	(0.34)
Permanent employment (season 2)	0.54	(0.35)
Permanent employment (season 3)	0.72	(0.34)
Permanent employment (season 4)	0.68	(0.34)
Student/military service (season 1)	0.77	(0.53)
Student/military service (season 2)	0.61	(0.55)
Student/military service (season 3)	0.28	(0.54)
Student/military service (season 4)	0.68	(0.53)

Table 3 (*Continued*).

Cyclical effect $d_b^i(x_i)$ *on threshold values*		
Non-French	0.87	(0.51)
Married	−0.76	(0.77)
Age 31–45	0.088	(0.60)
Age 46–65	0.54	(0.45)
High education	−0.91	(0.57)
Intermediate education	−0.39	(0.63)
Temporary employment	1.55	(0.93)
Permanent employment	0.92	(0.39)
Student/military service	−0.73	(0.62)
Log likelihood	−11122.38	

[a] *Note*: Standard errors in parentheses.

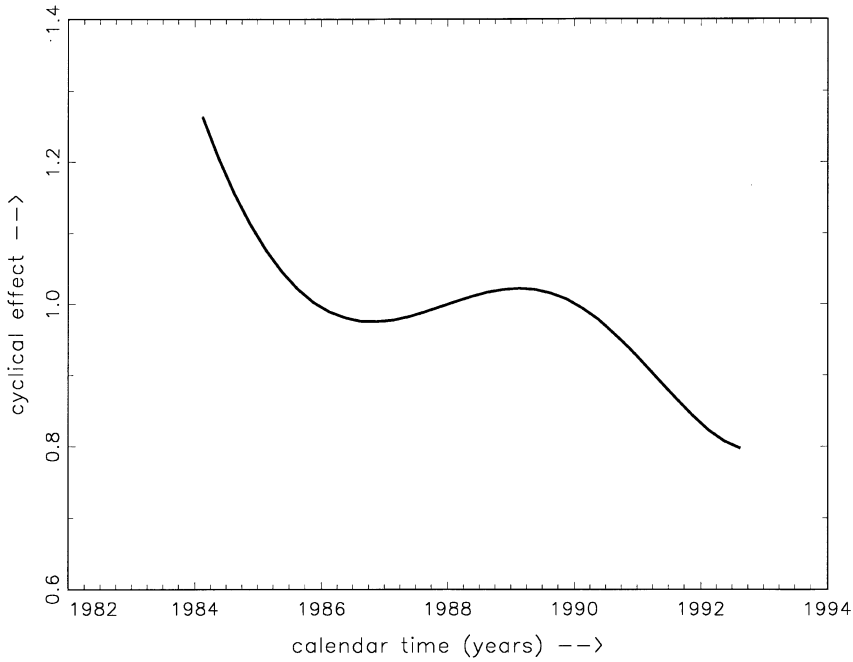

Fig. 3. The estimated cyclical variation of the individual exit probability ($\psi_{2,b}(\tau)$).

In Fig. 3 we depict how the estimated contemporaneous cyclical effect ($\psi_{2,b}(\tau)$) changes over calendar time.[22] The contemporaneous effect includes

[22] Recall that we use polynomials to specify this effect. Polynomials ultimately go to plus or minus infinity, and as a result of this the fit at the borders of the macro-data time interval can be bad. We therefore omit the parts of the graph near these borders.

a downward trend, so if there is no variation in the composition of the inflow then the exit probabilities have generally decreased between 1982 and 1993. We only observe a slight increase in the period that runs from 1986.II to 1989.III. It should be noted that the estimated function $\psi_{2,b}(\tau)$ closely follows the conventional macro-economic business cycle indicators for France, like for example real GDP growth per year or capacity utilization rate.

Before we discuss cyclical variation in the composition of the inflow, we first examine the estimated effect of the personal characteristics on the individual exit probability, and their joint distribution in the inflow. The estimates of β imply that older individuals have a lower individual exit probability, whereas individuals who have the French nationality, are married, or have intermediate education, have a higher individual exit probability, as have individuals who entered unemployment after a temporary job. Not surprisingly, the estimates of β are very similar to those obtained by a separate estimation with the micro duration data where we ignore cyclical effects. They are also very similar to those in D'Addio (1997) and Lollivier (1994a), who use French Labor Force Survey data to estimate unemployment duration models.

The estimated joint distribution of personal characteristics in the inflow fits the micro data well. We performed Chi-square goodness-of-fit tests by comparing the empirical distribution of X to the estimated distribution. A joint test that incorporates all possible cells (3 years times 4 seasons times 144 possible realizations of X) is unfeasible because of the large number of empty cells. We therefore performed separate tests for each possible pairwise combination of elements of X. These are performed separately for each of the 12 quarters in the micro sample period, as well as for the full micro sample period. The results are in Table 4. Most of the tests accept the null hypothesis of a correct specification.[23] The bivariate distributions of 'labor market state before inflow' with 'married' and 'age', respectively, are not well fitted. We over-estimate the extremely small numbers of married and older individuals who enter unemployment from the educational system or from military service. This may be solved by expanding the set V (see Eq. (8)) of the Hermite density, because then areas with almost zero probability can be generated (see Gabler et al., 1993). See also footnote 25.

A formal test of cyclical variation in the composition in the inflow amounts to a joint test of $d_b^i(x_i)=0$ for every i and for every x_i. The Likelihood Ratio test statistic equals 37.8. Since the model under the alternative hypothesis contains 9 additional parameters, we reject the null hypothesis at conventional

[23] In addition, the estimated distribution picks up the correlations between the characteristics in the data, and it captures the differences between the seasons.

Table 4
χ^2-tests for the goodness of fit of the composition of the inflow. These tests are performed for cross-tables of all combinations of two observed personal characteristics[a]

Non-French × Married	0.5	(8)	0
Non-French × Age	1.1	(13)	1
Non-French × Education	1.6	(13)	0
Non-French × Labor market state before inflow	5.6	(14)	1
Married × Age	1.1	(13)	2
Married × Education	11.0	(13)	2
Married × Labor market state before inflow	272.7	(22)	10
Age × Education	7.8	(16)	0
Age × Labor market state before inflow	353.7	(22)	9
Education × Labor market state before inflow	7.8	(22)	1

[a] *Note*: The first row number is the χ^2-test statistic over the full micro data period. The number within parentheses is the 95% critical value. The third number is the number of times the tests per quarterly interval result in a rejection, out of a total of 12 quarterly intervals.

levels of significance. We conclude that the effect of cyclical variation in the composition of the inflow is significant.

Now let us turn to the business cycle effect on aggregate durations that works through the composition of the inflow. The best indicator of this is the way in which the estimated mean covariate effect on the exit probability changes over the cycle. The mean covariate effect at calendar time τ equals

$$\mathrm{E}_{x|\tau}\left[\exp(X'\beta)\right] = \sum_x \exp(x'\beta)\Pr(X = x|\tau). \tag{16}$$

This can be estimated by substituting the estimated β and the estimated distribution of X in the inflow, including the way this changes with the cycle (we suppress seasonal variation here by imposing the average seasonal effect in the distribution of X in the inflow). Fig. 4 depicts how the indicator of the compositional effect varies over τ. Again we neglect the areas near the borders of the macro-data time interval. It is clear that, on average, individuals who enter unemployment in a boom are (a bit) more disadvantaged than the individuals who enter unemployment during a recession. Note that this goes against Darby et al. (1985) and Davis et al. (1996), who argue that individuals entering in a recession are more disadvantaged. The graphs of indicators for single covariates as functions of the moment of inflow are not very informative: the functions for covariates with a positive effect on exit are all marginally increasing on the macro-data time interval, and it is difficult to eyeball any cyclical effect.

We are now in a position to compare both cyclical effects in order to find out which one dominates. We examine the aggregate probability that someone who enters unemployment at the starting date τ of a quarter exits

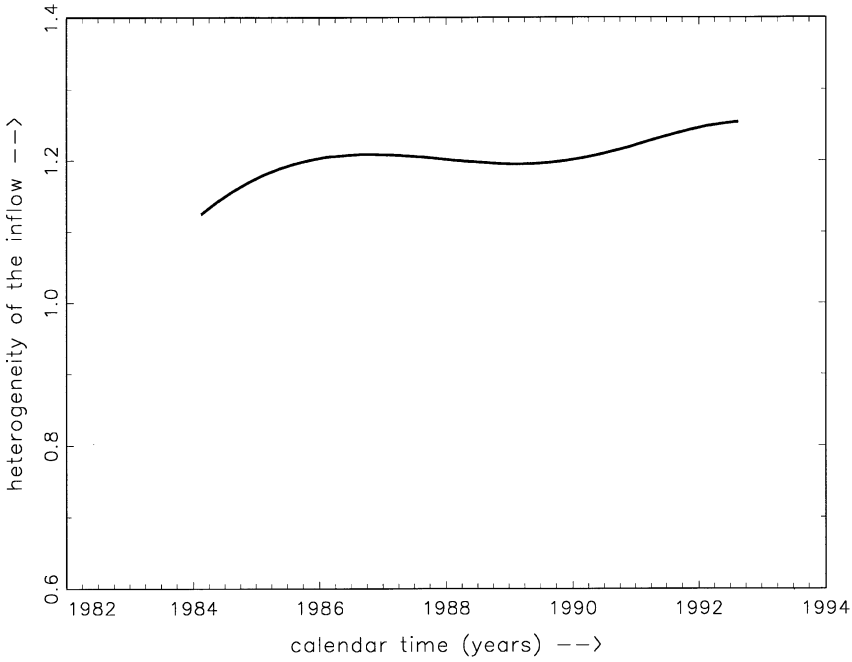

Fig. 4. The estimated cyclical variation of the indicator of the compositional changes in the inflow.

within 3 months. The solid line in Fig. 5 plots the estimate of this probability as a function of τ (again, we suppress seasonal variation by imposing average seasonal effects in the individual exit probability as well as in the distribution of X in the inflow). The dashed line plots the same probability, but now it is imposed that there is no contemporaneous cyclical effect (i.e., $\psi_{2,b}(\tau)$ is fixed at its mean level in $\psi_2(\tau)$ but not in $c_\tau^i(x_i)$). This means that the compositional effect is the only remaining cyclical effect left in the model. The dotted line again plots the aggregate probability, but now it is imposed that there is no variation in the composition of the inflow. In the latter case, the contemporaneous effect is the only cyclical effect left in the model. The figure clearly shows that the contemporaneous effect $\psi_{2,b}(\tau)$ explains almost all cyclical variation in the probability of leaving unemployment within 3 months. In contrast, the cyclical variation due to compositional changes in the inflow does not explain the variation in this exit probability at all. It should be noted that this result also holds for exit probabilities out of other duration classes than the class from zero to 3 months. We also examined the exit probabilities in

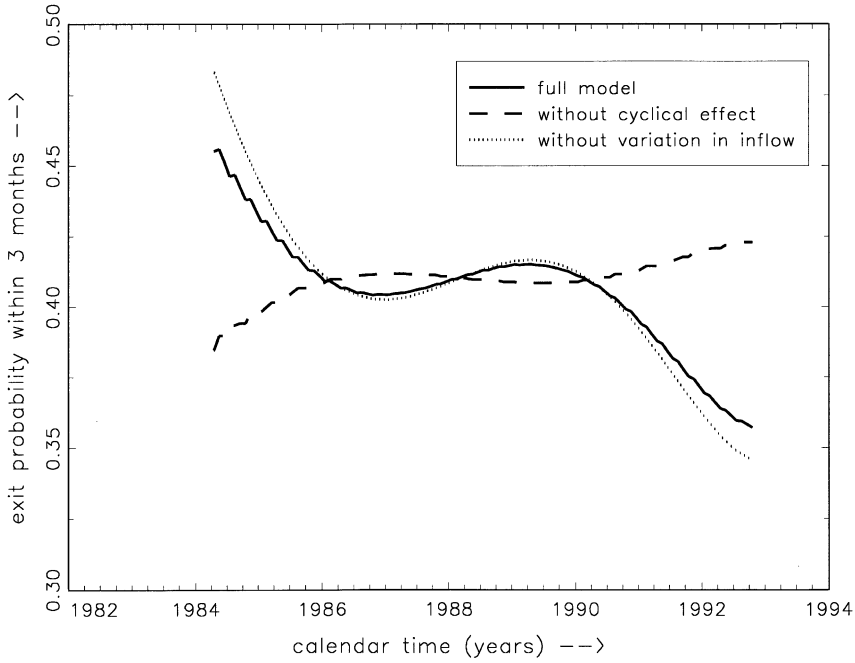

Fig. 5. The estimated probability of leaving unemployment within 3 months.

cases where only a subset of the personal characteristics is imposed to have a time-invariant inflow distribution. The results confirm the above conclusion.[24]

It thus seems that the effect of cyclical variation in the composition of the inflow is quantitatively unimportant. One may wonder whether this is due to the distribution of X in the inflow being flat over time, or due to the individual exit probabilities $\theta(t \mid \tau, x, v)$ being insensitive to x. Because of the non-linearity of the model, this cannot be straightforwardly answered. The magnitudes of the β estimates, which govern the sensitivity of $\theta(t \mid \tau, x, v)$ with respect to x, are plausible and in line with what

[24] Note that the model only allows for cyclical variation in the composition of the inflow if there is variation in $\psi_{2,b}(\tau)$ (see Eq. (15)). To investigate the sensitivity to this, we examined a more general model specification. In particular, the contemporaneous cyclical effect is specified as $\psi_2(\tau)=\psi_{2,s}(\tau)(\psi_{2,b}(\tau))^{\kappa}$. It is clear that if $\kappa=0$, then $\psi_2(\tau)$ does not display cyclical variation even if $\psi_{2,b}(\tau)$ varies over τ, which is necessary for variation in the composition of the inflow. However, we were not able to estimate this model. During the ML iterations, the values of κ, $d_b^i(x_i)$, and the parameters η_i of $\psi_{2,b}(\tau)$ did not converge even though the likelihood value did not improve in comparison to the value of the estimated model with $\kappa = 1$. This suggests that κ is not well identified, and the specification with unrestricted κ is too general.

is typically found in the literature. For example, the estimated individual exit probability of an individual aged 25 is twice as large as the probability of an individual aged 50 who is otherwise identical. If we multiply all the β estimates with a factor 5 then the compositional effect becomes of the same order of magnitude as the direct contemporaneous effect, in Figs. 3–5. If we multiply them with a factor 10 then the compositional effect dominates. (Of course, the compositional effect then still has the opposite sign as the direct contemporaneous effect and the over-all effect.) This suggests that one would need implausibly large values of β to obtain a sizeable compositional effect, or, in other words, that the weakness of the compositional effect is not due to the individual exit probabilities $\theta(t \mid \tau, x, v)$ being insensitive to x. The estimated marginal distributions of X in the inflow do not change dramatically over the cycle. Generally, the corresponding probabilities stay within a 10% range of the fractions given in Table 1. This suggests that the weakness of the compositional effect is due to the distribution of X in the inflow being flat over time. However, it should be stressed that these inferences are subject to a qualification. The importance of the over-all compositional effect may actually be under-estimated because of an inability to correctly model or identify the way in which the composition of the inflow changes over time. To obtain more evidence, much longer micro samples would need to be collected.

Keeping this qualification in mind, our results imply that the persistence in unemployment after a negative shock is not primarily due to an inflow of disadvantaged workers with low individual-specific exit probabilities. On the contrary, even workers with relatively good qualifications are hampered by a recession if they search for a job. This suggests that policies aimed at bringing the unemployed back to work during a recession should not focus exclusively on workers with the worst qualifications.

Now let us turn to the seasonal effects. Again we distinguish between a contemporaneous effect and an effect working through the composition of the inflow. Concerning the former, the individual exit probabilities are estimated to be highest in the second quarter of the year, when the seasonal effect $\psi_{2,s}(\tau)$ has its highest level, and lowest in the first quarter. Concerning the other effect, we examine the estimated mean covariate effect on the exit probability as a function of the season of inflow, analogous to (16) above. It turns out that this effect is highest in the second half of the year (1.25 for the third and 1.24 for the fourth quarter) and lowest in the first half of the year (1.16 for the first and 1.14 for the second quarter). The seasonal variation in the composition of the inflow mainly works through differences in the age distribution in the inflow. In the second half of the year, the proportion of young individuals in the inflow is on average higher, and these have higher individual exit probabilities.

The estimated standard deviation σ of the measurement errors in the macro data equals 0.035. This is relatively small, so the model fits the macro data well. As expected, the parameter $d_{<'87}$ capturing the change in 1986 in the policy towards youth unemployment is estimated to be smaller than one. Finally, we find significant unobserved heterogeneity on the micro level. This is important, because it means that omission of it from the model would have resulted in biased estimates of the duration dependence, and hence of the cyclical effects (recall the discussion in Section 3.5).[25]

We end this subsection with a test of whether the duration dependence and the contemporaneous seasonal effect are the same in the micro and the macro data. First, we allow the duration dependence in the macro data to differ from the duration dependence in the micro data. The Likelihood Ratio test statistic equals 17.9. Since we introduce 10 additional parameters, we do not reject the null hypothesis that the duration dependence patterns are the same. Second, we allow the contemporaneous seasonal effects to be different in the micro and macro parts of the model. The Likelihood Ratio test statistic equals 17.8 with only 3 additional parameters, so we reject null hypothesis that they are the same. The differences mostly concern the fourth quarter. At that quarter, the macro exit probability is larger than the micro exit probability. We conclude that most of the difference between the macro data and the micro data concerns the level of the exit probability.

6. Conclusion

We combine two types of unemployment duration data. The micro (individual survey) data enable identification of the determinants of the individual duration distribution and the composition of the inflow, but not of how the individual exit probabilities and the inflow composition change over the business cycle. The macro (aggregate time-series) data subsequently enable identification of these cyclical effects, from the way in which the aggregate exit probabilities vary over the time at outflow and the time at inflow. The micro and macro data use somewhat different unemployment definitions. We model this primarily by way of measurement and specification errors in the macro data, and we estimate the model by maximization of the joint likelihood. This approach to dealing with a discrepancy between two data sets may be useful in many instances where micro and macro data are combined, not just in duration examples.

[25] We also estimated the model without the explanatory variable 'labor market state before inflow'. This increases the estimated dispersion of v, but the estimates of the other parameters are insensitive to this. Results are available upon request.

It turns out that the macro and the micro data set are not in serious conflict with each other. The only (identifiable) systematic difference concerns the absolute level of the individual exit probabilities, which is higher for the macro data. In addition, the effect of the fourth season on the exit probability is different. However, the duration dependence pattern and the other seasonal effects are the same for both data.

The estimation results suggest that the countercyclicality of the aggregate mean unemployment duration originates from the fact that the individual exit probabilities vary over the cycle for all types of individuals. We estimate the effect of changes in the composition of the inflow on the cyclical behavior of the mean duration to be small. However, this result could depend on that fact that the micro data cover a relatively short period.

Acknowledgements

This research was stimulated by discussions with Francois Laisney. We thank him for his suggestions and encouragement. Also thanks to Mars Cramer for many comments, including those on the use of maximum likelihood estimation. Jaap Abbring, Denis Fougère, Andrew Chesher, Kenneth Carling, Geert Ridder, the associate editor and two anonymous referees also provided useful suggestions. We are grateful to INSEE (Paris) and to Jean-Marc Robin of the Laboratoire de Microéconometrie at INSEE-CREST for giving permission to use the micro data. The Direction de l'Animation de la Recherche des Etudes et des Statistiques of the Ministère du Travail et des Affaires Sociales of France kindly provided the macro data.

Appendix: The likelihood function

In this appendix we give the likelihood function of the full model. It consists of three parts. In the micro data the individual unemployment durations and the individual characteristics at the moment of inflow into unemployment contribute to the likelihood function. The macro data only contribute to the likelihood function by way of the measurement errors in the quarterly exit probabilities.

We start by describing the part of the likelihood function that corresponds to the micro data. Consider an individual with observed individual characteristics x and unobserved individual characteristics v, who enters unemployment at calendar time τ and leaves unemployment after t months. Conditional on x, v and τ, the likelihood contribution corresponding to the length of the unemployment spell is given by

$$\ell_1(t \mid \tau, x, v) = \theta(t \mid \tau + t, x, v) \prod_{u=0}^{t-1} (1 - \theta(u \mid \tau + u, x, v))$$

(see also Eq. (3)), where $\theta(t \mid \tau, x, v)$ is given in Eq. (4). Right-censoring of the spell of unemployment is considered as exogenous and can therefore be dealt with straightforwardly (see e.g. Lancaster, 1990). The likelihood contribution corresponding to data on observed individual characteristics $x = (x_1, \ldots, x_n)$, when entering unemployment at calendar time τ, is given by

$$\ell_2(x|\tau) = \Pr(X_1 = x_1, \ldots, X_n = x_n|\tau),$$

where $\Pr(X_1 = x_1, \ldots, X_n = x_n)$ is given in Eq. (6). Assume that we observe J spells for this given individual, denoted by $j = \{1, \ldots, J\}$. The length of the spells are given by t_j, the moment of inflow by τ_j, the observed individual characteristics by x_j and the unobserved characteristics by v. Since, we include 'labor market state previous to unemployment' as explanatory variable, x_j can differ between the spells of a single individual. The unobserved component v is assumed to remain constant over all spells of a single individual. The contribution of a given individual to the likelihood function is

$$\ell_{\text{micro}} = \mathrm{E}_v \left[\prod_{j=1}^{J} \ell_1(t_j|\tau_j, x_j, v) \ell_2(x_j|\tau_j) \right].$$

The contribution of the macro data to the likelihood function follows from the probability distribution of the measurement errors in Eq. (14). The measurement errors $e_{t,\tau}$ are normally distributed with mean 0 and variance $2\sigma^2$. The covariance between $e_{t,\tau}$ and $e_{t+3,\tau+3}$ equals $-\sigma^2$, all other covariances are 0. Now let e denote the vector of measurement errors $(e_{0,\tau_0}, e_{0,\tau_0+3}, e_{3,\tau_0+3}, e_{0,\tau_0+6}, e_{3,\tau_0+6}, e_{6,\tau_0+6}, \ldots)'$, where τ_0 is the initial calendar time in the macro data. This vector has mean 0 and let Σ denote the covariance. The contribution of the macro data to the likelihood function is given by multivariate normal density

$$\ell_{\text{macro}} = |\Sigma|^{-1/2} \exp\left\{ -\frac{e'\Sigma^{-1}e}{2} \right\}$$

ignoring the constant $(2\phi)^{-n/2}$, where n is the number of terms in the vector e.

The likelihood function is the product of the contributions of all individuals in the micro data and the contribution of the macro data.

References

Abramowitz, M., Stegun, I.A., 1970. Handbook of Mathematical Functions; With Formulas, Graphs and Mathematical Tables. Dover, New York.

Baker, M., 1992. Unemployment duration: compositional effects and cyclical variability. American Economic Review 82, 313–321.

Baker, M., Melino, A., 2000. Duration dependence and nonparametric heterogeneity: a Monte Carlo study. Journal of Econometrics 96, 357–393.

Bean, C.R., 1994. European unemployment: a survey. Journal of Economic Literature 32, 573–619.

CSERC, 1996. Inégalités d'emploi et de revenu: les années quatre-vingt-dix. CSERC, Paris.

D'Addio, A.C., 1997. Le chomage des jeunes en France. Working Paper, CORE, Louvain.

Darby, M.R., Haltiwanger, J.C., Plant, M., 1985. Unemployment rate dynamics and persistent unemployment under rational expectations. American Economic Review 75, 614–637.

Davis, S.J., Haltiwanger, J.C., Schuh, S., 1996. Job Creation and Destruction. MIT Press, Cambridge, MA.

Devine, T.J., Kiefer, N.M., 1991. Empirical Labor Economics. Oxford University Press, Oxford.

Dynarski, M., Sheffrin, S.M., 1990. The behavior of unemployment durations over the cycle. Review of Economics and Statistics 72, 350–356.

European Commission, 1994. Employment in Europe 1994. European Commission, Brussels.

Gabler, S., Laisney, F., Lechner, M., 1993. Semi-nonparametric maximum likelihood estimation of binary choice models with an application to labor force participation. Journal of Business and Economic Statistics 11, 61–80.

Gallant, A.R., Nychka, D.W., 1987. Semi-nonparametric maximum likelihood estimation. Econometrica 55, 363–390.

Gnedenko, B.V., Kolmogorov, A.N., 1954. Limit Distributions for Sums of Independent Random Variables. Addison-Wesley, Reading, MA.

Honoré, B.E., 1993. Identification results for duration models with multiple spells. Review of Economic Studies 60, 241–246.

ILO, 1982. Statistics of Labour Force, Employment, Unemployment and Underemployment. ILO, Geneva.

ILO, 1989. Statistical Sources and Methods; Volume 4: Employment, Unemployment, Wages and Hours of Work. ILO, Geneva.

Imbens, G.W., 1994. Transition models in a non-stationary environment. Review of Economics and Statistics 76, 703–720.

Imbens, G.W., Lancaster, T., 1994. Combining micro and macro data in microeconometric models. Review of Economic Studies 61, 655–680.

Imbens, G.W., Lynch, L.M., 1992. Labor market transitions over the business cycle. Working Paper. Harvard University, Cambridge.

Lancaster, T., 1990. The Econometric Analysis of Transition Data. Cambridge University Press, Cambridge.

Laisney, F., Lechner, M., 1996. Combining panel data and macro information for the estimation of a panel probit model. Zeitschrift für Wirtschaft- und Sozialwissenschaften 116, 339–358.

Layard, R., Nickell, S., Jackman, R., 1991. Unemployment: Macroeconomic Performance and the Labour Market. Oxford University Press, Oxford.

Le Monde, 1997. Le chômage a baissé de 0,2% en février mais le marché du travail reste atone, 29 March.

Lollivier, S., 1994a. Unemployment duration and business cycle. Working Paper, INSEE-CREST, Paris.

Lollivier, S., 1994b. Fabrication d'un calendrier mensuel d'activité à partir de l'enquête sur l'emploi. Working Paper, INSEE-CREST, Paris.

Magnac, T., 2000. Subsidised training and youth employment: distinguishing unobserved heterogeneity from state dependence in labour market histories. Economic Journal 110, 805–837.

Magnac, T., Visser, M., 1999. Transition models with measurement errors. Review of Economics and Statistics 81, 466–474.

Tanner, M., Wong, W., 1987. The calculation of posterior distributions by data augmentation. Journal of the American Statistical Association 82, 528–540.

Van den Berg, G.J., 1990. Search behaviour, transitions to nonparticipation and the duration of unemployment. Economic Journal 100, 842–865.

Van den Berg, G.J., 2001. Duration models: specification, identification, and multiple durations. In: Heckman, J.J., Leamer, E. (Eds.), Handbook of Econometrics, Vol. V. North-Holland, Amsterdam.

Van den Berg, G.J., van der Klaauw, B., 1998. Combining micro and macro unemployment duration data. Working Paper, Tinbergen Institute, Amsterdam, and INSEE-CREST, Paris.