Combining Micro and Macro Unemployment Duration Data

Gerard J. van den Berg *
Bas van der Klaauw †

August 1998

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.

*Free University Amsterdam, Tinbergen Institute and CEPR.
†Free University Amsterdam and Tinbergen Institute.
Address: Department of Economics, Free University Amsterdam, De Boelelaan 1105, NL–1081 HV Amsterdam, The Netherlands.

Keywords: unemployment composition, heterogeneity, duration dependence, business cycle, seasons, unemployment definition. JEL code: J64, C41, E24, C51.

This research was stimulated by discussions with François 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 and Andrew Chesher 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.

1 Introduction

Unemployment has been a top issue for economic research and policy for many decades. Macro-economic research on unemployment traditionally focuses on the macro unemployment rate and its behavior over the business cycle. However, the recently expanding macro 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, Nickell and Jackman (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, Haltiwanger and Plant (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] Clearly, for reliable estimation of business cycle effects, it is necessary to have data that include at least a complete cycle. 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 as additional regressors (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 specify a model that allows 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

[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 U.S..

by way of a product of a flexible high-order polynomial in calendar time (capturing business cycle effects) and dummy variables capturing seasonal effects. Dependence of individual exit probabilities on the elapsed duration captures genuine duration dependence due to e.g. stigma effects reducing the number of job opportunities of the long-term unemployed (see e.g. Vishwanath (1989) and Van den Berg (1990a)).

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 this 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 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.

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, national statistics count registrations at public employment agencies, whereas labor force surveys produce statistics that are based on self-reported unemployment in surveys of randomly sampled individuals. The latter statistics are usually more in line with the definition of unemployment as given by the International Labour Organization (ILO) (see ILO (1982); see also Section 2 below). In the past years, there has been a shift towards a greater role

for labor force survey data in national unemployment statistics.[2,3] As a result, in most European countries, in every month, two different unemployment statistics are published. The media then take pains to explain why they differ, and if the changes from last month are different for the two measures then that has to be explained as well (see e.g. Le Monde, 29 March 1997). This situation is mirrored in the empirical scientific literature. Macro studies often use time series data based on registered unemployment, whereas micro studies use longitudinal survey data. It is no exaggeration to state that the simultaneous use of the different measures is responsible for a substantial amount of confusion concerning unemployment, in public opinion[4] as well as in the scientific literature.[5]

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 (in Sections 2 and 3 we go into this in detail), and consequently it describes a different set of individuals. The most important difference is that the macro definition does not cover individuals looking for part-time or temporary jobs, whereas the micro definition does. In addition to this, the macro unemployment definition itself has not been time-invariant, and both concepts are imperfectly measured.

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

[2]International organizations like the OECD, the UN and the European Union (Eurostat) use unemployment statistics based on labor force survey data and the ILO definition, because this is the only way to permit comparisons across countries.

[3]Since surveys are often collected only once a year, the survey data are typically combined with monthly available data on registered unemployment in order to track short-term fluctuations around the yearly measure.

[4]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 on the publication of unemployment statistics (CSERC, 1996; Le Monde, 29 March 1997).

[5]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". Baker (1992), in his study on the effect of the business cycle on U.S. unemployment durations, states that "Clearly, a comparative study of inference from grouped and panel data is an important topic for further research".

a deeper level. By estimating the determinants of the duration distributions associated with both measures, we are able to describe and explain to what extent they are dissimilar. The data provide sufficient information on this. For example, we will aggregate the micro data, and compare the duration dependence of the exit rate out of unemployment in these data to that of the exit rate in the macro data from the same calendar time period. Also, we will compare the effects of the season at the moments of inflow and outflow in both data sets. The full model contains a number of overidentifying restrictions. Notably, the genuine duration dependence and the seasonal effects are assumed to be the same in the micro and macro part of the model. In general, we test for the equality of parameters that are supposed to describe the same phenomena.

As noted, there is a number of 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 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.[6]

As a result, empirical inference is non-standard in the sense that the stochastic elements in the micro and macro observations are from different distributions. Nevertheless, the model can be estimated by maximum likelihood methods, maximizing the product of the two likelihood functions associated with the micro and macro data. Indeed, under an alternative interpretation of the data-collection process, the standard conditions for application of the maximum likelihood principle are satisfied. In the Appendix to this paper we also show that our estimation procedure is fully equivalent to a Bayesian estimation procedure with a given loss function.

To estimate the model, it is not necessary to make parametric assumptions about the genuine duration dependence pattern. We simply estimate a parameter for each quarterly duration interval. As a consequence, the results are not

[6]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 (1994) for an application.

subject to the well-known biases originating from misspecification of this pattern. 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.

Note that in case of (unobserved) heterogeneity, individuals with the largest exit probabilities on average leave unemployment first. This leads to a decline in the "average quality" (i.e., the average individual-specific effect in the exit probability) of a cohort of unemployed in the course of time. Thus, negative duration dependence at the aggregate level may occur even in absence of genuine duration dependence at the micro level. This is of importance for policy analysis (see e.g. Layard, Nickell and Jackman (1991) and Van den Berg and Van Ours (1996)). Abbring, Van den Berg and Van Ours (1994) used macro data to distinguish between genuine duration dependence and unobserved heterogeneity. In their case, identification is crucially dependent on the multiplicative structure of the exit probabilities. Note that the macro data do not allow observation of the composition of the inflow into unemployment or the way it changes over time. Perhaps not surprisingly, the latter changes can be identified to a certain extent under rather arbitrary functional-form assumptions. However, as Abbring, Van den Berg and Van Ours (1994) show, any trend in the composition is fundamentally unidentified from macro data. Obviously, in the present paper these limitations are avoided because of the fact that we observe heterogeneity in the micro data. 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.

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 U.S. 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 U.S. 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 nonparticipation)

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, Haltiwanger and Plant (1985) examine U.S. 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.[7] They conclude that changes in the composition of the inflow are a primary determinant of cyclical variations in the exit probability. Baker (1992) and Davis, Haltiwanger and Schuh (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, Haltiwanger and Schuh (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 in a somewhat ad-hoc way 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.

Another branch of relevant literature concerns the empirical literature that uses micro data from the same source as ours (i.e., the French labor force survey) in order to estimate reduced-form unemployment duration models (Moreau and Visser, 1989; Lollivier, 1994a; Magnac and Robin, 1994; Magnac, Robin and Visser, 1995; Magnac and Visser, 1995; Magnac, 1996; D'Addio, 1997). Abbring,

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

Van den Berg and Van Ours (1994) use macro data from the same source as ours. In Section 4 we compare our empirical implementation to that in this literature.

To our knowledge, the only study in which both micro (survey) and macro (registered) unemployment duration data are used to estimate duration models is Albæk and Holm Larsen (1993). They have individual records from both types of data, concerning the same set of Danish individuals. This enables estimation of reduced-form duration models conditional on personal characteristics, with the micro data as well as with the macro data. It turns out that the average duration in the micro data is about twice as large as in the macro data, mostly because the macro data report transitions that are not reported in the micro data. In the estimation results, the main difference is in the constant term in the exit probability out of unemployment. The other estimates (duration dependence, covariate effects) do not differ significantly.

The outline of the paper is as follows. In Section 2 we examine how the raw micro data can be used to construct individual unemployment durations, and how the raw macro data can be used to obtain information on the aggregate unemployment duration distribution. In addition, we examine the definition of unemployment in both data sets in detail. These issues are of importance for the model specification, which is presented in Section 3. In Section 4 we then present summary statistics of the data, and we perform descriptive data analyses. In particular, we estimate separate reduced-form models for the micro data and for the macro data. Section 5 contains the estimation results for the joint model. Section 6 concludes.

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

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.

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, defined as follows:

1. Self-employed, or helping a family member with his or her work
2. Employed, receiving a salary, in a permanent position
3. Employed in a position with fixed duration, or a temporary appointment obtained by way of a commercial employment agency, or an apprenticeship, or seasonal work
4. Working as a paid trainee[8]
5. Unemployed
6. Student, or pupil, or unpaid trainee
7. Military service
8. Other situation: retired, in early retirement, disabled, housewife, and other

States 1–4 are forms of employment states, whereas states 6–8 are non-participation states. 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.

By comparing individual labor market states of consecutive months in the period from March 1990 to March 1993, individual unemployment durations can be constructed; they always consist of an integer number of calendar months. Personal characteristics of the respondent are recorded at the first interview. With this information, unemployment duration models can be estimated in which

[8] In training programs like TUC or SIVP; see Bonnal, Fougère and Sérandon (1997) for more information on these.

the individual monthly exit probability θ depends on the elapsed duration t, calendar time τ, and personal characteristics x.

The (implicit) definition of unemployment used here is similar to the ILO definition. First of all, note that individuals are asked to classify themselves. The ILO definition requires that the individual is (1) without employment, (2) seeking employment, and (3) currently available for employment (see ILO (1982)). The menu above does not explicitly refer to conditions (2) and (3). Indeed, the respondent does not receive any clarifying information (like a more detailed description) about the eight above-mentioned states, before or during the interview. However, the monthly labor market state questions are posed at the end of the survey, and at an earlier stage of the survey it is explicitly established whether the three ILO conditions are fulfilled for the respondent's situation at the interview date.[9] 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)). In any case, 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.

Finally, it should be noted that the 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. 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 individuals 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.

We now turn to the precise definition of unemployment in the macro data. When individuals (voluntarily) register at a public employment agency, they state that they want to work. At the moment of registration, the individual is classified

[9] The latter serves as input for unemployment statistics that are reported in the media and used by international organizations like the European Union (see Section 1).

by the agency into one of five categories:

1. Without employment, immediately available for employment, actively searching employment, seeking permanent full-time employment

2. Without employment, immediately available for employment, actively searching employment, seeking permanent part-time employment

3. Without employment, immediately available for employment, actively searching employment, seeking temporary or seasonal employment or employment for a given time, including very short durations

4. Without employment, not immediately available for employment, seeking employment

5. Employed, seeking other employment

It should be noted that "immediately" here means "within 15 days", and that "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 work. Individuals who have worked more than half of the time during the month can still be registered as being without work (at least, in our sample period; see below). Students seeking work during vacations and individuals who are temporarily laid off for more than four weeks are assigned to Category 3. The classification of a single individual may be revised at any time during the period of registration.

Responsibility for the loss of the last job does not affect registration as an unemployed individual. However, registration with ANPE is a necessary condition for the receipt of any unemployment benefits (with one exception; see below). As we shall see below, this has consequences for the comparability of the micro and the macro data.

The macro data cover only Category 1. Indeed, the number of individuals in Category 1 defines the official ("registered") unemployment statistic, which is the most commonly cited unemployment statistic in the media (see Section 1). However, Categories 2 and 3 presumably include many individuals who would classify themselves as being unemployed. Data on the over-all outflow of individuals from Categories 1–3 show that Categories 2 and 3 are quantitatively less important, in particular for men. For example, in 1994.IV, the male outflow out of Categories 1–3 consisted for 94% of Category 1, for 2% of Category 2, and for 4% of Category 3. For women in 1994.IV, these figures are 87%, 10% and 3%,

respectively. Categories 4 and 5 are quantitatively even less important. Note that Category 4 can be thought of as being a state of nonparticipation, since individuals in this category are not working and cannot make an immediate transition to employment. (This category includes students who look for a job that should start after finishing school.) Category 5 consists of individuals who are currently employed or self-employed.

Reasons for removing an individual out of Category 1 include (in addition to finding suitable work or movement to another category) (a) illness for more than 15 days, (b) failure to comply with register continuation requirements or job search guidelines, (c) participation in training schemes, and (d) being over 55 years of age and exempted from seeking work while continuing to receive unemployment benefits up until receipt of a retirement pension. Temporary unavailability for work due to holidays does not result in deletion.

Since 1982, the registration process and the operationalization of the definitions of the categories have been changed a number of times. This is of importance for our purposes. Below is a list of changes.

1. Before 1983, individual register continuation required reporting in person to the employment agency. Between 1983 and 1985, this updating method has been replaced by reporting by mail. At the same time, registration was computerized (ILO, 1989).

2. All changes in the unemployment benefits system affecting entitlement to benefits and the benefits level can be expected to affect the incentive to register. Major changes have occurred in June 1982 and in April 1983 (ILO, 1989; Ayong Le Kama, 1995).

3. From 1984, older unemployed persons in receipt of unemployment benefits were exempted from seeking employment and from registering. The age limit has since been decreased from 57.5 to 55 years. This currently affects about 250,000 individuals (ILO, 1989; Libération, 9 June 1997).

4. From October 1986, the timing of data collection and statistical processing of the raw data has improved substantially (see ILO (1989) for details). As a result, measurement errors in individual labor market transitions and durations can be expected to have been reduced.[10]

5. From mid-1987, job seekers who were deleted for failure to comply with register continuation requirements or job search guidelines could only be re-registered after a waiting period of three months (ILO, 1989).

[10]The data display a discontinuity around this date (see Abbring, Van den Berg and Van Ours (1994)). We return to this in Section 4.

6. From 1992, measures aimed at maintaining accurate registration of long term unemployed have been intensified, resulting in larger amounts of removals (Libération, 9 June 1997).

7. From 1994, individuals who forget to submit their registration update do not receive a reminder before being deleted from the register. This has reduced total registration by tens of thousands of individuals (Canard Enchaîné, 7 May 1997; Libération, 9 June 1997).

8. From the middle of 1995 onward, Category 1 excludes job seekers who have worked for more than 78 hours during the month. (A new Category 6 was created for such cases.) This has affected about 300,000 unemployed workers in two years time (ILO, 1989; Libération, 9 June 1997).

9. From late 1996, the actual first registration of an individual has to take place at an agency of the Assedic (the agency responsible for payment of unemployment insurance) instead of the ANPE (the employment agency). This seems to have reduced the motivation of individuals without any benefits entitlement (notably young schoolleavers) to register. The inflow into registered unemployment seems to have decreased by 10% because of this (Canard Enchaîné, 7 May 1997; Libération, 9 June 1997).

10. From December 1996, the control efforts aimed at cleaning up the registers have again been intensified (Canard Enchaîné, 7 May 1997; Libération, 9 June 1997).

There is evidence that a number of these changes were initiated by the government in order to decrease the published unemployment figures (Canard Enchaîné, 7 May 1997; Libération, 9 June 1997). Because this seems to be particularly true for the period after 1993, we do not use data after 1993.I. Because the classification into the five categories above 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). Since detailed duration information is unavailable for Categories 2 and 3, we only use data on Category 1. Because the micro unemployment data do include unemployed workers seeking part-time or temporary employment, and because women relatively often search for the latter types of jobs, we only use data on men.

3 The model

3.1 Modelling individual exit probabilities

In the micro data as well as in the macro data, unemployment durations 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 ended. Both t and τ are therefore taken to be discrete variables. Since the micro durations are measured in months and the macro durations in quarters, 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. As will become clear below, in that case the likelihood would be greatly complicated. In sum, our empirical model specification is genuinely discrete.

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 $\theta\left(t|\tau, x\right)$. 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). Suppose for the moment that all heterogeneity across individuals is observable. We assume that $\theta\left(t|\tau, x\right)$ can be written as

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

with ψ_1 and ψ_2 positive. This specification can be derived from a continuous-time Proportional Hazard (PH) model, under some assumptions. Consider a continuous-time PH model with, in obvious notation, individual exit rate $\lambda(t|\tau, x) = \psi_1(t)\psi_2(\tau)\exp(x'\beta)$. 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

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

If ψ_1 and ψ_2 are constant within monthly intervals, this probability equals the expression for $\theta(t|\tau,x)$ of (1). In fact, we 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

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

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

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

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

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

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 $\theta(\cdot|\cdot,x,v)$. For example, $\Pr(T=t|\text{inflow at }\tau;x;v)$ 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 $x_1, \ldots, x_n$ whose values are time-invariant for a given individual. This is not restrictive, because the micro data do not contain continuous explanatory variables and do not show how personal characteristics vary over time. We normalize the model by assuming 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) do not vary over time (for example, a dummy is always either zero or one, and not sometimes zero or one and sometimes one or two). 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. 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 $\{0, 1, 2, \ldots, \overline{x}_i\}$. We assume that the joint distribution of $X_1, \ldots, X_n$ in the inflow at cohort date τ can be written as

$$\Pr\left(X_1 = x_1, \ldots, X_n = x_n | \tau\right) = \int_{c_\tau^1(x_1)}^{c_\tau^1(x_1+1)} \cdots \int_{c_\tau^n(x_n)}^{c_\tau^n(x_n+1)} h\left(u_1, \ldots, u_n\right) du_1 \cdots du_n \tag{6}$$

There are two types of determinants of the right-hand side: the "core density" h on the one hand, and the "threshold values" $c_\tau^i\left(x_i\right)$ 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 $c_\tau^i\left(0\right) = -\infty$, $c_\tau^i\left(x_i\right) < c_\tau^i\left(x_i + 1\right)$, and $c_\tau^i\left(\overline{x}_i + 1\right) = \infty$. 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 prob-

abilities 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 Subsection 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,

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

where S is a normalizing constant ensuring that h integrates to one. We can then normalize further by fixing $\alpha_{0 \cdots 0} = 1$. Since the distributions of observables only depend on intervals of u, the unidentified scale of the density 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 means more flexibility. We take

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

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,

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

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

$$\Pr(X_1 = 0, X_2 = 0) = \Phi(c_1)\Phi(c_2) \quad +$$

$$+\frac{\alpha_{11}}{2\pi(1+\alpha_{11}^2)} \left[(2 + \alpha_{11}c_1c_2)e^{-\frac{1}{2}c_1^2 - \frac{1}{2}c_2^2} - \alpha_{11}\sqrt{2\pi}\left(c_1e^{-\frac{1}{2}c_1^2}\Phi(c_2) + c_2e^{-\frac{1}{2}c_2^2}\Phi(c_1)\right)\right]$$

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.

One may argue that this specification has the disadvantage that it models the probability distribution of discrete random variables using a continuous distribution function. However, this is not unusual in modelling ordered random variables. It would of course be more transparent to use a multivariate discrete probability distribution function like a distribution function that has a single probability parameter for every mass point of the joint distribution. However, such a distribution function has many additional parameters.[11]

[11] Another argument against such a discrete specification is that it a priori excludes personal characteristics that are continuously distributed. The framework used in this subsection can be straightforwardly extended to the case where there are both continuous and discrete personal characteristics.

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 Subsections 3.1 and 3.2 to describe these micro data. As a consequence, the parameters of interest are β, the function-values $\psi_1(t), \psi_2(\tau)$, the α-parameters and, finally, 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.

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

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 equation (10), $\Theta(t|\tau)$ can be rewritten as

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

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 Subsection 3.1. Let us ignore such heterogeneity v for the moment.

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

in which the expectations $\mathrm{E}_{x|\tau-t-i}$ are taken with respect to 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 $\theta(\cdot|\cdot, x)$; for example, $\Pr(T = t|\text{inflow at } \tau; x)$ is given by (3). As a special case, consider the denominator of the right-hand side of (12) for $t = 0$,

$$1 + \mathrm{E}_{x|\tau-1}[1 - \theta(0|\tau-1, x)] + \mathrm{E}_{x|\tau-2}[(1 - \theta(0|\tau-2, x))(1 - \theta(1|\tau-1, x))]$$

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 $\Theta(t|\tau)$.

Under certain additional assumptions (like absence of measurement errors), the macro data are deterministically equal to the corresponding model expressions.[12] 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. It is obvious 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. First of all, individuals who are unemployed according the micro definition may not bother to register at the employment agency if they expect to find a job in a different way, especially if they are not entitled to unemployment benefits or social assistance. Similarly, individuals who fail to comply with register maintenance requirements from the employment agency, or with its job search guidelines, do not satisfy the macro definition but may satisfy the micro definition. Individuals with a regular part-time job that fills less than 50% of the time satisfy the micro definition but not necessarily the macro definition. For

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

individuals aged over 55 years there are incentives not to register even though they may well satisfy the micro definition. Individuals who are not available for employment within 15 days (e.g. due to illness) or who currently do not search actively do not satisfy the macro definition but may satisfy the micro definition. The same holds for individuals who are on temporary lay-off. Finally, the macro definition does not cover individuals looking for part-time, temporary or seasonal jobs, while the micro definition does.

Other types of individuals satisfy the macro definition but not the micro definition. First of all, individuals who enjoy being unemployed may register at the employment agency in order to receive benefits, but may be unwilling to accept jobs. Such individuals do not classify themselves as unemployed. Secondly, if an individual is employed, but employment is known to terminate within 15 days, then the macro definition is satisfied but the micro definition is not. Thirdly, unemployed individuals who accidentally have worked more than 78 hours during the month satisfy the macro definition but not the micro definition. Fourthly, individuals who found employment but did not bother to notify the employment agency satisfy the macro definition until the agency finds out about this.

In most of the above 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. For example, registered individuals who leave the state of unemployment for a very short period of time (for example, in order to help a family member or to work) may not notify the employment agency of such events. Similarly, unemployed individuals may let their registration expire for reasons of negligence, and they may renew it after a while. Also, long-term unemployed individuals' job search activities may be redirected from permanent to temporary jobs.

In addition to this, the macro definition itself is not time-invariant. This is a result of changes in the register maintenance requirements for unemployed individuals, changes in the benefits system, changes in the data collection procedure etc. (see Subsection 2.2).

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 estab-

lish 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

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

with

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

Here, $\varepsilon_{t,\tau}$ captures measurement errors in $U(t|\tau)$ 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 $\widetilde{\Theta}(t|\tau)$ equals the right-hand side of equation (11) with U replaced by $\widetilde{U}$. By substituting equation (13) into this, we obtain

$$\log\left(1 - \widetilde{\Theta}(t|\tau)\right) = \log\left(1 - \Theta(t|\tau)\right) + e_{t,\tau} \tag{14}$$

where $e_{t,\tau} := \log \varepsilon_{t+3,\tau+3} - \log \varepsilon_{t,\tau}$. Equations (14) link the observed macro exit probabilities to the model. Note that $e_{t,\tau}$ is normally distributed with mean zero. The errors in equation (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,

$$\psi_1(t) \quad = \quad \sum_{i=1,2,\ldots} \psi_{1,i} \mathrm{I}\left(3i-3 \leq t < 3i\right),$$

$\mathrm{I}(\cdot)$ being the indicator function. However, the duration dependence is assumed to be constant after 30 months.

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,

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

The seasonal effect is written as

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

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 $\eta_i p_i(\tau)$, $i = 1, \ldots, 5$, where $p_0(\tau), p_1(\tau), \ldots, p_5(\tau)$ are mutually orthogonal polynomials of indexed degree.[13] The business cycle effect $\psi_{2,b}(\tau)$ at month τ is then specified as the value attained by

$$\exp\left\{\sum_{i=0}^{5} \eta_i p_i\left(\tau\right)\right\}.$$

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

[13] More specifically, we first linearly transform the calendar time domain to the domain of orthogonality of the Chebyshev polynomial, $[-1, 1]$, by means of

$\hat{\tau}\left(\tau\right) = 2\frac{\tau - \tau_0}{n_\tau - 1} - 1,$

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)

$p_0\left(\tau\right) = 1,$ and

$p_k\left(\hat{\tau}\right) = \frac{k}{2} \sum_{i=0}^{\left[\frac{k}{2}\right]} \left(-1\right)^i \frac{(k-i-1)!}{i!(k-2i)!} \left(2\hat{\tau}\right)^{k-2i}$ for $k = 1, 2, \ldots, 5.$

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 composition of the inflow on the business cycle by way of an observable business cycle indicator like the capital utilization ratio. However, the present approach is obviously more flexible. According to Abbring, Van den Berg and Van Ours (1994), a polynomial specification for the unemployment outflow is able to mimic the behavior of conventional business cycle indicators for France over time.

Calendar time affects the composition of the inflow by way of the threshold values $c_\tau^i(x_i)$ (see equation (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,

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

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. We take v to have two points of support (v_1,v_2) with associated probabilities $\Pr(v = v_1) = p_1 = 1 - \Pr(v = v_2)$, where $0 \leq p_1 \leq 1$. We reparameterize p_1 as $p_1 = \exp(p)/(1+\exp(p))$. Note that discrete mixture distributions are attractive from a computational point of view.

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 $\psi_1, \psi_{2,s}, \beta$ as well as $\psi_{2,b}$ on the time interval covered by the sample.[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 $\alpha_{...}$ parameters of the joint distribution of the covariates. Finally, the parameters $d_b^i(x_i)$ are identified from the micro inflow data, from a comparison of the inflow distribution of $X|\tau$ and $\psi_{2,b}(\tau)$ on the time interval covered by the micro-data sample.

Now 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 $e^{X'\beta}|(\tau - t)$. Identification of the effect of the business cycle on the distribution of $e^{X'\beta}$ 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

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

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 $e^{X'\beta}$, 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 $e^{X'\beta}$ 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 Descriptive data analyses

4.1 Micro data description

In this section we describe the micro and macro samples in detail. We report estimation results for reduced-form duration models for either sample. We conclude with a systematic account of the differences.

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. 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,[15]

[15]There are even spells observed to start in March 1991 and end in March 1992, where both

which is approximately 22% of the total number of spells in the sample.

Most of the studies that use the French Labor Force Survey refer to the existence of the recall errors. Lollivier (1994a) and Moreau and Visser (1989) exclude spells containing recall errors from the sample, whereas D'Addio (1997) and Magnac (1996) neglect the recall errors in the analysis.[16] It is clear that such approaches do not fully solve the problem. 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 (1995) 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 more ad hoc solution which is in line with Van den Berg (1990b). Like Magnac and Visser (1995), 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. Note that this probability distribution is chosen arbitrarily. However, 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.

The over-all monthly exit probability out of unemployment is given in Figure 1.[17] The figure shows some seasonal effects. The exit probability is higher at the

the moment of inflow and the moment of outflow are observed to contain a recall error.

[16]Magnac and Robin (1994) and Magnac, Robin and Visser (1995) use data from earlier interviews, which had a different setup for recording individual labor market histories. As a result, the type of recall errors in these earlier interviews differ from the type of recall errors in our database and their solution can not be applied here.

[17]Here we include respondents who were already unemployed in March 1990.

beginning of the year than at the end of the year.

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 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, Haltiwanger and Schuh (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: inflow after permanent employment (1,2), after temporary employment (3,4), after being a student or in military service (6,7) and after any other nonparticipation state (8). The numbers in parentheses correspond to the labor market states mentioned in Subsection 2.1. Table 1 provides a brief summary of the sample.

Now let us turn to the estimation of a descriptive reduced-form duration model using the micro data. We assume that the exit probability out of unemployment is specified as in Subsection 3.1. We allow for duration dependence, seasonal effects, and observed and unobserved heterogeneity. Note that due to the short period that the micro data cover, it does not make sense to estimate calendar time effects other than seasonal effects. In sum, we take equation (4) as the model for the exit probability, we adopt the parameterization of Subsection 3.4, and we impose there that $\psi_{2,b}(\tau)$ is constant. Since we have a so-called inflow sample of unemployment durations, there are no initial condition problems. We only have to deal with right-censoring if an individual is still unemployed in March 1993. We estimate the model by maximizing the likelihood function over the parameters, $\psi_{1,i}$ $(i = 2, \ldots, 11)$, ω_s $(s = 2, 3, 4)$, β, v_1, v_2 and p. The estimation results are in Table 2.

The parameter estimates show hardly any duration dependence during the first 15 months of unemployment. After that, the exit probability decreases. The seasonal effects show that the exit rate is highest during the second quarter and decreases over the year to the lowest level during the last quarter of the year. The

personal characteristics "living in Paris", "having children" and "profession" do not have a significant effect on the exit probability. Becoming older decreases the exit probability, whereas having the French nationality, being married, and having a intermediate level of education increase the exit probability. Individuals who flow into unemployment after temporary work have high exit probabilities, whereas individuals flowing into unemployment from any other state than the employment, education or military service have low exit probabilities. We observe significant unobserved heterogeneity. However, if we allow for three mass points for the distribution of v then one of them converges to one of the others during the iterations of the maximum likelihood procedure.

The estimation results above can best be compared to the results in D'Addio (1997) and Lollivier (1994a), as these two papers use a model framework that is rather similar to that used here. Both D'Addio and Lollivier find almost the same estimates for the parameters in the observed heterogeneity component of the hazard. D'Addio does not allow for seasonal effects and restricts the duration dependence to Weibull and log-logistic specifications. Lollivier uses piecewise constant duration dependence and includes seasonal dummies for every month. Lollivier ignores unobserved heterogeneity, and as a result the duration dependence is more negative. Because Lollivier excludes spells containing a recall error, many spells that end around March and April are excluded. This causes a large difference in the estimated seasonal effects.

4.2 Macro data description

In the empirical analyses, we use macro data on the first 12 quarterly duration classes, to obtain observations on $\widetilde{\Theta}(0|\tau), \widetilde{\Theta}(3|\tau), \ldots, \widetilde{\Theta}(30|\tau)$. The maximum monthly duration in the macro data is thus 35, which equals the maximum in the micro data.

Figure 2 shows the over-all quarterly exit probability of leaving unemployment over the macro sample period.[18] 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. It is also clear that seasonal effects dominate cyclical effects.

As noted in Subsection 2.2, around 1986, the procedure of collecting the data changed. As a result, the time series on $\widetilde{U}(t|\tau)$ exhibit ruptures at 1986.IV. This turns out to be particularly important for the series on $\widetilde{U}(0|\tau)$ (see Abbring, Van den Berg and Van Ours (1994)). We therefore add to the model a dummy

[18] Here we include individuals in duration classes corresponding to more than 12 quarters.

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.

Now let us turn to estimation of a reduced-form duration model using the macro data. To avoid identification problems we ignore heterogeneity in the inflow. Otherwise, the specification is in accordance to Section 3, with exit probabilities depending on duration dependence and calendar time effects.[19] We maximize the likelihood function over the parameters $\psi_{1,i}$ $(i = 2, \ldots, 11)$, η_i $(i = 1, \ldots, 5)$, ω_s $(s = 2, 3, 4)$, σ, $d_{<'87}$, and an intercept because both the duration dependence ψ_1 and the calendar time effects ψ_2 are normalized. The estimation results are in Table 3.

The parameter estimates show negative duration dependence for most of the time; only right after 4 and 8 quarters do we observe a slight increase. The seasonal effects show that the exit probabilities are highest in the second quarter and lowest in the third quarter. As expected, the dummy variable denoting the data series rupture at 1986.IV is smaller than 1, so, indeed this rupture in effect increases the exit probability for the first quarter of unemployment. Some of the business-cycle effect parameters are significantly different from zero.

The pattern of duration dependence differs from the duration dependence found in Abbring, Van den Berg and Van Ours (1994). They find that the duration dependence is relatively flat during the first 5 quarters and decreases afterwards. This difference in duration dependence can be explained by the absence of unobserved heterogeneity in the analysis of this subsection. The estimated calendar time effects are almost similar to theirs.

4.3 Comparison of the micro and macro data

We start this subsection by comparing the results of the previous two subsections. Not too much weight should be put on such a comparison, as both sets of results are based on models that are misspecified in the light of the general model specification in Section 3. In particular, the micro model ignores business cycle variation, while the macro model ignores heterogeneity.

[19] In particular, equation (14) specifies the relation between the observed exit probabilities and the corresponding model expressions, and the latter are expressed in terms of the parameters according to Subsection 3.3, where we impose that x and v are fixed constants. We adopt the parameterization of Subsection 3.4.

Except for the fourth quarter, the seasonal effects in the micro and the macro estimates are the same. According to the micro data, the exit probability decreases after the third quarter, while according to the macro data the exit probability increases slightly after the third quarter. The pattern of duration dependence differs between the micro and the macro data. This may be caused by ignoring unobserved heterogeneity in the macro data. Abbring, Van den Berg and Van Ours (1994), who correct for unobserved heterogeneity in the macro data, find a duration dependence pattern which is similar to the duration dependence found when we only use the micro data (recall however that their results on duration dependence and unobserved heterogeneity are based on strong identifying functional form restrictions). In Section 5 we formally test for the similarity of the duration dependence and seasonal effects.

Now let us compare the raw duration distributions in both data sets for the individuals flowing in at a certain 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 Figure 3. 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.

To compare the two datasets from another angle, we aggregate the micro data in the same way as the macro data were aggregated. First, we compute the number of individuals in the micro data who are unemployed at the end of a quarter τ. From these aggregated micro data we compute the over-all quarterly exit probability[20], which is plotted together with the macro over-all exit probability in Figure 4. Clearly, the exit probability in the micro data is smaller than in the macro data, over the entire period. Also note that the micro exit probability is more variable than the macro exit probability.

For a more formal analysis 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 $\widetilde{U}(t|\tau)$ values that are observed in the macro data, and they can be used to calculate the counterparts of the exit probabilities $\widetilde{\Theta}(t|\tau)$. We regress the difference between the macro exit probability and the micro exit probability on an intercept and the quarterly duration t, and we include dummies for the season of exit out of unemployment and the season of inflow into unemployment. The estimation results are in Table 4. The parameter

[20]For this purpose, we include individuals who were already unemployed in March 1990.

estimates are jointly insignificant and relatively small. "Leaving unemployment during the third season" is the only significant variable, although its effect is small. Note that the predicted third-season exit probability is larger in the macro data than in the micro data. The results from the duration analysis in the previous subsection, however, show the opposite.

So far we have not mentioned (or exploited) the fact that the macro data allow for stratification by age group. The reason that we do not use this information for the estimation of the full model is that the $\widetilde{U}(t|\tau)$ data are stratified by age group at τ rather than by age group at the moment of inflow. We do not know whether exit out of the stock of unemployed in a certain age group is due to re-employment or to a movement to another age group.

Keeping this in mind, the stratified data can be used for an informal comparison of the data sets. The macro data stratified by age provide the quarterly distribution of individuals with duration less than three months over three age-groups: (*i*) under 25 years, (*ii*) from 25 years to 50 years and (*iii*) over 50 years. For the micro data we can compute similar numbers. We compare these to the numbers in the macro data over the period from 1990.II to 1993.I. It turns out that the results are very similar. The youngest age group contains 37% of the total inflow in the micro data and 36% in the macro data. In both datasets the inflow by the youngest age group is highest during the third quarter, and it then decreases over the year to its lowest level at the second quarter. The inflow by the middle age group contains 52% of the total inflow in the micro data and 56% in the macro data. Again the seasonal pattern is the same, the proportion of the inflow being highest in the second quarter. Finally, the oldest age-group contains 11% of the total inflow in the micro data and 8% in the macro data. And again the seasonal pattern is similar, the proportion of unemployed in the oldest age-group being highest during the second and the fourth quarter. An explanation for the fact that the micro data contain more old individuals is that unemployed over 55 do not have to register at the public employment agency to be eligible for unemployment benefits, so they may therefore not show up in the macro data. In the micro data, 57% of the unemployed in the oldest age group are older than 56 years, which is 6.2% of the total inflow. Note that this suggests that at least some individuals aged over 55 do register at the employment agency.

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 the reduced-form duration analysis of the micro data in Subsection 4.1. As a result, x consists of indicators of nationality, age, being married, education, and the state before inflow into unemployment.

Recall from Subsection 4.3 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 $\theta(t|\tau, x, v)$ appearing in the macro expressions to differ from those in the micro expressions, as follows,

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

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

The unknown parameter δ gives the relative difference in the exit rates of the underlying continuous-time PH models. Note that θ_{micro} above is specified as in Subsection 3.1.[21]

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 $d_s^i(x_i)$ and $d_b^i(x_i)$, v_1, v_2, p, 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 datasets. Note that, as a result of the latter, the likelihood contributions consist of 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.[22]

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

[22] The usual asymptotic results for ML estimators hold in many cases where the separate contributions are not independently and identically distributed. It is important that asymptot-

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. For example, this approach is often used in so-called meta-analysis of different datasets (see e.g. DuMouchel (1990)). In the Appendix to this paper 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 datasets. For a given (zero-one) loss function, the best Bayesian point estimate is equal to the value that maximizes the likelihood function, and the corresponding Bayesian summary dispersion measure equals the ML estimate of the variance-covariance matrix.

5.2 Estimation results

The parameter estimates are in Table 5.[23] 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 depicted in Figure 5. 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.

In Figure 6 we depict how the estimated contemporaneous cyclical effect ($\psi_{2,b}(\tau)$) changes over calendar time.[24] The contemporaneous effect includes

ically 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 100 respondents in the micro survey data. Such an interpretation of the data-collection process is in fact not unreasonable if both datasets are collected for the single purpose of studying unemployment in all of its facets.

[23] Estimation of the full model requires about 100 hours, on a Pentium II 200 MHz PC with 16 MB RAM. We use GAUSS maximum likelihood routines.

[24] 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 parameter estimates of β are actually very similar to those obtained by the separate estimation with the micro data (see Subsection 4.1). Again, 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.

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 pairwise combinations of the elements of X, for different seasons of inflow (merging the three years in the micro data). In addition, we performed these tests for each year separately. All of these tests accept the null hypothesis of a correct specification. In addition, the estimated distribution picks up the correlations between the characteristics in the data, and it captures the differences between the seasons. It thus seems that a distribution based on Hermite series provides a useful (and computationally feasible) specification of the empirical distribution of explanatory variables.

The estimated marginal distributions of personal characteristics in the inflow do not change dramatically over the cycle. The estimated inflow fractions for the dummy variables in x stay within a 10% range of the micro sample averages given in Table 1. The only exception to this concerns the fraction of workers who were permanently employed before inflow: this fraction decreases from about 0.6 in the early eighties to about 0.4 in the early nineties. The decrease is halted during the boom in the late eighties, so one could say that its behavior is somewhat procyclical. The fractions of workers who flow in from other states all increase during the macro-data time interval. In the absence of information on the reason for inflow, one can only speculate about a possible relation between a large (small) inflow of permanently (temporarily) employed workers and a small (large) flow

of permanent (temporary) layoffs.

Now let us turn to the business cycle effect 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). Figure 7 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, Haltiwanger and Plant (1985) and Davis, Haltiwanger and Schuh (1996), who argue that individuals entering in a recession are more disadvantaged. The graphs of the estimated separate covariate effects 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 within 3 months. The solid line in Figure 8 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). 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 cases where only a subset of the personal characteristics is imposed to have a time-invariant inflow distribution. The results confirm the above conclusion.[25]

A formal test of cyclical variation in the composition in the inflow amounts to a joint test of $d_i^c(x_i) = 0$ 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 levels of significance. We conclude that the effect of cyclical variation in the composition of the inflow can not be ignored, even though it is quantitatively unimportant.

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,

[25]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 equation (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.

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 Subsection 3.5).

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. Albæk and Holm Larsen (1993) obtained the same result in their comparison of survey and administrative duration data concerning the same individuals. 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. On the one hand, the micro survey may overlook short spells. In particular, respondents may forget about such jobs if they are preceded and followed by long spells of unemployment, and they may forget about short unemployment spells if they are preceded and followed by long job spells. On the other hand, the macro data may contain spurious transitions out of unemployment because of unemployed workers who out of negligence let their registration expire. The fact that macro exit probability is particularly large at the fourth quarter could be in accordance to both explanations. Another cause of that seasonal effect could be that the composition of the inflow is structurally different under the macro unemployment definition than under the micro definition. However, our data do not allow for identification of this.

6 Conclusion

The macro and the micro dataset are not in serious conflict with each other. The only 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 clearly show that the countercyclicality of the aggregate mean unemployment duration can *not* be attributed to changes in the com-

position of the inflow over the cycle. Instead, it originates from the fact that exit probabilities vary over the cycle for all types of individuals. In France, the cyclical variation in unemployment durations affects all types of individuals likewise. The quality of the composition is somewhat countercyclical, but the effect of this on the cyclical behavior of the mean duration is small. These results go against the view on unemployment dynamics put forward in Darby, Haltiwanger and Plant (1985).

Some previous studies with U.S. data found that the distribution of the individual-specific reason of inflow into unemployment changes substantially over the cycle. Davis, Haltiwanger and Schuh (1996) even argue that the latter is an important determinant of the cyclical variation in durations, although Baker (1992) reaches a different conclusion. We do not observe the reason of inflow, but the individual-specific labor market state prior to inflow is related to this reason of inflow. We find that the distribution of this variable changes somewhat over the cycle, but this does not provide a quantitatively important explanation of cyclical variation in durations.

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 the most disadvantaged workers.

References

Abbring, J.H., G.J. van den Berg and J.C. van Ours (1994), The anatomy of unemployment dynamics, Working Paper, Tinbergen Institute, Amsterdam.

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

Albæk, K. and A. Holm Larsen (1993), Unemployment data, from surveys and administrative registers, in: H. Bunzel et al., eds., *Panel Data and Labour Market Dynamics*, North-Holland, Amsterdam.

Ayong Le Kama, A. (1995), Évolution de l'emploi et du chômage en France, Working Paper, CDC, Paris.

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

Bean, C.R. (1994), European unemployment: A survey, *Journal of Economic Literature* 32, 573–619.

Bonnal, L., D. Fougère, and A. Sérandon (1997), Evaluating the impact of French employment policies on individual labour market histories, *Review of Economic Studies* 64, 683–713.

Canard Enchaîné (7 May 1997), La guerre au chômage tourne à la chasse aux chômeurs.

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

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

Devine, T.J. and N.M. Kiefer (1991), *Empirical Labor Economics*, Oxford University Press, Oxford.

DuMouchel, W. (1990), Bayesian Metaanalysis, in: D. Berry, ed., *Statistical Methodology in the Pharmaceutical Sciences*, Marcel Dekker, New York.

Dynarski, M. and S.M. Sheffrin (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.

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

Gnedenko, B.V. and A.N. Kolmogorov (1954), *Limit Distributions for Sums of Independent Random Variables*, Addison-Wesley Publishing Company, Reading.

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

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. and T. Lancaster (1994), Combining micro and macro data in microeconometric models, *Review of Economic Studies* 61, 655–680.

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

Laisney, F. and M. Lechner (1994), Combining panel data and macro information: an application to the estimation of a participant model, Working Paper, University of Strasbourg.

Layard, R., S. Nickell, and R. Jackman (1991), *Unemployment: Macroeconomic performance and the labor market*, Oxford University Press, Oxford.

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

Libération (9 June 1997), ANPE, nouveau mode d'emploi.

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

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

Magnac, T. (1996), State dependence and heterogeneity in youth employment histories, Working Paper, INSEE-CREST, Malakoff/Paris.

Magnac, T. and J.M. Robin (1994), An econometric analysis of labour market transitions using discrete and tenure data, *Labour Economics* 1, 327–346.

Magnac, T., J.M. Robin and M. Visser (1995), Analysis of incomplete individual employment histories using indirect inference, *Journal of Applied Econometrics* 10, S153–169.

Magnac, T. and M. Visser (1995), Transition models with measurement errors, Working Paper, INSEE-CREST, Malakoff/Paris.

Moreau, A. and M. Visser (1989), Duration of youth unemployment in France, Working Paper, INSEE-CREST, Malakoff/Paris.

O'Hagan, A. (1994), *Kendall's advanced theory of statistics, Volume 2B Bayesian inference*, Edward Arnold, London.

Van den Berg, G.J. (1990a), Nonstationarity in job search theory, *Review of Economic Studies* 57, 255–77.

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

Van den Berg, G.J. and J.C. van Ours (1996), Unemployment dynamics and duration dependence, *Journal of Labor Economics* 14, 100–125.

Vishwanath, T. (1989), Job search, stigma effect, and escape rate from unemployment, *Journal of Labor Economics* 7, 487–502.

Appendix: An equivalent Bayesian estimation method

In this appendix we show that the ML estimation approach that we use to estimate the model is equivalent to a Bayesian estimation method. Gourieroux and Monfort (1995) and O'Hagan (1994) give useful and detailed descriptions of the Bayesian approach. The prior distribution of the complete parameter vector γ is defined on the set of all possible parameter values Γ, and is written as $\pi(\gamma)$. The idea of Bayesian updating is that adding information on the set of parameters is used to update the prior distribution. According to Bayes' rule, a dataset z generates a posterior distribution of γ given this dataset

$$\pi(\gamma|z) = \frac{\ell(z|\gamma)\,\pi(\gamma)}{\ell(z)} = \frac{\ell(z|\gamma)\,\pi(\gamma)}{\int_\Gamma \ell(z|\gamma)\,\pi(\gamma)\,d\gamma}$$

where $\ell(z|\gamma)$ is the likelihood of the dataset z, given γ. We adopt a noninformative or diffuse prior distribution, which means that the prior distribution is proportional to 1: $\pi(\gamma) \propto 1$. (This choice is by no means necessary for our estimation procedure.) We have two datasets available, the macro data z and the micro data y. It is easy to see that the posterior distribution given both datasets is proportional to

$$\pi(\gamma|z,y) \propto \ell(z|\gamma) \prod_{i=1}^{n} \ell(y_i|\gamma) \tag{17}$$

where n is the number of individuals in the micro data.

The best point estimate (or location summary) d of γ in the Bayesian sense is the value of γ which minimizes the expected loss. This value thus depends on the choice of the loss function. Consider the zero-one loss function, $L(d,\gamma) = 0$ if $|d-\gamma| \leq b$ and $L(d,\gamma) = 1$ if $|d-\gamma| > b$. This loss function does not give importance to the shape of the tails of the posterior distribution. (Note that, analogously to classical ML estimation, this loss function is less attractive if there is no global concavity.) The optimal d is the center of the set of width $2b$ having maximum probability (see O'Hagan (1994)), and the minimized expected loss is simply the posterior probability that γ is not in this set. In the limit as $b \downarrow 0$, the optimal estimate tends to the posterior mode. The latter will be our point estimate of γ,

$$\hat{\gamma} = \arg\max_{\gamma\in\Gamma} \log(\pi(\gamma|z,y)) = \arg\max_{\gamma\in\Gamma} \log(\ell(z|\gamma)) + \sum_{i=1}^{n} \log(\ell(y_i|\gamma))$$

The natural summary measure of dispersion corresponding to this loss function is the "modal dispersion" of the posterior, which is defined as minus the inverse of the hessian of the log posterior density evaluated at the posterior mode (see O'Hagan (1994)). This measure captures the local width of the peak around the posterior mode.

A quadratic loss function results in the posterior mean as best point estimate, $\hat{\gamma} = \mathrm{E}[\gamma|z, y]$. The corresponding summary measure of dispersion is the posterior variance $\mathrm{var}(\gamma|z, y)$. (Note that here as well as in the previous paragraph, the summary measure of dispersion can also be thought to represent the precision of the corresponding point estimate.) If we approximate the true posterior density by a normal density then the posterior mean and variance coincide with the posterior mode and the modal dispersion. The latter are much easier to calculate than the former, and this simple approximation approach for the quadratic loss function case has been rather common. Approximate normality of the posterior distribution can be justified with asymptotic arguments.[26]

Recall that the posterior mode is equal to the value of γ that maximizes the product of the likelihood function of the macro data and the likelihood function of the micro data. In other words, our point estimate of γ equals the value of γ provided by ML routines if the likelihood function to be maximized is specified as the product of the likelihood contributions of our two data sets. Also, our summary dispersion measure equals the estimate of the variance-covariance matrix as provided by ML routines. Note that these routines provide the exact value of our summary dispersion measure in case of a zero-one loss function, whereas in the classical context they only provide an estimate of the precision of the ML estimate.

In sum, the Bayesian estimates are equal to the estimates obtained with standard ML estimation routines for nonlinear models.

[26]Suppose we interpret (17) as follows: our prior distribution is the posterior distribution that we obtain by using only the macro data, $\ell(z|\gamma)$, and this prior distribution is updated by the likelihood function of the micro data $\prod_{i=1}^{n} \ell(y_i|\gamma)$. Then asymptotic results for i.i.d. random variables can be applied by letting the number of observations in the micro data become large (see Gourieroux and Monfort (1995) and O'Hagan (1994)). Alternatively, if we start with the noninformative prior and let the number of observations in both the micro and the macro data become large then we have to apply other limit theorems to justify asymptotic normality; see O'Hagan (1994).

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

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

Duration dependence $\psi_1(t)$		
$\psi_{1,1}$	1	
$\psi_{1,2}$	1.06	(0.081)
$\psi_{1,3}$	0.99	(0.10)
$\psi_{1,4}$	1.08	(0.13)
$\psi_{1,5}$	1.13	(0.16)
$\psi_{1,6}$	0.86	(0.17)
$\psi_{1,7}$	0.71	(0.18)
$\psi_{1,8}$	0.68	(0.22)
$\psi_{1,9}$	0.55	(0.23)
$\psi_{1,10}$	0.73	(0.38)
$\psi_{1,11}$	0.33	(0.35)
Contemporaneous seasonal effect $\psi_{2,s}(\tau)$		
ω_1	0	
ω_2	0.13	(0.078)
ω_3	−0.0050	(0.075)
ω_4	−0.25	(0.076)
Observed personal characteristics β		
Paris	0.010	(0.092)
Non-French	−0.27	(0.11)
Married	0.28	(0.091)
Age 31–45	−0.35	(0.087)
Age 46–65	−0.71	(0.11)
High education	−0.0073	(0.13)
Intermediate education	0.27	(0.13)
Having children	−0.00014	(0.071)
Intermediate skill	0.0042	(0.11)
Low skill and farmers	0.0023	(0.093)
Labor market state before inflow:		
Temporary employment	0.55	(0.15)
Permanent employment	0.18	(0.15)
Student / military service	0.28	(0.17)
Unobserved heterogeneity		
v_1	−2.65	(0.19)
v_2	−1.32	(0.24)
p	1.50	(0.39)
log likelihood	−4693.60	
N	1536	

Explanatory note: Standard errors are in parentheses.

Table 2: Estimation results for the duration model using the micro data only.

Duration dependence $\psi_1(t)$		
$\psi_{1,1}$	1	
$\psi_{1,2}$	0.89	(0.078)
$\psi_{1,3}$	0.77	(0.050)
$\psi_{1,4}$	0.62	(0.057)
$\psi_{1,5}$	0.68	(0.055)
$\psi_{1,6}$	0.65	(0.056)
$\psi_{1,7}$	0.60	(0.056)
$\psi_{1,8}$	0.46	(0.053)
$\psi_{1,9}$	0.48	(0.053)
$\psi_{1,10}$	0.48	(0.050)
$\psi_{1,11}$	0.40	(0.032)
Contemporaneous cyclical effect $\psi_{2,b}(\tau)$		
η_1	-0.053	(0.017)
η_2	-0.0070	(0.012)
η_3	-0.038	(0.013)
η_4	0.039	(0.0077)
η_5	0.020	(0.0078)
Contemporaneous seasonal effect $\psi_{2,s}(\tau)$		
ω_1	0	
ω_2	0.13	(0.035)
ω_3	-0.084	(0.033)
ω_4	-0.024	(0.037)
Intercept	-1.78	(0.052)
$d_{<'87}$	0.96	(0.33)
σ	0.050	(0.0018)
log likelihood	989.62	

Explanatory note: Standard errors are in parentheses.

Table 3: Estimation results for the duration model using the macro data only.

Intercept	0.045	(0.040)
Quarterly duration	-0.0019	(0.0058)
Season at the moment of exit		
Second season	0.0031	(0.032)
Third season	0.098	(0.039)
Fourth season	-0.018	(0.036)
Season at the moment of inflow		
Second season	-0.021	(0.032)
Third season	-0.013	(0.035)
Fourth season	-0.037	(0.037)

Explanatory note: Standard errors are in parentheses.

Table 4: Estimation results for the OLS regression on the difference between the quarterly exit probability in the macro data and the corresponding quarterly exit probability in the aggregated micro data.

Duration dependence $\psi_1(t)$		
$\psi_{1,1}$	1	
$\psi_{1,2}$	0.86	(0.039)
$\psi_{1,3}$	0.75	(0.034)
$\psi_{1,4}$	0.76	(0.041)
$\psi_{1,5}$	0.89	(0.054)
$\psi_{1,6}$	0.90	(0.066)
$\psi_{1,7}$	0.90	(0.083)
$\psi_{1,8}$	0.92	(0.095)
$\psi_{1,9}$	1.01	(0.12)
$\psi_{1,10}$	1.13	(0.14)
$\psi_{1,11}$	1.19	(0.15)
Contemporaneous cyclical effect $\psi_{2,b}(\tau)$		
η_1	−0.29	(0.017)
η_2	0.061	(0.011)
η_3	−0.088	(0.0084)
η_4	0.020	(0.0066)
η_5	0.036	(0.0052)
Contemporaneous seasonal effect $\psi_{2,s}(\tau)$		
ω_1	0	
ω_2	0.15	(0.012)
ω_3	0.050	(0.021)
ω_4	0.016	(0.025)
Observed personal characteristics β		
Non-French	−0.36	(0.10)
Married	0.23	(0.064)
Age 31–45	−0.29	(0.058)
Age 46–65	−0.74	(0.093)
High education	−0.0091	(0.0057)
Intermediate education	0.22	(0.088)
Labor market state before inflow:		
Temporary employment	0.54	(0.039)
Permanent employment	0.18	(0.052)
Student / military service	0.22	(0.087)
Unobserved heterogeneity		
v_1	−3.87	(0.29)
v_2	−2.10	(0.089)
p	−2.86	(0.19)
δ	0.27	(0.031)
$d_{<'87}$	0.80	(0.025)
σ	0.035	(0.0012)

Table 5: Estimation results for the full model.

Joint distribution of the observed heterogeneity $X\|\tau$		
Joint-dependence parameters		
α_{11000}	0.24	(0.037)
α_{10100}	0.094	(0.035)
α_{10010}	-0.081	(0.025)
α_{10001}	-0.020	(0.030)
α_{01100}	0.69	(0.046)
α_{01010}	0.018	(0.021)
α_{01001}	-0.017	(0.016)
α_{00110}	0.017	(0.0088)
α_{00101}	-0.10	(0.018)
α_{00011}	0.071	(0.027)
Seasonal effect $d_s^k(x_i)$ on threshold values		
Non-French (season 1)	1.91	(0.17)
Non-French (season 2)	2.04	(0.19)
Non-French (season 3)	2.17	(0.13)
Non-French (season 4)	2.03	(0.17)
Married (season 1)	-0.29	(0.098)
Married (season 2)	-0.40	(0.035)
Married (season 3)	-0.018	(0.018)
Married (season 4)	-0.29	(0.088)
Age 31–45 (season 1)	-0.0015	(0.013)
Age 31–45 (season 2)	-0.081	(0.060)
Age 31–45 (season 3)	0.33	(0.022)
Age 31–45 (season 4)	0.24	(0.073)
Age 46–65 (season 1)	1.68	(0.33)
Age 46–65 (season 2)	1.56	(0.39)
Age 46–65 (season 3)	2.13	(0.29)
Age 46–65 (season 4)	1.93	(0.34)
High education (season 1)	0.27	(0.16)
High education (season 2)	0.18	(0.19)
High education (season 3)	0.30	(0.17)
High education (season 4)	0.30	(0.18)
Intermediate education (season 1)	1.12	(0.25)
Intermediate education (season 2)	1.18	(0.29)
Intermediate education (season 3)	1.12	(0.28)
Intermediate education (season 4)	1.20	(0.30)
Temporary employment (season 1)	-0.29	(0.097)
Temporary employment (season 2)	-0.11	(0.19)
Temporary employment (season 3)	-0.18	(0.14)
Temporary employment (season 4)	-0.65	(0.062)
Permanent employment (season 1)	0.58	(0.18)
Permanent employment (season 2)	0.53	(0.20)
Permanent employment (season 3)	0.71	(0.20)
Permanent employment (season 4)	0.68	(0.18)
Student / military service (season 1)	0.77	(0.12)
Student / military service (season 2)	0.61	(0.14)
Student / military service (season 3)	0.28	(0.11)
Student / military service (season 4)	0.67	(0.14)

Table 5: (Continued).

Cyclical effect $d_b^i(x_i)$ on threshold values		
Non-French	0.88	(0.15)
Married	−0.73	(0.063)
Age 31–45	0.098	(0.049)
Age 46–65	0.55	(0.38)
High education	−0.91	(0.21)
Intermediate education	−0.39	(0.34)
Temporary employment	1.52	(0.15)
Permanent employment	0.91	(0.22)
Student / military service	−0.73	(0.12)
log likelihood	−11122.39	

Explanatory note: Standard errors in parentheses.

Table 5: (Continued).

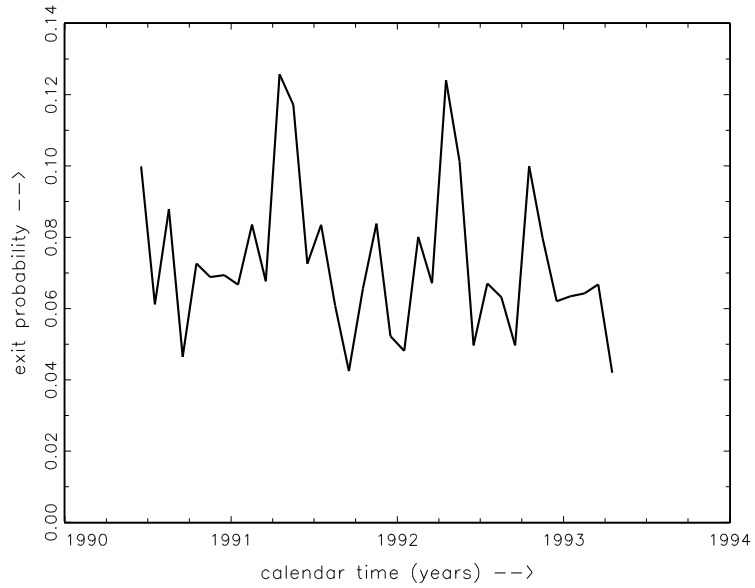

Figure 1: The monthly over-all exit probability in the micro data.

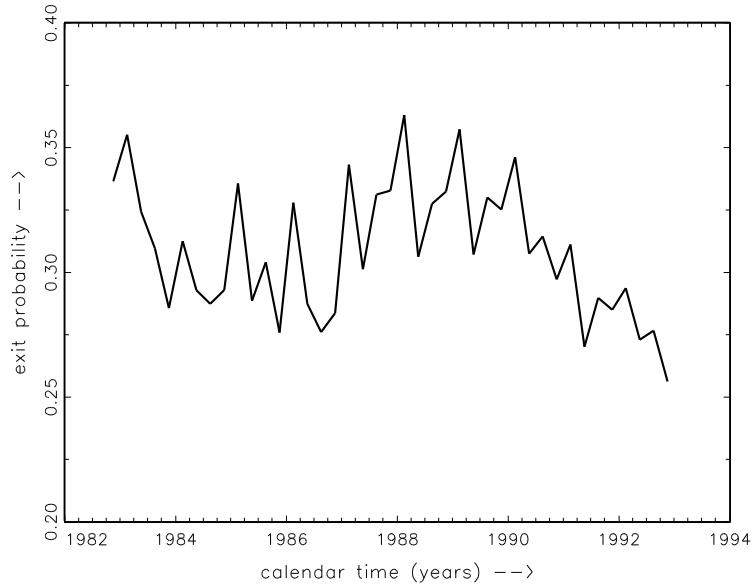

Figure 2: The quarterly over-all exit probability in the macro data.

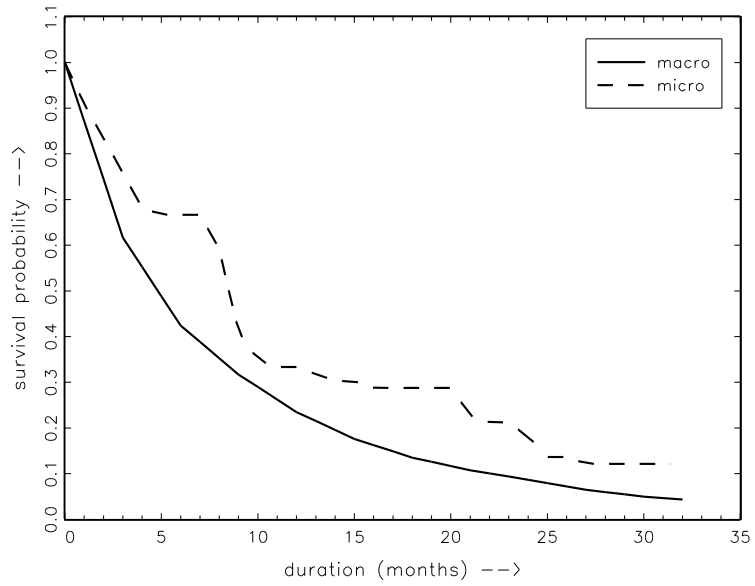

Figure 3: Kaplan-Meier estimate of the outflow in the micro and the macro data of individuals who were unemployed in June 1990 for less than 3 months.

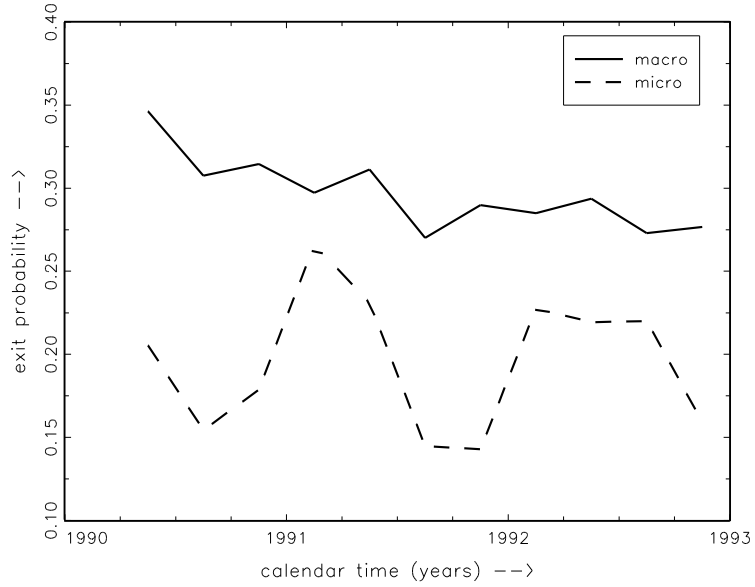

Figure 4: The quarterly over-all exit probabilities in the macro data and in the micro data.

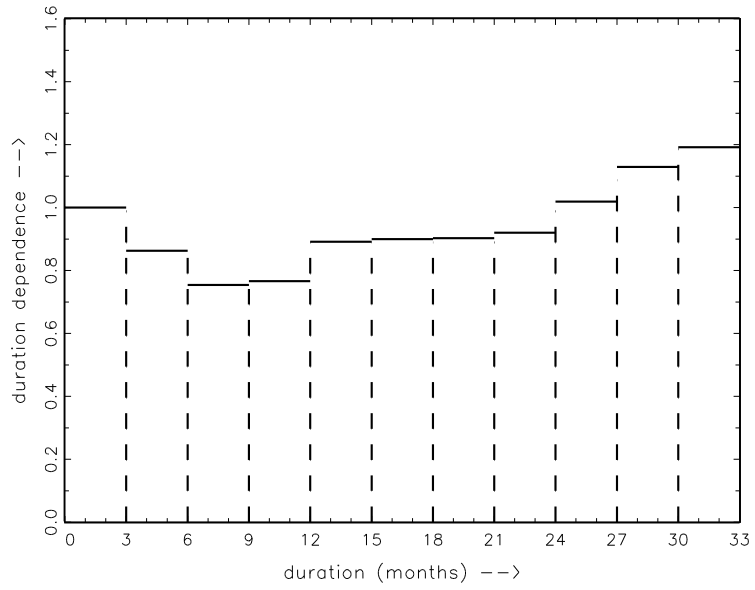

Figure 5: The duration dependence ($\psi_1(t)$).

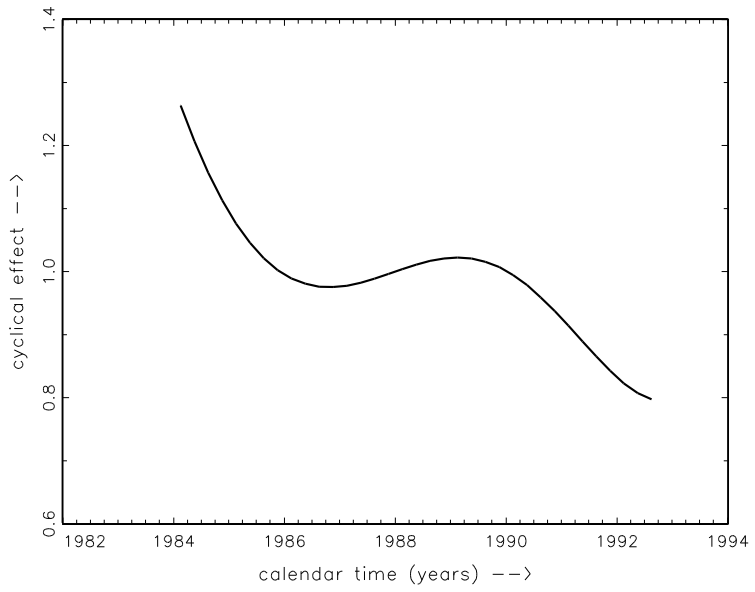

Figure 6: The baseline calendar time effects ($\psi_2(\tau)$).

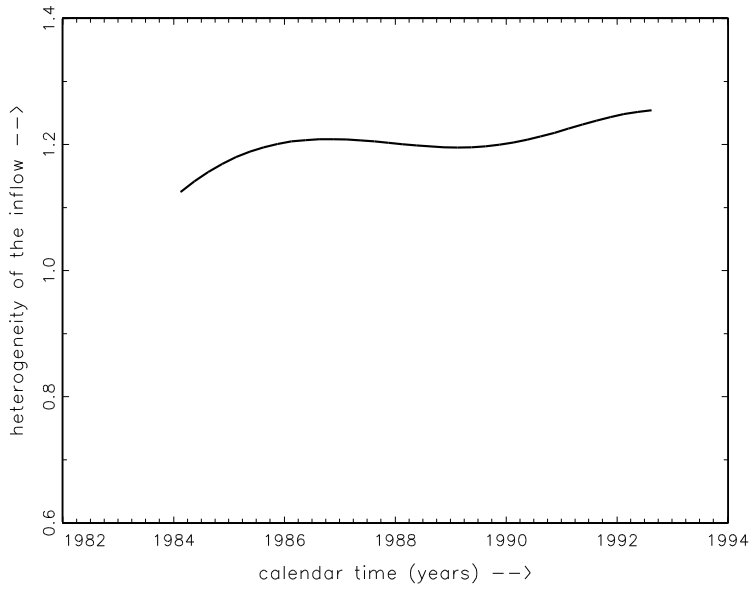

Figure 7: The variation in the composition of the inflow.

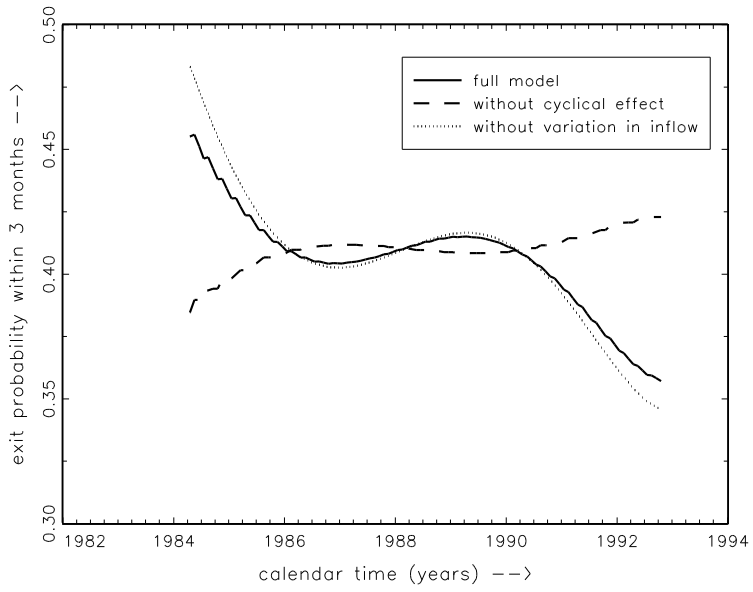

Figure 8: The probability of leaving unemployment within 3 months.