THE SURVEY OF INCOME AND PROGRAM PARTICIPATION

The Discourage Worker Effect: A Reappraisal Using Spell Duration Data

No. 57

Alberto Martini
University of Wisconsin-Madison

June 1988

U.S. Department of Commerce U.S. CENSUS BUREAU

Survey of Income and Program Participation

The Discourage Worker Effect:
A Reappraisal Using Spell Duration Data

No. 8810 57

Alberto Martini
University of Wisconsin-Madison

SIPP Working Paper #57

June 1988

A previous version of this paper was presented at the XIII meeting of the Eastern Economic Association, Washington, D.C., March 3-7, 1987, and at the Labor Economics Workshop of the University of Wisconsin-Madison. I wish to thank Martin David, Hank Griffs and Jim Walker for helpful comments and suggestions: the responsiblity for any remaining error is entirely mine. The empirical work contained in this paper greatly benefited from the use of the data and computing resources made available to me by the SIPP ACCESS project, supported by NSF grant #8411785.

TABLE OF CONTENTS

1. INTRODUCTION

Evidence accumulated over the last two decades has shown the dynamic character of the U.S. labor market, where large flows of individuals move each month between employment, unemployment and non-participation (Marston, 1976). [redacted] is [redacted] are [redacted] (Clark and [redacted]).

The purpose of this paper is to formulate and estimate a microdynamic reduced form model for the transition between unemployment and non-participation (hereafter UO transition). [redacted] constant a vector of personal characteristics, and to examine whether this empirical relationship is robust across different model specifications.

The tools utilized to perform such analysis are those recently developed by the econometric literature on duration data (hazard models). Moreover, this paper exploits an entirely new data set, the Survey of Income and Program Participation (SIPP), which contains weekly information on the labor force status of a large sample of individuals for a period as long as 36 months.

[redacted]

proc[redacted]). The discouraged worker hypothesis has been extensively tested in the 1960's using aggregate data, both cross-sectional (Bowen and Finegan, 1969) and time-series (Tella, 1965). This paper extends such previous work by utilizing individual-level spell duration data. [redacted] If one were interested in how labor market conditions affect the probability that an individual is *found* out of the labor force, then cross-sectional data and discrete-choice econometric techniques would be appropriate. However, the dynamic dimension of the problem is likely to be the more relevant from a labor policy point of view.

The plan of the paper is as follows. Section 2 briefly presents the use of duration models applied to transitions in the labor market. The discussion touches upon the issue of initial conditions (left censoring) and on the use of the *competing risks* model to allow for a three-state transition process. Section 3 describes the Survey of Income and Program Participation (SIPP) from which the sample used in the empirical estimation has been drawn. Summary statistics are presented on average duration of unemployment spells, together with non-parametric estimates of the survivor function for these spells. In section 4 estimation results are presented for a variety of parametric hazard models for the UO transition. Section 5 summarizes the results.

2. DURATION MODELS

Hazard models have many advantages compared with traditional regression techniques when dealing with duration data (i.e. when investigating the relationship between the outcome of a process that takes place over time and a set of explanatory variables, some of which might vary during the process). [redacted]). Moreover, the hazard model utilized in this study allows one to control for explanatory variables thet are not fixed during the spell, while the independent variables in a regression can assume only a single value for each observation. The unemployment rate, which is crucial in testing the discouraged worker hypothesis, is one example of a variable that can vary over the course of the spell.

A detailed discussion of econometric duration analysis is beyond the scope of this paper. This section sets out the essential ideas needed in the rest of the paper. For an exhaustive survey of standard survival analysis the reader is referred to Kalbfleisch and Prentice (1980), [redacted].

The hazard function

Let T be a continuous random variable representing duration in state i. The probability density function of T is:

$$f_i(t) = \lim_{\Delta t \to 0} \frac{P(t < T \leq t + \Delta t)}{\Delta t} \tag{1}$$

The hazard function $h_i(t)$ specifies the instantaneous rate of escape from i at time t, conditional upon survival to t. It is defined as:

$$h_i(t) = \lim_{\Delta t \to 0} \frac{P(t < T \leq t + \Delta t \mid T > t)}{\Delta t} = \frac{f_i(t)}{S_i(t)} \tag{2}$$

where

$$S_i(t) = 1 - F_i(t)$$

is defined as *survivor function.* Since

$$-\frac{d \log S_i(t)}{dt} = \frac{f_i(t)}{S_i(t)} = h_i(t) \tag{3}$$

integrating, we obtain

$$\int_0^t h_i(u) du = \int_0^t -\frac{d \log S_i(u)}{du} = -\log S_i(t) \tag{4}$$

from which

$$S_i(t) = \exp[-\int_0^t h_i(u)\, du] \tag{5}$$

is derived.

The latter result is of fundamental importance: (5) together with (2) allow the complete characterization of the density function in terms of the hazard function:

$$f_i(t) = h_i(t)\, exp[-\int_0^t h_i(u)\, du] \tag{6}$$

Duration dependence is said to exist if $dh(t)/dt \neq 0$. The only duration density with no duration dependence is the exponential distribution. For this case $h(t)=h$, a constant, and T above is an exponential random variable. If $dh(t)/dt < 0$ at $t = t_o$ there is said to exist *negative duration dependence* at t_o. One of the most widely used duration distribution is the the *Weibull*, with hazard function

$$h(t) = \alpha\lambda(\alpha t)^{\lambda-1} \tag{7}$$

which exhibits monotonically negative duration dependence if $\lambda < 1$, positive if $\lambda > 1$, and collapses to an exponential in case of equality. Other distributions, e.g. the lognormal, allow for non-monotonic duration dependence.

A useful method to evaluate non-parametrically some of the feature of the hazard is to estimate the sample survivor function via the Kaplan-Meier or product limit estimator:

$$S(t) = \prod_{j|t_j<t} (1 - \frac{d_j}{n_j}) \tag{8}$$

where d_t is the number of items which fail at t and n_t is the number who survives up to at least time t.

A non-parametric check for the adequacy of the Weibull distribution is obtained by plotting $\log(-\log(S(t)))$ against $\log(t)$. Given the Weibull survivor function

$$S(t) = \exp[-(\alpha t)^{\lambda}] \tag{9}$$

clearly $\log(-\log(S(t))) = \lambda(\log t + \log\alpha)$. The plot should give approximately a straight line, the slope of which provides a rough estimate of λ. An example of such plot is presented in figure 3 in the next section.

The conditional hazard function

By analogy with conventional regression models, in econometric duration analysis it is common to consider conditional duration distributions, or equivalently conditional hazard functions, where the conditioning is with respect to observed ($x(t)$) and unobserved ($\theta(t)$) variables.

The conditional hazard can be defined as:

$$h_i(t \mid x(t)\ \theta(t)) = \lim_{\Delta t \to 0} \frac{P(t < T \leq t + \Delta t \mid T > t\ x(t)\ \theta(t))}{\Delta t} \tag{10}$$

Duration models have many advantages with respect with traditional regression techniques. One reason is that regression techniques are unable to deal with *censored observations* (spells which are not completely observed because they are in progress either at the beginning or at the end on the sampling period). Moreover, hazard models allow one to control for explanatory variables thet are not fixed during the spell, while the independent variables in a regression can assume only a single value for each observation.

Once a parametric functional form is assumed for the conditional hazard function, its parameters can be estimated by maximum likelihood, utilizing spell duration data. The likelihood function is formed in the following way: right censored spells would contribute only the survivor function to the likelihood, since the only information they convey is that the spell is at least t periods long. Completed spells enter the likelihood with the entire right hand side of (6). Discussion of left censoring in deferred to the end of this section.

The main problem with the estimation of duration models is that usually the economic theory does not produce a "structural" functional form for the dependence of the hazard function on observed and unobservable variables. The common practice in duration analysis is to use a *reduced form* specification for the hazard. The behavioral interpretation of the reduced form estimated parameters becomes obviously problematic, since they represent "net effects" of the explanatory variables on the hazard.

The parametric functional form utilized in this work is of the proportional hazard function class, with a flexible Box-Cox specification of duration dependence (see Flinn and Heckman, 1982a).

$$h_{ij}(t) \;=\; \exp\left[\, X(\tau + t)\,\beta_{ij} \;+\; \gamma_{1ij}\frac{t^{\lambda_{1ij}} - 1}{\lambda_{1ij}} \;+\; \gamma_{2ij}\frac{t^{\lambda_{2ij}} - 1}{\lambda_{2ij}}\,\right] \qquad (11)$$

where τ represents calendar time, X() is a vector of (possibly) time varying explanatory and control variables, t represents the duration of the spell, β_{ij}, γ_{ij} and λ_{ij} are transition specific parameters to be estimated by maximum likelihood procedure.

The above specification has some very convenient properties. Exponentiation guarantees non-negativity of the estimated hazard. The log of the likelihood function is separable in the transition specific hazard functions, so each parameter vector $[\, \beta_{ij}, \gamma_{ij}\,, \lambda_{ij}\,]$ can be estimated using type i spells only. The Box-Cox specification of the duration term encompasses a variety of duration dependence forms frequently found in the literature: restricting $\lambda_{1ij} = 0$ and $\gamma_{2ij} = 0$ produces a Weibull, or logarithmic specification; $\lambda_{1ij} = 1$ and $\gamma_{2ij} = 0$ produces a Gompertz specification. Restricting the λ_{ij}'s to be integers, produces a quadratic specification. The most common forms of duration dependence employed are the logarithmic (Weibull) and the quadratic: estimation results for these specifications are reported in the following section.

The issue of unobserved heterogeneity

The issue of the potential impact of unobserved heterogeneity on the estimation needs to be briefly addressed. A number of techniques have been recently developed to handle the difficulties created by the presence of unobserved heterogeneity in duration models. One method is to assume that the unobserved heterogeneity component in drawn from a (flexible) parametric distribution (such as gamma or log-normal). Heckman and Singer (1984) demonstrate that an incorrect assumption about the parametric form of the distribution of the unobserved heterogeneity component can lead to grossly incorrect inference about duration dependence, as well as about the effect of other covariates. The two authors strongly recommend the use of non-parametric methods. Although some progress has been made to design such methods, they are still in their infancy, especially with respect to models with time-varying covariates and with a competing risks specification. For this reason no attempt is made in this paper to explicitly control for unobserved heterogeneity.

The competing risks model

The results illustrated in the first paragraph of this section apply only to a two-state model, or to an higher dimensional model if transitions from state i are possible to only one of the remaining states. The latter restriction is not plausible in the context of three-state model of labor force dynamics, where an unemployed person can be considered *at risk* of both getting a job and dropping out of the labor force. This is the rationale for the application of the *competing risks* model, developed by the biostatistics literature to take into account the possibility of multiple causes of death (see Kalbfleish and Prentice, 1980). In this model, an observed transition to state j (i.e. a completed spell of type ij) represents at the same time a ***right censored*** spell of type ik, since the individual was also at risk of transiting to state k.

The density of (completed) spells of type ij is then

$$f_{ij}(t) = h_{ij}(t) \exp\left[-\int_0^t h_{ij}(u)\,du\right] \exp\left[-\int_0^t h_{ik}(u)\,du\right] \qquad (12)$$

where the last exponential term represents the survivor function for a spell of type ik.

When right censored, a spell of type i enters the likelihood in the following way:

$$S_i(t) = \exp\left[-\int_0^t h_{ij}(u)\,du\right] \exp\left[-\int_0^t h_{ik}(u)\,du\right] \tag{13}$$

Here the rationale for the competing risks model is even more intuitive: a spell for which the destination state is not observed is potentially of both type ij and ik.

It can be noted, however, that the log likelihood factors into *transition specific* components: the parameters of the h_{ij} hazard can be estimated by maximizing only the following log likelihood function:

$$L_{ij} = \sum_{i\ spells} \left[\delta \log h_{ij}(t_i) - \int_0^{t_i} h_{ij}(u)\,du\right] \tag{14}$$

where δ is equal to zero if the spell is right censored *or* if the spell terminates in state k, and equal to one otherwise. The summation runs over *all* spells of type i.

The initial condition (left censoring) problem

Most of the duration analysis literature assumes that the origin date of each sampled spell coincides with the start date of the sample, i.e. there are no left censored spells. However, this condition is not met by most of the longitudinal survey sampling schemes, including SIPP.

Left censored spells do not have the same distribution as spells that start after the beginning of the sample, since they are not a random sample from the population of spells. The reason is intuitively clear. While individuals that enter the state at each instant after the start date represent ***flows***, individuals with a spell in progress at the onset of the survey represent a ***stock***. Such stock is formed by the "survivors" of all the preceding cohorts of entrants (flows), and has, in general, a different composition from the "typical" flow (even assuming time stationarity). Spells sampled at a particular moment in time are defined as ***length biased*** since the probability of being sampled is proportional to their length [1]. The only case where flows and stock have the same composition is when the distribution of spell lengths is exponential (i.e. the hazard exibits no duration dependence) and the population is homogeneous, conditions which are unlikely to be met in most situations arising in the social sciences.

Under certain conditions, the solution of the initial conditions problem can be that of excluding left censored spells from the estimation, which is equivalent to "sampling the flows". Heckman and Singer (1984) have shown that, in a time homogeneous environment with no unobservable heterogeneity, using only spells that begin after the start date of the sample gives inefficient but consistent parameter estimates. This is the solution adopted in this paper. Excluding left censored spells here has the additional advantage of being able to observe, for all cases, the labor force state the individuals occupies before the current unemployment spell. The loss of efficiency is not really an issue, given the size of the remaining sample.

[1] A tangible example of length bias is offered in table 3.1 below

3. THE DATA

The Survey of Income and Program Participation (SIPP) is a longitudinal survey designed primarily to collect information on income and transfers recipiency from various government sources. It contains also detailed information on labor force status, wage and other characteristics relevant to labor market studies.

The survey is organized according to the following design. Starting with 1983, and in every year thereafter, a probability sample of the US non-institutional population is selected: each sample is called a "panel". A panel is interviewed every four months for a period of about three years (32 months). Each interview is defined as a "wave". Within each wave, interviews are conducted according to a staggered design: one fourth of the sample is interviewed every month. Each such fraction is called a "rotation group". The data utilized in this study are drawn from the 1984 panel, first interviewed in October 1983. The data cover the period from June 1983 to March 1986 (first 8 waves). The initial sample size of the 1984 panel was of 20,000 households.

Labor force status variables is recorded for each single week in the reference period (i.e. the 17 or 18 weeks preceding the interview). From the raw data I constructed weekly labor force histories, where in each week individuals are classified as having a job, looking for work (or on layoff from a job), or, residually, as out of the labor force. After selecting out children (aged 15 or less) and those individuals that were not present in the survey for at least three subsequent interviews, a sample of size 35,300 was obtained. Of these, 16,200 were continuosly employed during the entire time on the survey, 9,500 were continuosly out of the labor force. The remaining 9,600 changed labor force status at least once during the survey period, or were continuously unemployed.

The weekly labor force histories were then utilized to construct *spell duration data.* For each spell of employment, unemployment or non participation, a number of items were computed:

i) length of the spell in state i;

ii) a censoring indicator;

iii) destination state j or k, when the spell is completed within the sampling frame;

iv) previous state j or k, when the spell begins after the start date of the sample;

v) the serial order of the spell since the start of the survey;

vi) the calendar week of start;

vi) a vector of covariates, which can be either fixed (i.e. at the value they had at the onset of the spell) or time-varying (in which case the entire, usually monthly, time path of the variable was recorded).

[redacted]
Table 3.1 reports average durations (expressed in weeks) of unemployment spells for the male subsample. Table 3.2 reports the same results for women. In the upper panels of these tables the spells are broken down according to their censoring status (rigth censored, left censored, completed ending in employment (UE) and completed ending in non-participation (UO)), and then according to age group and race. In panel (a) all the observed spells are utilized, while panel (b) represents a restricted sample, containing only the first non left censored spell for each individual ever observed to be unemployed over the sampling frame. Purpose of this restriction is to to avoid oversampling of individuals with multiple spells (discussion of this issue is deferred to section 4 below). Durations are uniformly higher when only one spell per individual in considered, as expected, since, over a fixed time frame, repeated spells are on average associated with shorter durations.

Average durations of right and left censored spells clearly show the effect of *length bias*: (i.e. spells that are in progress at a particular point in time have higher mean duration that those observed over a period of time). The figures reported for censored spells represent averages of their observed portions only: nevertheless, these average incomplete durations are between two and three times longer than the average duration of completed spells. Spells that are both right and left censored have been considered as left censored (their number is relatively small, 37 for men and 11 for women): most of these spells belong to individuals who left the sample before the end of the survey. Their average duration is in fact 77 and 64 weeks respectively, while the total sampling frame is 140 weeks.

The overall pattern of duration across genders and age and racial groups confirm the qualitative results obtained in the past from other data sets. The novelty here is that spells are broken down by *"destination state"*, i.e. the labor force state entered when the unemployment spell ends. Blacks, men and mature individuals tend to have longer spells than their counterparts, and this pattern is more pronounced for UE spells than for UO spells. UO spells are relatively more frequent for blacks, women and young people, i.e. for groups with traditionally a lower degree of labor market attachement. ~~Completed UE spells are longer than UO spells: a result common to all groups~~.

The lower panels of tables 3.1 and 3.2 contain a breakdown of mean durations according to destination state and *"previous state"* (i.e. the labor force state from which the individual entered unemployment). The previous state has a relatively scarce impact on the duration of UE spells, while for UO spells, previous employment produces an average duration between two and three times longer than previous non-participation. These results, paired with those reported in the following two tables, suggest that previous state is an important predictor of the outcome of an unemployment spell, especially when analyzing UO spells. These results suggest also that the Markovian assumption widely used in labor market studies (the future depends on the past only through the present state) is grossly inappropriate, at least when population heterogeneity is not controlled for.

Table 3.1 AVERAGE DURATION OF UNEMPLOYMENT SPELLS - MALE SUBSAMPLE

(a) all spells

subpopulation	mean	std dev	cases
for entire population	11.4	14.8	7835
right censored spells	19.5	20.2	743
age 17-24	19.2	19.5	293
non black	17.0	17.1	225
black	26.5	24.7	68
age 25-64	19.7	20.7	450
non black	18.2	18.7	371
black	26.8	27.3	79
left censored spells	20.4	20.9	1149
age 17-24	16.7	17.7	467
non black	15.1	15.0	370
black	22.6	24.5	97
age 25-64	22.9	22.6	682
non black	21.9	21.6	569
black	28.0	26.3	113
completed spells, ue	9.5	11.1	4014
age 17-24	9.7	10.5	1628
non black	9.2	9.5	1468
black	14.6	16.3	160
age 25-64	9.4	11.4	2386
non black	9.4	11.3	2134
black	9.6	12.2	252
completed spells, uo	6.7	10.7	1929
age 17-24	6.1	9.6	993
non black	5.8	9.1	738
black	7.0	10.8	255
age 25-64	7.3	11.7	936
non black	7.5	11.9	729
black	6.4	11.2	207

(b) first spell only

subpopulation	mean	std dev	cases
for entire population	12.5	14.6	3097
right censored spells	25.9	24.1	279
age 17-24	23.9	23.3	88
non black	21.4	19.3	65
black	30.9	31.5	23
age 25-64	26.8	24.5	191
non black	25.7	24.0	152
black	31.2	26.0	39
completed spells, ue	11.4	12.2	2059
age 17-24	10.7	10.7	821
non black	10.2	10.0	742
black	15.5	15.5	79
age 25-64	11.9	13.1	1238
non black	11.7	12.8	1120
black	13.8	15.1	118
completed spells, uo	10.5	13.4	759
age 17-24	8.7	11.7	397
non black	8.1	11.2	287
black	10.4	12.7	110
age 25-64	12.3	14.9	362
non black	12.0	14.8	297
black	14.1	15.1	65

(c) all spells

	mean	std dev	cases
to employed	10.8	12.1	4804
from employed	9.1	10.8	3281
from out of l.f.	11.5	11.9	733
from left censored	16.9	15.0	790
to out of l.f.	8.9	13.9	2249
from employed	12.5	13.5	425
from out of l.f.	5.0	9.1	1504
from left censored	22.0	21.6	320
to right censored	22.4	24.4	782
from employed	17.5	19.2	554
from out of l.f.	25.2	22.0	189
from left censored	77.5	31.4	39

(d) first spell only

	mean	std dev	cases
to employed	11.4	12.2	2059
from employed	11.4	12.3	1652
from out of l.f.	11.3	11.7	407
to out of l.f.	10.5	13.4	759
from employed	15.0	14.9	247
from out of l.f.	8.3	12.1	512
to right censored	25.9	24.1	279
from employed	25.9	25.4	205
from out of l.f.	25.7	20.3	74

Table 3.2 AVERAGE DURATION OF UNEMPLOYMENT SPELLS - FEMALE SUBSAMPLE

(b) first spell only

subpopulation	mean	std dev	cases
right censored spells	20.3	19.9	214
age 17-24	18.5	17.3	74
non black	17.1	17.4	57
black	23.0	16.5	17
age 25-64	21.3	21.1	140
non black	21.1	22.2	116
black	22.3	15.3	24
completed spells, ue	9.9	10.6	1504
age 17-24	9.0	9.0	578
non black	8.7	8.9	524
black	11.9	9.7	54
age 25-64	10.4	11.4	926
non black	10.0	10.5	824
black	13.6	16.9	102
completed spells, uo	8.4	10.7	1334
age 17-24	7.4	9.5	498
non black	6.9	8.2	382
black	8.86	12.7	116
age 25-64	9.1	11.3	836
non black	8.7	10.7	678
black	10.5	13.5	158

(a) all spells

subpopulation	mean	std dev	cases
for entire population	9.3	12.0	7167
right censored spells	18.9	18.0	512
age 17-24	18.8	18.4	193
non black	18.0	19.3	146
black	21.2	15.5	47
age 25-64	19.0	17.7	319
non black	18.3	18.1	263
black	22.0	15.5	56
left censored spells	16.5	15.9	806
age 17-24	14.5	15.0	292
non black	14.1	14.9	230
black	16.0	15.5	62
age 25-64	17.7	16.2	514
non black	17.7	16.4	420
black	17.8	15.2	94
completed spells, ue	9.1	9.8	2671
age 17-24	8.4	8.9	1077
non black	8.1	8.6	958
black	10.9	10.9	119
age 25-64	9.5	10.4	1594
non black	9.3	9.7	1395
black	11.3	14.3	199
completed spells, uo	6.1	9.5	3178
age 17-24	5.8	8.8	1125
non black	5.6	8.0	788
black	6.1	10.3	337
age 25-64	6.3	9.9	2053
non black	6.3	9.5	1554
black	6.1	10.8	499

(d) first spell only

to employed	9.9	10.6	1504
from employed	10.1	10.4	894
from out of l.f.	9.6	10.9	610
to out of l.f.	8.4	10.7	1334
from employed	13.0	12.8	295
from out of l.f.	7.1	9.6	1039
to right censored	20.3	19.9	214
from employed	19.2	20.0	112
from out of l.f.	21.6	19.7	102

(c) all spells

to employed	9.8	10.6	3117
from employed	8.9	9.6	1643
from out of l.f.	9.4	10.3	1028
from left censored	14.2	13.5	446
to out of l.f.	7.3	10.9	3527
from employed	12.1	12.5	470
from out of l.f.	5.0	8.5	2708
from left censored	17.9	15.8	349
to right censored	19.9	19.2	523
from employed	17.0	18.3	262
from out of l.f.	20.9	17.4	250
from left censored	64.9	23.7	11

Tables 3.3 and 3.4 contain a breakdown of spells on the basis of the previous and destination spell. Panels (a) contain data for the entire sample, while (b) and (c) are again restricted to the first non left-censored spell for each individual. In panel (c) the counts are relative to completed spells only. From the row margin of panel (c) we obtain that 73 per cent of completed unemployment spells end up in employment for men, vs. 53 per cent for women. This result bears a close resemblance to that obtained by Clark and Summers using CPS gross flows (Clark and Summers, 1979). This fact is particularly interesting, since CPS flows data derive from monthly discrete observations, while SIPP spell data are based on continuous measurement (moreover, the reference period is quite different in terms of the overall state of the economy).

Table 3.3 BREAKDOWN OF UNEMPLOYMENT SPELLS BY PREVIOUS AND DESTINATION STATE - MALE SUBSAMPLE

(A)

DESTINATION STATE

PREVIOUS STATE (ROW PCT / COL PCT)	EMPLOYED	OUT OF L.F.	RIGHT CENSORED	ROW TOTAL
EMPLOYED	77.0 68.3	10.0 18.9	13.0 70.8	4260 54.4
OUT OF L.F.	30.2 15.3	62.0 66.9	7.8 24.2	2426 31.0
LEFT CENSORED	68.8 16.4	27.9 14.2	3.4 5.0	1149 14.7
COLUMN TOTAL	4804 61.3	2249 28.7	782 10.0	7835 100.0

(B)

DESTINATION STATE

PREVIOUS STATE (ROW PCT / COL PCT)	EMPLOYED	OUT OF L.F.	RIGHT CENSORED	ROW TOTAL
EMPLOYED	78.5 80.2	11.7 32.5	9.7 73.5	2104 67.9
OUT OF L.F.	41.0 19.8	51.6 67.5	7.5 26.5	993 32.1
COLUMN TOTAL	2059 66.5	759 24.5	279 9.0	3097 100.0

(C)

DESTINATION STATE

PREVIOUS STATE (ROW PCT / COL PCT)	EMPLOYED	OUT OF L.F.	ROW TOTAL
EMPLOYED	87.0 80.2	13.0 32.5	1899 67.4
OUT OF L.F.	44.3 19.8	55.7 67.5	919 32.6
COLUMN TOTAL	2059 73.1	759 26.9	2818 100.0

Table 3.4 BREAKDOWN OF UNEMPLOYMENT SPELLS BY PREVIOUS AND DESTINATION STATE - FEMALE SUBSAMPLE

(A)

DESTINATION STATE

ROW PCT COL PCT PREVIOUS STATE	EMPLOYED	OUT OF L.F.	RIGHT CENSORED	ROW TOTAL
EMPLOYED	69.2 52.7	19.8 13.3	11.0 50.1	2375 33.1
OUT OF L.F.	25.8 33.0	67.9 76.8	6.3 47.8	3986 55.6
LEFT CENSORED	55.3 14.3	43.3 9.9	1.4 2.1	806 11.2
COLUMN TOTAL	3117 43.5	3527 49.2	523 7.3	7167 100.0

(B)

DESTINATION STATE

ROW PCT COL PCT PREVIOUS STATE	EMPLOYED	OUT OF L.F.	RIGHT CENSORED	ROW TOTAL
EMPLOYED	68.7 59.4	22.7 22.1	8.6 52.3	1301 42.6
OUT OF L.F.	34.8 40.6	59.3 77.9	5.8 47.7	1751 57.4
COLUMN TOTAL	1504 49.3	1334 43.7	214 7.0	3052 100.0

(C)

DESTINATION STATE

ROW PCT COL PCT PREVIOUS STATE	EMPLOYED	OUT OF L.F.	ROW TOTAL
EMPLOYED	75.2 59.4	24.8 22.1	.1189 41.9
OUT OF L.F.	37.0 40.6	63.0 77.9	1649 58.1
COLUMN TOTAL	1504 53.0	1334 47.0	2838 100 0

The sample survivor function

The sample survivor function is a useful representation of the duration pattern of the data. In fig. 1 the Kaplan-Meier or product limit estimator of the survivor function is reported for the sample of non left censored unemployment spells, with no selection on destination state (i.e. UO and UE spells are pooled together). The dashed lines represent a 95% confidence interval, computed according to *Greenwood's formula* (see Kalbfleish and Prentice, 1980). By the 26th week (traditionally used as a cutoff point to indicate long-term unemployment) only about 15 per cent of men are still unemployed, and 10% of women.

Fig. 2 contains estimates of the sample survivor function for UO spells only, based on to two different assumptions. The upper estimate represents the product limit estimate computed according to the competing risks specification: spells ending in employment are considered as right censored. The lower estimate is computed *excluding* UE spells. The large proportion of UE spells causes the two estimates to diverge dramatically, especially for men. The upper estimate assumes that the two causes of "death" are independent, while the lower assumes an extreme form of dependence: the unemployed who end up getting a job were never at risk of dropping out of the labor force, one risk excludes the other.

One distinctive feature of these estimated survivor functions is the extremely irregular shape, with sharp "steps" at regular intervals. These "steps" would correspond to spikes in the hazard function. Such irregularities reveal a measurement problem with spell data obtained from SIPP, that has been defined as the "seam-transition" problem. Possibly due to recall bias, interviewees tend to place the start date of spells in the first week of the each reference period. As a consequence, an abnormal number of transitions are observed at the seam between two waves. This produces "heaping" in the frequency distribution of spell duration, at values that are multiples of the length of the reference period. This is what is observed in fig. 2 around the 18th, 34th and 52nd week. The consequence of heaping is also visible in fig.3, which reports a plot of the log-log survivor function: linearity (Weibull distribution) holds between the "steps", but not overall.

Fig 1 KAPLAN-MEIER ESTIMATE OF THE SURVIVOR FUNCTION:
all unemployment spells, male subsample

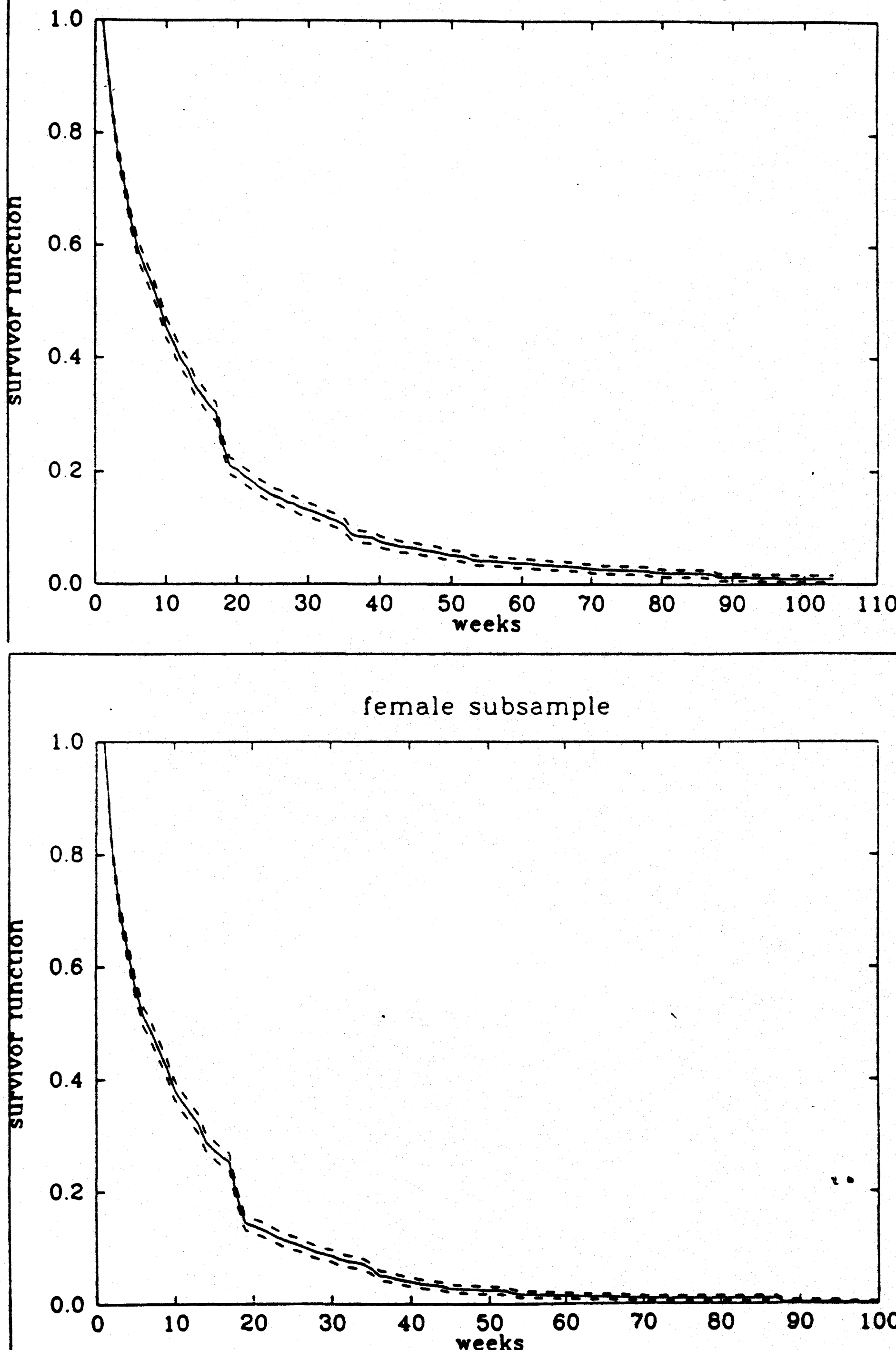


Fig.2 KAPLAN-MEIER ESTIMATES OF THE SURVIVOR FUNCTION:
UO spells, male sample

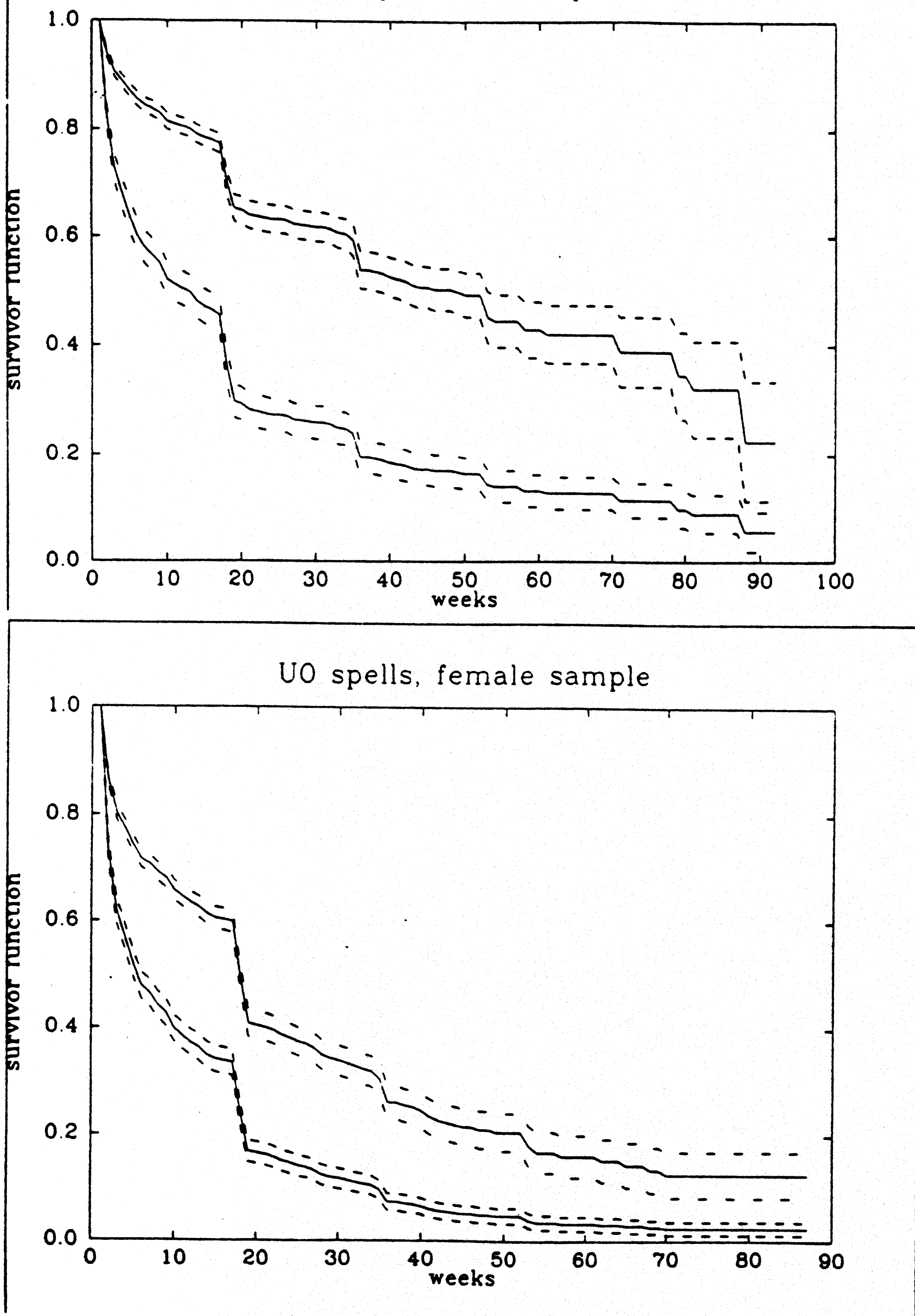
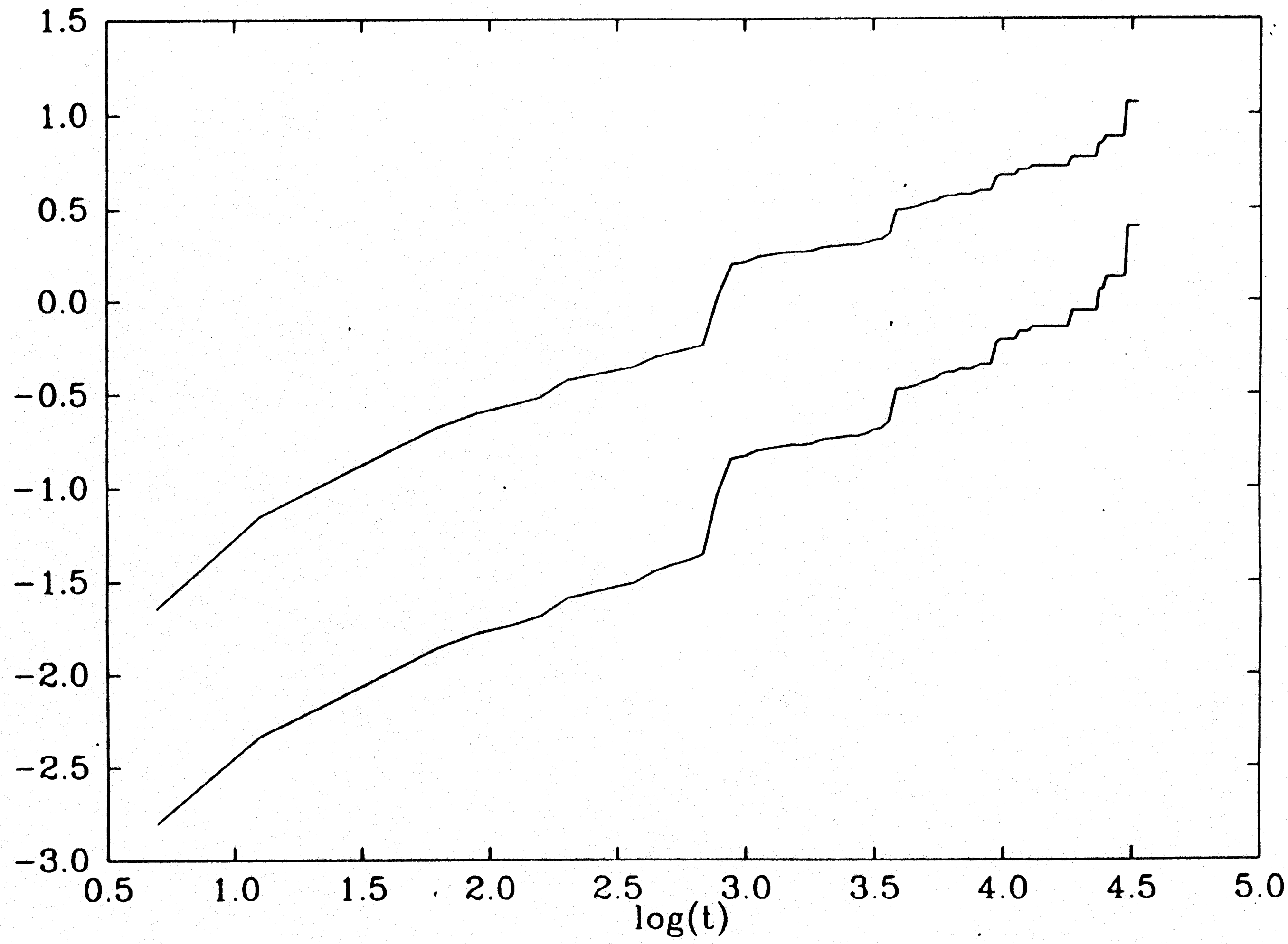


Fig 3 LOG(−LOG(SURVIVOR FUNCTION)) vs. LOG(t): UO spells, male sample



4. ESTIMATION RESULTS

In this section estimation results are presented for several parametric specifications of the UO hazard. All the estimates presented here are obtained from a sample containing only one spell per individual, the first non left censored spell observed. While some individuals appears with only one spell in the sample, others (almost 50%) have multiple spells. The rationale for the above selection is the following. In the parametric hazard specification used here, no attempt is made to control for the effect of unobservable heterogeneity on the probability of transition. If such unobservable component is significant and (as it is likely the case) is correlated across spells for the same individual, utilizing all the observed spells implies oversampling the "unobserved type" with multiple spells. The alternative could be that of weighting the data, a solution not pursued here.

a) Variables selection criteria and interpretation

Time-invariant covariates

A common set of explanatory variables is introduced in all specifications to control for the effect of observable characteristics on the probability of labor force withdrawal: this set includes age, education, race, marital status and residence in a metropolitan area. These variables are treated as time-invariant, in the sense that they are given the value they assume at the onset of the spell.

Age squared is introduced in the hazard together with age to control for non linearities in the age effect. Individuals over 64 and below 17 are selected out of the sample in order to exclude demographic groups with idiosyncratic labor force experience. Despite such restriction, the age effect maintains a "convex" pattern, with a minimum around age 35-40 (see table 4.1 and 4.2). This suggests that the "employers think too young or too old" listed by the CPS (and also SIPP) questionnaire as a reason for not looking for work, is a relevant cause of discouragement, holding constant labor market conditions. It is interesting to notice how such convex pattern is more pronounced for men that it is for women. The interpretation could be that for women, since they are already disproportionately represented among secondary workers, the fact of being at the extremes of the age distribution has a less strong effect on the probability of discouragement than it has for men.

Race is expressed as a dummy variable, equal to one for blacks, zero otherwise: the effect on withdrawal is generally positive, reflecting the scarcer labor market opportunities that blacks on average face. Marital status (equal to one if married with spouse present, zero otherwise) rarely shows a significant effect for men, although the sign is always negative. On the contrary, for women being married has a positive and significant effect on the probability of withdrawal. This is broadly consistent with the empirical evidence on the effect of marital status on the labor supply of women.

Education is expressed in years of schooling completed. Together with marital status, education is the variable that shows the most relevant difference between the two genders. While for men the effect of education is positive but almost never significant [1], withdrawal is less likely for more educated women: the interpretation could be that for women education is a better proxy for labor force attachement than it is for men.

One additional time-invariant explanatory variable is introduced in some of the specifications of the conditional hazard function: a dummy variable that takes a value of one if the individual was previously (i.e. in the spell immediately before the current one) employed, zero if s/he was out of the labor force. The motivation for such inclusion is two-fold. On one hand, the purpose is to test the Markovian assumption frequently made in labor force studies, according to which future states depend on the past only through the present state. The estimated coefficient of such "previous state" variable shows a very strong and significant negative effect on the probability

[1] The model of table 4.1 was reestimated substituting education as a continuous variable with two dummies, one representing attained college degree and the other high school diploma. For the male subsample, the estimated coefficients (standard errors) were 0.23 (0.13) for college and -0.92 (0.089) for high school. Among women they were both negative and significant (respectively -0.27 (0.10) and -0.11 (0.058)

of withdrawal, suggesting the presence of "lagged occurrence dependence": those that enter unemployment from non-participation have a much higher probability of interrupting their search efforts than those that previously held a job. Paraphrasing Heckman and Borjas (1981), non-participation seems to cause future non-participation.

This results has also an alternative (and more plausible) interpretation, in terms of unobserved heterogeneity: some individuals have a lower labor force attachment than others, and this unobservable component is strongly correlated with previous labor market experience. The order of causality here is different: it is not the previous experience which permanently "scars" the individual, causing subsequent withdrawal, but is the heterogeneity component which "causes" both previous labor force state and future transition behavior. In the models estimated in b) below, previous state is controlled for with a dummy variable. In the following section, instead, the model is reestimated *selecting* on previous state.

Time-varying covariates

The motivation for one of the time-varying covariates has been discussed in section 3: in order to control for the abnormal number of transitions that take place at the wave-to-wave seams, a dummy variable is utilized, which takes a value of one on the last week of each reference period, zero at any other time. The estimated coefficient of this variable always shows a very strong positive "effect", which merely reflects the existence of the measurement error. The estimated coefficient is always around 2.5, indicating that at the seam the probability of transition is three and a half times higher than in all remaining weeks.

Recipiency of unemployment insurance benefits has been introduced in the conditional hazard function as a dummy variable, equal to one if benefits were received during the month, zero otherwise. The choice of a dummy, instead of the more common *replacement rate*, is motivated by the fact that a fraction of unemployed in the sample were not previously employed, and for this group the replacement rate in not defined. Since spell duration is expressed in weeks, monthly values of the dummy variable have been imputed to each week in the month. This introduces some "noise" in the data, but the alternative solution (aggregating duration data using months as a unit of measurement) would have been even more problematic, since a substantial proportion of spells are less than four weeks long. The same imputation procedure has been used for the local unemployment rate. Seasonally unadjusted monthly figures at the State level have been utilized for this variable.

TABLE 4.1 ESTIMATED HAZARD PARAMETERS FOR TRANSITION BETWEEN UNEMPLOYMENT AND NON-PARTICIPATION

DURATION SPECIFICATION	MEN		WOMEN	
	LOGARITHMIC	QUADRATIC	LOGARITHMIC	QUADRATIC
CONSTANT	- 1.894 (5.09)	- 2.237 (6.18)	- 1.764 (6.46)	- 2.066 (7.97)
AGE /10	- 0.840 (4.49)	- 0.886 (4.84)	- 0.191 (1.36)	- 0.220 (1.65)
AGE SQUARED /100	0.119 (5.18)	0.124 (5.39)	0.027 (1.42)	0.031 (1.72)
EDUCATION /10	0.087 (0.60)	0.095 (0.67)	- 0.336 (3.20)	- 0.341 (3.30)
RACE	0.237 (2.52)	0.206 (2.26)	0.249 (3.66)	0.233 (3.38)
MARITAL STATUS	- 0.124 (1.17)	- 0.084 (0.83)	0.106 (1.71)	0.110 (1.86)
RESIDENCE IN METRO AREA	0.155 (1.84)	0.189 (2.27)	0.025 (0.42)	0.037 (0.64)
PREVIOUS SPELL	- 1.593 (16.1)	- 1.615 (16.5)	- 1.209 (16.7)	- 1.237 (17.4)
WAVE-TO-WAVE SEAM DUMMY	2.669 (30.6)	2.610 (29.0)	2.484 (40.7)	2.427 (36.7)
RECEPIENCY OF UI BENEFITS	- 0.441 (2.67)	- 0.459 (2.78)	- 0.506 (3.77)	- 0.527 (3.96)
LOCAL UNEMPLOYMENT RATE/100	5.148 (2.61)	5.585 (2.87)	3.699 (2.66)	3.946 (2.95)
LOG(DURATION)	- 0.545 (15.5)	-	- 0.604 (23.7)	-
DURATION /10	-	- 0.726 (12.7)	-	- 0.849 (16.3)
DURATION SQUARED /100	-	0.068 (10.3)	-	0.077 (11.4)
	(absolute value of asymptotic t statistic in parentheses)			
N	3109	3109	3532	3532
Right censored spells	279	279	254	254
LOG LIKELIHOOD	- 2989.9	- 3037.6	- 5199.5	- 5339.4

b) Estimation results with alternative duration specifications.

A baseline model is estimated with two alternative specifications for the duration term, quadratic and logarithmic. The results are shown in table 4.1 . The size and significance of the estimated coefficients prove to be fairly robust across the two specifications. The unemployment rate variable shows a positive effect, for both genders, confirming the existence of a discouraged worker effect. The size of the effect is such that a one percent increase in unemployment, *ceteris paribus*, would cause about a 5 percent increase in the probability of withdrawal among men, 4 percent among women.

Blacks are about 25 percent more likely to withdraw from the labor force than non-blacks, other things being equal. For men, education and marital status do not seem to have any discernible effect: for women one additional year of education reduces the probability of withdrawal by 3 percent, while being married increases it by 10 percent. Receipt of unemployment insurance benefits reduces the probability of leaving the labor force by more than 40 percent for men, 50 for women. This result matches with the negative effect traditionally found for UI recepiency on the probability of reemployment. It should be stressed that in the current specification of the hazard the fact of being previously employed (which usually is a condition for UI eligibility) is already controlled for by the "previous state" variable: hence that of UI benefits is a net effect. Being previously employed by itself reduces the probability of withdrawal by two and a half times: this very strong effect motivates the separate estimation conducted in c) below, where the sample is *selected* on previous state.

The duration term shows a negative sign in the logarithmic specification, and a convex pattern in the quadratic, with a minimum around 50 weeks. A plot of the two hazards for the male subsample is shown in fig. 4.1. While it is not possible to formally discriminate between the two models with a likelihood ratio test, since they are non-nested [2], an heuristic argument can be formulated by inspecting the plot. The rising portion of the quadratic

[2] The likelihood ratio test is able to discriminate between the quadratic and the linear, or Gompertz, specifications, since they are nested. The Gompertz is in effect rejected by

hazard is almost entirely beyond the empirically relevant range: in fact by the 50th weeks almost 95 percent of the unemployment spells that end up in withdrawal are completed. One could then argue that the logarithmic specification is a more parsimonious representation of the duration pattern.

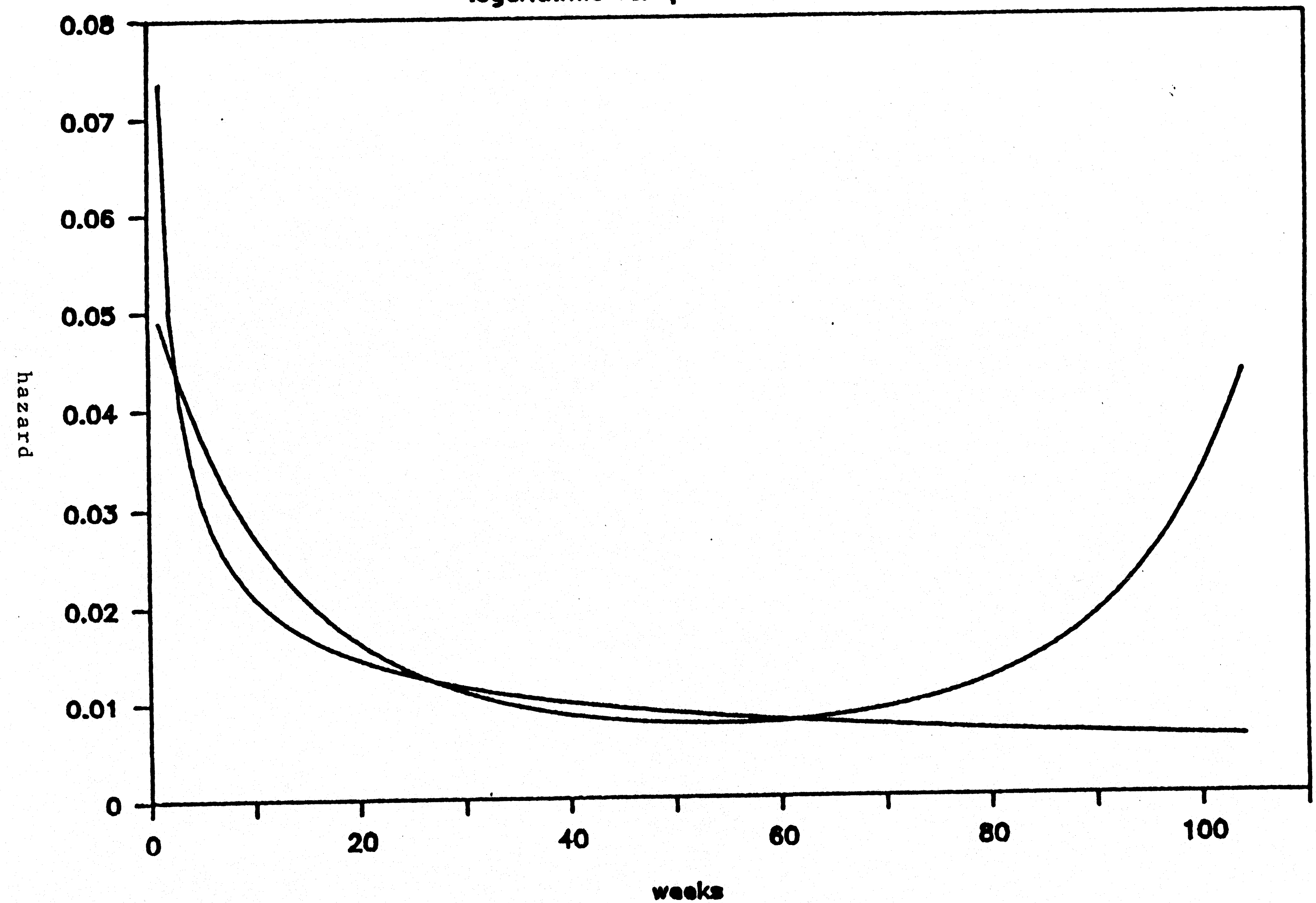
A general comment is necessary on the significance of a negative sign on the duration term. It is a well known result in the duration analysis literature that a such negative sign is not necessarily an indication of true negative duration dependence (i.e. of the fact that the probability of withdrawal decreases with elapsed time in unemployment). This result could be also due to the presence of unobservable heterogeneity. Take an extreme example. There are two types of individuals, one with low probability of transition, the other with high probability, and this component is not observable. The high probability type tends to leave first, leaving a sample which is more and more disproportionately composed of low probability individuals. When estimation is performed on the aggregate sample, this changing composition shows up *biasing* the duration terms toward negative values.

---

the data

fig 4.1 ESTIMATED UO HAZARD

logarithmic vs. quadratic duration



The second specification tested contains an additional interaction term between duration and the unemployment rate variable. The estimation results are shown in table 4.2. To motivate this extension of the model, it is useful to give some *structural* interpretation to the use of the unemployment rate variable. Unemployment rate represents a proxy for the impact of labor market conditions on the job prospects of the unemployed. In "informal" search theoretic terms, the unemployment rate could proxy for the rate of arrival of job offers (which depends as well on personal characteristics). An unemployed worker withdraws from the labor force when the discounted expected utility of search falls below that of staying out of the labor force: the rate of arrival of offers and is crucial in determining the direction of this inequality.

Allowing only for a direct, contemporaneous effect of the unemployment rate on the probability of withdrawal, independent of duration, is equivalent to assuming that the rate of arrival of wage offers is known to the unemployed from the start of the spell. This might overlook some important feature of the job search process. It probably takes time to the unemployed worker to build up an estimate of his/her job market possibilities. A test of this "learning" hypothesis can be performed by allowing for an interaction term between duration and the unemployment rate. The testable implication is that the effect of the interaction term is positive, at the same time reducing in size the main effect (when only the latter is specified, it represents an average effect over the spell). The logarithmic model shown in table 4.2 confirms this prediction. The estimates from the quadratic model offers a mixed picture (positive for men, negative for women) but they are in general scarcely reliable, due to the high collinearity between the two duration terms and the interaction term.

The sign and significance level of all the remaining coefficients is not substantially altered by the introduction on the interaction term.

TABLE 4.2 ESTIMATED HAZARD PARAMETERS FOR TRANSITION BETWEEN UNEMPLOYMENT AND NON-PARTICIPATION

MODEL WITH INTERACTION

	MEN		WOMEN	
DURATION SPECIFICATION	LOGARITHMIC	QUADRATIC	LOGARITHMIC	QUADRATIC
CONSTANT	− 1.654 (4.47)	− 2.042 (5.57)	− 1.534 (5.57)	− 2.073 (8.03)
AGE /10	− 0.880 (4.88)	− 0.942 (5.06)	− 0.217 (1.60)	− 0.198 (1.63)
AGE SQUARED /100	0.124 (5.39)	0.127 (5.29)	0.030 (1.66)	0.026 (1.62)
EDUCATION /10	0.112 (0.80)	0.091 (0.62)	− 0.328 (3.12)	− 0.372 (3.89)
RACE	0.241 (2 53)	0.308 (3.55)	0.230 (3.43)	0.223 (3.48)
MARITAL STATUS	− 0.121 (1.16)	− 0.039 (0.39)	0.107 (1.81)	0.128 (2.37)
RESIDENCE IN METRO AREA	0.163 (1.96)	0.159 (1.98)	0.034 (0.58)	0.010 (0.18)
PREVIOUS SPELL	− 1.587 (18.1)	− 1.602 (19.5)	− 1.199 (16.8)	− 1.222 (17.9)
WAVE-TO-WAVE SEAM DUMMY	2.673 (31.0)	2.634 (30.2)	2.477 (41.2)	2.426 (37.3)
RECEPIENCY OF UI BENEFITS	− 0.425 (2.57)	− 0.398 (2.41)	− 0.501 (3.76)	− 0.543 (4.05)
LOCAL UNEMPLOYMENT RATE/100	3.403 (1.67)	6.412 (2.87)	1.944 (1.36)	4.724 (3.23)
LOG(DURATION)	− 0.688 (12.1)	−	− 0.747 (19.1)	−
DURATION /10	−	− 1.035 (13.4)	−	− 0.874 (19.7)
DURATION SQUARED /100	−	0.119 (33.3)	−	0.104 (23.9)
INTERACTION DURATION • UNEM.RATE /100	0.202 (3.38)	0.017 (0.17)	0.236 (5.63)	− 0.065 (0.68)
(absolute value of asymptotic t statistic in parentheses)				
N	3109	3109	3532	3532
Right censored spells	279	279	254	254
LOG LIKELIHOOD	− 2983.9	− 3273.1	− 5186.8	− 5353.3

c) Selecting on previous state

In order to allow for a full interaction between the previous state dummy with all the other coeficients, the model is reestimated selecting on previous state. Only the logarithmic specification is tested here, with and without the interaction term (table 4.3). The results are relative to the male subsample only. The effect of this type of selection is quite substantial. The convexity of the age pattern is increased for previously employed individuals, and greatly reduced for those previously out of the labor force (this difference now closely reproduces the one found between men and women, discussed before). The effect of education is positive only for the previously employed, while the race effect totally disappears for this subgroup. Recepiency of UI benefits has almost no effect for previous non-participants (the effect actually should be zero, since they are not eligible for UI: the residual effect could be caused by measurement errors). The duration effect disappears for previously employed individuals: we saw in table 3.1 that for this group the average time to withdrawal was substantially longer than for the other group. These individuals show relatively fewer very short spells, which account for most of the negative duration dependence.

Another significant impact of this selection is on the unemployment rate variable. Its effect almost totally disappears for previously employed individuals, while it is still very strong for the other subgroup. Hence, the discouraged worker effect seems limited to a subset of the unemployed, those who already have shown a lower labor force attachment.

TABLE 4.3

ESTIMATED HAZARD PARAMETERS FOR TRANSITION BETWEEN UNEMPLOYMENT AND NON-PARTICIPATION

SELECTING ON PREVIOUS SPELL TYPE

	PREVIOUSLY EMPLOYED		PREVIOUSLY OUT OF THE LABOR FORCE	
CONSTANT	- 3.331 (4.82)	- 3.367 (4.61)	- 1.850 (5.02)	- 1.636 (4.48)
AGE /10	- 1.772 (5.53)	- 1.768 (5.56)	- 0.374 (1.63)	- 0.432 (1.93)
AGE SQUARED /100	0.243 (5.92)	0.242 (5.90)	0.054 (1.86)	0.061 (2.10)
EDUCATION /10	0.561 (2.19)	0.560 (2.11)	- 0.166 (0.91)	- 0.134 (0.77)
RACE	0.001 (0.00)	0.002 (0.00)	0.362 (3.46)	0.375 (3.64)
MARITAL STATUS	- 0.210 (1.22)	- 0.210 (1.23)	- 0.013 (0.07)	- 0.005 (0.03)
RESIDENCE IN METRO AREA	0.255 (1.64)	0.254 (1.63)	0.015 (0.16)	0.022 (0.23)
PREVIOUS SPELL	-	-	-	-
WAVE-TO-WAVE SEAM DUMMY	2.922 (21.4)	2.920 (21.0)	2.564 (26.7)	2.557 (26.6)
RECEPIENCY OF UI BENEFITS	- 0.669 (2.92)	- 0.671 (2.91)	- 0.314 (1.20)	- 0.291 (1.12)
LOCAL UNEMPLOYMENT RATE/100	1.477 (0.39)	1.686 (0.42)	4.976 (2.25)	3.488 (1.54)
LOG(DURATION)	- 0.053 (0.83)	- 0.037 (0.33)	- 0.757 (18.9)	- 0.896 (13.5)
INTERACTION DURATION • UNEM.RATE /100	-	- 0.168 (0.17)	-	0.231 (2.55)
	(absolute value of asymptotic t statistic in parentheses)			
N	2148	2148	1211	1211
LOG LIKELIHOOD	- 1198.6	- 1198.5	- 2126.4	- 2121.6

5. CONCLUSIONS

In this paper [redacted] made to estimate the effect of [redacted] the probability that an unemployed worker [redacted] This effect is found to be strong, at least for a substantial fraction of the [redacted] Moreov[redacted] to increase during the course of the spell. The interpretation offered is that the un[redacted] about his/her labor market [redacted]ies as [redacted]

The effect of some personal characteristics on the probability of withdrawal was also found to be very strong. While age, race and marital status have a[redacted] education [redacted]le.

E[redacted] the [redacted]quent withdrawal. When the [redacted] selected on the basis [redacted]und to vir[redacted]ing very strong [redacted].

REFERENCES

Bowen, W., and Finegan, A., 1969, *The Economics of Labor Force Participation*, Princeton University Press, Princeton.

Clark, K., and Summers, L., 1979, "Labor Force Dynamics and Unemployment: a Reconsideration", *Brookings Papers on Economic Activity*, 1.

Flinn, C., and Heckman, J., 1982, "New Methods for Analyzing Structural Models of Labor Force Dynamics", *Journal of Econometrics*, 8.

—, 1982b, "New Methods for the Analyzing Structural Models of Labor Market Dynamics", *Journal of Econometrics*, 18.

Flinn, C., and Heckman, J., 1983, "Are Unemployment and Out of the Labor Force Behaviorally Distinct States?", *Journal of Labor Economics*, 1.

Heckman, J., and Singer, B., 1984, "Econometric Duration Analysis", *Journal of Econometrics*, 24.

Kalbfleish, J., and Prentice, R., 1980, *The Statistical Analysis of Failure Time Data.* Wiley, New York.

Marston, S., 1976, "Employment Instability and High Unemployment Rates", *Brookings Papers on Economic Activity*, 1.

Mincer, J., 1965, "Labor Force Participation and Unemployment", in Gordon and Gordon, eds., *Prosperity and Unemployment*, J. Wiley and Sons, New York.

Tella, A., 1965, "Labor Force Sensitivity to Employment by Age and Sex", *Industrial Relations.*