THE SURVEY OF INCOME AND
PROGRAM PARTICIPATION

THE DISCOURAGED WORKER EFFECT: A REAPPRAISAL USING SPELL DURATION DATA

No. 57

A. Martini
University of Wisconsin-Madison

U. S. Department of Commerce BUREAU OF THE CENSUS

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

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.

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.*

5. CONCLUSIONS

In this paper a still exploratory attempt was made to estimate the effect of labor market conditions on the probability that an unemployed worker withdraws from the labor force. This effect is found to be strong, at least for a substantial fraction of the unemployed. Moreover, this effects seems to increase during the course of the spell. The interpretation offered is that the unemployed worker is learning about his/her labor market opportunities as the spell progresses.

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 an effect broadly consistent with empirical evidence from other areas of labor economics, education shows a positive effect for men and a negative one for women, a result that is not easily interpretable.

Entering the unemployment spell from employment strongly reduces the probability of subsequent withdrawal. When the sample was selected on the basis of such previous experience, the discouraged worker effect was found to virtually disappear for previously employed individuals, while remaining very strong for those who entered from out of the labor force.

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	0.581 (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

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 4.2

ESTIMATED HAZARD PARAMETERS FOR TRANSITION BETWEEN UNEMPLOYMENT AND NON-PARTICIPATION

MODEL WITH INTERACTION

DURATION SPECIFICATION	MEN		WOMEN	
	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)	0.306 (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 (1.98)	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

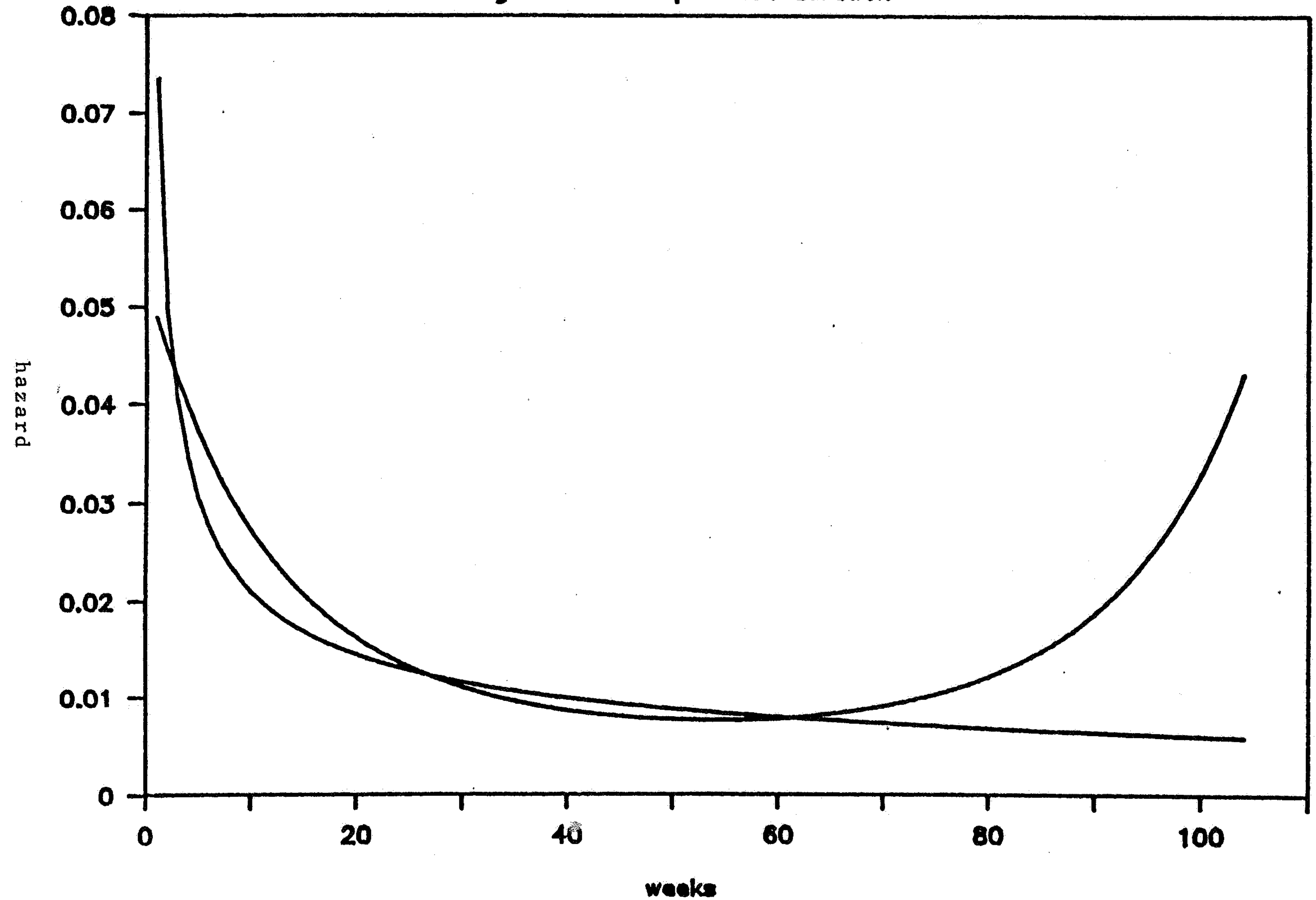
The second specification tested contains an additional interaction term between duration and the unemployment rate variable. The estimation results are shown in table 4.2. To motivate this extension of the model, it is useful to give some *structural* interpretation to the use of the unemployment rate 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.

fig 4.1 ESTIMATED UO HAZARD

logarithmic vs. quadratic duration



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

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

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

DURATION SPECIFICATION	MEN LOGARITHMIC	MEN QUADRATIC	WOMEN LOGARITHMIC	WOMEN QUADRATIC
CONSTANT	- 1.894 (5.09)	- 2.237 (6.18)	- 1.764 (6.46)	- 2.086 (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 (18.1)	- 1.615 (18.5)	- 1.208 (18.7)	- 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)	- 0.508 (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)
N	3109	3109	3532	3532
Right censored spells	279	279	254	254
LOG LIKELIHOOD	- 2989.9	- 3037.6	- 5199.5	- 5339.4

(absolute value of asymptotic t statistic in parentheses) 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.

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 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)

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

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

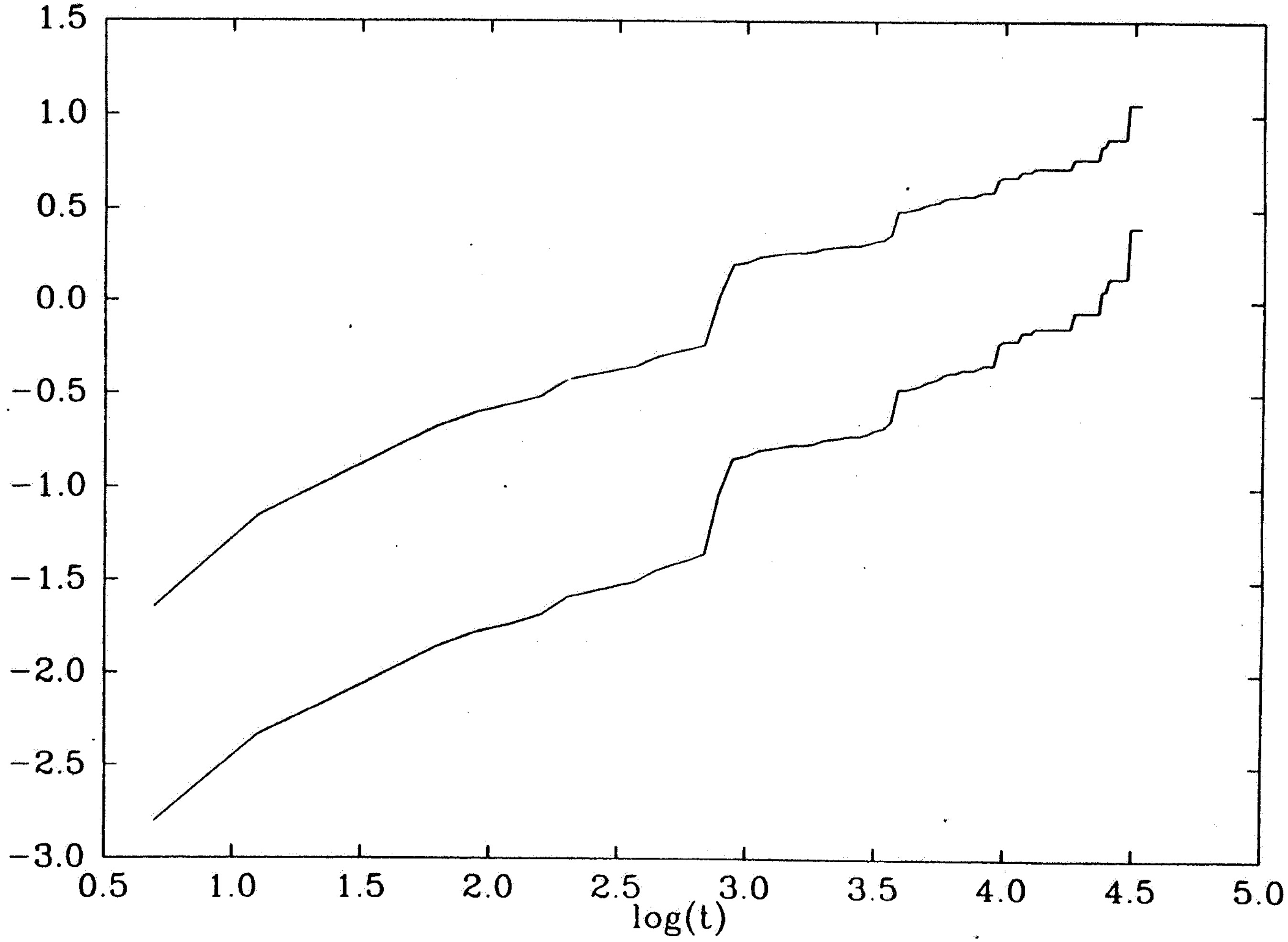


Fig.2 KAPLAN-MEIER ESTIMATES OF THE SURVIVOR FUNCTION:
UO spells, male sample

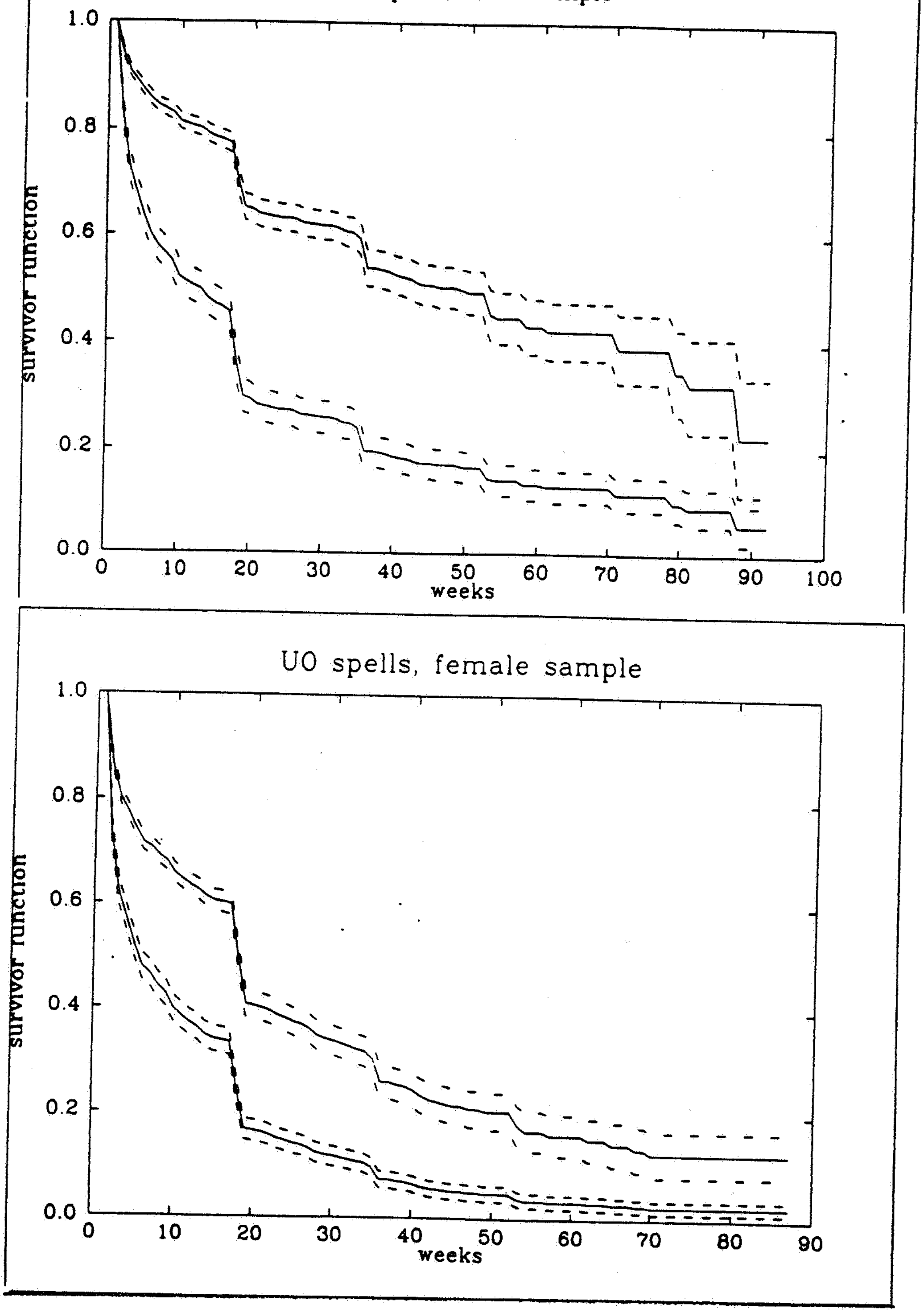
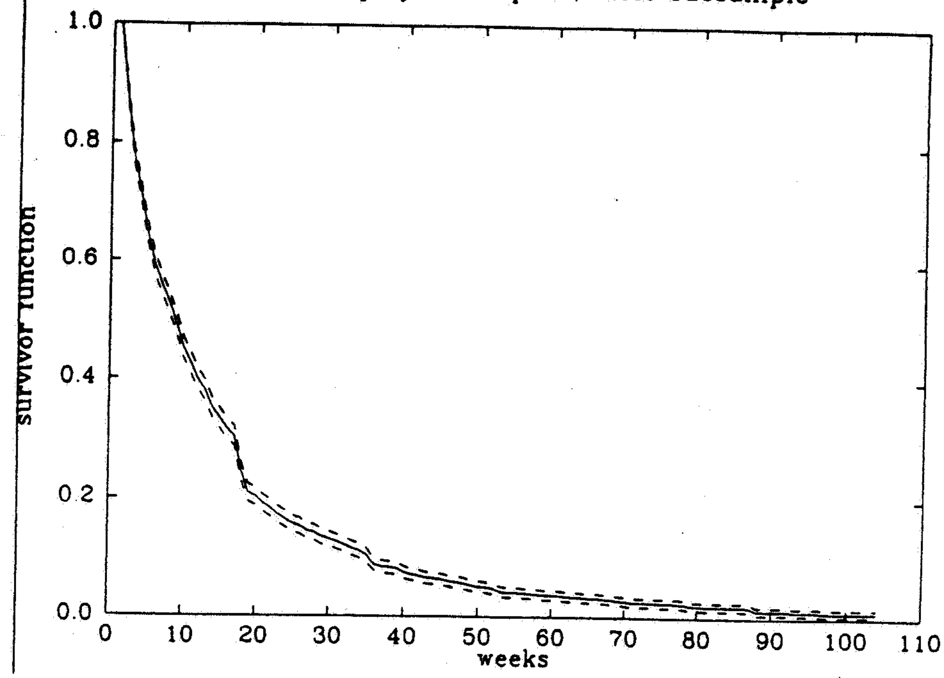
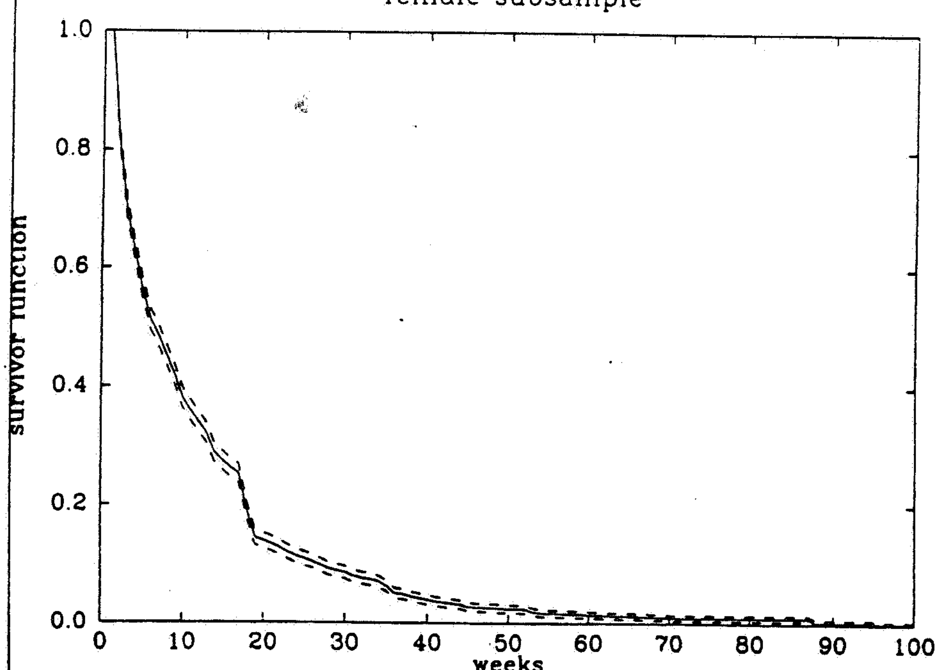


Fig 1 KAPLAN-MEIER ESTIMATE OF THE SURVIVOR FUNCTION: all unemployment spells, male subsample



female subsample



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.

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

(A)

DESTINATION STATE

PREVIOUS STATE (ROW PCT / COL PCT)	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

PREVIOUS STATE (ROW PCT / COL PCT)	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

PREVIOUS STATE (ROW PCT / COL PCT)	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

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

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 3.2 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

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

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

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

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.

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).

A total of about 15,000 unemployment spells were obtained. Tables 3.1 to 3.4 contain descriptive statistics on the subsamples of men aged 17 to 64 (7835 unemployment spells) and women of the same age group (7167 spells). 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

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:

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

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:

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

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:

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

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 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

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

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).

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

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

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 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.

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

where

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

is defined as *survivor function.* Since

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

integrating, we obtain

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

from which

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

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:

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

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

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). 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, 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), and, for applications to socioeconomic phenomena, to Heckman and Singer (1984).

The hazard function

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

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

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:

procyclical movement of the labor force participation rate (Mincer, 1965). 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. The advantage of using individual-level data over aggregate data is by now widely recognized. The use of spell duration data (coupled with that of hazard rate analysis) allows one to look at the *dynamic* aspect of the discouraged worker effect, i.e., whether the labor market conditions affect the rate at which individuals drop out of the labor force. 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.

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). However, most of the empirical and theoretical work on labor market dynamics has focused on two of these flows. On-the-job search and matching models analyze the job-to-job transition. The unemployment search model, and the related empirical work on the duration of unemployment, focuses almost exclusively on the stationary "single risk" situation, where job search is carried on until an acceptable wage offer is received, and a transition to employment occurs. This ignores the fact that a substantial proportion of unemployment spells, especially among women, are reported as terminating in *withdrawal from the labor force* (Clark and Summers, 1979).

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). The main objective is to empirically determine the strength of the relationship between the probability of withdrawal from the labor force and local labor market conditions (as imperfectly proxied by the local aggregate unemployment rate), holding 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.

The results of the analysis show a positive and significant relationship between the unemployment rate and the probability of withdrawal from the labor force. This result is consistent with the *discouraged worker effect* hypothesis, the familiar explanation in labor economics for the observed

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.

Dept. of Economics
1180 Observatory Dr.
Madison, WI 53706

May 16, 1988

Dan Kasprzyk
Population Division
Bureau of the Census

Dear Dan:

please find enclosed the revised copy of my paper. Sorry for the delay. Things get hectic around here at the end of the semester.

57-8810

Sincerely

Alberto Martini

Alberto Martini

Enclosure: paper

THE DISCOURAGED WORKER EFFECT: A REAPPRAISAL USING SPELL DURATION DATA

Alberto Martini
Department of Economics
University of Wisconsin-Madison
1180 Observatory Drive
Madison, WI 53706

June 1987

A previous version of this paper was presented at the *XIII* meeting of the Eastern Economic Association, Washington, D.C., March 5-7, 1987, and at the Labor Economics Workshop of the University of Wisconsin-Madison. I wish to thank Martin David, Hank Griffis and Jim Walker for helpful comments and suggestions: the responsibility 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.

Preliminary
Not for quotation
or citation

THE WEALTH OF THE AGED AND NONAGED, 1984

by

Daniel B. Radner

Social Security Administration

March 1987

Prepared for

NBER Conference on Income and Wealth

Conference on the Measurement of Saving, Investment, and Wealth

Baltimore, Maryland, March 27-28, 1987

The Wealth of the Aged and Nonaged, 1984*

I. Introduction

This paper presents estimates of wealth for 1984 from the Survey of Income and Program Participation (SIPP) and discusses wealth data requirements for the analysis of economic status. We are interested in the economic resources available to households other than the very wealthy, and our particular focus is on age groups. The degree of concentration of wealth and/or power, the question that wealth data traditionally have been used to address will not be discussed. We are interested in the economic resources available to households other than the very wealthy, and our particular focus is on age groups.

In this paper we have a short time horizon. We are not directly concerned with life cycle issues of saving and accumulation, but with the amounts of resources available to units of different ages at a particular time. One question that is important to examine is how many of the aged have "substantial" wealth that could be used in special circumstances --e.g., to help pay for high medical expenses to offset income loss. Detailed age groups are examined because the broad aged and nonaged groups often used are not homogeneous. For example, it is useful to distinguish between the "young old" (e.g., age 65-74) and the "old old" (e.g., age 75 and over).

Two types of estimates from wealth data are of interest to us. The first is wealth at a particular time. This includes amounts of wealth, the distribution of wealth, and the concerns about the data. Thus, a household survey that did not do a good job of capturing the upper tail of the wealth distribution could be of use to us.

Several specific implications of our perspective for the characteristics of the wealth data needed are discussed below. First, we need a data source that covers the entire wealth distribution (or the entire distribution except for the upper tail). Thus, data sources such as estate tax returns that are confined to the upper tail are not appropriate for our purpose. Second, wealth data are needed for all age groups of the population. This follows from the fact that we want to examine and compare both the aged and nonaged. This requirement means that data sources that are confined to particular age groups (e.g., the Social Security Administration's Retirement History Study) are not appropriate. Third, it is necessary that several types of information other than wealth be available for the unit. Information on income is of crucial importance, and information on socioeconomic characteristics (e.g., unit size, sex and marital status of the unit head) is very important. Data from estate tax returns are inappropriate for this reason also. Fourth, the wealth data should be available for units other than persons. Families and unrelated individuals (often called family units) or households are acceptable units. Fifth, the data need to be comprehensive enough so that a reasonable definition of net worth can be formed. Although both sides of the comparison if possible. Because aggregate amounts of some asset types are highly concentrated in the upper tail (e.g., corporate stock), a substantial adjustment to the control aggregate is necessary if the upper tail is not included. Perhaps such an adjustment could be performed using a better estimate of aggregates held by the upper tail (e.g., from estate tax multiplier estimates). 2/ Of course, comparisons of aggregates are only partial tests of the accuracy of the estimates. Even if the aggregates were correct, the estimated distribution could be very inaccurate.

In discussing how useful the data are for our purpose, one issue is the general accuracy of the wealth data that are available. Although wealth data obtained in household surveys often have been criticized as inaccurate, the problems with accuracy probably are worst in the upper tail of the distribution. We suspect that comparisons that exclude the upper tail would be more favorable to survey wealth data, at least for some types of assets. But the data for the remainder of the distribution also have problems; item nonresponse rates can be substantial and answers given can be inaccurate. 3/ On the other hand, more inaccuracy in the data is acceptable for our purpose than is acceptable for the measurement of inequality or the change in inequality. 4/

Our perspective is as a user of public use microdata files. Amounts in such files are often top-coded to prevent disclosure. Also, the amounts are restricted by the size of amounts that could be coded in the survey. Fortunately, such problems are far less important if the upper tail of the distribution is excluded from the analysis.

the highest constant consumption path (Nordhaus 1973, Irvine 1980, Beach 1981).

Comparing different age groups using the annuity approach has been criticized on the grounds that the method does not take into account the likelihood that the incomes of young units will rise and that those units ordinarily will be able to increase their wealth as they age (Projector and Weiss 1969). Some researchers have tried to take this into account essentially by estimating future earnings (Nordhaus 1973, Taussig 1973, Irvine 1980).

Some researchers have combined income and wealth by converting income flows into stocks of wealth and adding that wealth to other types of wealth. For example, in looking at the aged, Hurd and Shoven (1982) capitalized several sources of income and added those values to estimates of wealth. Also, for limited purposes some researchers have taken a simpler approach to combining income and wealth and summed current income and liquid assets (David 1959, Steuerle and McClung 1977), or income and net worth (Steuerle and McClung 1977).

Radner and Vaughan (1984, 1987), in looking at a short time horizon, did not combine income and wealth. They considered income and wealth jointly as a two-dimensional classification and examined such characteristics of the joint distribution as the percentage of each age group that had relatively low wealth and relatively low income.

estimates shown here, includes home equity, vehicle equity, business equity, financial assets, real estate, and IRA and Keogh accounts, minus debts. The value of household durables, equities in pension plans, and the cash value of life insurance are not included in the estimates.

The 1983 Survey of Consumer Finances (SCF) obtained information on wealth, income, and socioeconomic characteristics (Avery et al. 1984a, 1984b; Avery and Elliehausen 1986). The survey contained two portions, a multistage probability sample and a high-income frame. Estimates are shown here for the probability sample alone and for the probability sample plus the high-income frame. The estimates shown here for the probability sample are based on information for about 3,700 family units, while the estimates that include the high frame are based on about 4,100 family units. The high-income supplement was obtained by drawing about 5,000 family units from tax information. Interviews were completed with 438 of those family units (9 percent). Net worth, as defined in the estimates including the high supplement, includes home equity, real estate, business equity, financial assets, and retirement assets (which includes IRA's, Keogh accounts, the cash-value of life insurance, and employer-sponsored thrift, profit-sharing, and tax-deferred savings plans), minus debts. The net worth concept used for the estimates that do not include the high-frame excludes the cash value of life insurance and at least some business equity. Both definitions exclude automobile equity, the value of attitudes about retirement in September 1979 (Cartwright and Friedland 1985). Personal interviews were completed with about 3,600 households. The sample was a multistage area probability sample; there was no oversampling of the upper part of the distribution. Estimates were presented for units that differ from those presented for other surveys--the units are similar to Census families and unrelated individuals except that family members age 18 or older in general are considered to be separate units. Estimates are presented for about 4,300 of these "family units." In these estimates, net wealth includes home equity, personal property, vehicle equity, business equity, liquid and investment assets, and the imputed value of employer-based pensions, IRA's, Keogh plans, and annuities.

The Greenwood "synthetic" estimates were made using data from income tax returns, estate tax returns, and a household survey (Greenwood 1983a). The basic microdata file used was constructed by statistically matching survey information from the Current Population Survey and income tax returns from the 1973 Individual Income Tax Model. Corporate stock, debt instruments, and real estate held were estimated primarily by capitalizing amounts from income tax return data. Then net wealth was estimated by regression from a sample of 1972 estate tax returns, using the capitalized corporate stock, debt instrument, and real estate amounts. The regression parameters were used to assign an amount of net wealth to each family unit in the basic file. Net wealth, as used in these estimates, is a more comprehensive definition than used in most surveys.

seven different data sources described; it should be noted that the definitions of "wealth" used are not strictly comparable. Also, the wealth-holding units are not comparable in all cases. The 55-64 age group is used as the base for these relative means. Six of the estimates are from household surveys, while the other two (Greenwood and Wolff) are "synthetic" estimates.

The different estimates of relative means do not appear to be very similar. The 55-64 age group has the highest mean for three estimates (SIPP, ISDP, SFCC), although the SFCC might show a peak at an older age if more age detail were available. The two SCF estimates peak in aged age groups, while the Pension Commission estimate peaks in the 45-54 age group. The two synthetic estimates peak in the aged age group.

The ranges of relative means for specific age groups are quite broad. For the 65 and over age group, the range is from 0.73 to 1.24. The range for the 45-54 age group is from 0.68 to 1.04, and the range for the 35-44 age group is from 0.42 to 0.83. Even if we confine the comparison to SIPP, SCF, ISDP, and SFCC (data sources for which relative medians are available in table 2), differences are still substantial, although smaller. The ranges then are 0.75 to 1.24 for the 65 and over group, 0.68 to 0.96 for the 45-54 group, and 0.42 to 0.61 for the 35-44 group.

When we examine relative medians (table 2), we find that the differences are quite a bit smaller. Those estimates are available only for SIPP, SCF, ISDP, and SFCC. In every case the peak is in the 55-64 age group. The ranges are substantially smaller than for relative means; 0.75 to 0.82 for the 65 and over group, 0.76 to 0.83 for the 45-54 group, and passbook savings accounts, money market deposit accounts, certificates of deposit, interest earning checking (e.g., NOW) accounts, money market funds, U.S. government securities, municipal or corporate bonds, stocks and mutual fund shares, U.S. savings bonds, IRA and Keogh accounts, regular checking accounts, mortgages held for sale of real estate, amount due from sale of business or property, other interest earning assets, and other financial assets.

Unsecured debt includes credit card and store bills, doctor, dentist, hospital, and nursing home bills, loans from financial institutions and individuals, and educational loans. It should be noted that, although the value of household durables is not included in wealth, debt incurred to purchase those items is included in unsecured debt.

It is useful to comment on the accuracy of the wealth data contained in the 1984 SIPP. Most of the information about accuracy that does exist is in the form of comparisons between SIPP aggregates and control aggregates. At this time we do not have comparisons that exclude the upper tail of the wealth distribution. The Bureau of the Census has compared aggregates from the 1984 SIPP with Federal Reserve Board balance sheet data (U.S. Bureau of the Census 1986b, table D-3). These comparisons show that home equity is overstated in SIPP by 30 percent, and that vehicle equity is overstated by 43 percent. On the other hand, equity in business and rental property and financial assets are understated by about 25 percent. Unsecured debt is underestimated by about 35 percent. Although

The sensitivity of the age-wealth relationship to the definition of wealth used is shown in tables 5 and 6. Table 5 shows medians and table 6 shows relative medians. When we look at net worth, the medians for the aged groups are in a range of roughly $10,000, from $54,600 for the 75 and over age group to $65,800 for the 65-69 age group. There is a decline as age increases within the aged group. The aged medians are similar to the median for the 45-54 age group, and below the median for the 55-64 group. These relationships are evident in table 6, which shows relative medians. 7/ 8/

When vehicle equity is excluded from net worth, the median falls by relatively small amounts (by $2,300 to $6,300). The youngest age group now has a median of zero, and the peak is still in the 55-64 age group ($66,600). Relative medians rise very slightly for the aged groups, and fall substantially for the youngest groups. When home equity is also excluded from net worth, there is a much larger impact. However, that impact differs widely among the age groups. The youngest group shows no change and the 25-34 group shows a decline of only $3,900. In contrast, the 55-64 group shows a fall of $51,100. All age groups under 55 now have medians under $10,000, while all age groups are under $20,000. The peak is now in the 65-69 group at $16,200. The relationship as age rises is not smooth, with an increase through the 55-64 group followed by small increases and decreases. Relative medians rose substantially for most aged groups, and fell substantially for groups under 55. It should be noted that mean amounts for this definition (not

Median Net Worth by Net Worth Quintile

Median net worth by age and net worth quintile (within age group) is shown in table 7 and relative medians within those quintiles are shown in table 8. Median net worth is very low in the bottom quintile for all age groups, ranging from minus $1,200 in the under 25 group to $2,400 in the 55-64 group. Medians for the second quintile are also not very high; the median for each age group is below $36,000. In every case except one, the median for the second quintile is less than one half the overall median for the age group. The one exception is the 65-69 age group, in which the ratio of the two medians is slightly over one half. In contrast, the top quintile shows medians above $150,000 for all age groups 35 and over.

Within each quintile the age pattern is roughly similar--low amounts at the young ages, a peak in the 55-64 group, and declines among the aged groups. It is interesting to note that, for each of the top four quintiles, median net worth declines within the aged group as age rises. The decline between the 65-69 and 75 and over age groups is 27 percent for the second quintile, 17 percent for the third quintile, 15 percent for the fourth quintile, and 19 percent for the top quintile. The decline is about 35 percent for the bottom quintile, but the amounts are very small in that group. An examination of the relative medians shows that, for all groups under 55 relative medians rise as net worth increases. The pattern is not as strong for the aged groups, where the top three quintiles show similar relative medians. For the 70-74 is highest in the 25-54 age groups (78-81) and falls to 38 percent in the 65 and over group.

Financial assets are held by more than 90 percent of all groups age 35 and over. The percentages of the middle 60 percent holding selected components of financial assets are shown in table 10. Savings accounts are held by roughly two-thirds of all households, with relatively little variation among age groups. Money market accounts are more prevalent among the aged (23 percent) than among the nonaged, as are certificates of deposit (38 percent for the aged). Interest earning checking accounts show less variation among age groups, with the aged showing a slightly higher percentage (29 percent) than the nonaged. Stocks and mutual funds are most prevalent in the 35-64 age groups (21-22 percent), but the aged percentage is not much lower (17 percent). U.S. savings bonds are also most prevalent in the 35-64 age groups (19-20 percent); twelve percent of the aged hold such bonds. The 55-64 age group shows the highest percentage with an IRA (40 percent), while only 6 percent of the aged have an IRA.

Mean amounts of the various asset types are shown for the middle 60 percent in table 11. These means are for all households in the middle 60 percent of the age group, not just for those with the specific asset type. We can see that, for each age group, mean amounts of vehicle equity, business equity, and real estate are all quite low--below $7,000. The sum of these three asset types minus unsecured debt is below $11,000 for each age group. Thus, in an absolute sense, these about double counting of income and assets; such concerns are important in an analytical use of the data. Thus, in the estimates we will use income including asset income and wealth including income-producing assets.

The income classifications used require some explanation. The income definition is total household money income for the four-month period preceding the interview. These amounts are then adjusted for household size using an equivalence scale based on the scale implicit in the U.S. poverty lines. 9/ Then, within each age group, households are separated into quintile groups based on the size of their adjusted total money income. There is a presumption that, within each age group, households in higher income quintiles are "better off" than those in lower quintiles. We will examine the wealth of households in these different income quintiles. Although all age groups will be examined, there will be more emphasis on the aged than the nonaged.

Median Amounts

Table 13 shows median net worth by adjusted income quintile and age. Median net worth is low for the bottom income quintile for each age group. The peak ($20,500) occurs in the 55-64 age group. For the under 35 age groups, median net worth is low for all income groups. The second and third income quintiles also show peaks in the 55-64 age group, but the 70-74 age group has the highest median in the fourth income quintile and the 65-69 age group has the highest median in the top quintile. This table shows that aged households with low quintile. The percentage with unsecured debt shows a relatively small increase as income rises, with a range from 32 to 45 percent.

The percentage with financial assets exhibits a strong increase as income rises, with most of the increase occurring between the first and third quintiles. Table 16 shows the percentage of aged households holding selected financial assets. The percentage holding each of these assets rises sharply as income rises. Savings accounts are held by 39 percent of the bottom quintile and 76 percent of the top quintile. Savings accounts are the only financial asset shown here that is held by a substantial proportion of the bottom income quintile. The percentages held by the bottom and top quintiles respectively are 6 and 47 for money market accounts, 12 and 55 for certificates of deposit, 10 and 53 for interest earning checking accounts, 2 and 51 for stocks and mutual funds, 2 and 21 for U.S. savings bonds, and 1 and 21 for IRA's. The second income quintile holds primarily savings accounts and certificates of deposit. U.S. savings bonds and IRA's are not very prevalent, even among households in the top income quintile.

Table 17 shows mean amounts of specific assets for the aged, by adjusted income quintile. It should be noted that the estimates in this table suffer from the problems associated with estimating the upper tail of the wealth distribution. Aside from home equity, the bottom income quintile has small amounts of all other asset types. The second and third income quintile have substantial amounts ($10,000 or more) only of home equity and financial assets. The fourth and fifth financial assets less than median annualized income. For the youngest age groups the ratios are quite small; the ratios are below 0.25 in all quintiles under age 45. The ratios exceed 1.00 for the higher income aged groups. However, the bottom quintile for each aged group shows a low ratio, and the ratios for the second quintile are only in the 0.37 - 0.53 range. The top quintile in the aged groups has ratios just above or below 2.00.

A third way of examining the age-wealth- income relationship is by looking at the distribution of households by their ratio of wealth to income. Here we will use the ratio of financial assets to income. Those distributions by age are shown in table 21. Only 2 percent of the youngest age group had financial assets exceeding annualized income, and only 5 percent had financial assets that were more than one half of income. For that age group, 26 percent had no financial assets and 55 percent had a positive ratio less than 0.10. The percentages for the aged are quite different than for the young, but do not differ much within the aged group. For that group as a whole, 25 percent had ratios under 0.10 (including zero) and 48 percent had ratios of at least 1.00. One third of the group had ratios of 2.00 or more.

Table 22 shows the estimates for the 65 and over group by adjusted income quintile. Not surprisingly, the percentages differ greatly by income quintile. For the bottom quintile, 53 percent had either zero very low in the bottom quintile in each age group.

An examination of the middle 60 percent of the net worth distribution in each age group shows that, except for the under 25 group, home equity is by far the most important asset for each age group. Home equity accounts for 57 percent of the net worth of the aged, while financial assets account for 34 percent.

When wealth is examined for income quintiles (based on income adjusted for household size) within age groups, we find that median net worth is low for the bottom income quintile for each age group. Median financial assets is low for the bottom three quintiles in every age group. For the bottom income quintile in the aged group, home equity constitutes 73 percent of net worth and financial assets account for 15 percent. For the top income quintile of the aged group home equity accounts for only 30 percent, with financial assets accounting for 51 percent of net worth.

Ratios of median financial assets to median annualized income are below 1.00 for all income quintiles in each nonaged group. The ratio exceeds 1.00 for higher income aged households. More than 80 percent of households in the under 25 age group have financial assets that are less than 10 percent of their annualized income. For the aged, the corresponding figure is 25 percent. For the aged, that percentage ranges from 53 percent for the bottom income quintile to only 7 percent in the top income quintile.

estimates were made to augment the survey data that have recently become available. For some purposes, a combination of survey and synthetic (or estate tax multiplier) estimates might be the most useful.

Table 1 - Sources

SIPP: U.S. Bureau of the Census 1986b, table 3

SCF: excluding high frame: Avery et al. 1984b, table 7
including high frame: Avery et al. 1986, table 2

ISDP: Radner and Vaughan 1984, table 2

SFCC: Projector and Weiss 1966, table A8

Pension Commission: Cartwright and Friedland 1985, table 2

Greenwood: Greenwood 1983b, table 2

Wolff: Wolff 1983, table 5

Table 2 - Sources

SIPP: U.S. Bureau of the Census 1986b, table 5

SCF: excluding high frame: Avery et al. 1984b, table 7
including high frame: Avery et al. 1986, table 2

ISDP: Radner and Vaughan 1984, table 2

SFCC: Projector and Weiss 1966, table 8

Table 4.--Median and mean net worth by age, 1984

Age of Head	Thousands of dollars Median	Thousands of dollars Mean	Relative values Median	Relative values Mean
Under 25............	2.3	7.3	.03	.06
25-34...............	8.4	25.1	.12	.22
35-44...............	35.8	63.5	.49	.55
45-54...............	56.7	100.7	.78	.87
55-64...............	72.5	115.9	1.00	1.00
65 and over.........	59.7	91.1	.82	.79
65-74.............	62.5	99.8	.86	.86
65-69...........	65.8	107.6	.91	.93
70-74...........	59.9	90.9	.83	.78
75 and over.......	54.6	78.7	.75	.68
All ages............	32.7	69.9	.45	.60

Source: Preliminary tabulation from the 1984 SIPP, Wave 4.

Table 6.--Relative medians for alternative definitions of wealth, by age, 1984 [a]

Age of Head	Net Worth	Net worth excluding vehicle equity	Net worth excluding vehicle and home equity	Financial assets minus debt	Wealth	Financial assets
Under 25...........	.03	0	0	0	.05	.03
25-34..............	.12	.06	0	0	.14	.08
35-44..............	.49	.46	.12	.06	.52	.22
45-54..............	.78	.76	.37	.23	.82	.41
55-64..............	1.00	1.00	1.00	1.00	1.00	1.00
65 and over.........	.82	.84	.96	1.31	.81	1.09
65-74.............	.86	.89	.97	1.38	.85	1.17
65-69..........	.91	.92	1.05	1.50	.90	1.30
70-74..........	.83	.84	.86	1.31	.81	1.07
75 and over.......	.75	.77	.93	1.25	.74	1.01
All ages...........	.45	.43	.17	.14	.47	.26

a/ Age 55-64 is used as the base.

Source: Derived from table 5.

Table 8.--Relative median net worth by age and net worth quintile, 1984 [a]

	Quintile [b]				
Age of Head	1	2	3	4	5
Under 25............	-.50	.01	.03	.05	.08
25-34...............	-.21	.05	.12	.20	.27
35-44...............	.02	.32	.49	.57	.62
45-54...............	.23	.66	.78	.82	.85
55-64...............	1.00	1.00	1.00	1.00	1.00
65 and over.........	.32	.75	.82	.83	.82
65-74.............	.33	.82	.86	.87	.86
65-69...........	.44	.92	.91	.92	.91
70-74...........	.22	.69	.83	.82	.81
75 and over.......	.29	.67	.75	.78	.74
All ages............	0	.22	.45	.60	.68

a/ Relative medians are defined within each quintile, using age 55-64 as the base.

b/ Defined within each age group.

Source: Derived from table 7.

Table 10.--Percentage holding selected financial assets, households with medium net worth, by age, 1984 [a]

Age of Head	Type of asset						
	Savings accounts	Money market accounts	Certificates of deposit	NOW accounts	Stocks or mutual funds	U.S. savings bonds	IRA
Under 25.........	54	2	2	15	4	9	2
25-34............	63	6	7	21	13	13	10
35-44............	71	11	12	23	21	19	18
45-54............	72	13	18	23	21	19	29
55-64............	71	21	30	27	22	20	40
65 and over......	67	23	38	29	17	12	6
65-74..........	67	25	39	29	17	14	8
65-69........	67	26	37	30	19	15	12
70-74........	67	23	40	28	17	13	4
75 and over....	66	21	39	29	18	10	.3
All ages.........	67	12	17	23	17	16	16

a/ Medium net worth is defined as the middle 60 percent of the net worth distribution in each age group.

Source: Preliminary tabulation from the 1984 SIPP, Wave 4.

Table 12.--Percentage composition of net worth, households with medium net worth, by age, 1984 [a]

	Type of asset or debt						
Age of Head	Net Worth	Home equity	Vehicle equity	Financial assets	Business equity	Real estate	Unsecured debt
Under 25.........	100	14	85	30	1	2	32
25-34............	100	51	34	24	3	5	17
35-44............	100	67	13	17	3	7	7
45-54............	100	66	11	18	3	8	6
55-64............	100	61	8	25	2	7	3
65 and over......	100	57	5	34	1	5	1
65-74..........	100	58	6	32	1	5	1
65-69........	100	57	6	31	1	6	1
70-74........	100	57	6	33	1	5	1
75 and over....	100	55	3	37	0	5	1
All ages.........	100	62	12	23	3	6	5

1/ Medium net worth is defined as the middle 60 percent of the net worth distribution in each age group.

Source: Preliminary tabulation from the 1984 SIPP, Wave 4.

Table 14.--Median financial assets by adjusted income quintile and age, 1984 a

(thousands of dollars)

Age of Head	Quintile 1	2	3	4	5
Under 25............	0	.1	.3	.6	1.4
25-34..............	0	.3	.8	1.7	5.2
35-44..............	.1	.8	2.1	4.3	12.8
45-54..............	0	1.7	3.9	7.9	24.7
55-64..............	.1	4.0	10.0	18.2	46.0
65 and over.........	.4	3.2	15.0	24.2	63.3
65-74.............	.1	4.0	12.4	25.5	63.9
65-69...........	.2	5.6	10.2	31.0	68.0
70-74...........	.1	3.0	12.5	26.0	60.7
75 and over.......	.6	2.7	13.0	30.0	62.7
All ages............	0	1.0	2.5	4.8	16.8

a/ Income quintiles are based on income adjusted for household size and are defined within age groups.

Source: Preliminary tabulation from the 1984 SIPP, Wave 4.

Table 16.--Percentage holding selected financial assets, by adjusted income quintile, age 65 and over, 1984 [a]

Quintile	Type of financial asset						
	Savings accounts	Money market accounts	Certificates of deposit	NOW accounts	Stocks or mutual funds	U.S. savings bonds	IRA
1.............	39	6	12	10	2	2	1
2.............	54	12	28	19	9	6	3
3.............	68	23	42	30	17	12	5
4.............	74	32	44	38	25	16	10
5.............	76	47	55	53	51	21	21
Total...........	63	24	36	30	21	11	8

a/ Income quintiles are based on income adjusted for household size and are defined within the age group.

Source: Preliminary tabulation from the 1984 SIPP, Wave 4.

Table 18.--Percentage composition of net worth, by adjusted income quintile, age 65 and over, 1984 a

	Type of asset or debt						
Quintile	Net worth	Home equity	Vehicle equity	Financial assets	Business equity	Real estate	Unsecured debt
1..............	100	73	4	15	2	8	1
2..............	100	60	4	27	3	7	1
3..............	100	51	4	34	2	10	1
4..............	100	46	5	35	3	12	1
5..............	100	30	3	51	3	12	1
Total............	100	42	4	41	3	11	1

a/ Income quintiles are based on income adjusted for household size and are defined within the age group.

Source: Prelimary tabulation from the 1984 SIPP, Wave 4.

Table 20.--Ratio of median financial assets to median annualized income, by adjusted income quintile and age, 1984 [a]

Age of Head	Quintile 1	2	3	4	5	Total
Under 25...........	0	.01	.02	.03	.04	.02
25-34..............	0	.02	.04	.06	.12	.04
35-44..............	.01	.04	.07	.11	.23	.08
45-54..............	0	.08	.13	.18	.39	.13
55-64..............	.02	.28	.43	.54	.78	.42
65 and over.........	.08	.41	1.22	1.32	1.90	.87
65-74.............	.02	.43	.89	1.28	1.79	.82
65-69...........	.03	.53	.66	1.38	1.79	.82
70-74...........	.02	.37	1.01	1.48	1.89	.86
75 and over.......	.13	.42	1.30	1.99	2.19	.99
All ages...........	0	.07	.12	.16	.33	.12

a/ Income quintiles are based on income adjusted for household size and are defined within age groups.

Source: Preliminary tabulation from the 1984 SIPP, Wave 4.

Table 22.--Percentage distribution of households by the ratio of financial assets to annualized income, by adjusted income quintile, age 65 and over, 1984 [a]

Quintile	Zero financial assets	Ratio						Total b/
		Under 0.1	0.1-0.3	0.3-0.5	0.5-1.0	1.0-2.0	2.0 and over	
1.........	35	18	12	6	6	9	13	99
2.........	16	17	11	7	11	12	25	100
3.........	6	13	10	7	10	16	38	100
4.........	3	11	10	7	14	17	39	100
5.........	1	6	7	6	11	23	47	100

a/ Income quintiles are based on income adjusted for household size and are defined within the age group.

b/ A few households with zero or negative income are not shown.

Source: Preliminary tabulation from the 1984 SIPP, Wave 4.

6/ It should be noted that the file used in this paper is a preliminary file. Estimates from the final file would be expected to be slightly different, especially for business equity and vehicle equity, and thus for net worth. One household with extremely high net worth is omitted from the estimates shown in this paper.

7/ The inclusion of other asset types in net worth also can affect the age-wealth relationship. The 1979 ISDP contained an estimate of the value of consumer durables. Unpublished tabulations from that file showed that moving from a definition of net worth that excluded consumer durables to one that included consumer durables produced small increases in the relative medians for age groups under age 45 and a small decrease in the relative median for the 65-74 age group.

8/ In a recent paper, Wolff (1985) examined mean wealth by age group for alternative broad definitions of wealth. The most comprehensive definition included pension and social security wealth and human capital.

REFERENCES

Avery, Robert B., and Elliehausen, Gregory E. (1986), "Financial Characteristics of High-Income Families," Federal Reserve Bulletin, Vol. 72, No. 3 (March), 163-177.

Avery, Robert B., Elliehausen, Gregory E.; Conner, Glenn B.; and Gustafson, Thomas A. (1984a), "Survey of Consumer Finances, 1983," Federal Reserve Bulletin, Vol. 70, No. 9 (September), 679-92.

__________ (1984b), "Survey of Consumer Finances, 1983: A Second Report," Federal Reserve Bulletin, Vol. 70, No. 12 (December), 857-68.

Beach, Charles M. (1981), Distribution of Income and Wealth in Ontario: Theory and Evidence, Toronto.

Cartwright, William S., and Friedland, Robert B. (1985), "The President's Commission on Pension Policy Household Survey 1979: Net Wealth Distributions by Type and Age for the United States," Review of Income and Wealth, Series 31, No. 3 (September), 285-308.

David, Martin (1959), "Welfare, Income, and Budget Needs," The Review of Economics and Statistics, Vol. XLI (November), 393-99.

Irvine, Ian (1980), "The Distribution of Income and Wealth in Canada in a Lifecycle Framework," Canadian Journal of Economics, Vol. XIII (August), 455-74.

Joint Economic Committee (1986), "The Concentration of Wealth in the United States," July, United States Congress.

Moon, Marilyn (1977), The Measurement of Economic Welfare, New York.

Murray, Janet (1964), "Potential Income from Assets: Findings of the 1963 Survey of the Aged," Social Security Bulletin, Vol. 27 (December), 3-11.

Nordhaus, William D. (1973), "The Effects of Inflation on the Distribution of Economic Welfare," Journal of Money, Credit, and Banking, Vol. 5, 465-504.

Projector, Dorothy S., and Weiss, Gertrude S. (1966), Survey of Financial Characteristics of Consumers, Board of Governors of the Federal Reserve System.

________ (1969), "Income-Net Worth Measures of Economic Welfare," Social Security Bulletin, Vol. 32 (November), 14-17.

Radner, Daniel B., and Vaughan, Denton R. (1984), "The Joint Distribution of Wealth and Income for Age Groups, 1979," ORS Working Paper No. 33, Office of Research, Statistics, and International Policy, Social Security Administration, March.

U.S. Bureau of the Census (1986b), Household Wealth and Asset Ownership: 1984, Current Population Reports, Series P-70, No. 7 (July), U.S. Government Printing Office.

Weisbrod, Burton A., and Hansen, Lee W. (1968), "An Income-Net Worth Approach to Measuring Economic Welfare," American Economic Review, Vol. LVIII (December), 1315-29.

Wolff, Edward N. (1983), "The Size Distribution of Household Disposable Wealth in the United States," Review of Income and Wealth, Series 29, No. 2 (June), 125-146.

Wolff, Edward N. (1985), "The Effects of Pensions and Social Security on the Distribution of Wealth in the U.S.," Economic Research Report #85-39, C.V. Starr Center for Applied Economics, New York University (December).

Wolfson, Michael C. (1979), "Wealth and the Distribution of Income, Canada 1969-70," Review of Income and Wealth, Series 25 (June), 129-40.

DRAFT

THE DISCOURAGED WORKER EFFECT: A REAPPRAISAL USING SPELL DURATION DATA

Alberto Martini
University of Wisconsin-Madison
1180 Observatory Drive
Madison, WI 53706

March 1987

Paper presented at the *XIII* meeting of the Eastern Economic Association, Washington, D.C., March 5-7, 1987. I wish to thank the participants of the Labor Economics Workshop at the University of Wisconsin-Madison for helpful comments on a previous version of this work: the responsability 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.

1. INTRODUCTION

According to the familiar decomposition, the response of the labor force participation rate (LFPR) to a demand-induced decrease in employment consists of two components, the the *discouraged worker* effect (DWE) and the *added worker* effect (AWE). According to the DWE hypothesis, a deterioration in employment opportunities causes a fraction of the unemployed to stop looking for work: they would then be classified as out of the labor force by official statistics. At the same time, increased labor force participation is likely to occur among those belonging to households where another member has become unemployed. In this case the demand-induced increase in unemployment triggers further increases, due to new labor force entrants.

Earlier studies, well represented by Tella (1965), could only identify the net effect on the size of the labor force: the DWE seemed to dominate, with the effect of smoothing the impact of the business cycle on the measured unemployment rate and causing the labor force participation rate to move procyclically. These empirical findings had an important impact on public policy discussions: attempts were made, in the US as well as in other countries, to adjust the measure of potential output in order to account for the "hidden unemployment", and to adapt the survey instruments in order to reflect the existence of discouraged workers.

Tella's approach consisted mainly of fitting an equation of the form:

$$(L/P)_{it} = a_0 + a_1 U_t + a_2 T$$

where $(L/P)_t$ is the labor force participation rate of subgroup i, U is the aggregate unemployment rate and T is a time trend, intended to represent the gradual change in attitude toward work within certain subgroups, married women for example. A negative estimated value for a_1 was interpreted as an indication of the existence and prevalence of the DWE. This empirical work mainly utilized aggregate time-series data on the *level* of the LFPR and the unemployment rate: no attention was yet given (most likely because of lack of reliable data) to the *flows* within the labor market.

Interest in the dynamic aspects of the labor market increased in the 1970's, stimulated by the growing body of theoretical work on job search and by the increased reliability of CPS gross flow data. Economists came to view unemployment as a dynamic phenomenon: instability of employment, brevity of unemployment spells, flows into and out of unemployment have been central themes of the empirical work done during this period (see for example Marston, 1976 and Clark and Summers, 1979). Two findings emerging from this body of literature are particularly relevant for studying the *DWE*. First, almost half of all unemployment spells end with the unemployed person leaving the labor force. Clark and Summers report that in 1974 about 40 percent of spells end in withdrawal among men, 60 percent among women: these figures are remarkably similar to those shown in table 2.1 below and relative to the period 1983-85. A second issue, especially stressed by Clark and Summers, has to do with the problematic nature of the distinction between unemployment and non-participation. Many observed transitions possibly arise from inconsistent reporting of quite consistent behavior. Repeated spells of unemployment separated only by brief periods outside the labor force are not infrequent. The issue of measurement error has to be kept in mind when designing and evaluating empirical tests on the *DWE*.

On the theoretical side, the large amount of research done during the 1970's in modelling search behavior of the unemployed only marginally touched upon the issue of discouragement, and labor force withdrawal in general, despite the empirical relevance of the flow between unemployment and non-participation. The reason for this scarcity of analytical work has to be found in some of the methodological features of job search models. Most of this theoretical work is centered on the use of dynamic optimization techniques, which, in order to yield some kind of testable implications, usually require stringent assumptions of stationarity. These assumptions are acceptable in the context of a two state model of employment and unemployment, where it is assumed that unemployed workers keep looking for work until an acceptable wage offer is received. In the standard two-state model, the probability of reemployment is made to depend parametrically on the rate of arrival and on the distribution of wage offers. If stationarity is assumed (i.e. nothing changes in the environment, so the stochastic process that generates wage offers is "driving the model") the model has the reservation wage property: there exists a unique value of the reservation wage that makes the individual indifferent between accepting an offer and continuing to search. Any received offer above this reservation wage will cause a transition to employment.

Transitions between unemployment and non-participation cannot be easily accomodated within this framework. Allowing for the possibility of withdrawal greatly complicates the standard search model, and to date there are very few fully developed three-state models where reemployment and withdrawal are modelled simultaneously using dynamic optimization techniques. In the few examples of three-state models (Toikka, 1976, or Flinn and Heckman, 1982b) withdrawal from the labor force is assumed to be uniquely caused by changes in the value of non-market time, where new values arrive according to a stochastic process analogous to that which generates new wage offers. This approach, however, does not capture the essence of the *DWE*, i.e. the hypothesis that labor market conditions have a measurable impact on search behavior and intentions of the unemployed.

To these theoretical difficulties, we have to add the fact that empirical tests of the *DWE*, and of three-state models in general, impose very stringent data requirements: in particular, transitions between unemployment and non-participation have to be recorded. This requirement is not met by most of the longitudinal surveys. The Survey of Income and Program Participation, which is utilized in this study, is one notable exception.

The focus of this paper is mainly empirical: it formulates and estimates a microdynamic reduced form model of unemployment/non-participation transition. The main objective is to empirically determine the strength of the relationship between the probability withdrawal from the labor force and local labor market conditions (as imperfectly proxied by the local aggregate unemployment rate), holding constant a vector of personal characteristics, and to examine whether this empirical relationship is robust across different model specifications.

Such an empirical test is performed by estimating the *hazard function* for the transition between unemployment and non-participation (a brief account of hazard functions is given in section 3). 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). 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, 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 DWE, is one example of a variable that can vary over the course of the spell.

The plan of the paper is as follows. Section 2 describes the Survey of Income and Program Participation (SIPP) from which the sample used in the empirical estimation has been drawn. SIPP contains weekly information on labor force status of each individual, who is classified as either employed, looking for work or not looking for work. From this information continuous labor force histories can be constructed over the time period covered by the survey. Section 3 briefly presents the use of hazard modeling techniques 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. In section 4 estimation results are presented for a variety of model specifications. The DWE hypothesis (a positive effect of the level of the unemployment rate on the probability that the workers will stop searching) is tested. A model where the effect of labor market conditions interacts with the duration of the spells is also discussed and estimated. Section 5 summarizes the results.

2. 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 and on individual and household characteristics relevant to labor market studies.

The survey is organized according to the following scheme. 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 reinterviewed 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 84 panel, first interviewed in October 1983. The data presently available cover the period from June 1983 to March 1986. The initial sample size of the 84 panel was of 20,000 households.

Labor force status variables are collected with reference to 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. Not all the information contained in the questionnaire has been utilized. For example, those individuals who classify themselves as having a job are also asked whether they have been absent from work without pay: in case of affermative answer, they are asked whether the reason was layoff or something else. This information has not been utilized in constructing the work histories, since the "reason for being absent" question is not asked for specific weeks, but generally for the entire reference period. These individuals were then classified as continuously employed, even if they were out of work for short periods.

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 recorded: duration, calendar time of start, previous and subsequent spell type, whether it was completed within the sampling frame or censored, and the serial order of the spell. These data were augmented with a vector of characteristics pertaining to the individual, 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 on the variable was recorded). Among the latter the local unemployment rate was also included.

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, while only 25 were continuously unemployed. The remaining 9,600 changed labor force status at least once during the survey period. Out of the latter subsample, 17,535 unemployment spells were obtained, of which 8,915 belonged to women. Only one unemployment spell per person was actually used in estimation, in order to avoid oversampling of individuals with multiple spells (a discussion of this issue is in section 4 below). In addition, individuals older than 64 or younger than 17 were eliminated. This left with a sample of 4492 unemployment spells for women and 4002 for men.

Table 2.1 contains means for some of the demographic characteristics of the individuals in this last sample, together with some descriptive statistics on their unemployment spells. The reported average durations of right and left censored spells clearly show the effect of the so-called *length bias*: spells that are in progress at a particular point in time have higher mean duration that those observed over a period of time. This is due to the fact that spells in progress at a particular time have a probability of being sampled proportional to their length. 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. Notice that spells that are both right and left censored have been excluded from these computations (since their length is just equal to the length of the sampling frame).

The lower part of table 2.1 contains a breakdown of spells on the basis of the previous and destination spell. The simple breakdown by destination spell (first two lines) bears a close resemblance to the results obtained 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. The remaining four lines in the table represent a breakdown by previous spell of those unemployment spells that end up, respectively, in employment and nonparticipation. These figures suggest that, especially for men, previous spell is a fairly good predictor of subsequent spell.

	WHITE MEN	BLACK MEN	WHITE WOMEN	BLACK WOMEN
AGE	29.7	28.5	30.1	30.6
EDUCATION	12.1	10.7	12.2	10.5
% MARRIED	40.8	45.2	44.6	43.6
% BLACKS		14.1		15.6
N.OF SPELLS	1.81	1.51	1.64	1.70

AVERAGE SPELLS DURATION

	WHITE MEN	BLACK MEN	WHITE WOMEN	BLACK WOMEN
RIGHT CENS.SPELLS MEAN DURATION	14.9	18.1	13.6	16.3
LEFT CENS.SPELLS MEAN DURATION	14.1	17.3	13.9	14.8
COMPL.SPELLS (ENDING IN EMPL'NT) M. DURATION	7.46	5.90	7.13	6.75
COMPL.SPELLS (ENDING IN OLF) M. DURATION	3.88	4.02	3.88	3.27

BREAKDOWN OF UNEMPLOYMENT SPELLS BY PREVIOUS AND SUBSEQUENT SPELL TYPE

	WHITE MEN	BLACK MEN	WHITE WOMEN	BLACK WOMEN
% OF TYPE U->E	64.6	54.1	48.1	24.2
% OF TYPE U->O	35.4	45.9	51.9	75.8
	100.0	100.0	100.0	100.0
% OF TYPE E->U->E	82.4	82.3	62.6	58.8
% OF TYPE O->U->E	17.6	17.7	37.4	41.2
	100.0	100.0	100.0	100.0
% OF TYPE E->U->O	16.5	10.1	13.3	6.9
% OF TYPE O->U->O	83.5	89.9	86.7	93.1
	100.0	100.0	100.0	100.0

Another type of descriptive device that can be utilized when dealing with duration data is the application of life-table actuarial methods, they consists in a non-parametric description of the duration pattern of the data. Since these methods are widely used in demography, the terminology used here is that of mortality tables. The data are broken down into duration intervals (of arbitrary length), then for each interval the number of individuals exposed to the risk of "dying" is estimated as:

$$r_i = n_i - c_i/2$$

where n_i is the number of individuals entering the interval and c_i the number censored in the interval. It is assumed that censored observations occur randomly (from a uniform distribution) in the interval. The conditional probability that the individual "dies" in the interval, given that he enters the interval, is computed as

$$q_i = d_i/r_i$$

where d_i is the number dying in the interval. From these quantities the *empirical hazard function* can be computed: it expresses the instantaneous probability of dying conditional upon survival. It constitutes the non-parametric equivalent of a parametric hazard function estimated without covariates (i.e. with a constant and the duration terms alone). The empirical hazard function is computed according to the formula:

$$\lambda_i = \frac{1}{h_i} \frac{d_i}{r_i - d_i/2}$$

where h_i is the length of interval i. In table 2.2 the duration pattern for the male unemployment spells is reported. Only one spell per individual is utilized. Left censored spells are excluded and spells that end up in employment are considered as right censored (a motivation for these choices is given in the next section).

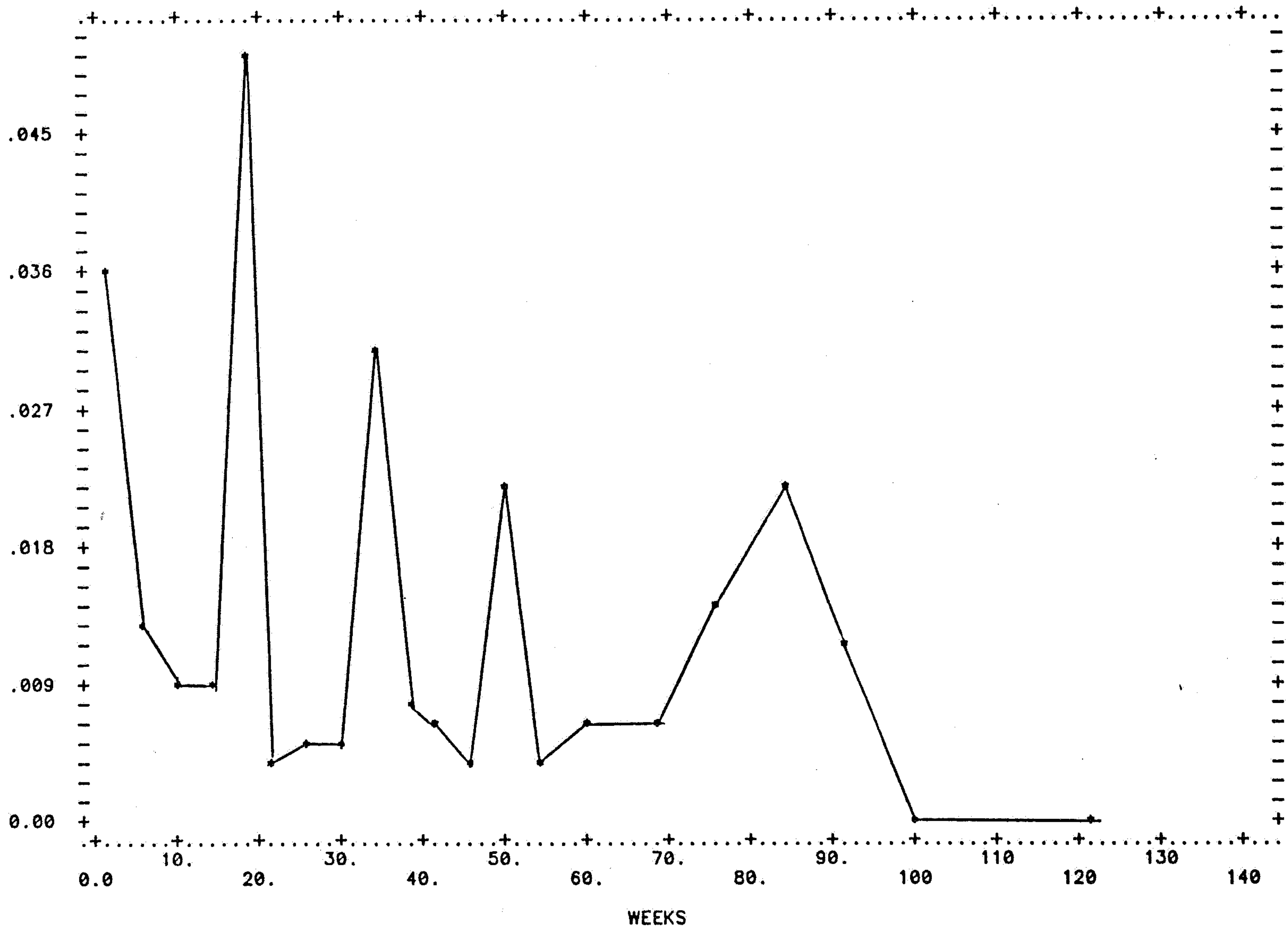
TABLE 2.2 LIFE-TABLE COMPUTATION OF EMPIRICAL HAZARD FUNCTION - MALE SUBSAMPLE

h_i	n_i	c_i	d_i	r_i	q_i	p_i	hazard
0.00 - 4.00	3097	768	369	2713.0	0.1360	0.8640	0.0365
4.00 - 8.00	1960	400	86	1760.0	0.0489	0.9511	0.0125
8.00 - 12.00	1474	303	49	1322.5	0.0371	0.9629	0.0094
12.00 - 16.00	1122	201	36	1021.5	0.0352	0.9648	0.0090
16.00 - 20.00	885	249	139	760.5	0.1828	0.8172	0.0503
20.00 - 24.00	497	91	6	451.5	0.0133	0.9867	0.0033
24.00 - 28.00	400	54	7	373.0	0.0188	0.9812	0.0047
28.00 - 32.00	339	47	6	315.5	0.0190	0.9810	0.0048
32.00 - 36.00	286	76	29	248.0	0.1169	0.8831	0.0310
36.00 - 40.00	181	24	5	169.0	0.0296	0.9704	0.0075
40.00 - 44.00	152	17	4	143.5	0.0279	0.9721	0.0071
44.00 - 48.00	131	22	2	120.0	0.0167	0.9833	0.0042
48.00 - 52.00	107	25	8	94.5	0.0847	0.9153	0.0221
52.00 - 56.00	74	7	1	70.5	0.0142	0.9858	0.0036
56.00 - 64.00	66	15	3	58.5	0.0513	0.9487	0.0066
64.00 - 72.00	48	14	2	41.0	0.0488	0.9512	0.0062
72.00 - 80.00	32	8	3	28.0	0.1071	0.8929	0.0142
80.00 - 88.00	21	4	3	19.0	0.1579	0.8421	0.0214
88.00 - 96.00	14	5	1	11.5	0.0870	0.9130	0.0114
96.00 - 104.00	8	5	0	5.5	0.0000	1.0000	0.0000
104.00 - 140.00	3	3	0	1.5	0.0000	1.0000	0.0000

(symbols are defined in the text)

FIG. 2.1 EMPIRICAL HAZARD FOR UNEMPLOYMENT SPELLS - MALE SUBSAMPLE

(21 intervals)



Note: the upper bounds for the time intervals are 4,8,12,16,20,24,28,32,36,40,44,48,52,56,64,72,80,88,96,104,140. Hazard probabilities are evaluated at mid-points.

The empirical hazard reported in table 2.2 is plotted in fig 2.1. The overall tendency of the hazard is declining over time: huge spikes, however, are easily noticeable at regular intervals. Such spikes reveal a measuremet error 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 reference period. As a consequence, an abnormal number of transitions are observed at the seam between two waves. When individuals report themselves in an identical labor force state for one (or more) entire wave, and then declare a change at the start of the following one, this produces spikes in the frequency distribution of spell duration (and in the hazard), at values that are multiples of the length of the reference period. This is what is observed in fig 2.1 , especially around the 18th, 34th and 52nd week. Such spikes, however, are only the most visible consequence of this phenomenon: the entire frequency distribution of durations is affected, also at values not multiples of the reference period.

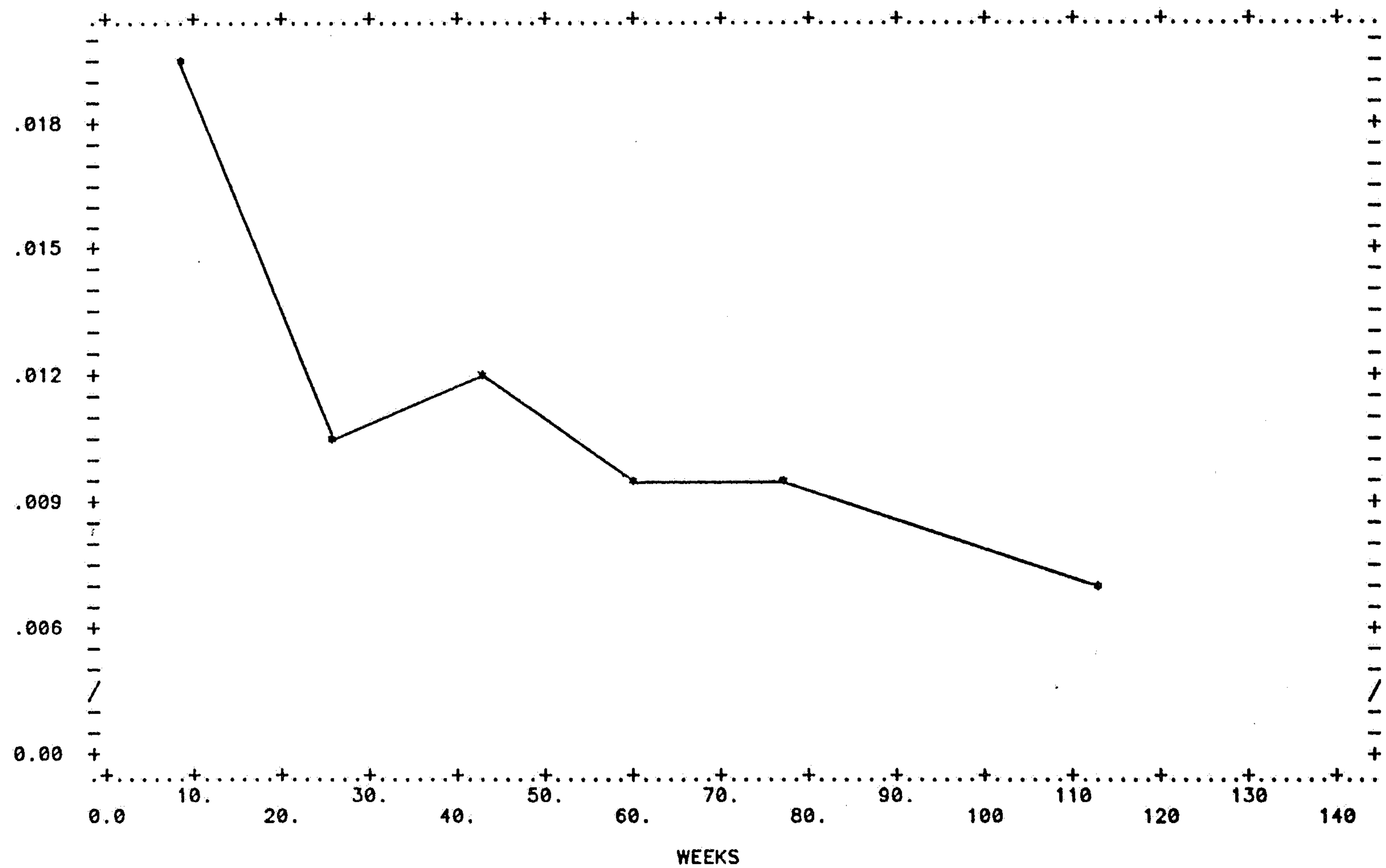
In this non-parametric framework, a more "legible" hazard function can be obtained by changing the interval length in such a way that each interval corresponds to a wave (i.e. 17 weeks): the effect is that of "smoothing" the empirical distribution of the variable affected by the measurement error. The result of this operation is shown in fig 2.2 . However, the degree of arbitrariness implicit in this kind of procedure is evident.

When parametric estimation is performed, this type of correction is not feasible. The solution adopted in this paper is that of including among the time-varying explanatory variables a dummy variable, that assumes a value of one for the final week of each reference period, when the "risk" of transition is (abnormally) high.

FIG 2.2

EMPIRICAL HAZARD FOR UNEMPLOYMENT SPELLS - MALE SUBSAMPLE

(6 intervals)



Note: the upper bounds of the six time intervals are 17,34,51,68,85,140.
Hazard probabilities are evaluated at mid-points.

3. PARAMETRIC HAZARD MODELS

A detailed discussion of hazard functions estimation techniques is beyond the scope of this paper. For an exhaustive survey of survival analysis the reader is referred to Kalbfleisch and Prentice (1980), and, for applications to socioeconomic phenomena, to Heckman and Singer (1984). The issues immediately relevant for the type of estimation performed are here briefly presented and discussed.

The estimation methods described in this section assume the availability of the following information for each observed spell (a three-state model is considered here):

i) length of spell in state i.

ii) destination state j or k, when the spell is completed within the sampling frame.

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

iv) serial order of the spell .

v) a vector of covariates, which can be fixed or vary over the spell.

a) 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:

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

Once a parametric functional form is assumed for the hazard function, its parameters can be estimated by maximum likelihood. The likelihood function would be 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 d) below.

The results obtained so far 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 would 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.

b) The competing risks model

The biostatistics literature has developed a set of techniques to take into account the possibility of multiple outcomes, or *competing risks*, in the analysis of duration data (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

$$f_{ij}(t) = h_{ij}(t) \exp\left[- \int_0^t h_{ij}(u)\, du \right] \exp\left[- \int_0^t h_{ik}(u)\, du \right] \qquad (7)$$

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:

$$S_i(t) = \exp[-\int_0^t h_{ij}(u)\,du]\,\exp[-\int_0^t h_{ik}(u)\,du] \qquad (8)$$

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:

$$L_{ij} = \sum_{i\ spells} [\delta \log h_{ij}(t_i) - \int_0^{t_i} h_{ij}(u)\,du] \qquad (9)$$

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.

c) Reduced form hazard function

Since the "true" functional form of the duration model is here unknown (i.e. is not produced by a structural economic model) a reduced form approach is adopted, as is actually done in most of the literature concerned with duration of unemployment. This approach consists of specifying a parametric form for the conditional hazard function, where the conditioning is with respect to observable variables. This allows one to estimate the impact of key economic variables on the duration of unemployment and on the probability of transiting to a different labor force state.

The 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).

$$h_{ij}(t) = \exp [\, X(\tau + t)\,\beta_{ij} + \gamma_{1ij}\frac{t^{\lambda_{1ij}} - 1}{\lambda_{1ij}} + \gamma_{2ij}\frac{t^{\lambda_{2ij}} - 1}{\lambda_{2ij}}\,] \qquad (10)$$

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 [β_{ij}, γ_{ij} , λ_{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. Attempts to fit more general specifications failed to converge to a stable solution.

d) 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. A tangible example of length bias was shown in table 2.1. 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. The assumption of exponentiality is usually contradicted by the data. Moreover, it is very unlikely that heterogeneity can be entirely controlled for by conditioning on observable variables.

Under certain conditions, the solution of the initial conditions problem can be that of excluding left censored spells from the analysis, 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. Excluding left censored spells is not, however, the general solution to the initial condition problem. This is particularly true in the presence of the "mover-stayer" type of (unobservable) heterogeneity. Take as an example non-participation or employment: a large fraction of individuals never "move out" of their labor force state for very long periods of time, while a smaller fraction show short and repeated spells. These very different probabilities of transition cannot be explained by observable characteristics alone. When the length of the sampling frame is relatively short, as in SIPP, excluding left censored spells has the consequence of selecting out almost the totality of such "stayers".

On the other hand, for this "stayers-type" spells, in absence of other information, it is very difficult to construct a likelihood function (unless spell lengths are assumed to have an exponential distribution, i.e. no duration dependence). Presample values of covariates are typically not available, at least not with the same detail and precision of sample values. Even if the entire history could be recovered from the start of each spell, incorporating spells in progress at the onset of the survey would still require correcting for length bias (and such correction often entails formidable computational problems).

When dealing with unemployment spells only, the above problem is less severe. The proportion of those that could be defined as "stayers" among the unemployed is very limited. For example, in the sample selected from SIPP there are only 25 cases where the individual is in the same spell of unemployment at the start and at the end of the survey. The concept of "stayers" itself is somehow misused when referred to unemployment: no matter how long the spell, unemployment is a temporary, unstable situation.

The solution adopted here is that of excluding left censored spells from estimation: the loss of efficiency is not really an issue, given the size of the remaining sample.

4. TESTING THE DWE: EMPIRICAL RESULTS AND INTERPRETATION

In this section estimation results are presented for several specifications of the transition function between unemployment and non-participation (hereafter UO transition). All the results presented here are obtained from a sample that contains only one spell per individual, and specifically the first non-left censored spell observed. While some individuals appears with only one such spell in the sample, others 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" of individual with multiple spells. Another solution could be that of reweighting the spells, in such a way to give less weight to multiple spells. This solution is 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. Descriptive statistics were presented in table 2.1 .

Age squared is introduced in the hazard together with age to control for non linearities in the age effect. It should also be noticed that 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 indeed 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 spell" variable shows a very strong and significant negative effect on the probability 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.

According to an alternative interpretation, this result could be due to unobservable 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 that causes both previous occupancy and future transition behavior. In the models estimated in b) below, previous spell is controlled for with a dummy variable. In c) instead, one specification is reestimated *selecting* on previous spell, in order to gain some additional insight on this issue.

Time-varying covariates

The motivation for one of the time-varying covariates has been discussed in section 2 above: 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. This matches with the spikes in the empirical hazard of fig. 2.1 .

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 large proportion 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, whose estimated coefficient is of central importance in testing the *DWE*.

Discussion of the results relative to the latter variable, as well as to the duration terms, is postponed to the following paragraphs.

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, with 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 spell" 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 spell.

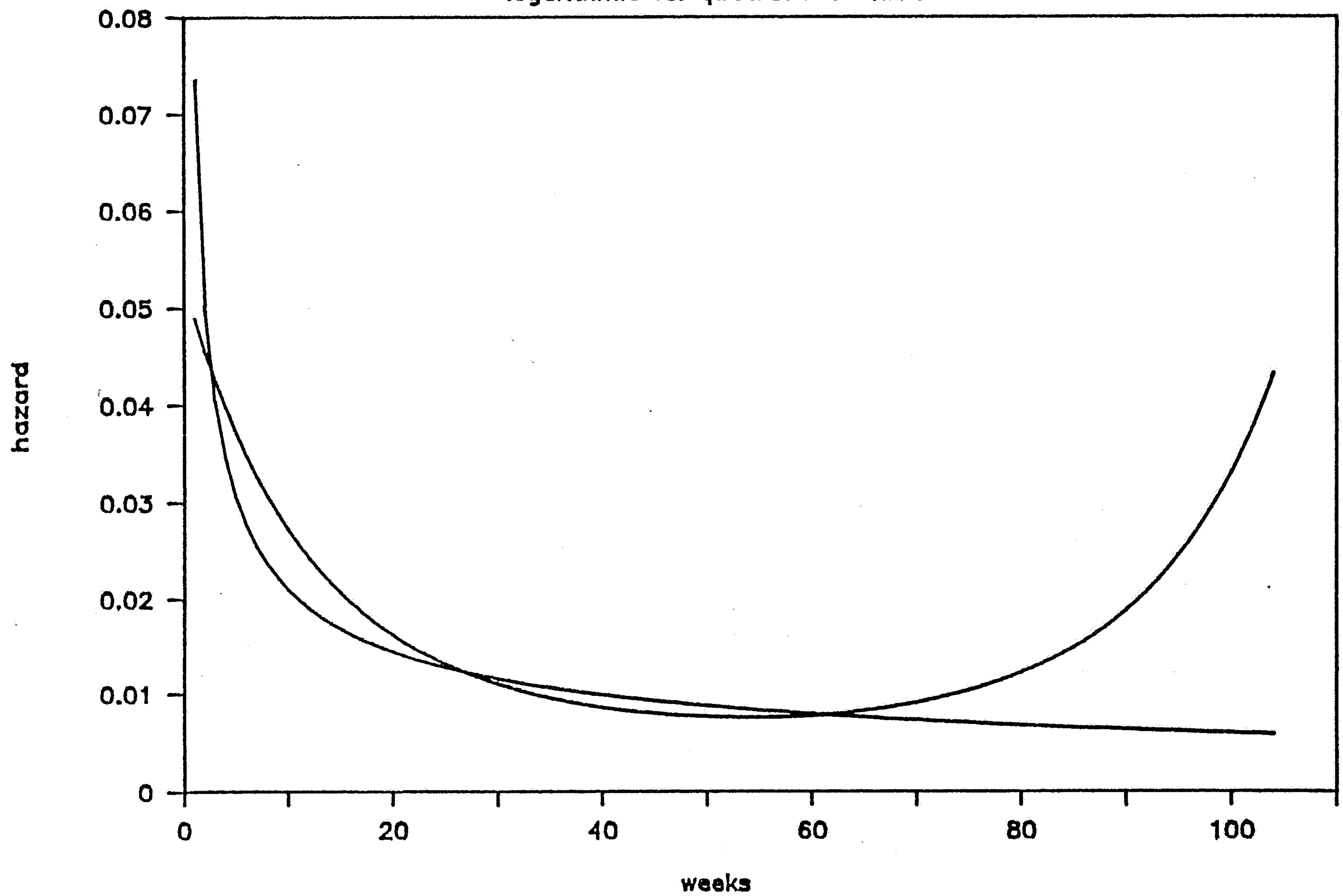
The duration term shows a negative sign in the logarithmic specification, and a convex pattern in the quadratic, with a minimum around 50 weeks. A plot of the two hazards for the male subsample is shown in fig. 4.1. While it is not possible to formally discriminate between the two models with a likelihood ratio test, since they are non-nested (2), an heuristic argument can be formulated by inspecting the plot. The rising portion of the quadratic hazard is almost entirely beyond the empirically relevant range: in fact by the 50th weeks almost 97 percent of the unemployment spells that end up in withdrawal are completed (see table 2.2). One could then argue that the logarithmic specification is a more parsimonious representation of the duration pattern. Moreover, the empirical hazards of fig 2.1 and 2.2 do not show any significant rising portion.

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.

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 (18.1)	− 1.615 (18.5)	− 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	5.146 (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

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 necessary 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 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 from search falls below that from staying out of the labor force: the rate of arrival of job offers 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 is reasonable to assume that it takes some 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 of this model 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 confirmes 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.

MODEL WITH INTERACTION

DURATION SPECIFICATION	MEN		WOMEN	
	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)	0.306 (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 spell

Another variation of the main model is obtained by selecting on the previous spell variable, in order to evaluate its impact on the other coefficients. Only the logarithmic specification is tested here, with and without the interaction term (table 4.3). The results are relative to the male subsample. 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 is analogous to the one 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 should be zero, since they are not eligible for UI: the residual effect could be caused by measurement errors).

However, the most 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 *DWE* seems limited to a subset of the unemployed, those who already have a lower labor force attachment.

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 a still exploratory attempt was made to estimate the effect of labor market conditions on the probability that an unemployed worker withdraws from the labor force. This effect is found to be strong, at least for a substantial fraction of the unemployed. Moreover, this effects seems to increase during the course of the spell. The interpretation offered is that the unemployed worker is learning about his/her labor market opportunities as the spell progresses.

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 an effect broadly consistent with empirical evidence from other areas of labor economics, education shows a positive effect for men and a negative one for women, a result that is not easily interpretable.

Entering the unemployment spell from employment strongly reduces the probability of subsequent withdrawal. When the sample was selected on the basis of such previous experience, the discouraged worker effect was found to virtually disappear for previously employed individuals, while remaining very strong for those who entered from out of the labor force.

NOTES

(1).
In an attempt to get some understanding of this puzzling result, 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).
(2).

The likelihood ratio test is able to discriminate between the quadratic and the linear, or Gompertz, specifications, since they are nested. The results reported are sufficient to reject the Gompertz, since the estimated coefficient on the duration squared term is always significantly different from zero.

REFERENCES

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

Flinn, C., and Heckman, J., 1982a, "Models for the Analysis of Labor Force Dynamics", in *Advances in Econometrics*, edited by G. Rhodes and R. Basmenn, JAI Press.

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

Heckman, J., and Borjas, G., 1980, "Does Unemployment Causes Future Unemployment?", *Economica*, 47.

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.

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

Toikka, R., 1976, "A Markovian Model of Labor Market Decisions by Workers", *The American Economic Review*, 66.