A list of Working Papers on the last pages

No. 198, 1988

JOB SEARCH AND YOUTH UNEMPLOYMENT: Analysis of Swedish Data

by

James W. Albrecht*, Bertil Holmlund**,
and Harald Lang***

* Department of Economics, Georgetown University, Washington D.C.

** Department of Economics, Uppsala University, Uppsala

*** The Industrial Institute for Economic and Social Research (IUI), Stockholm

The paper was presented at the Annual Congress of the European Economic Association, Bologna, Italy, August 1988. Albrecht acknowledges research support for this project from the National Science Foundation and from the C V Starr Center for Applied Economics, New York University. Holmlund acknowledges support from the Swedish Ministry of Labor and the Swedish Council for Research in the Humanities and Social Sciences (HSFR).

Nov., 1988

1. Introduction

Youth unemployment is a serious problem in the western economies. Some young people have little difficulty in finding and keeping a job; others find the transition into full–time labor force participation a difficult passage. This paper reports on a project addressing some of the questions relevant to this transition. Specifically, how can we understand variations in unemployment and employment durations across youth? How much of the observed variation can be ascribed to individual choice; how much can be ascribed to variations in labor market opportunities?

The basis for our investigation is a "continuous–time" dataset drawn as a random sample from youth registered as unemployed with the labor market office in Stockholm. This dataset contains week–by–week information on labor force status for up to 830 individuals over a 78–week period starting in January, 1981. Three interviews conducted during this period provide background data plus information on job search behavior. We use these data to investigate sources of variation across individuals in search intensities and in reservation wages, and we examine how job search behavior and labor market opportunities affect time spent without a job.

We analyze these data in the framework of the model of individual job search behavior developed in Albrecht, Holmlund, and Lang [1988]. This model uses a two–state (employed/not employed) continuous–time Markovian setup to characterize individual choice with respect to (i) how much effort to devote to job search (both on– and off–the–job) and (ii) whether or not to accept new job offers. We use this model to carry out a series of qualitative comparative statics calculations, examining how search intensities and the reservation wage can be expected to vary across individuals.[1]

---

[1]Our model is similar in spirit to that of Burdett and Mortensen [1978]. In addition to extending their menu of results, we have developed a general methodology for carrying out qualitative comparative statics calculations in dynamic programming models.

the two states, not employed (n) and employed (e).[2] Our model treats these transitions as "memoryless" in the sense that behavior in a state and the outcome realized in that state do not depend on how the state was reached; neither do behavior nor outcome depend on the length of time spent in the state to date. Despite this exponential structure, there is an important link between states. An individual's job–search behavior depends not only on the environment of his current state, but also on the environment he expects to face in states to be reached in the future. In particular, optimal job search behavior for an individual without a job will depend, among other factors, on how likely he is to lose his job once employed.

Our main focus in the data analysis is on the transition out of the non–employment state. The combination of data on search intensities and reservation wages, together with a model of how these variables are determined, allows us to take a structural approach to explaining duration in non–employment. Our starting point, in common with other studies, is the hazard from non–employment to employment, $h_n = \alpha(s)[1-F(r)]$. Here $\alpha(s)$ is the offer arrival rate, a function of search effort, and $F(r)$ is the distribution of wage offers, evaluated at the reservation wage. Variation in non–employment duration across individuals is seen as the result of the combination of environmental variation, ie, variation across individuals in the functions $\alpha(\cdot)$ and $F(\cdot)$, and variation across individuals in the choice of s and r.

Non–structural studies of duration use the fact that optimal job search behavior will be influenced by any factors affecting $\alpha(\cdot)$ and $F(\cdot)$. This allows a reduced–form approach in which the hazard is modelled simply as depending on all exogenous variables of the system. However, a reduced–form approach precludes sorting out **how** the exogenous variables determine duration, ie, whether the various

---

[2]We do not distinguish between "unemployed" and "not in the labor force." This is appropriate in a model in which search intensity, including the possibility of choosing not to search at all, is endogenous. In our empirical work we exclude individuals who are "unavailable for work," eg, those who are in school, manpower training programs, and the military.

effects operate primarily through variation in the opportunities open to job–seekers or through variation in job search behavior. A structural approach has precisely this "sorting out" ambition.

Lancaster [1985] has developed some useful econometrics for structural duration analysis. The basic idea is that, under certain conditions, the joint analysis of the determinants of job search behavior and duration can be carried out in the standard simultaneous equations framework.[3] We use his setup, supplemented by some modifications dictated by the nature of our data, to carry out a joint analysis of search effort and the reservation wage and their effects on duration. In the next section we sketch our econometric specification and estimation procedure; in the final section we present some preliminary results.

2. Specification and Estimation

According to our model, duration in employment (d_e) and in non–employment (d_n) are both exponential random variables. We specify the hazard for the j^{th} employment spell as

(1) $$\ell n\ h_{ej} = Z_j\eta + \epsilon_{ej}.$$

The vector Z_j gives the values of all exogenous variables in the system associated with the j^{th} employment spell, ie, general labor market conditions and the personal characteristics of the individual experiencing the spell,[4] and the term ϵ_{ej} represents

---

[3]Lancaster [1985] estimated a model in which duration depends on r, while at the same time r depends on elapsed duration. His model allowed for "true duration dependence" at the cost of being "quasi–structural" in the sense that the dependence of r on elapsed duration was simply postulated, as opposed to derived from a non–stationary model of optimal search. An exponential model based on Lancaster's methodology is presented in Jones [1988]. As in the Lancaster and Jones papers, most structural search models have not treated search intensity as a choice variable. An exception is Jensen and Westergård–Nielsen [1987].

[4]It is useful to partition the exogenous variables of the system into two components: (i) those variables that may influence individuals' labor market opportunities, eg, general labor market conditions and individuals' "human capital characteristics" such as education and experience, and (ii) those variables that may influence individuals' job search behavior but that are of no direct relevance to employers, eg,

"unobserved heterogeneity" in the employment hazard.

Duration in non–employment is also treated as an exponential random variable, but we take a structural approach to the specification of this hazard. The hazard for the i^{th} spell of non–employment is given by

(2) $$h_{ni} = \alpha_i(s_i)[1-F_i(r_i)].$$

The hazard for the individual experiencing the i^{th} spell of non–employment equals the product of the offer arrival rate and the probability that an offer, should one be received, will be accepted.

Both the offer arrival rate and the acceptance probability are individual–specific. Since the offer arrival rate is non–negative we specify

(3) $$\ln \alpha_i(s_i) = \alpha_{0i} + \alpha_{1i}s_i.$$

Arrival rates will vary across individuals for two reasons. Some job–seekers will search more vigorously than others; some will attract more offers at any given level of effort than others. We specify the predictable components of the offer arrival rate parameters as

(4) $$\alpha_{0i} = X_i\theta_0; \qquad \alpha_{1i} = X_i\theta_1.$$

Note that these individual–specific parameters depend only on general labor market conditions and "employer–relevant" personal characteristics; personal characteristics such as receipt of unemployment compensation are presumed not to affect the ease with which an individual locates a job offer. Note also that this specification involves an interaction between search intensity and the factors determining the parameters of the offer arrival rate. That is,

(5) $$\ln \alpha_i(s_i) = X_i\theta_0 + s_iX_i\theta_1.$$

The specification we have chosen for the wage offer distribution is the Rayleigh,

(6) $$1-F_i(r_i) = \exp\{-\xi_i r_i^2\} \; ; \; r_i \geq 0.$$

This functional form is both econometrically convenient and appealing from the

---

receipt of unemployment compensation or non–wage income. We denote the first set of variables by X and the second by W; thus $Z_j = [X_j, W_j]$.

point of view of realism.[5] The predictable component of the individual–specific parameter of the wage offer distribution is specified as

(7) $-\xi_i = X_i\theta_2$;

that is,

(8) $\ell n[1-F_i(r_i)] = r_i^2 X_i \theta_2$.

We have argued that the components of the non–employment hazard will vary from individual to individual. Some of the sources of this parameter variation can be observed by the econometrician; others cannot. Differences between the actual values for an individual's hazard parameters and the values we would expect conditional on his observed characteristics are a type of unobserved heterogeneity. Similar to the employment hazard, we denote this unobserved heterogeneity by ϵ_{ni} and write the log hazard as

(9) $\ell n\ h_{ni} = X_i\theta_0 + s_i X_i \theta_1 + r_i^2 X_i \theta_2 + \epsilon_{ni}$.

Note that this is not the "true log hazard plus an error term"; rather, it is the true log hazard, consisting of two components, one of which depends on variables we can observe, one of which does not.

The econometric specification is closed by equations for search intensity and the reservation wage,

(10) $s_i = \max[0, \beta_0\alpha_{0i} + \beta_1\alpha_{1i} + \beta_2\xi_i + \beta_3 \ell n h_{ei} + \beta_4 W_i + \epsilon_{si}]$

(11) $\ell n\ r_i = \gamma_0\alpha_{0i} + \gamma_1\alpha_{1i} + \gamma_2\xi_i + \gamma_3 \ell n h_{ei} + \gamma_4 W_i + \epsilon_{ri}$.

Individual job search behavior depends on the labor market environment (ie, on the individual–specific parameters of the offer arrival rate and of the wage offer distribution and on the individual's risk of losing his job, once employed) as well as on

---

[5]An alternative is the Pareto specification,

$1-F_i(r_i) = (w_{0i}/r_i)^{\xi_i}$

used by Lancaster [1985] and Jones [1988]. Neither had data on search intensity, so the problem of distinguishing the effect of the offer arrival rate from that of the "base wage" (w_{0i}) on duration was not a concern. However, the Pareto specification would pose serious identification problems in a model such as ours in which both s and r are treated explicitly.

variables such as receipt of unemployment compensation that are of no direct relevance to employers. Finally, the error terms, ϵ_{si} and ϵ_{ri}, reflect our inability to observe all of the determinants of job search behavior.

Equations (1) and (9)–(11) are the essence of our specification. These equations present an obvious simultaneity problem. An individual with a favorable unobserved heterogeneity component in the non–employment hazard is one whose durations will tend to be shorter than would be expected given his observable characteristics. The simultaneity problem arises because, according to our theory, a favorable non–employment hazard has implications for the optimal choice of s and r. Likewise, an individual with a favorable (low) draw of ϵ_e is one whose employment durations will tend to be longer than would be expected. Again, this has implications for the optimal choice of s and r. In short, we have strong grounds to suspect correlation among ϵ_e, ϵ_n, ϵ_s, and ϵ_r.

Following Lancaster [1985], our strategy for dealing with this problem is straightforward: we convert our setup to the standard simultaneous equations framework. The assumption that durations are exponential makes this conversion easy. If duration is exponential, then expected log duration equals the negative of Euler's constant ($c \approx 0.577$) minus the log hazard. This means we can estimate the parameters of the log hazards by regression. In particular, the parameters of (1) can be estimated from

(12) $$-\ell n\, d_{ej} = c + Z_j\eta + \epsilon_{ej} + \nu_{ej},$$

and the parameters of (9) can be estimated from

(13) $$-\ell n\, d_{ni} = c + X_i\theta_0 + s_iX_i\theta_1 + r_i^2X_i\theta_2 + \epsilon_{ni} + \nu_{ni}.$$

The error ν_{ej} in (12) is the discrepancy between expected and actual log duration in the j^{th} employment spell. This error is independent of expected log employment duration with mean 0 and variance $\pi^2/6$. The interpretation of ν_{ni} is analogous.

Simultaneity bias is not our only problem: we also have a problem of missing data and selectivity bias. The most serious aspect of this problem is that we lack data for s and r corresponding to some spells of non–employment. Since the data for s and r come from interviews conducted at three particular points during the sample period, a particular spell of non–employment need not "cover" an interview time. (An individual interviewed in March 1981, September 1981, and March 1982 might have a 6–week spell without work during November and December 1981. We lack direct information about his job search behavior during this 6–week spell.) Further, the data we do observe for s and r are not randomly selected: we are more likely to observe s and r for individuals with long durations in non–employment and/or short durations in employment since these are precisely those individuals whose spells are most likely to cover the interview times. This has implications for estimating the log duration equation (13) for non–employment.

We have developed an "OLS selectivity bias correction" technique to handle this problem. The key idea is to recognize that although the problem sketched in the preceding paragraph is very similar to the one dealt with in Heckman [1979], the selection equation(s) are not probits (as in Heckman), but rather, ordinary regressions. The implication is that the required selectivity bias corrections can be effected by inserting the estimated residuals from the selection equations into the regression contaminated by selectivity bias. As in Heckman [1979], the fact that the selectivity bias correction terms are only estimates makes it necessary to correct all estimated standard errors. In our context, these standard error corrections are conceptually simple but computationally tedious.

3. The Data and Some Results

The data we use were collected through local employment exchange offices in Stockholm. These data are built around interviews with a number of youth registered as unemployed in the county of Stockholm at the end of January of 1981. A sample of 890 in the age group 16–24 was drawn, constituting one out of every four registered as unemployed. These youth were interviewed in the Spring of 1981, the Fall of 1981, and in the Spring of 1982. A fourth, follow–up interview was conducted in 1985. Among the 890 youth, 830 took part in at least one of the four interviews and 527 took part in all four.

These interviews provide a "continuous time" recording of the youths' labor market status. During the 1981–82 period this recording was week–by–week for periods of up to 78 weeks; thereafter, the recording was month–by–month. We concentrate on the week–by–week information, using the month–by–month data only for the purpose of completing spells. Labor market status is recorded in 10 categories. We identify 2 of these categories, permanent and temporary employment, with "employment" (state e); and we identify 2 categories, unemployment and "other," with "non–employment" (state n). The inclusion of the "other" category ensures that those who are "not in the labor force" but who lack any clear reason to eschew sufficiently attractive employment are included in the analysis.

Along with the "event history" data, much information is available from the interviews on the youths' characteristics and on their labor market behavior. Age, sex, citizenship, experience, and education, along with the ratio of vacancies to unemployment, our measure of local labor market conditions, are taken as the variables influencing individual labor market opportunities. In the notation of footnote 4, these are our X–variables. Unemployment compensation, non–wage income, and family dependents are the variables we expect to influence job search behavior but not individual labor market opportunities. These are our W–variables. Finally, the reservation wage and search intensity are our measures of job search behavior.

Table 1 presents a part of our preliminary analysis of these data, namely, estimates of the parameters of equation (13), the log duration of non–employment regression. These parameters were estimated using a two–step procedure. In the first step we generated instruments for s and r^2, correcting for the selectivity bias/missing data problem mentioned at the end of the preceding section. In the second step we estimated (13) by OLS, using the instrumental variables generated in the first step. Thus, we have essentially carried out a variation on 2SLS, the variation arising from our "OLS selectivity bias correction."[6]

The dependent variable in Table 1 is log duration (ie, not the negative of log duration). We begin in the first two columns with simple OLS reduced form estimates. Looking at the second column, we can see that the results make sense. Increasing work experience, education at the gymnasium and university level, and improved labor market conditions (an increase in the ratio of vacancies to unemployment in the local labor market) tend to decrease duration. Age and receipt of UI tend to increase duration.

Columns 3–5 present the structural form estimates. Column (3) includes the (non–interacted) effect of the job search behavior variables on log duration. As expected, an increase in search intensity tends to decrease duration in non–employment, and an increase in the reservation wage tends to increase duration. The estimates in column (4) can be related to the individual–specific parameters of the offer arrival rate function and wage offer distribution. The coefficients on the first (non–interacted) set of variables give the effect of variations in personal characteristics and labor market conditions on α_0; for example, our estimates suggest that, holding all else equal, a university graduate can expect to receive more job offers per week, independent of search effort, than a job–seeker with the base level of education (less than gymnasium). The coefficients on the second (interacted with s) set of variables give the corresponding effects on a_1. For example, the marginal efficiency of search seems to be increasing in education. Finally, the third (interacted with r^2) set of variables give the effects of variations in personal characteristics and general labor market conditions on $-\xi$. Increased education (at least at the university level)

---

[6]The estimated standard errors presented in Table 1 do not take account of the selectivity bias correction.

Table 1. Estimates of the log Duration Equation

	Reduced Form		Structural Form		
	1	2	3	4	5
Intercept	0.784 (2.614)	0.658 (2.156)	1.561 (3.744)	2.557 (2.442)	2.340 (4.493)
Age	0.101 (7.250)	0.103 (7.235)	0.080 (5.089)	0.067 (1.475)	0.080 (4.996)
Woman	–0.096 (–1.612)	–0.106 (–1.778)	–0.154 (–2.487)	–0.062 (–0.300)	–0.163 (–2.633)
Foreign Cit.	0.079 (0.958)	0.070 (0.853)	0.042 (0.507)	0.049 (0.185)	0.051 (0.616)
Experience	–0.456 (–2.509)	–0.055 (–2.979)	–0.029 (–1.276)	–0.014 (–0.281)	–0.023 (–1.005)
Gymnasium	–0.490 (–7.221)	–0.492 (–7.214)	–0.431 (–6.140)	–0.359 (–1.506)	–0.350 (–1.523)
University	–0.762 (–5.268)	–0.767 (–5.314)	–0.488 (–2.859)	–1.001 (–1.412)	–0.937 (–1.347)
Vac./Unempl.	–0.546 (–7.866)	–0.542 (–7.544)	–0.513 (–7.374)	–1.276 (–5.127)	–1.241 (–5.195)
UI		0.242 (2.195)			
Income		0.097 (1.441)			
Family Dependents		0.107 (1.620)			
s			–0.081 (–2.953)	–0.246 (–1.168)	–0.221 (–3.462)
$r^2/1000$			0.195 (3.074)	0.608 (1.230)	0.539 (3.611)
<u>s interactions</u>					
Age				0.002 (0.158)	
Woman				–0.015 (–0.369)	

<u>Notes</u>. Non–employment spells that end in employment account for 56 percent of all spells. All variables except Age, Experience, Vac./Unempl., s and r^2 are dummies. Duration is measured in weeks, age and experience are measured in years, and s is measured

Table 1 — Continued

	Reduced Form		Structural Form		
	1	2	3	4	5
s interactions, continued					
Foreign Cit.				0.007 (0.128)	
Experience				–0.003 (–0.223)	
Gymnasium				–0.032 (–0.663)	–0.033 (–0.732)
University				–0.235 (–2.278)	–0.226 (–2.265)
Vac./Unempl.				0.143 (2.912)	0.139 (2.946)
r^2 interactions					
Age				–0.005 (–0.205)	
Woman				0.040 (0.422)	
Foreign Cit.				–0.011 (–0.083)	
Experience				0.010 (0.300)	
Gymnasium				0.064 (0.573)	0.068 (0.635)
University				0.533 (2.216)	0.513 (2.207)
Vac./Unempl.				–0.348 (–3.020)	–0.335 (–3.038)
R^2	0.165	0.174	0.177	0.196	0.194
Adjusted R^2	0.160	0.166	0.170	0.179	0.182
Mean Sq Error	0.947	0.940	0.935	0.925	0.921
# observations	1070	1070	1070	1070	1070

Notes cont. as hours of job search per week. The reservation wage is measured in real terms, using the CPI as deflator (CPI=1 in January 1981). Gymnasium is equivalent to senior high school. Income refers to income from spouse or cohabitant.

tends to increase ξ, and improved local labor market conditions tend to decrease ξ. That is, increased education tends to exacerbate the effect of an increased (log) reservation wage on duration; improved local labor market conditions tend to mitigate the effect.

4. Conclusion

In this paper we presented an econometric specification for a structural search model, together with some preliminary empirical results. Although our results are tentative, they are encouraging. Our estimation procedures, correcting for simultaneity and the selectivity bias/missing data problem that arises in our data, produce intuitively sensible results. The possibility that a more complete analysis of these data will shed some light on the sources of variation in youth labor market experiences seems good.

References

1. Albrecht, J., Holmlund, B. and H Lang, 1988, "Job Search and the Transition to Employment: Theory," mimeo.

2. Burdett, K. and D. Mortensen, 1978, "Labor Supply under Uncertainty," in: R Ehrenberg, ed., **Research in Labor Economics**, Vol 2, JAI Press.

3. Heckman, J., 1979, "Sample Selection Bias as a Specification Error," **Econometrica** 47, 153–161.

4. Jensen, P. and N. Westergård–Nielsen, 1987, "A Search Model Applied to the Transition from School to Work," **Review of Economic Studies** 54, 461–472.

5. Jones, S., 1988, "The Relationship Between Unemployment Spells and Reservation Wages as a Test of Search Theory," mimeo, forthcoming in **Quarterly Journal of Economics.**

6. Lancaster, T., 1985, "Simultaneous Equations Models in Applied Search Theory," **Journal of Econometrics** 28, 113–126.