

# Human Capital, Unemployment Duration and Individual Heterogeneity \*

Richard W. Evans  
University of Texas at Austin  
evans@eco.utexas.edu

Thomas G. Koch  
University of Texas at Austin  
koch@eco.utexas.edu

January 2007

## Abstract

We estimate the effect of human capital on the unemployment hazard using longitudinal survey data. We are able to account for individual unobserved heterogeneity in two ways. First, we model the individual heterogeneity as being randomly distributed. Second, we account for fixed unobserved individual heterogeneity using traditional panel data methods. In our specifications that control for individual heterogeneity, we find that the level of human capital has a negative effect on unemployment duration—contrary to the findings of many other studies.

*keywords:* Unemployment duration; Human capital; Duration analysis

*JEL classification:* J64; C41; J24

---

\*We especially thank Jocelyn Wikle for her significant contributions to the early drafts of this work. We also thank Daniel Slesnick and Daniel Hamermesh for their helpful comments. Special thanks as well to Jonathan Gruber, for his code for imputing UI benefit levels. All errors are our own. Corresponding authors: Tel.: +1 512 471 3211; fax: +1 512 471 3510. Department of Economics, University of Texas at Austin, 1 University Station C3100, Austin, TX 78712-0301.

# 1 Introduction

Most papers studying unemployment duration focus on the effect of unemployment insurance (UI). But often overlooked is the estimated effect of human capital. Furthermore, these estimates of the effect of human capital on unemployment duration can be biased by not accounting for individual heterogeneity. Most studies estimate a positive or non-significant relationship between education and unemployment duration. However, education is endogenous, and the previous literature does not adequately account for this potential source of bias. The goal of this study is to estimate the effect of education on unemployment duration, accounting for individual heterogeneity.

The basic labor search models surveyed by Rogerson, Shimer, and Wright (2005) suggest that human capital affects unemployment through two main channels. More human capital increases the rate of job offers, thereby decreasing unemployment duration. But more human capital can also increase an agent's reservation wage, thus increasing unemployment duration. So the effect of human capital on unemployment duration depends on which effect dominates.

Most previous studies on unemployment duration have provided evidence of either a positive or non-significant effect of human capital on unemployment duration, or rather, a reservation-wage effect that weakly dominates. In all eight of his unemployment hazard model specifications, including those that control for individual heterogeneity, Meyer's (1990) estimates of the effect of human capital (years of schooling) on unemployment duration are all positive with some being statistically significant. Meyer uses the Continuous Wage and Benefit History (CWBH), and he models the heterogeneity as being randomly distributed in his specifications that control for individual heterogeneity. Kiefer (1985) uses a duration model on job training data, and also finds a positive relationship between education and unemployment duration. However, these studies do not account for the potential bias due to endogenous education decisions. Pavoni (2003) simply assumes a positive relationship between human capital and the rate of job offers in his structural model.

Some more recent papers do find a negative effect. Hunt (1999) looks at unemployment duration among the transitioning East German economy from 1990 to 1996 using data from the German Socio-Economic Panel (GSOEP). Hunt finds that different measures of human capital (whether completed equivalent of high school, whether completed university, whether had apprenticeship) all have either a negative effect or no effect on unemployment duration. However, she does not use the panel dimension of the GSOEP and has no control for individual heterogeneity.

More recently, Tatsiramos (2006) looked at the effects of UI benefits on both unemployment duration and the subsequent employment duration using data from the European Community Household Panel (ECHP) on eight countries. Similar to Bover, Arellano, and Bentolila (2002), Tatsiramos identifies the effect of UI by comparing recipients' and non-recipients' outcomes. He controls for individual heterogeneity in the same way as Meyer (1990)—by modeling it as a stochastic variable with known distribution. Human capital is represented by indicator variables for secondary education and higher education. Education level only has a significant monotone negative effect on unemployment duration in three out of eight European countries (Germany, Italy, and the U.K.). In the other five countries, the estimated effect is either mixed, positive, or not significant.

As documented previously, some papers control for individual heterogeneity while others do not and some of the papers use panel data while others use cross-sectional data. We argue that accounting for individual heterogeneity is extremely important for obtaining unbiased estimates, especially when using panel data. The simple example is the case in which an individual has unobservably low ability that translates more and longer unemployment spells. If a simple Cox hazard model is used, all the individuals are pooled over the panel and each unemployment spell of the low ability individual becomes multiple unemployment spells for multiple individuals, each of which is longer in duration than the average spell. It is clear how this would bias any estimates in an unemployment duration hazard model.

Our goal in this paper is to estimate the effect of human capital on unemployment duration using panel data and accounting for both observed and unobserved

individual heterogeneity. This is the first unemployment hazard study we know of that takes advantage of the panel dimension of individual data in order to account for unobserved heterogeneity. As described above, other papers (Meyer (1990) and Tatsiramos (2006)) have attempted to control for unobserved heterogeneity by modeling it as a random variable, according to an assumed distribution. But no papers have taken advantage of the panel dimension of the data to account for the individual unobserved heterogeneity.

Another advantage of using longitudinal survey data is its reduced degree of right censoring in unemployment duration data. Rather than using administrative data that are censored when benefits are exhausted, the unemployment duration data in the PSID are only censored at 52 weeks; the number of censored individuals is small relative to the CWBH.

Our findings offer a contrast to what has conventionally been found in the UI literature regarding the effect of education on the probability of becoming employed. We find that the level of human capital has a positive effect on the unemployment hazard (decreases unemployment duration) in specifications that use panel data to account for unobserved individual heterogeneity. Our positive coefficients on human capital are particularly interesting because they reinforce the idea that the job offer effect in standard labor search theory dominates the reservation wage effect.

## 2 Data

The data for this project come primarily from the PSID. Our sample consists of heads of household between 1976 and 1992. We examine only heads of household who claim to have been unemployed in the previous year at least once during the sample period. Because of the biased nature of survey-reported benefit amounts, we impute the UI benefit level the individual would be eligible for using the methodology of Gruber (1994), instead of what they reported or eventually claimed.<sup>1</sup>

---

<sup>1</sup>We thank Jonathan Gruber for sharing his code that incorporated the different UI laws across different states in different years to impute each individual's UI benefit level.

**Table 1: Summary Statistics, Restricted PSID: 1976-1992 ( $N=8,071$ )**

Variable	Mean	Std. Dev.	Min.	Max.	$\rho(\text{un. dur.})^a$
Unemployment duration (weeks)	13.4	12.7	1	52	1.000
Real UI benefit level (weekly \$) <sup>b</sup>	227.97	355.18	6.73	3,185.87	-0.055
Replacement rate	0.587	0.149	0.085	0.999	0.106
Real annual income (\$) <sup>b</sup>	14,572	10,553	528	116,936	-0.157
Years of education	11.6	2.6	0	17	-0.040
Father's years of education	13.3	3.2	0	18	-0.014
Mother's years of education	12.9	2.7	0	18	0.002
Age	32.4	10.7	17	79	0.024
White	0.55	0.50	0	1	-0.129
Black	0.42	0.49	0	1	0.133
Married	0.58	0.49	0	1	-0.076
Male	0.77	0.42	0	1	-0.051
Number of children	1.11	1.29	0	9	-0.003
Tenure at previous job (weeks)	45.2	94.1	0.0	999.0	-0.043
State unemployment rate	7.41	2.14	2.20	18.00	0.159
County unemployment rate	7.74	2.82	0.80	34.00	0.086
UI Extension year	0.46	0.50	0	1	0.045
Long-term spells <sup>c</sup>	0.10	0.29	0	1	0.739

<sup>a</sup> Correlation coefficient of the variable and unemployment duration.

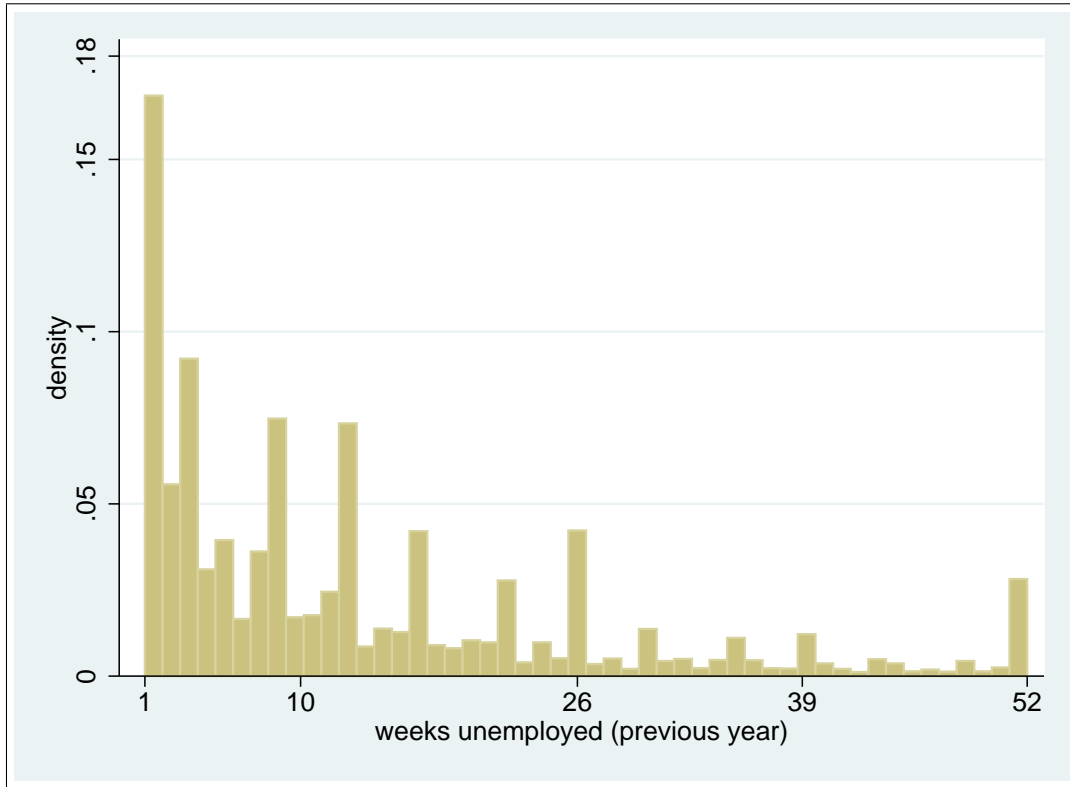
<sup>b</sup> Weekly UI benefit level and annual income of the head of household are transformed into real variables by using the CPI core price index with base year 1992 = 100.

<sup>c</sup> These are spells that are longer than the maximum UI benefit period, i.e. greater than 26 weeks in a year in which the maximum UI-benefit period was not extended by the U.S. Congress or greater than 39 weeks in a year in which the benefit period was extended by Congress.

Our data set consists of 8,071 observations on 3,850 individuals. Summary statistics are presented in Table 1. The average number of weeks unemployed in our sample is 13.4 weeks. Truncation of unemployment spell duration in the PSID is at 52 weeks. Included in our analyses is an indicator variable for years in which the U.S. Congress federally mandated extension of UI benefits, which is the case for most recession years in our sample. Figure 1 is a histogram of unemployment spell durations across the entire sample. Most individuals have spells shorter than ten weeks. It also demonstrates an irregularly large spike of spells ending at 26 weeks; UI benefits, with a few exceptions, expire after 26 weeks of enrollment.

Using the panel dimension of the data allows us to control for individual heterogeneity in a way that has been neglected in previous studies. Table 2 reports the number of years over the sample period in which the head of household reported an

**Figure 1: Histogram of Unemployment Spell Durations Over Entire Sample: PSID 1976-1992**



unemployment spell as well as the number and level of education changes in those years. Individuals who experienced more than one unemployment spell or who had their education level change during the sample period provide the “depth” of the panel.

Just over half of our sample has some panel dimension in number of unemployment spells, in that they experienced more than one spell over the sample period. However, only about six percent of individuals in our sample experienced a change in their education level during the sample period. This small number of individuals who experienced changes in education level during the sample will become important for our fixed-effects estimation (specification (5) in Table 3).

**Table 2: Number of Unemployment Spells and Education Changes, Restricted PSID: 1976-1992**

Number of spells			Years of educ. chg.		
	Freq.	Pct.		Freq.	Pct.
1	1,982	51.48	0	3,982	49.34
2	860	22.34	1	116	1.44
3	432	11.22	2	73	0.90
4	251	6.52	3	25	0.31
5	140	3.64	4	20	0.25
6	75	1.95	5	2	0.02
7	48	1.25	7	3	0.04
8	19	0.49			
9	21	0.55	missing <sup>a</sup>	3,850	47.70
10	9	0.23			
11	8	0.21			
12	2	0.05			
13	2	0.05			
14	1	0.03			
Total	3,850	100.00		8,071	100.00

<sup>a</sup> These missing observations are simply the first observation for each individual. Note that the total of missing education changes equals the total of individuals with at least one unemployment spell (3,850).

### 3 Hazard Estimation

The standard Cox proportional hazard time-dependent model makes the individual hazard rate (here, the probability of exiting unemployment or becoming employed) a function of the baseline hazard rate and the values of the covariates.<sup>2</sup> Define the hazard rate  $\lambda_{i,t}$  as the probability of exiting unemployment for individual  $i$  at time  $t$  given that the individual is still unemployed.

$$\lambda_{i,t} \equiv \lim_{h \rightarrow 0^+} \frac{\text{Prob}(t+h > T_i \geq t | T_i \geq t)}{h} \quad (1)$$

<sup>2</sup>Good references for Cox's partial likelihood estimation are Cox (1972) and Cox (1975).

where  $T_i$  is the *ex post* length of individual  $i$ 's unemployment spell. Then the hazard rate is a function of the baseline hazard and covariates in the following way:

$$\lambda_{i,t} = \lambda_{0,t} \exp \{ \boldsymbol{\beta}' \mathbf{x}_{i,t} \} \quad (2)$$

where  $\lambda_{i,t}$  is the probability that individual  $i$  exits unemployment (becomes employed) at time  $t$ ,  $\lambda_{0,t}$  is the baseline hazard rate at time  $t$ , and  $\boldsymbol{\beta}' \mathbf{x}_{i,t}$  is the vector of time-varying covariates and their respective coefficients. The model is then estimated by maximum likelihood.

Meyer (1990) proposes an extension to an estimation method of Prentice and Gloeckler (1978) that estimates a proportional hazard model as in (2) but also accounts for the censoring issue of his administrative data. We will refer to this methodology hereafter as PGM (Prentice, Gloeckler, Meyer). Meyer's baseline empirical model is simply the Cox proportional hazard model in (2) estimated by maximum likelihood in such a way as to account for censored data. Meyer then adds randomly distributed unobserved individual heterogeneity to his baseline model.

$$\lambda_{i,t} = \theta_i \lambda_{0,t} \exp \{ \boldsymbol{\beta}' \mathbf{x}_{i,t} \} \quad (3)$$

where  $\theta_i$  is a random variable of known distribution that is independent of  $\mathbf{x}_{i,t}$ .

The specifications of the respective hazard functions in equations (2) and (3) are clearly not linear. In order to use panel data estimation methods, we need our estimating equation to be linear. Amemiya (1985, §11.2) derives a linear relation from equation (2) in which the log of unemployment spell duration  $T_i$  is the dependent variable (contrary to the hazard rate models in which a probability of exit is the dependent variable).

$$\log(T_i) + c = -\boldsymbol{\beta}' \mathbf{x}_{i,t} + u_{i,t} \quad (4)$$

where  $T_i$  is the expected duration of unemployment and  $c$  is Euler's constant ( $c \cong 0.577$ ). This allows us to control for unobserved individual heterogeneity using the panel dimension of the data by estimating (4) by fixed-effects panel estimation, which

differences out the unobserved heterogeneity.<sup>3</sup> However, one drawback of this fixed-effects model is that it does not account for the censored data. The variation we do observe in about six percent of individuals in our sample allows us to get a point estimate, although they are imprecisely estimated.<sup>4</sup>

Table 3 contains results from four of our specifications as well as from Meyer's (1990) primary specification. Column 1 comes directly from Meyer (1990), and is his estimation using the CWBH data and estimating a model like equation (3) that accounts for censoring and in which unobserved individual heterogeneity is a random variable. Column 2 contains our results from estimating a simple Cox proportional hazard function as in equation (2). Column 3 contains results from estimating a Cox proportional hazard model but following Meyer (1990) in accounting for the censored data in the maximum likelihood estimation. Column 4 is our replication of Meyer's (1990) primary specification in which we estimate a hazard model that accounts for unobserved individual heterogeneity by modeling it as a random variable as in equation (3) and accounts for censoring. Lastly, column 5 is our estimation of the model in equation (4) using fixed-effects panel estimation.

The key results from Table 3 come from the estimated coefficients on years of education in the first row. Note that specifications (4) and (5) control for unobserved individual heterogeneity by either modeling it as a random process or by differencing it out, respectively. On the other hand, specifications (2) and (3) ignore unobserved heterogeneity. In the specifications that ignore the unobserved individual heterogeneity, the coefficients on years of education are negative and significant, indicating that more education decreases the probability of getting a job (increases unemployment duration). However, the positive coefficients on our specifications that control for unobserved individual heterogeneity suggest that more education increases the probability of getting a job and decreases unemployment duration. These results are consistent with the theoretical claim that the job offer effect dominates the reservation wage effect. A problem with the fixed-effect estimates in column 5 is that hardly

---

<sup>3</sup>For a detailed discussion of the potential spell aggregation bias, see Appendix A-3.

<sup>4</sup>We don't use a random effects model because of the endogenous nature of education.

**Table 3: Unemployment Hazard Model Estimates**

Variable	Specification				
	Meyer 1990 (1)	Cox Hazard (2)	PGM Baseline (3)	PGM Hetero. (4)	Fixed Effects (5)
Years of education	-0.018 (0.012)	-0.016 (0.005)	-0.022 (0.009)	0.049 (0.008)	0.036 (0.032)
Log of real UI benefit level (after tax)	-0.869 (0.204)	-0.050 (0.034)	-0.089 (0.060)	0.027 (0.057)	0.015 (0.048)
Father's years of education	0.234 (0.083)	0.003 (0.004)	0.002 (0.006)	0.006 (0.006)	
Mother's years of education	0.101 (0.072)	-0.003 (0.004)	0.006 (0.008)	-0.011 (0.007)	
Log of real annual income at previous job	0.729 (0.142)	0.280 (0.025)	0.328 (0.046)	0.044 (0.042)	-0.044 (0.041)
Age	0.108 (0.107)	-0.008 (0.001)	-0.010 (0.002)	-0.007 (0.002)	-0.019 (0.005)
White	0.234 (0.083)	0.189 (0.024)	0.245 (0.044)	0.036 (0.039)	
Married	0.101 (0.072)	0.080 (0.033)	0.057 (0.060)	0.146 (0.054)	0.085 (0.063)
Male		-0.089 (0.037)	-0.030 (0.063)	-0.334 (0.058)	
Number of children	-0.039 (0.024)	-0.002 (0.010)	-0.021 (0.017)	-0.091 (0.017)	0.001 (0.021)
Tenure at previous job		0.0002 (0.0001)	-0.0001 (0.0002)	0.0005 (0.0002)	-0.0001 (0.0002)
State unemployment rate	0.097 (0.022)	-0.074 (0.007)	-0.074 (0.012)	0.013 (0.011)	-0.099 (0.010)
County unemployment rate		-0.005 (0.005)	0.007 (0.009)	-0.019 (0.008)	0.005 (0.008)
UI extension year		0.033 (0.160)	0.213 (0.258)	-0.662 (0.269)	0.019 (0.209)
UI benefit level × extension year		-0.001 (0.032)	-0.035 (0.055)	0.137 (0.054)	0.004 (0.042)
Accounts for unobserved heterogeneity	yes	no	no	yes	yes
Uses panel dimension	no	no	no	no	yes
Accounts for censoring	yes	no	yes	yes	no
Data source	CWBH	PSID	PSID	PSID	PSID
Number of observations ( $N$ )	3,365	8,071	8,071	8,071	8,071
Log-likelihood value	-8,928	-63,687	-907	-5,029	

any of the coefficients are significant due to the small number of individuals with variation in education during the sample.

In the case of Meyer (1990), we propose that the negative coefficient on education is a result of pooling bias. If a low-ability worker with a given education level has multiple unemployment spells, Meyer (1990) treats those spells as separate individuals and, thereby, the effect of education on the probability of exiting unemployment is biased downward. In our specifications that do not account for unobserved heterogeneity, we have the same type of bias. However, in our specifications (4) and (5), which do control for unobserved individual heterogeneity, we get positive coefficients on education. Using an average of specifications (4) and (5), four more years of education cut one's average unemployment duration by about a day. The size of this effect is small, but it is positive across specifications and significant in specification (4).

It is also important to note which specifications account for the censoring present in the duration data. Although our fixed-effects model—specification (5)—utilizes the panel dimension of the data, it does not account for the censoring. Meyer's (1990) PGM methods, which are reported in specifications (3) and (4) of Table 3, both account for the censored data, but the standard Cox proportional hazard in specification (2) does not. As with specifications (3) and (4), Meyer (1990) in column (1) does control for the censoring in the data.

## 4 Conclusion

Our findings offer a striking contrast to conventional results in the literature regarding the effect of education on the probability of becoming employed. We contradict the result from the administrative-data literature that higher levels of education actually tend to increase the average unemployment duration. Our results suggest that a pooling bias is introduced when unobserved individual heterogeneity is not accounted for in duration model estimation. Our specifications that account for individual heterogeneity (columns 4 and 5 of Table 3) give the result that a higher level of education

or human capital decreases the expected duration of unemployment. Conversely, the opposite sign results in all our specifications that neglect unobserved individual heterogeneity.

The sign of the effect of education on unemployment duration is important in the debate that surrounds unemployment and unemployment insurance policy. Further development of linear duration models that can account for unobserved individual heterogeneity would allow researchers to better take advantage of the information in the numerous panel datasets worldwide. Also missing is work on duration models using panel data that account for censored observations. These advances would help in the accurate assessment how unemployment duration is influenced by other economic factors.

## References

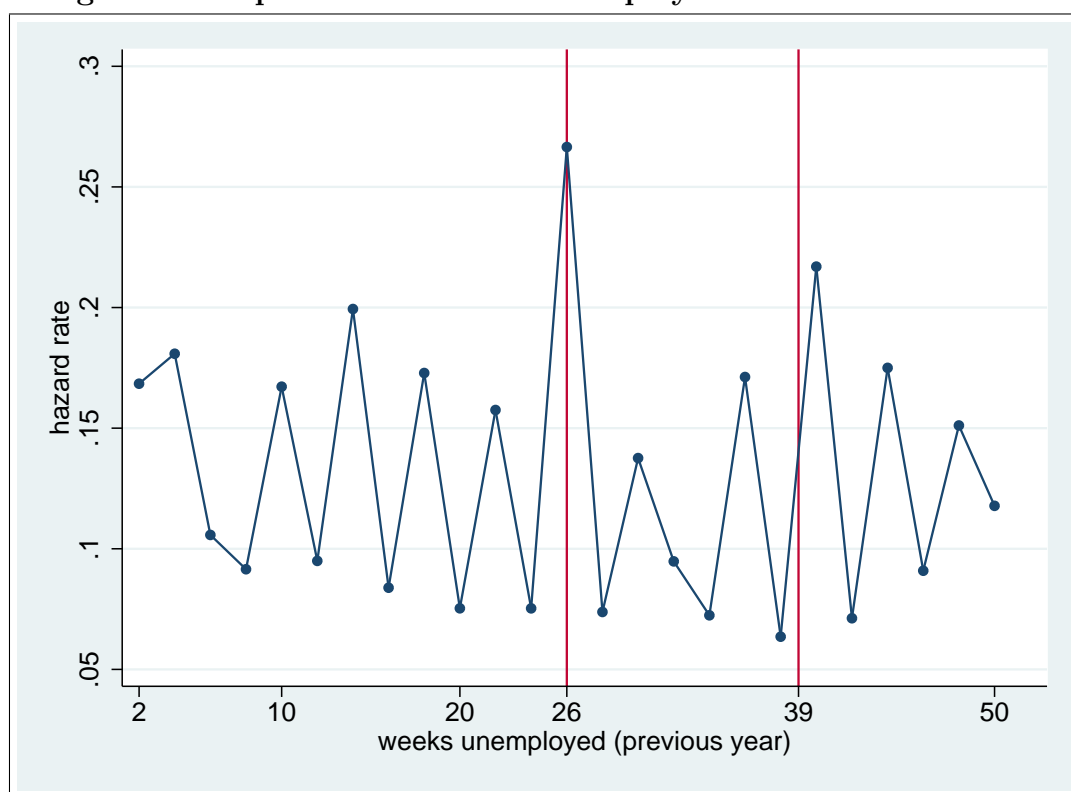
- AMEMIYA, T. (1985): *Advanced Econometrics*. Harvard University Press.
- BOVER, O., M. ARELLANO, AND S. BENTOLILA (2002): “Unemployment Duration, Benefit Duration and the Business Cycle,” *Economic Journal*, 112(479), 223–65.
- COX, D. R. (1972): “Regression Models and Life-Tables,” *Journal of the Royal Statistical Society, Series B (Methodological)*, 34(2), 187–220.
- (1975): “Partial Likelihood,” *Biometrika*, 62(2), 269–276.
- GRUBER, J. (1994): “The Consumption Smoothing Benefits of Unemployment Insurance,” NBER Working Papers 4750, National Bureau of Economic Research.
- HUNT, J. (1999): “Determinants of Non-employment and Unemployment Durations in East Germany,” NBER Working Papers 7128, National Bureau of Economic Research.
- KIEFER, N. (1985): “Evidence on the Role of Education in Labor Turnover,” *Journal of Human Resources*, 20(3), 445–452.
- MEYER, B. D. (1990): “Unemployment Insurance and Unemployment Spells,” *Econometrica*, 58(4), 757–782.
- MOFFITT, R. (1985): “Unemployment Insurance and the Distribution of Unemployment Spells,” *Journal of Econometrics*, 28(1), 85–101.
- PAVONI, N. (2003): “Optimal Unemployment Insurance with Human Capital Depreciation and Duration Dependence,” Discussion paper, University College of London.
- PRENTICE, R. L., AND L. A. GLOECKLER (1978): “Regression Analysis of Grouped Survival Data with Application to Breast Cancer Data,” *Biometrics*, 34(1), 57–67.
- ROGERSON, R., R. SHIMER, AND R. WRIGHT (2005): “Search-Theoretic Models of the Labor Market: A Survey,” *Journal of Economic Literature*, 63(4), 959–988.
- TATSIRAMOS, K. (2006): “Unemployment Insurance in Europe: Unemployment Duration and Subsequent Employment Stability,” IZA Discussion Paper 2280, IZA Institute for the Study of Labor.

# APPENDIX

## A-1 Data Description

In this appendix, we include some descriptions of the data that are instructive, but were not necessary in the main text of the paper. A useful representation of the data comes from the empirical hazard function. Figure 2 shows the empirical hazard rate for the entire sample; it shows increasing variance over spell length. Its irregular shape in the right tail is due to the small number of observations with spells that long.

**Figure 2: Empirical Hazard of Unemployment: PSID 1976-1992**



As mentioned earlier, one strength of the PSID is that we observe the number of weeks unemployed in a year, which goes beyond the censoring imposed by administrative UI data sets. Table 4 presents the means of the relevant variables from Table 1 separated by whether the individual spell was greater than the point at which the UI benefits are exhausted or less than the exhaustion point. We define unemployment spells that exceed the benefit exhaustion point as long-term unemployment. For the long-term unemployed, we include persons unemployed at least 26 or 39 weeks, the number of weeks an unemployed worker is typically eligible for benefits depending on when the federal government extended UI benefits. Since the actual length of individual spells is not observed, this is not necessarily the same as saying that the spell would be censored in an administrative data set.

**Table 4: Means by Spell Duration, Restricted PSID: 1976-1992**

Variable	Exhausted benefit means	Non-exhausted benefit means	Total means
Unemployment duration (weeks)	42.3	10.3	13.4
Real UI benefit level (weekly \$)	165.85	234.49	227.97
Replacement rate	0.621	0.583	0.587
Real annual income (\$)	11,295	14,917	14,572
Years of education	11.5	11.7	11.6
Father's years of education	13.2	13.3	13.3
Mother's years of education	12.8	12.9	12.9
Age	32.6	32.4	32.4
White	0.40	0.56	0.55
Black	0.58	0.40	0.42
Married	0.49	0.59	0.58
Male	0.72	0.78	0.77
Number of children	1.11	1.11	1.11
Tenure at previous job (weeks)	38.4	45.9	45.2
State unemployment rate	7.96	7.35	7.41
County unemployment rate	8.13	7.70	7.74
UI Extension year	0.32	0.48	0.46
Number of observations ( $N$ )	767	7,304	8,071

In our sample, 767 unemployment spells are defined as long-term, while 7,304 do not reach the exhaustion point. Those who were long-term unemployed lived in states and counties with higher unemployment rates, suggesting that local labor market conditions are important. Education, one measure of human capital, does not change much across these two sub-samples, though the mean number of weeks on the previous job (the tenure variable, another candidate to proxy for human capital) is over seven weeks shorter for those who would become long-term unemployed. The long-term unemployed are less white and more black than the short-term unemployed. The mean of real total income of those who are long-term unemployed is nearly \$4,000 less than for those who are short-term unemployed.

Table 4 also presents the breakdown of average UI replacement rate statistics of the imputed replacement rate by duration of unemployment spell. Since it is common in the literature to find increased unemployment spells with higher benefit levels, it should come as no surprise to find the long-term unemployed have a higher mean replacement rate, because higher-income unemployed individuals (who are unemployed for shorter spells on average) typically receive the state-specific maximum level of benefits.

We also ran a logit regression on the long-term unemployed variable. The results are presented in the first specification in Table 5. The coefficients on the log of the real UI benefit level and on years of education are both positive and statistically significant, and we get the intuitive result that the coefficient on income is nega-

tive. A state's unemployment rate has a statistically strong and positive association with long-term unemployment even when controlling for county unemployment rate. The substantial and significant negative coefficient on the race indicator for being white again suggests that race also plays a big role in determining the duration of unemployment.

**Table 5: Logit and Ordered Logit Estimation of Spell Duration and Number of Spells, Restricted PSID**

Variable	Logit long-term spell 1976-1992		Ordered Logit number of spells 1976-1977	
	coef.	(std. err.)	coef.	(std. err.)
Log of real UI benefit level (weekly \$)	0.212	(0.077)	0.139	(0.458)
Years of education	0.034	(0.017)	-0.032	(0.039)
Father's years of education	-0.007	(0.012)	0.021	(0.030)
Mother's years of education	0.004	(0.015)	0.029	(0.037)
Log of real annual income (\$ prev. job)	-0.692	(0.085)	-0.035	(0.422)
Age	0.012	(0.004)	0.014	(0.010)
White	-0.495	(0.084)	0.003	(0.217)
Married	-0.190	(0.112)	-0.145	(0.262)
Male	0.139	(0.118)	0.541	(0.359)
Number of children	0.009	(0.033)	0.038	(0.064)
Tenure at previous job (weeks)	-0.0005	(0.0005)	-0.003	(0.004)
State unemployment rate	0.192	(0.022)	-0.033	(0.069)
County unemployment rate	0.013	(0.017)	-0.029	(0.063)
UI Extension year	-1.024	(0.093)	0.016	(0.201)
Number of observations ( $N$ )	8,071		476	
Log-likelihood	-2,349.0		-407.1	
$\chi^2_{(14)}$	371.1		12.1	

The effect of UI extension in the logit regression has two components. First, if an individual's unemployment spell is in an extension year, the duration defining long-term unemployment as used here is raised to 39 weeks, thus making it less likely that a individual is long-term unemployed. In the other direction, this extension increases the duration of benefits available, which could increase the duration of unemployment. The former effect dominates, as the probability of long-term unemployment is lower in the extension years.

Table 6 gives the means of the variables used in the analyses below, and reveals how individuals who have multiple unemployment spells in the sample are different from those who report just one unemployment spell. Average unemployment spell duration does not change much across number of spells, although it rises slightly for individuals with three or more spells over the sample. The benefit level, replacement rate, and income variables show that individuals with the most income have the fewest number of unemployment spells. Also, education is correlated with fewer

unemployment spells. Age, marriage, and number of children are correlated with more unemployment spells, and white people tend to have fewer spells.

**Table 6: Means by Number of Spells, Restricted PSID: 1976-1992**

Variable <sup>a</sup>	Number of spells			
	1	2	3+	All
Unemployment duration (weeks)	13.47	13.47	13.54	13.49
Real UI benefit level (weekly \$)	328.22	210.22	185.81	264.58
Replacement rate	0.574	0.594	0.591	0.583
Real annual income (\$ prev. job)	15,009	13,942	14,528	14,645
Years of education	11.40	7.82	6.54	9.33
Father's years of education	13.33	13.53	13.34	13.38
Mother's years of education	12.73	12.82	12.94	12.80
Age	31.96	31.00	32.94	32.00
White	0.56	0.53	0.54	0.55
Black	0.39	0.43	0.43	0.41
Married	0.55	0.52	0.60	0.56
Male	0.71	0.71	0.81	0.74
Number of children	0.98	0.98	1.18	1.03
Tenure at previous job (weeks)	52.4	44.2	39.2	47.1
State unemployment rate	7.13	7.32	7.55	7.28
County unemployment rate	7.32	7.57	7.94	7.54
UI extension year	0.49	0.43	0.46	0.47
Long-term spells	0.104	0.099	0.094	0.101
Number of individuals	1,982	860	1,008	3,850

<sup>a</sup> These variables actually represent the average values of the variables across all spells.

One might also think that individuals who have consecutive unemployment spells are different from individuals whose spells are more spread out. Table 7 shows the distribution of unemployment spells across their position in a consecutive string of spells. Table 8 reports the means for the sample, according to the unemployment spell's order within a series of consecutive years of unemployment. On average, "last" spells have fewer weeks of unemployment than "first" spells and "middle" spells. Therefore, it is not apparent that overflow of weeks into the year ending consecutive years with unemployment is important.

It might also be the case that consecutive years of unemployment are due in part to aggregate labor market conditions—that is, an improving state unemployment rate leads to a higher unemployment hazard. This, in turn, decreases expected number of weeks employed, masking the number of weeks "gained" by overflow. Such effects do not seem to exist, as the state unemployment rate for "first" spells and "last" spells are virtually the same on average.

We do not observe individual unemployment spells, but the number of weeks unemployed in a given year. It may be the case for some observations in consecutive

**Table 7: Order of Consecutive Unemployment Spells, Restricted PSID: 1976-1992**

Consecutive order	Frequency	Percent	Cumulative percent
First	1,646	20.39	20.39
Middle	1,553	19.24	39.64
Last	1,852	22.95	62.58
Singleton	3,020	37.42	100.00
Total	8,071	100.00	

years that an unemployment spell in the first year “overflows” into the second year. Using a method described in Appendix A-2, we can consider the number of weeks gained and lost because of this overflow as random variables, with estimable expectations and variances. These calculations are very sensitive to the unemployment hazard, so we calculate them with a variety of reasonable hazard rates.

## A-2 Overflow Methodology

Define  $g$  as the number of unemployment-spell weeks gained due to calendar overflow of a spell from the previous year, and define  $y$  as the number of unemployment-spell weeks lost to calendar overflow of a current spell into the next year. We can write the respective probabilities for  $g$  and  $y$  in the following way:

$$\begin{aligned}
 Pr(g|L) &= Pr(\text{spell begins in week } 1^{t-1} \text{ and lasts } 52 + g \text{ weeks}) + \dots \\
 &\quad + Pr(\text{spell begins in week } 52^{t-1} \text{ and lasts } g + 1 \text{ weeks}) \\
 &= \sum_{j=1}^{52} Pr(\text{spell begins in week } j^{t-1} \text{ lasts } g + 52 - (j - 1) \text{ weeks}) \\
 Pr(y|L) &= \sum_{i=1}^{52} Pr(\text{spell begins in week } i^t \text{ and lasts } 52 - (i - 1) \text{ weeks}).
 \end{aligned}$$

Each probability is conditional on either being the first ( $F$ ) or the last ( $L$ ) year in consecutive pair of years with unemployment. The superscripts for the initial week of the spell denote whether the spell began in the year  $t$  or  $t - 1$ .

In general, if an unemployment spell is in the middle of a three-or-more-year stretch of years with unemployment (i.e., both  $F$  and  $L$ ), then  $\neg g \perp y$ . This is because of a crowding out effect—there are only 52 weeks in a year, and so a large  $g$  will increase the expected  $y$ . However, given that 52 weeks is a large span compared to the average duration of unemployment, the convenience of independence should not be alarming. Table 9 shows the probabilities of duration gain from overflow  $g$  for different values of the hazard rate  $\lambda$

The groundwork has been laid to calculate  $E(g - y|F, L)$ . If we assume independence of  $g$  and  $y$ , and that the employment hazard functions are the same for

**Table 8: Means by Consecutive Spell Position, Restricted PSID: 1976-1992**

Variable	Consecutive spell position				
	First	Middle	Last	Singleton	All
Unemployment duration (weeks)	14.3	15.2	13.3	12.0	13.4
Real UI benefit level (weekly \$)	230.09	206.01	187.88	262.69	227.97
Replacement rate	0.578	0.606	0.606	0.570	0.587
Real annual income (\$ prev. job)	16,018	13,246	12,458	15,763	14,572
Years of education	11.51	11.07	11.59	12.05	11.64
Father's years of education	13.29	13.14	13.44	13.39	13.33
Mother's years of education	12.84	12.85	12.93	12.83	12.86
Age	32.38	31.93	31.92	33.02	32.43
White	0.55	0.53	0.53	0.56	0.55
Black	0.42	0.44	0.43	0.40	0.42
Married	0.58	0.63	0.58	0.56	0.58
Male	0.77	0.82	0.78	0.75	0.77
Number of children	1.19	1.23	1.09	1.02	1.11
Tenure at previous job (weeks)	44.4	45.5	41.2	47.9	45.2
State unemployment rate	7.46	7.69	7.56	7.15	7.41
County unemployment rate	7.89	8.18	7.81	7.40	7.74
UI extension year	0.40	0.48	0.54	0.43	0.46
Long-term spells	0.11	0.12	0.09	0.08	0.10
Number of spells ( $N$ )	1,646	1,553	1,852	3,020	8,071

all individuals in all years under consideration, then  $E(g - y|F, L) = 0$ . The second assumption requires some understanding. Since we want to estimate how the employment hazard shifts as individual characteristics shift, this may be a poor assumption, especially in our fixed-effects estimations of the effects of increased education. However, the marginal effect of education as measured here requires at least a year out of the labor force, and thus, would not be eligible to be unemployed and looking for work in two consecutive years.

Using the above specification, we can treat  $g$  and  $y$  as random variables for the first and last years of consecutive years of unemployment, and construct expected values. We assume that the week in which a spell starts is independent of employment hazard and is uniformly distributed across weeks. This latter assumption should provide an upper bound for these values, as seasonality might dictate a higher hazard of employment and lower probability of a spell beginning at the end of the year.

### A-3 Spell Aggregation and Bias in the Linear Model

With the Amemiya fixed-effects model from equation (4), we can derive explicit forms for the bias induced by aggregating the unemployment spells experienced dur-

**Table 9: Values for Overflow Weeks Lost/Gained, by Employment Hazard**

	E(duration)=1/λ		
	λ=8	λ=10	λ=12
P(g>0)	11.76%	15.45%	18.97%
E(g)	0.9346	1.511	2.168
Var(g)	12.83	24.64	39.45

ing a year. Consider two estimators,  $\widehat{\beta}^j$  and  $\widetilde{\beta}^j$ , which represent the (negative of the) marginal effect of the  $j^{\text{th}}$  explanatory variable in  $\mathbf{x}_i$ . The first variable has no aggregation—that is, each unemployment spell is treated as a different spell. The second variable does have annual aggregation, as we have in our PSID data. We can write the bias for each component as follows. Superscripts on coefficients correspond to explanatory variables, while subscripts are for observations; i.e.,  $x_i^j$  is the  $j^{\text{th}}$  component of observation  $i$ . Lower-case  $t_i^{k(i)}$  represents the length individual  $i$ 's  $k(i)^{\text{th}}$  spell.  $T_i = \sum_{k(i)} t_i^{k(i)}$  is the aggregated number of weeks unemployment of individual  $i$ . Overbars denote the mean of the variable. Deviations from mean are denoted with asterisks.

$$\widehat{\beta}^j - \widetilde{\beta}^j = \frac{\sum_i x_i^{j*} \left( \sum_{k(i)} \log t_i - \log T_i + \overline{\log T_i} - \overline{\log t_i} \right)}{\sum_i x_i^{j*}}$$

We can read the numerator above as the covariance between the explanatory variable, and how the total time spent unemployed is distributed across the spells. It can be shown that for an individual:

$$\sum_{k(i)} \log t_i^{k(i)} < \log T_i \Leftrightarrow \prod_{k(i)} t_i^{k(i)} < \sum_{k(i)} t_i^{k(i)}$$

That is, excepting one-week spells, if an explanatory variable is correlated with increased variance in spell lengths for an individual relative to the mean, then we would expect a negative bias. Systematic correlation of this kind hard to consider *ex ante*.