The effects of intermittent employment histories on earnings

Job market paper

Anna Nesterenko

Stony Brook University

February 7, 2009

Abstract

This paper examines the effects of prior employment histories on subsequent labor market outcomes. Recognizing the limitations of the traditional experience measure introduced by Mincer (1958, 1974), I include additional variables in the earnings function to capture gaps in individuals' employment histories. While this ameliorates the omitted variables problem, it introduces an endogeneity concern, as experience gaps are likely correlated with other covariates and the unobserved error term. I overcome these concerns by implementing an IV approach. Additionally, to mitigate the problem of unobserved individual heterogeneity, I use a fixed effects specification. I also correct for sample selection as a means to adjust for non-random participation in the labor market, which is especially important for the female subsample. The analysis is performed using the National Longitudinal Survey of Youth (1979-2004). The use of the detailed data on employment histories for both men and women allows to better account for the heterogeneity of labor experience arising from different career interruptions. The fit of the model improves when accumulated previous employment gaps and most recent unemployment spells are included into the specification. Moreover, the negative effect of previous years not working are more prominent for younger adults, while the magnitude diminishes for older adults who have spent more time in the labor market, with females more heavily penalized for each year not working. Applying improved measures of work experience helps explain the differences in earnings based on the nature and quality of labor market experience acquired, and not necessarily based on gender and race differences alone.

Keywords: employment histories, career interruptions, earnings

JEL: J24, J31

1 Introduction

In this paper, I am revisiting the Mincerian earnings equation. The traditional experience measure, which reflects post-schooling investment in human capital, is calculated as the difference between an individual's actual age and his (estimated) age at completion of schooling. While a potential experience measure approximates the amount of time an individual could have been working, it does not necessarily reflect actual acquired experience, especially for those with interrupted careers. If an individual is unemployed during long periods of time, he or she may need more years to reach the "same" level of experience compared to someone with uninterrupted careers. While employment histories are different across individuals, a conventional measure, relying on the age and completed education, might be insufficient to capture such heterogeneity. However, this traditional experience measure has been widely used in empirical labor studies, it is widely available in many data sets, is easy to construct and explain in any analysis. With this in mind, I suggest to augment the potential experience measure with additional variables to better account for heterogenous employment histories.

Properly measured actual employment and acquired experience is important for the variety of applications relying on the Mincerian equation. In the analysis of the returns to education, it is important to separate the effects of human capital acquired through schooling from the one obtained in the labor market. For this, an accurate measure of true labor market experience is needed. The analysis of gender wage differentials implies that women with otherwise similar characteristics, including the experience, earn less than men. Given that women have more interrupted careers, it is especially important to use a comprehensive experience measure that would approximate the experience of comparable continuity and quality across men and women.

Accurate experience measures of acquired experience are essential not only for the empirical estimations of returns to education, gender wage gaps and other types of discrimination in the labor markets, but they are also important to better understand the hiring decisions in real life. Almost every employer is interested in hiring new employees with relevant experience in the field. More importantly, employers usually pay attention to the continuity of the careers of their potential employees, and ask to explain the gaps in their resumes. Once the individual is out of school and enters the job market, any significant non-working spells become questionable and give a negative signal for the employers, since skills acquired in school or during prior employment might become obsolete and human capital can depreciate. The latter is particularly important for highly educated individuals who invested a lot in the human capital to begin with. Moreover frequent or lengthy non-working spells might be a signal of poor work attitudes and lack of motivation of the potential employee.

While Mincer (1974) recognized the necessity of using direct information on experience in the earnings function, and especially for those whose careers are not continuous, it is only with the increased data availability over the recent years, that different authors have proposed various ways to approximate actual experience. Building upon the existing literature I will examine augmented earnings function for both men and women, looking at 28 years of individual employment histories and comparing early and established careers of the individuals in my sample.

This paper is organized as follows. Section 2 briefly discusses the relevant literature and the alternative experience measures used. I discuss the data I use and present some descriptive statistics in section 3. I then suggest a theoretical and econometric framework for my model in section 4, and discuss the estimation strategies and results in section 5. Section 6 concludes.

2 Literature Review

The literature that discusses the issues of employment histories and labor earnings is rich and multifaceted. Most of the econometric models adopt some form of the "earnings function" with various measures of labor market experience and personal characteristics as some of the independent variables. Depending on the theoretical question and empirical model at hand, the differences in wages among the workers are attributed to (or mostly explained by) their personal characteristics (gender, education, etc.) or labor market behavior (work experience, training received, search strategies, unemployment experienced).

One common feature of the models that analyze the effects of prior labor market experiences on the subsequent outcomes is the use of measures of labor market experience. One of the most simple and widely used one is *potential experience*, which is usually constructed as individual's current age, minus number of years of education, minus six years (Mincer, 1958, 1974). One of the disadvantages of this measure is its inability to account for the heterogenous quality of the acquired experience. Namely, different individuals with the same number of years of *potential experience* might have followed quite different employment paths after completion of school, which have resulted into different types of experience acquired.

In their analysis of the effects of human-capital accumulation on the female wages and earnings, Mincer and Polachek (1974) stated that the "potential work experience" measure is inadequate for those labor market participants whose work histories have been interrupted leading to a different length and quality of actual work experience. To account for the intermittent labor market career, the authors distinguish among several chronological stages over the life cycle of married women: prematernal (mainly continuous employment), childbearing and child care (non-participation and intermittent participation at the labor market), and permanent return to the labor force after children reach school age. Each stage is characterized by different degrees of incentives to invest in human capital, hence resulting in different rates of return to human capital. For the empirical estimation, they used NLS data from 1967 with the recall information. To circumvent the problem of endogeneity of the "work experience" variable , they separately estimated female labor-supply equation with husband's income and husband's education as exclusion restrictions, and then used *estimated* work experience variable in the main wage regression. Results of such 2SLS estimation are reported to be similar with those from the OLS procedure (women are "penalized" by lower wages at the next job after some period of non-working, but their wages rebound relatively quickly.) As one of the applications, they use their estimation to explain the wage gap between married men and women, and between married men and single women. Miller (1993) discusses the limitations of using potential experience measure in the estimations of the probability of participation, and suggests that potential experience or age are likely to reflect reduced productivity and skill obsolescence over time, especially when participation is noncontinuous.

Recognizing the noisiness of the *potential experience* measure, some authors suggest using *actual experience*, which is taken as actual number or weeks/hours worked, or as an actual number of years of full employment since the age of 18 (Kim and Polachek, 1994; Miller, 1993), or a number of years with part-year full employment or full-year part-time employment. Other measures of experience include combination of "full-time", "part-time" employment and "non-working time" variables. Simpson (2000) analyzed the effects of intermittency on earnings for both men and women, using the Canadian Survey of Labour and Income Dynamics (SLID). He included full-time and part-time labor market experience, and also non-working time along with "potential experience" in the estimation of the earnings function to find that the patterns of intermittency do have effect on the earnings later in life, and these effects are less favorable for women and for younger workers. Hotchkiss and Pitts (2005) in their study based on the HRS data introduce an *index for intermittency* to analyze the women's intermittent behavior at the labor market. After having controlled for selection into the labor force, and then into the "intermittent sector", they estimate a 16% wage penalty among women who report some substantial spells of no labor market activity, compared to the women who work continuously.

For the estimation of a more comprehensive work history model, Light and Ureta (1995) and Spivey (2005) include an array of the variables to measure labor market experience and career interruptions: fraction of the weeks worked during each observed year since the career start, dummy variables for "career in progress" and tenure with the most recent employer. Light and Ureta (1995) justified the use of a more flexible specification for a wage equation in their study of gender wage differentials. The analysis was performed on a somewhat restricted sample of white men and women, who were observed between ages of 24 and 30 and reported at least some work experience during the time of the survey. The authors estimated and analyzed a work history model, which included a set of experience variables to describe individual employment history in greater detail, as opposed to the more traditional models that use potential or cumulative actual experience. Inspired by their work, Spivey (2005) employed NLSY79 data from 1979–2000 for both men and women, and used different sets of variables to account for potential experience, time of non-employment, career interruptions and fraction of weeks worked in the wage equations. Based on the various model specifications she used, Spivey (2005) concluded that total time on non-employment, recent career interruptions (of almost a year), as well as those in the very beginning of one's career, all negatively effect current wages, with wage losses being more pronounced for men than for women.

In the study of the relationship between different early labor market experiences and consecutive adult labor market outcomes, Gardecki and Neumark (1998) realize the necessity of looking beyond just few years after finishing school and entering the labor market by young people, and extend their analysis until the individuals in the sample are in their 30s. Early labor market experiences are defined as those that occurred during first five post-schooling years, and include different types of on-the-job and off-the-job training, enrollment in community colleges and lengths of labor market experience and tenure. They examine the effect of these experiences on wages, and find positive effects of the earlier spells of training, but not a significant effect of other types of experiences. Undoubtedly, different career paths early on will lead to quite different labor market outcomes later in the career. This hypothesis is supported by the findings of Alon et al. (2001), who suggest that along with the different amount of accumulated labor market experience, its timing and volatility explain different levels of labor market attachment.

The differences in career paths considered in this paper stem from the volatility of the employment history, as well as timing of and reasons for the various job changes. Job mobility events can be associated with voluntary separations, which result from quits for family-related and non-family-related (economic) reasons, or involuntary separations, which result from lay-offs and discharges. As workers move from one job to another, they can change their occupation, industry, or both. Additionally, we can distinguish among job-to-job and job-to-unemployment events, hence allowing for the unemployment spells in the employment history. Several researchers analyzed the effects of these factors on wages, mainly focusing on one group of factors at a time.

Focusing on the types of separations, Keith and McWilliams (1995) study the effects of cumulative job mobility, which they define as sum of the job separations of different types. Using 10 years of the data from the NLSY, they conclude that cumulative prior job histories, and not only the most recent separation event, affect the wages for both men and women, although these job histories are statistically different by gender. By having disaggregated mobility events by different categories, they find different effects of different types of separations, which are in most cases similar for men and women. While all voluntary separations are positively related to the subsequent wages for both men and women, their separate categories, i.e., quits for economic vs. family-related reasons, have different effects in all the specifications used. Women appear to quit for family-related reasons more, and then they experience lower wages at their next job, which authors explain by their sorting into occupations which allow for intermittent careers but pay less.

Since career interruptions are more frequent among women, some authors focus only on the female subsample in their analysis. Baum (2002) examines the effect of work interruptions to give birth on female wages. To control for individual heterogeneity and non-random selection, he performs fixed effect estimations and adjusts for sample selection. He finds that while work interruptions in general reduce women's wages, the effect is less significant when women return to the same job as they held before maternity leave. Swaffield (2007) focuses on female wages, but presents the estimations for both men and women in order to derive the implications for the female wage differential. She concludes that the unexplained part of the gender wage differential is reduced when detailed employment history is used. The research is performed based on the British Household Panel Survey, 1991–1997.

In the analysis of the employment histories, job mobility, career interruptions and earnings, some authors use cross-sectional OLS and/or fixed effects specifications (Albrecht et al., 1999; Kim and Polachek, 1994; Swaffield, 2007). Other studies additionally account for non-random participation in the labor market, and correct their estimations for sample selection using Heckman selection model (Baum, 2002). Some authors recognize the endogeneity of work experience measures in the earnings equation (Mincer and Polachek, 1974; Kim and Polachek, 1994), but also admit the difficulty of finding valid instruments (Swaffield, 2007).

3 Data

The data set used for this analysis is the National Longitudinal Survey of Youth 1979, with the original sample of 12,686 young men and women first interviewed in 1979. The survey was administered annually from 1979 to 1994, and biennially after that. In 2004, round 21 of the interviews was conducted. There are three subsamples in the NLSY79 data set: cross-sectional (representative of noninstitutionalized civilian youths), supplemental (oversample of civilian Hispanic, black and economically disadvantaged non-black/non-Hispanic youth), and military (enlisted in the active military forces). The respondents from the military subsample will not be considered for the purpose of current analysis on labor market activity, since the incentives and decisions affecting their employment history are quite different from those in the civilian labor force, and should rather be modeled and analyzed separately.

Even though all members of the original cross-sectional and supplemental subsamples have been eligible for interviewing during each round of the NLSY79, funding constraints limited the number of the supplemental sample members interviewed after 1990, and there has been some non-interviewed individuals every year starting from 1980.[1] Hence, there is a different number of respondents in each wave, and for the regression analysis I will consider only those individuals who have responded to each wave of the survey.

Selected summary statistics for each wave of the study are presented in Table 1. These are the mean values of the respective variables among all respondents in each wave, unless otherwise noted. Potential experience is calculated as respondent's age minus years of education,

[1] Refer to the explanation of the reasons for non-interview in the Appendix.

minus six, while adjusted experience is the sum of all weeks worked divided by 52 (reflects an equivalent of the years of full-time employment). Motivation for the use of potential and adjusted experience measures will be provided later.

Table 1: Selected summary statistics (mean values)

Year	**Age**	**Education**	**Potential experience**	**Adjusted experience**	**Total income**†	**Hourly wage**‡	**% of workers**	**N**§
1979	17.577	10.441	1.159	.652	4170.113	3.894	79.36	4767
1980	18.515	11.063	1.467	1.057	5057.557	4.174	84.67	5275
1981	19.514	11.609	1.917	1.563	5295.379	4.068	82.81	5329
1982	20.507	12.040	2.480	2.132	6028.712	4.269	82.20	5376
1983	21.473	12.346	3.137	2.705	6931.639	4.655	80.98	5421
1984	22.482	12.541	3.946	3.320	7867.069	5.026	82.01	5435
1985	23.465	12.699	4.770	3.970	8975.456	5.258	84.34	5493
1986	24.534	12.807	5.730	4.658	10366.96	5.742	85.19	5509
1987	25.701	12.900	6.801	5.391	11895.06	6.458	85.44	5496
1988	26.946	12.970	7.973	6.127	12736.27	6.657	86.78	5520
1989	27.862	13.018	8.844	6.878	13345.45	6.802	87.46	5528
1990	28.964	13.069	9.895	7.650	14191.6	7.215	87.57	5528
1991	29.844	13.098	10.746	8.404	14425.85	7.364	87.86	5536
1992	30.864	13.130	11.734	9.162	14468.09	7.333	86.90	5520
1993	31.818	13.189	12.629	9.948	15455.89	7.583	87.09	5529
1994	32.904	13.218	13.686	10.712	15974.99	7.738	87.17	5532
1996	34.768	13.286	15.481	12.305	17310.3	8.657	87.29	5532
1998	36.723	13.313	17.410	13.918	18464.88	9.266	88.46	5537
2000	38.880	13.367	19.513	15.570	20628.68	9.976	88.79	5538
2002	40.794	13.407	21.387	17.207	21689.82	10.930	88.88	5515
2004	43.097	13.453	23.644	18.825	22708.11	11.283	86.58	5433

† Total real income (deflated by CPI, 1982–84 = 100) for workers (work == 1) with non-missing income.

‡ Real hourly wage (deflated by CPI, 1982–84 = 100) calculated as a ratio of total real income to hours worked, for workers (work == 1) with non-missing income.

§ Number of respondents in each round.

Before any formal econometric analysis is performed, the data set will be first analyzed to uncover some relationships between different variables, and bring additional insights toward the understanding of the link between employment histories and labor earnings. As the descriptive statistics suggest (Table 1), there is some discrepancy between potential and adjusted experience, which tends to increase slightly with respondents' age. It is more interest-

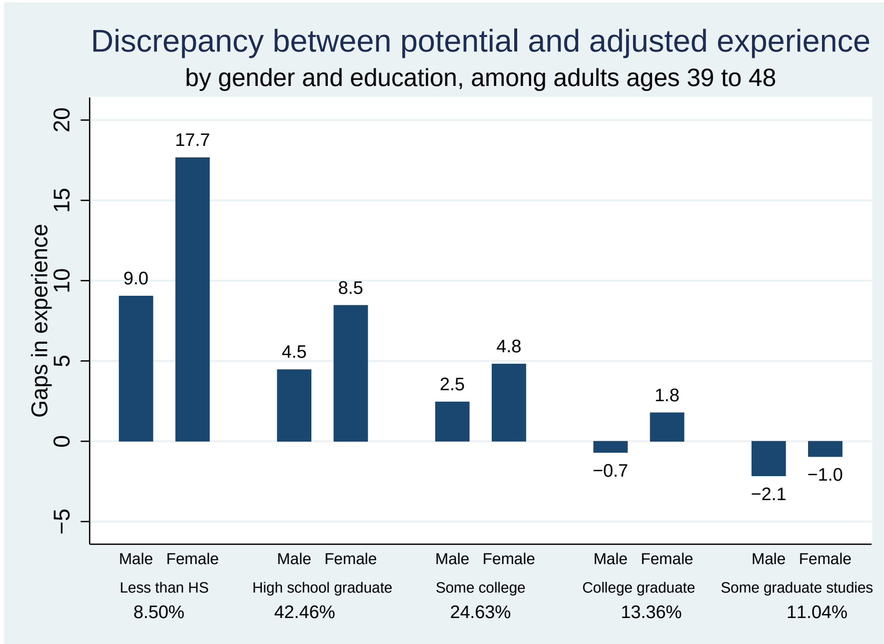
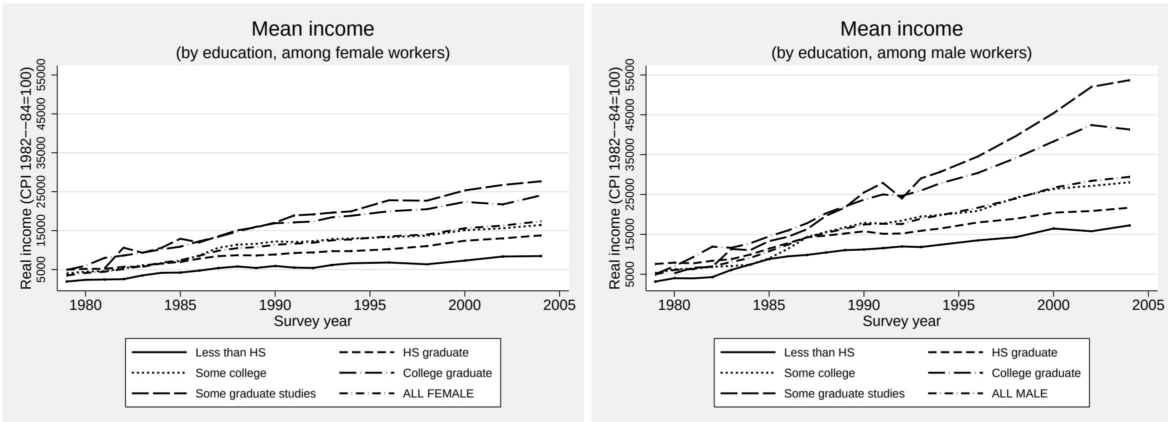


Figure 1: Employment gaps, by gender and education

ing, however, to look at the variation of such differences in labor market potential attachment and actual acquired experience across different demographic groups. If people worked all or most of the time after having completed their education, potential and adjusted experience measures would almost coincide. However, there are noticeable differences between these two measures across different education groups (Figure 1). High school drop-outs have, on average, the largest difference between potential and adjusted experience. The almost negligible difference between these two measures among college graduates suggests that they worked almost full-time all the time after finishing their education. Individuals with some graduate studies must have combined their education and work, since they have accumulated, on average, more adjusted (actual) than potential experience. Hence, for the individuals with less education, on average, potential experience will tend to overestimate their actual time of employment, while for the individuals with more education, potential experience will be an underestimate of their actual employment. Moreover, this negative correlation between gaps in experience and education implies that highly educated individuals earn more (Figure 2), not only due to their higher investments in human capital through schooling, but also due to their higher attachment to the labor market.

All the figures (Figures 1, 3, and 4) indicate that women have less adjusted experience, since potential experience should be very similar within different demographic groups in this representative sample. Married men (Figure 3) have the smallest discrepancy between these two experience measures, meaning that such individuals work almost full-time right after completing their education. For single, never married individuals this difference is relatively small, as well. Married women seem to experience more disruptions to their employment



(a) Working females

(b) Working males

Figure 2: Income growth over time, by gender and education.

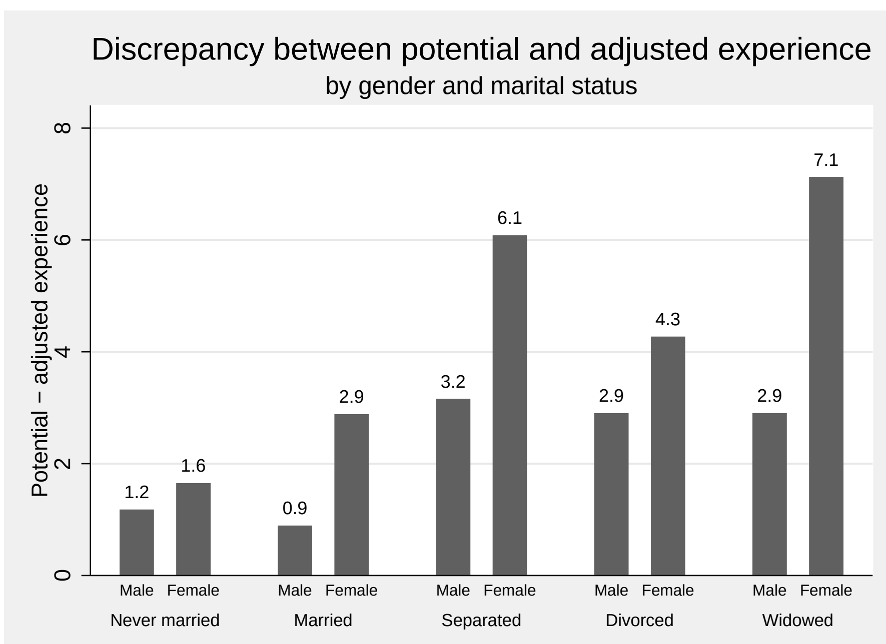
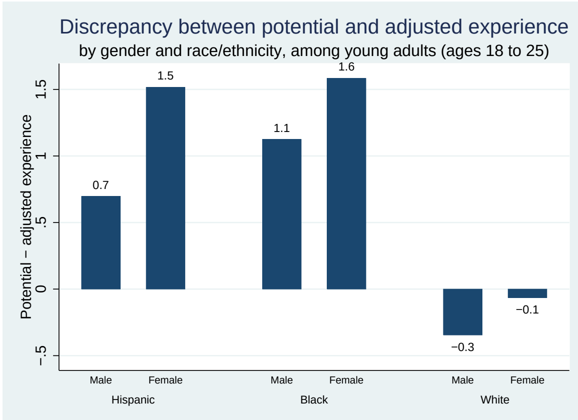
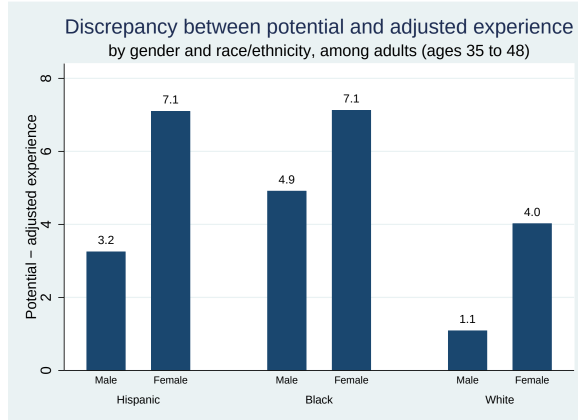


Figure 3: Employment gaps, by gender and marital status



(a) Young adults



(b) Older adults

Figure 4: Employment gaps, by race, gender and age groups.

history, and they are even more noticeable for previously married women (separated, divorced or widowed), and they are significantly higher than those of men with the same marital status. Consistent with other literature, women tend to experience more employment interruptions.

There is also some evidence in the data that men who work all year round and reported working over 1750 hours per year,[2] experience higher income growth over their careers than working women or men who do not work all year round (Figure 5). To establish a more accurate relationship and reach valid conclusions, a multivariate analysis has to be performed, which is the subject of the following sections.

4 The Model

To explain the link between different episodes in the employment history and consecutive labor market outcomes, with the focus on labor earnings, I will start by using a standard earnings function, and then augment it to account for the peculiarities of the individuals' employment history. Following Mincer (1974), I will consider the log of labor earnings to be a function of experience and its square, education and other exogenous individual characteristics:

$$\ln Y_{it} = \alpha + \beta_1 Experience_{it} + \beta_2 Experience_{it}^2 + \gamma Z_{it} + \varepsilon_{it}, \tag{1}$$

[2]The cutoff point of 1750 hours was chosen as an equivalent of fulltime employment according to the BLS definition of full-time workers (persons who work 35 hours or more per week) and assuming full-year employment of at least 50 weeks. Since this variable is constructed based on the total hours worked during past calendar year, it will also capture those individuals who worked less than 50 weeks but reported a lot of hours. For example individuals who worked 14 hours per day for 5 days a week during only 26 weeks in the past year, will also be classified as "full-time" workers. Hence, this variable should be used in conjunction with the one measuring the number of weeks worked.

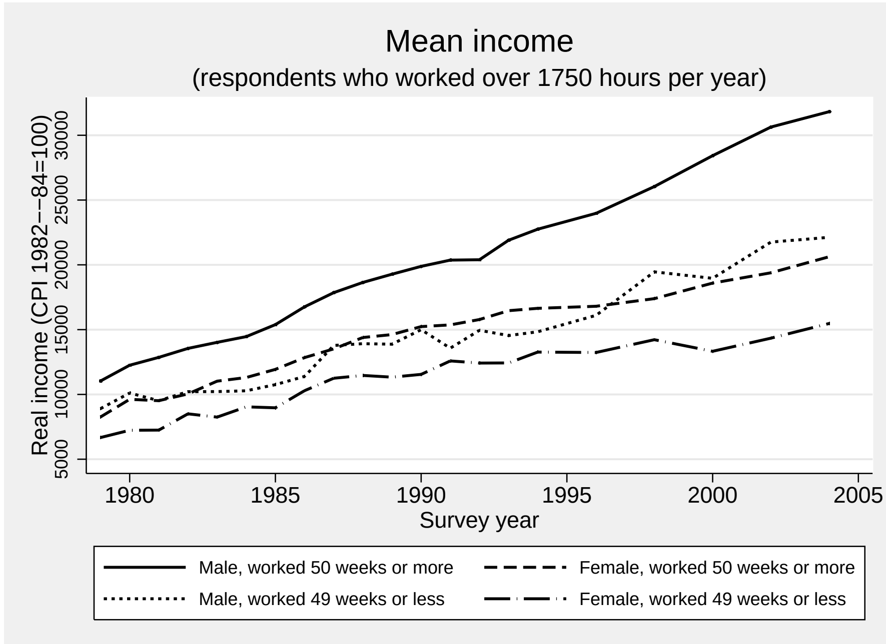


Figure 5: Income change over time (by gender and number of weeks worked)

where Y_{it} are individual labor earnings in every period, $Experience_{it}$ is the measure of potential experience (taken as a difference between individual actual age and an estimated age at completion of schooling), and Z_{it} is a vector of individual characteristics, including completed education. There is a number of shortcomings associated with the estimation of such benchmark model. As it is acknowledged by different authors, potential experience can be a rather noisy measure of true labor market experience acquired. Additionally, we can expect there might be non-random selection into the labor market (sample selection bias), as well as unobserved individual effects might be correlated with the regressors.

The fact that the potential experience inaccurately reflects actual working experience among heterogenous labor market participants, and various non-employment spells are not accounted for in the traditional measure used, can be treated as an omitted variables problem. For this to be a source of biases, we need these variables to be correlated with the other covariates in the earnings function. As I mentioned earlier, I find striking differences in employment gaps (time spent not working, in year equivalents) across the educational categories, gender and race. To alleviate possible biases caused by the omitted variables, I introduce additional variables in the Mincerian earnings function, like gaps in experience, which reflect the time spent not working during the course of the potential labor market attachment. I consider more remote accumulated gaps (constructed as potential experience minus adjusted experience[3] minus non-working spells during past year), as well as recent gaps in employment (weeks unemployed and weeks out of labor force during past year). I use these various gaps

[3] Adjusted experience is an equivalent of full years of employment, calculated as sum of all weeks worked divided by 52.

in employment along with the potential experience measure to allow for the heterogeneity of work experience among individuals, especially when career interruptions are significant. The combination of potential experience and gaps in employment will reflect actually acquired experience.

To simplify the notation and further discussion of the econometric model used, I will collapse employment history variables into the X_{it} vector, and the Z_{it} vector will include all other exogenous variables. In my model, I also allow for the presence of individual heterogeneity (c_i) along with idiosyncratic error component (u_{it}) in the unobserved error term. Following this notation (which adopts the notation used by Chamberlain (1984) and Wooldridge (2002)), I will re-write my earnings equation as:

$$\ln Y_{it} = \alpha + \beta X_{it} + \gamma Z_{it} + c_i + u_{it} \tag{2}$$

Exogeneity assumption implies that $E(Z'_{is}u_{it}) = 0$, for any s, t. However, there is possible endogeneity of some elements in the X_{it} vector. Given that it now includes measures of employment gaps along with the potential experience, there could be contemporaneous correlation between the elements of X_{it} and u_{it}, which occurs due to previous omission of important time-varying explanatory variable. Moreover, there is simultaneity in the dependent variable Y_{it} (labor earnings) and elements of X_{it}: individual motivation can affect both the individual earnings and amount of working time (and acquired experience). Additionally, in the presence of unobserved individual heterogeneity c_i, elements of X_{it} can be correlated with some individual time-invariant characteristics that affect earnings (like family upbringing, attitudes and motivation).

To address the problem of contemporaneous correlation, I am using the instrumental variables approach. The challenge of this technique is to find valid instruments, which are, in this case, correlated with individual choice on the amount of (non)working time, but uncorrelated with the individual earnings. I suggest to use local labor market conditions (unemployment rates in the region of residence[4]) that affect supply and demand of labor, but can be considered exogenous to individual earnings, which are determined by the contracts and employee's qualifications.

To purge the effects of unobserved individual heterogeneity, I am exploiting the longitudinal nature of the data and using panel data techniques. Given my previous assumption on the correlation between the individual effects and other covariates, it is hard to justify the use of the random effects for the estimation, which rely on the assumption of zero covariance: $cov(X_{it}, c_i) = 0$.[5] Fixed effects, on the other hand, allow for the correlation between the error term and covariates, and use the within estimator.

[4]The NLSY79 unemployment rate variables are constructed using state and metropolitan area labor force data from the *Employment and Earnings* publication for each survey year.

[5]Refer to the Estimation section for the formal tests for the choice between random and fixed effects models.

In the data, labor income is observed only for the labor market participants, who represent some non-random sample of all individuals. Improving the benchmark model further, I account for the possible non-random selection on unobservables, which implies that some unobserved individual characteristics that affect one's decision to work in any given period also affect the income earned. Following Heckman (1979), the participation equation for my model will be:

$$work_{it} = a_t + b_t WA_{it} + c_t Z_{it} + e_{it}, \tag{3}$$

where *WA* is a vector of work attitude variables and local labor market conditions that are likely to affect individual decision to work, but not the wages. Given the panel nature of the data used, each individual is making a decision whether or not to work repeatedly in each period, but we observe labor income only for the periods when an individual is employed for some positive number of weeks. Since the error terms of the participation and earnings equations are correlated, I will control for the sample selection using an estimation method proposed by Wooldridge (1995) and discussed by Dustmann and Rochina-Barrachina (2007), among others. As the first step, I will estimate a probit on $work_{it}$ for each period, and will estimate λ_{it} for each period, respectively.[6] I will then include constructed inverse Mill's ratio (λ_{it}) as an additional regressor in the main equation (2).

5 Estimation and Results

Following the estimation strategy outlined above, I examine different model specifications and discuss my main findings. For all the specifications, the dependent variable is the natural logarithm of real labor income (nominal income[7] adjusted for regional inflation, with CPI $1982-1984 = 100$). In all of the models, I control for the level of completed education as of the survey year (with *high school graduate* as an omitted category). To account for the convexity of the log earnings equation in schooling (Belzil and Hansen, 2002), I include dummy variables for different levels of education instead of the number of years of schooling. I also control for gender, marital status (with *married* as an omitted category), and race/ethnicity. Additionally, I use AFQT score variable as a proxy for ability, as well as include the interactions between the AFQT scores and ethnicity. To better understand the effects of experience on earnings, I split the sample into two age groups. Young adults (between ages 23 to 29), who are in the early stages of their careers, are likely to experience high income growth, while for the individuals older than 30 years of age, whose careers are established, income growth

[6] $\lambda_{it}(\cdot) = \phi(\cdot)/\Phi(\cdot)$, where $\phi(\cdot)$ is the standard normal density function and $\Phi(\cdot)$ is the standard normal cumulative distribution function.

[7] In the questionnaire, the question about income was asked as follows: "During past calendar year, how much did you receive from wages, salary, commissions, or tips from all jobs, before deductions for taxes or anything else?"

patterns are less steep. In my sample, by the time respondents turn 30, they are done with schooling and have at least 4 years of potential experience. Within these two age groups, I estimate the models separately for men and women.

As the first step, I estimate the benchmark model (1) with pooled OLS, including potential experience and its squared term, and controlling for the demographic characteristics, and region/type of residence (columns OLS(1) in tables 2 and 3). The returns to potential experience turn out to be of different magnitude for the two age groups: over 20% returns for young males, and only 5% returns for older males (compared to 15% in the combined sample for all ages[8]). While returns to potential experience are high for younger females (about 17%), they are statistically insignificant and very small for women over 30.

Recognizing the shortcomings of the traditional experience measure suggested by Mincer (1958, 1974), I include additional variables in the earnings function to capture gaps in the employment history. These non-working spells, omitted in the traditional experience measure, cause some human capital depreciation and skills atrophy, and have negative effects on the earnings. With this in mind, I estimate an expanded model with pooled OLS, where I include previous (more remote) gaps in experience (in years),[9] as well as the number of weeks unemployed and weeks out of labor force during the year (columns OLS(2) in tables 2 and 3). Controlling for the gaps in experience, both previously accumulated and most recent, decreased the magnitude of the returns to potential experience among younger men and women, but not so much the men over 30, and they are still statistically insignificant for women over 30.

These model specifications are rather parsimonious, but they still provide some very useful insights on the use of the different measures of labor market experience, and their effect on the estimated coefficients for other explanatory variables used, as well as alter the interpretation of the results. Employment gaps variables are statistically significant and have the expected negative signs in the estimations for the two age groups, both for men and women. More remote labor market experience, however intermittent, has less weight on the loss of earnings, than the more recent episodes in the employment history. While each year spent not working decreases the earnings by anywhere from 5% to 9%, each additional week spent not working last year reduces yearly income by 3-4%. In the multivariate analysis, the benchmark model produces rather high (and statistically significant) estimates on the returns to education, suggesting that a higher educational degree alone (compared to high school completion) can guarantee higher labor earnings for the individuals with the same duration of labor market attachment. In other words, if one has completed the schooling and can present the bachelor's or master's diploma, that will ensure higher income. Expanded OLS model specifications with different measures of employment gaps, on the other hand, produce lower (and still statistically significant) estimates on the returns to education. In the latter case, the individuals

[8]Estimation results for male and female subsamples of all ages are available from the author upon request.

[9]Previous gaps in experience do not include most recent non-working spells.

Table 2: Different cross-sectional model specifications: MEN ONLY

	23 to 29 years old			OVER 30 years old		
Variable	**OLS (1)**	**OLS (2)**	**IV**	**OLS (1)**	**OLS (2)**	**IV**
Dependent variable: ln *of real labor income, deflated by CPI* 1982 − 84 = 100						
potential experience	0.2294**	0.1447**	0.1372**	0.0566**	0.0551**	0.0511**
potential experience^2	-0.0104**	-0.0052**	-0.0056**	-0.0009**	-0.0008**	-0.0009**
previous gaps		-0.0687**	-0.0471**		-0.0526**	-0.0261**
weeks unemployed		-0.0370**	-0.0647**		-0.0300**	-0.1246**
weeks out of LF		-0.0423**	-0.0420**		-0.0342**	-0.0317**
less than HS	-0.3438**	-0.1120**	-0.0930**	-0.2130**	-0.0388*	-0.0377
some college	0.0892**	-0.0132	-0.0272†	0.1544**	0.0756**	0.0786**
college graduate	0.5211**	0.2312**	0.2132**	0.4702**	0.2771**	0.2899**
some grad studies	0.5481**	0.2584**	0.2582**	0.5691**	0.3205**	0.3543**
hispanic	-0.1538**	-0.0521†	-0.0249	-0.1334**	-0.0620*	-0.0717*
black	-0.3203**	-0.1865**	-0.1626**	-0.2808**	-0.1487**	-0.0895**
never married	-0.3636**	-0.2028**	-0.1836**	-0.4200**	-0.2897**	-0.2243**
separated	-0.2454**	-0.1167**	-0.0901*	-0.2817**	-0.1715**	-0.1206**
divorced	-0.2578**	-0.1226**	-0.0850*	-0.2344**	-0.1190**	-0.0406
widowed	0.2006	0.3285†	0.2526	0.0539	0.0114	0.0686
AFQT score	0.0029**	0.0021**	0.0020**	0.0038**	0.0032**	0.0031**
AFQT, Hispanics	0.0024**	0.0008	0.0005	0.0021**	0.0012*	0.0012*
AFQT, blacks	0.0048**	0.0043**	0.0041**	0.0032**	0.0025**	0.0014*
urban residence	0.0201	0.0179	0.0107	-0.0359**	-0.0227†	-0.0194
SMSA resident	0.2161**	0.1543**	0.1469**	0.2228**	0.1845**	0.1542**
northeast	0.1480**	0.0948**	0.0665**	0.0583**	0.0388**	0.0260
south	0.0936**	0.0265†	-0.0039	-0.0216	-0.0405**	-0.0656**
west	0.0717**	0.0539**	0.0319	-0.0127	0.0039	-0.0085
constant	8.1131**	8.7051**	8.8775**	8.8766**	9.0517**	9.2527**
N of obs	13698	13430	13430	18088	17877	17873
R^2_{adj}	0.203	0.485	0.435	0.283	0.429	0.082
First-stage $F_{(4,N-L)}$			28.58			7.56
p-value			0.0000			0.0000
Underidentification test			111.276			29.980
$\chi^2_{(4)}$ p-value			0.000			0.000
Hansen J statistic			21.509			1.156
$\chi^2_{(3)}$ p-value			0.0001			0.7636

Significance levels: † : 10% $*$: 5% $**$: 1%

Instrumented variables: weeks unemployed. Excluded instruments: unemployment rates in the region of residence

Table 3: Different cross-sectional model specifications: WOMEN ONLY

	23 to 29 years old			OVER 30 years old		
Variable	**OLS (1)**	**OLS (2)**	**IV**	**OLS (1)**	**OLS (2)**	**IV**
Dependent variable: ln *of real labor income, deflated by CPI* $1982-84=100$						
potential experience	0.1731**	0.0996**	0.0665**	0.0108	0.0050	0.0027
potential experience^2	-0.0075**	-0.0021**	-0.0020*	0.0003	0.0007**	0.0007**
previous gaps		-0.0918**	-0.0545**		-0.0662**	-0.0604**
weeks unemployed		-0.0387**	-0.1191**		-0.0311**	-0.0650
weeks out of LF		-0.0462**	-0.0468**		-0.0414**	-0.0410**
less than HS	-0.4985**	0.0368	0.0641†	-0.4158**	-0.0322	-0.0328
some college	0.2951**	0.0615**	0.0373†	0.2135**	0.0516**	0.0482**
college graduate	0.7066**	0.2974**	0.3026**	0.5237**	0.2305**	0.2350**
some grad studies	0.7703**	0.2963**	0.3069**	0.6639**	0.2772**	0.2845**
hispanic	-0.1704**	0.0167	-0.0055	-0.0094	0.1096**	0.0987**
black	-0.3500**	-0.0591*	0.0494	-0.2018**	-0.0517*	-0.0355
never married	0.0933**	0.0019	0.0466*	0.0984**	0.0565**	0.0814*
separated	-0.0906*	0.0368	0.1383**	-0.0310	0.0798**	0.0929**
divorced	0.0931**	0.1032**	0.1692**	0.1574**	0.1577**	0.1751**
widowed	0.1402	0.1900*	0.0755	-0.0142	0.0170	0.0133
AFQT score	0.0041**	0.0027**	0.0018**	0.0037**	0.0017**	0.0015**
AFQT, Hispanics	0.0043**	0.0008	0.0013	0.0036**	0.0007	0.0009
AFQT, blacks	0.0057**	0.0027**	0.0016*	0.0070**	0.0045**	0.0042**
urban residence	0.0571*	0.0294	0.0081	0.0256	0.0134	0.0152
SMSA resident	0.1856**	0.1008**	0.0869**	0.2315**	0.1578**	0.1489**
northeast	0.1146**	0.0981**	0.0578*	0.1157**	0.0684**	0.0672**
south	0.1286**	0.0491**	0.0170	0.0993**	0.0283*	0.0201
west	0.0098	0.0633**	0.0527*	0.0283	0.0393*	0.0349†
constant	7.4976**	8.3575**	8.7780**	8.3176**	8.9287**	9.0053**
N of obs	14535	14231	14231	19807	19514	19510
R^2_{adj}	0.159	0.534	0.310	0.168	0.461	0.432
First-stage $F_{(4,N-L)}$			9.65			3.49
p-value			0.0000			0.0075
Underidentification test			38.481			13.887
$\chi^2_{(4)}$ p-value			0.0000			0.0077
Hansen J statistic			12.632			8.636
$\chi^2_{(3)}$ p-value			0.0000			0.0345

Significance levels: † : 10% $*$: 5% $**$: 1%

Instrumented variables: weeks unemployed. Excluded instruments: unemployment rates in the region of residence

with similar length of actual work experience are compared, and their returns to education appear to be more realistic and in line with the existing literature on the returns to education. Additionally, high negative returns to incomplete high school education are diminished in the extended models with employment histories. As high school drop-outs spend the most time not working, compared with those with higher educational attainments, they are also penalized, in terms of lower earnings, for their non-working time, and not only for lower education.

While potential experience variable could be considered exogenous, since it depends on age and completed education, actual experience (a combination of the employment histories variables) reflects individual labor market participation choices over the years, which makes it an endogenous variable. Hence, different gaps in the employment can be considered endogenous as well. To deal with the endogeneity, I apply instrumental variables approach and estimate IV models for cross-sectional data. A Hausman test confirms the endogeneity only of the "weeks unemployed" variables, while previous gaps in experience (potential minus adjusted experience) and weeks out of labor force are treated as exogenous.[10] The set of excluded instruments for all IV specifications includes dummy variables for different unemployment rates for labor market in the place of current residence. Treating the data as cross-sectional for the IV estimations allows including race and AFQT score variables in the main equation. All cross-sectional IV specifications are estimated using 2-step efficient GMM with robust standard errors, and estimation results are presented in the columns IV, in tables 2 and 3. To test the validity of the instruments, I report the results of the underidentification test (Kleibergen-Paap rk *LM*-statistic[11]) and overidentification test (Hansen *J* statistic[12]). The instruments perform well only in some subsamples.

Compared with the OLS results, accounting for the endogeneity of the recent unemployment spells, slightly decreased the returns to potential experience for younger adults, almost did not change them for men over 30, and the estimated coefficients remained insignificant for women over 30. Estimated returns to experience are still noticeably higher for young adults, which is consistent with the proposition that labor income increases more in the beginning

[10] I instrument all three variables and test for the endogeneity of all three of them jointly to conclude that the null hypothesis that the regressors are exogenous could be rejected; I then instrument all three variables but test for the endogeneity of each of them separately, and find that the null hypothesis of exogeneity of "gaps in experience" and "weeks out of labor force" could not be rejected.

[11] To test whether the equation is identified and excluded instruments are correlated with the endogenous regressors, I am considering Kleibergen-Paap rk *LM*-statistic. In this test, the null hypothesis is that the equation is underidentified, and the matrix of reduced form coefficients on the $L1$ excluded instruments has $rank = K1 - 1$ where $K1$ is the number of endogenous regressors. Under the null, the statistic is distributed as χ^2 with $(L1 - K1 + 1)$ degrees of freedom. A rejection of the null indicates that the matrix is full column rank, and the model is identified.

[12] Hansen *J* statistic is used in the test of overidentifying restrictions. In this test, the joint null hypothesis is that the instruments are valid (uncorrelated with the error term and excluded instruments are correctly excluded from the estimation equation). Under the null, the test statistic is distributed as χ^2 in the number of $(L - K)$ overidentifying restrictions.

of the individual career. Coefficients on the previously accumulated gaps in experience are slightly higher for women, suggesting that women are more heavily penalized for the interrupted careers than men. Moreover, at the same time considering statistically insignificant coefficients on the potential experience for women over 30, it means that it is more difficult for these women to make up for years not working, because the length of their labor market attachment does not matter that much. Accounting for the endogeneity and instrumenting weeks unemployed, increase the estimated coefficient on this variable for all men and younger women. So, treating time spent unemployed as individual choice affected by local labor market conditions, increased the negative effect of such non-working spells on individual earnings. The other coefficients in the IV models are generally comparable to those from the respective OLS specifications that include employment gaps variables.

As descriptive evidence from the data set suggests, there is a fraction of respondents who did not work during each survey year. Very small number of respondents have not worked at all during all years the survey was conducted, while most of them have been in and out of the labor force over this time. Every period an individual is making a decision of whether or not to work, and the cumulative history of prior decisions to work along with the earnings in each past period, will affect the decision to work in the current period. Certainly, family upbringing, current family situation and personal motivation are important factors in the decision to work. Since labor earnings are observed only for labor market participants, there is clearly a problem of sample selection present in the estimation of the earnings function, where some unobserved individual characteristics that affect one's decision to work also affect the income earned.

I first use Heckman sample selection maximum likelihood specification to estimate my model. Following my previous analysis, I estimate the model separately for men and women within each of the two age groups (Table 4). Recognizing that different factors determine the choice to work for men and women, I use different specifications for the selection equations. For men, I use the variable to reflect whether an adult male present in the household when respondent was 14, worked for pay. For women, such "role model" would be a working women in the household when respondent was 14. I also include the number of the (biological, step and adopted) children in the household at each year of the interview in the selection equation for women. This variable was statistically insignificant in the selection equation for men, hence excluded from the final specification. In terms of attitudes toward work, the following two questions were asked during the initial rounds of the interviews, "Now I would like to talk with you about your future plans. What would you like to be doing when you are 35 years old?"[13] and "If, by some chance, you (and your (husband/wife)) were to get enough money to

[13]This question was asked during the first six rounds of the interviews. Response options included "present job", "some occupation", "married, family" or "other". I grouped those people responding "present job" or "some occupation" into one category – those who would like to work at the age of 35, and everybody else – into another. Since this question was asked during the first six rounds, respondents could answer it while they were 14 to 19 (the youngest cohort) up to 22 to 27 (the oldest cohort). Since there is no one age, at which any respondent

live comfortably without working, do you think you would work anyway?"[14] Indicators that a respondent would like to work at the age of 35 (as opposed to just having a family or doing something else), and that a respondent would be willing to work anyway, are used as variables affecting one's decision to work in the participation equation. Since both of these variables reflect respondents' attitudes early in life and do not have a panel component, I use them for the cross-sectional data estimations. Other variables used as exclusion restrictions include the unemployment rate in the region of residence. In the selection equation, I also control for the marital status, ethnicity and type of residence. Estimated coefficients in the participation equations all have expected signs. Having a working adult in the household when respondents were young increased the probability to work when respondents became older. Positive attitudes toward work are significant predictors for female labor market participation. However, with relatively small variation in work attitudes among men, and relatively high labor force participation, these two variables were insignificant predictors for men's decision to work (and also excluded from final selection equation specification). Number of the children in the household reduced women's probability to work in each period. As for the local labor market conditions, lower unemployment rates (with the unemployment rate over 12% as an omitted category) positively affect probability to work.

Table 4: Heckman selection model (ML)

	23 to 29 years old		OVER 30 years old	
Variable	**Male**	**Female**	**Male**	**Female**
Earnings equation				
Dependent variable: ln *of real labor income, deflated by CPI* $1982-84=100$				
potential experience	0.1516**	0.1003**	0.0562**	0.0212**
potential experience^2	-0.0057**	-0.0021**	-0.0008**	0.0003
previous gaps	-0.0656**	-0.0722**	-0.0504**	-0.0580**
weeks unemployed	-0.0357**	-0.0370**	-0.0284**	-0.0275**
weeks out of LF	-0.0403**	-0.0431**	-0.0336**	-0.0359**
less than HS	-0.1123**	0.0183	-0.0586**	-0.0475*

Significance levels : † : 10% $*$: 5% $**$: 1%

Continued on next page

had a chance to answer this question, I decided to look at their responses at the age of 18 or the earliest age available. Based on such age considerations and following two types of responses constructed above, I created a dummy variable "choice to work at18" to reflect respondents' choice to work when they become older. About 81% of respondents stated that they would like to work at the age of 35, when asked at the age of 18 (or earliest age available).

[14]This question was asked only during 1979 interview, when the respondents were 14 to 22 years old. 82% of my sample positively responded to this question. Based on the "yes–no" responses to this question, I created a dummy variable "work anyway".

Table 4 – *continued from previous page*

	23 to 29 years old		OVER 30 years old	
Variable	**Male**	**Female**	**Male**	**Female**
some college	-0.0150	0.0664**	0.0735**	0.0539**
college graduate	0.2312**	0.3143**	0.2944**	0.2507**
some grad studies	0.2913**	0.3272**	0.3517**	0.3048**
hispanic	-0.0157	0.1307**	-0.0365	0.1567**
black	-0.1656**	0.0643*	-0.1139**	0.0258
never married	-0.1693**	-0.0043	-0.2247**	0.0794**
separated	-0.0490	0.0708*	-0.1191**	0.0999**
divorced	-0.1168**	0.0663**	-0.0905**	0.0913**
widowed	0.3271*	0.2073*	0.0221	0.0840
AFQT score	0.0020**	0.0031**	0.0029**	0.0023**
AFQT, Hispanics	0.0003	-0.0000	0.0011*	0.0002
AFQT, blacks	0.0042**	0.0014*	0.0025**	0.0034**
urban residence	0.0190	-0.0219	-0.0227†	-0.0090
SMSA resident	0.1690**	0.0961**	0.1534**	0.1451**
northeast	0.0945**	0.0649**	0.0439**	0.0575**
south	0.0372*	0.0092	-0.0386**	-0.0001
west	0.0600**	0.0458*	0.0187	0.0500**
constant	8.6893**	8.4949**	9.0839**	8.9486**
Selection equation: participation in the labor market				
Dependent variable: Work (positive number of weeks)				
adult male worked	0.2353**		0.3838**	
adult female worked		0.2130**		0.1526**
choice to work		0.0674**		0.1431**
work anyway		0.1168**		0.0820**
children		-0.3831**		-0.1754**
hispanic	-0.2426**	-0.1991**	-0.3992**	-0.0545†
black	-0.4147**	-0.2043**	-0.5206**	-0.0895**
never married	-0.2588**	-0.2793**	-0.6318**	-0.2842**
separated	-0.3850**	-0.1378**	-0.4197**	-0.1463**
divorced	-0.1008	0.0233	-0.3771**	0.1232**
widowed	5.4214**	-0.5544**	-0.1219	-0.2509**
urban residence	0.0827	0.1062**	-0.0462	0.0660*
SMSA resident	-0.1588*	0.0016	0.2439**	0.0430
unemployment rates:				

Significance levels : † : 10% $*$: 5% $**$: 1%

Continued on next page

Table 4 – *continued from previous page*

Variable	23 to 29 years old Male	Female	OVER 30 years old Male	Female
less than 3%	0.7685**	0.8287**	0.1434	0.2081**
3% to 5.9%	0.2736**	0.3689**	0.2706**	0.1437**
6% to 8.9%	0.1829**	0.2735**	0.2152**	0.1179**
9% to 11.9%	0.0959	0.1366**	0.1243	0.0417
constant	1.6460**	0.9639**	1.3816**	0.8262**
athrho	-0.9814**	-1.0309**	-0.7732**	-1.2125**
lnsigma	-0.4495**	-0.2805**	-0.4687**	-0.2184**
λ	-0.481	-0.585	-0.406	-0.673
Std.err. (for λ)	0.020	0.029	0.026	0.014
χ^2_1 (Wald test)	105.135	143.338	47.727	650.581
$p-value$	0.000	0.000	0.000	0.000
N of obs	11826	16795	16004	22567

Significance levels : † : 10% $*$: 5% $**$: 1%

I also make an attempt to correct for the sample selection using panel nature of the data. Among the variety of the estimators proposed by different authors, I will use an extension of the estimator proposed by Wooldridge (1995). As it was mentioned above, I estimate participation equation (3) for each period by a probit model, generate respective period-by-period $\lambda's$, and then estimate earnings equation (2) by instrumental variables, fixed effects, and fixed effects IV, including λ estimated in the first step as an additional regressor (Tables 5 and 6). In cross-sectional and panel IV specifications, corrected for by-period selection, I use all of the exclusion restrictions as above, apart from the unemployment rates, since the latter are used as excluded instruments for the endogenous "weeks unemployed" variable.

In Heckman ML estimations, estimated λ coefficients were statistically significant for all gender-age groups, implying the importance of sample selection into the labor market for all respondents. However, using by-period sample selection correction (as in Tables 5 and 6), estimated λ coefficients are statistically significant in all specification for women, but significant only in some specifications for younger men. Negative λ coefficients suggest that there is some non-random selection into the labor market, and some of the unobserved characteristics that determine one's decision to work, are also the characteristics that negatively affect one's income.

Once sample selection is accounted for and additional measures of labor market experience are used, the estimated coefficients on potential experience become statistically significant for women, hence can be compared with the respective estimates for men. While returns to experience are somewhat smaller for women in the respective age groups, their losses of earnings associated with the interrupted careers are similar or even slightly bigger than those of men.

In addition to comparing estimated returns to experience and depreciation rates due to the lost time, it is also interesting to see how the gender wage differential changes, when model specifications are extended to account for heterogeneous employment histories.

Following Jann (2008), I applied Blinder-Oaxaca decomposition technique (Oaxaca, 1973; Blinder, 1973) to estimate the gender wage differential. I am interested to see whether improving the specification of the wage equation decreases the "unexplained" part of the wage differential. Better accounting for the employment histories should decrease the effects of gender discrimination, as both men and women with weak labor force attachment should be paid less.

I present the results of Blinder-Oaxaca decomposition for the two age groups (Tables 7 and 8) based on the model specifications discussed above. In the benchmark model, I use only potential experience, while expanded model also has gaps in employment included. For the sample selection model, I adjust for the selection for each gender group separately using Heckman maximum-likelihood estimation.

In the sample of younger adults (Table 7), the mean of the log income is between 9.33 and 9.42 for men across different specifications, while it is between 8.88 and 9.06 for women. The gender difference in predicted means of log income is 0.44 for the OLS and IV specifications, and slightly lower 0.35 for the Heckman selection model. In the three-fold decomposition, the income gap is divided into three parts. The first part reflects the mean increase in women's income if they had the same characteristics (endowments) as men. Quite interestingly, in the benchmark specification, where I control for the education and other demographics, if women had the same endowments as men, they would be paid less. A possible explanation for this would be higher educational attainments of women: there are more women with some college education, while there are more men who are high school drop-outs. So if women indeed had the characteristics of men, with slightly lower education for at least some categories, women would earn less. However, in the extended models controlling for employment history, if women had the same (higher) degree of labor force attachment as that of men, there earnings would increase by 8-14%. The second part estimates the change in women's income when applying the men's coefficients to the women's characteristics (hence reflects the effects of gender discrimination). The third part is the interaction term that measures the simultaneous effect of differences in endowments and coefficients. In the two-fold decomposition, income differential is divided into two parts: explained (due to the differences in the characteristics) and unexplained. For the two-fold decomposition, the reference coefficients are obtained from the pooled model, which has an indicator for each gender group. Based on the decomposition results, the unexplained part of the income differential decreased as I introduced additional variables to reflect the gaps in employment, and decreased even more when sample selection

Table 5: IV and FE, corrected for (by-period) selection: MEN ONLY

	23 to 29 years old			OVER 30 years old		
Variable	**IV**	**FE**	**FE IV**	**IV**	**FE**	**FE IV**
Dependent variable: ln *of real labor income, deflated by CPI* $1982-84=100$						
potential experience	0.1453**	0.1768**	0.1650**	0.0512**	0.0717**	0.0642**
potential experience^2	-0.0059**	-0.0075**	-0.0068**	-0.0009**	-0.0012**	-0.0011**
previous gaps	-0.0531**	-0.0081	-0.1057*	-0.0296**	-0.0455**	-0.0780**
weeks out of LF	-0.0420**	-0.0343**	-0.0355**	-0.0331**	-0.0297**	-0.0301**
weeks unemployed	-0.0537**	-0.0297**	-0.0685**	-0.1059**	-0.0247**	-0.0946**
less than HS	-0.1079**	-0.1057	-0.0411	-0.0668**	0.0990*	0.1764*
some college	-0.0255	-0.2167**	-0.3378**	0.0677**	-0.0485	-0.0820
college graduate	0.2171**	0.2440**	0.0813	0.3019**	0.1840**	0.0762
some grad studies	0.2782**	0.4016**	0.1237	0.3596**	0.2535**	0.0794
hispanic	0.0064			-0.0720*		
black	-0.1434**			-0.1161**		
never married	-0.1362**	-0.0360†	-0.0410	-0.2270**	-0.0568*	-0.0555†
separated	-0.0192	0.0507	0.0104	-0.1041**	-0.0403	-0.0184
divorced	-0.1107**	-0.0500	-0.0358	-0.0572*	-0.0462**	-0.0237
widowed		0.0000		0.2817*	0.0000	
AFQT score	0.0014**			0.0028**		
AFQT, Hispanics	0.0000			0.0012*		
AFQT, blacks	0.0039**			0.0019**		
urban residence	0.0165	0.0061	0.0142	-0.0140	-0.0320*	-0.0117
SMSA resident	0.1564**	0.0760*	0.0494	0.1472**	0.0261	-0.0240
northeast	0.0851**	0.0878	0.0213	0.0211	0.0002	-0.0372
south	0.0231	0.0346	-0.0009	-0.0562**	-0.0263	-0.0357
west	0.0386†	0.0963†	0.0598	-0.0031	-0.1053*	-0.0883
λ (male)	-0.6748**	-0.3517*	-0.1420	-0.0999	0.1272	0.1124
constant	8.8433**	8.6771**		9.2374**	9.1368**	
corr(c_i, xb)		0.1605			0.2165	
LM test statistic	79.5147		25.6438	29.7160		16.0016
χ^2_4 *p*-value	0.0000		0.0000	0.0000		0.0030
Hansen *J* statistic	18.1771		3.6645	2.7518		3.0290
χ^2_3 *p*-value	0.0004		0.3000	0.4315		0.3872
N of obs	11266	11266	11231	15101	15103	15060

Significance levels : † : 10% $*$: 5% $**$: 1%

Instrumented variables: weeks unemployed. Excluded instruments: unemployment rates in the region of residence

Table 6: IV and FE, corrected for (by-period) selection: WOMEN ONLY

	23 to 29 years old			OVER 30 years old		
Variable	**IV**	**FE**	**FE IV**	**IV**	**FE**	**FE IV**
Dependent variable: ln *of real labor income, deflated by CPI* $1982-84=100$						
potential experience	0.0802**	0.1319**	0.1124**	0.0090	0.0209**	0.0183**
potential experience^2	-0.0027**	-0.0053**	-0.0039**	0.0005**	0.0003*	0.0004†
previous gaps	-0.0572**	-0.0596**	-0.1281†	-0.0618**	-0.0518**	-0.0634**
weeks out of LF	-0.0465**	-0.0381**	-0.0397**	-0.0407**	-0.0348**	-0.0349**
weeks unemployed	-0.0989**	-0.0354**	-0.0818†	-0.0539	-0.0270**	-0.0619
less than HS	0.0734*	-0.1148	-0.0134	-0.0127	0.0214	0.0329
some college	0.0390*	-0.1491**	-0.2441*	0.0445**	0.0693†	0.0733
college graduate	0.2939**	0.3187**	0.1920	0.2288**	0.2088**	0.1965**
some grad studies	0.2983**	0.3332**	0.1178	0.2735**	0.2002**	0.1791*
hispanic	0.0294			0.1043**		
black	0.0693†			-0.0193		
never married	0.0444*	0.0543*	0.0686*	0.0996**	0.1015**	0.1153**
separated	0.1436**	0.0812*	0.1001*	0.1276**	0.0687**	0.0704*
divorced	0.1479**	0.0935**	0.1246**	0.1627**	0.0596**	0.0668**
widowed	0.0880	0.3706**	0.3044*	0.0519	0.0174	-0.0353
AFQT score	0.0019**			0.0009*		
AFQT, Hispanics	0.0009			0.0009		
AFQT, blacks	0.0012			0.0039**		
urban residence	-0.0063	0.0356	0.0385	0.0094	-0.0180	-0.0239
SMSA resident	0.0927**	0.0788*	0.0442	0.1522**	0.0275	0.0230
northeast	0.0548*	-0.0742	-0.0886	0.0549**	-0.0448	-0.0754
south	0.0186	-0.0022	0.0042	0.0109	-0.1349**	-0.1525*
west	0.0498*	-0.0611	-0.0527	0.0315	-0.1474**	-0.1493*
λ (female)	-0.2592**	-0.4492**	-0.4688**	-0.3856**	-0.6508**	-0.6363**
constant	8.7313**	8.5919**		9.0540**	9.2367**	
corr(c_i, xb)		0.1833			0.1797	
LM test statistic	36.0974		6.7931	13.0817		7.8826
χ^2_4 *p*-value	0.0000		0.1472	0.0109		0.0960
Hansen *J* statistic	12.0890		4.5708	11.0142		6.4810
χ^2_3 *p*-value	0.0071		0.2061	0.0116		0.0904
N of obs	13548	13548	13397	18473	18477	18370

Significance levels : † : 10% $*$: 5% $**$: 1%

Instrumented variables: weeks unemployed. Excluded instruments: unemployment rates in the region of residence

was accounted for. The decomposition results are similar for the older adults (Table 8), but gender income differences become larger, and so are the unexplained part of this differential. Overall, accounting for the heterogeneous employment histories, intermittent careers, and sample selection into the labor market allows to reduce the unexplained part of the wage differential by around 50% in some specifications.

Table 7: Blinder-Oaxaca decomposition results: Respondents aged 23 to 29

	Benchmark OLS	IV with selection	Heckman, ML
Differential			
Prediction (males)	9.3296**	9.3817**	9.4155**
Prediction (females)	8.8843**	8.9362**	9.0640**
Difference	0.4454**	0.4455**	0.3515**
Three-fold decomposition			
Endowments	-0.0371**	0.0841**	0.1370**
Coefficients	0.4845**	0.3615**	0.2304**
Interaction	-0.0021	-0.0000	-0.0159**
N of obs (males)	13698	10626	11320
N of obs (females)	14535	11181	13548
Significance levels: † : 10% ∗ : 5% ∗∗ : 1%			

6 Conclusions

Adequate measures of employment histories and acquired work experience are essential for the range of empirical applications, especially those relying on the earnings equation. In this paper I augment the traditional Mincerian measure of potential experience with additional variables to reflect time not working. The estimations are performed separately for men and women of different age groups, since employment histories and earnings patterns are different for each subsample.

I estimate about 5% returns to experience among men over 30 years old, and these results are robust across all specifications I use: benchmark OLS, extended OLS, IV specifications and selection corrected models. On the contrary, for younger men potential experience measure used alone (as in the benchmark OLS) overestimates the returns to experience by around 35%. Accounting for the time spent not working, for the endogeneity of the unemployment spells and for the unobserved heterogeneity, reduces estimated returns to experience from 23% to 14–17%. Similarly, for the younger women, potential experience in the traditional specification overestimates the returns to experience as well, and the estimated coefficient goes down from 17% to 6–13% in the alternative specifications. For women over 30, estimated returns to

Table 8: Blinder-Oaxaca decomposition results: Respondents over 30 years of age

	Benchmark OLS	IV with selection	Heckman, ML
Differential			
Prediction (males)	9.8120**	9.8622**	9.8953**
Prediction (females)	9.2493**	9.2900**	9.4648**
Difference	0.5627**	0.5722**	0.4305**
Three-fold decomposition			
Endowments	-0.0227**	0.1707**	0.1648**
Coefficients	0.5514**	0.4601**	0.2526**
Interaction	0.0339**	-0.0586*	0.0131*
N of obs (males)	18088	14203	15153
N of obs (females)	19807	15258	18473

Significance levels: † : 10% $*$: 5% $**$: 1%

experience became statistically significant only in the models adjusted for non-random selection into the labor market. Returns to experience among older women are noticeably smaller (only around 2%) as compared to the men in the same age group.

In all of the specifications, I separate time not working into more old employment gaps and more recent unemployment and out of labor force periods. The estimated coefficients on all types of non-working time are negative and statistically significant. Previous, more remote gaps have smaller negative effect on the earnings than most recent non-working spells. The effects of previous gaps are more prominent for the younger adults (as compared to those over 30), as well as for women (as compared to men in the respective age group) in most of the specifications. Accounting for the endogeneity of the weeks unemployed and instrumenting this variable with the unemployment rates in the region of residence, increases the negative effect of the estimated coefficients on weeks unemployed.

Using more comprehensive measures of employment histories affects not only the estimated coefficients on the returns to experience, but also the estimated returns to education and the coefficients on the race/ethnicity variables. While in the traditional (benchmark) specification high earnings are explained by higher levels of education, in the augmented specifications with detailed employment histories measures, higher yearly earnings are explained by a combination of higher levels of educations and lower unemployment gaps among highly educated individuals. The negative effects of the race/ethnicity variables decrease in the specifications using more comprehensive experience (and gaps in experience) measures. This suggests that the effect of race discrimination in the labor market decreases when actually acquired experience, as opposed to potential labor force attachment, is accounted for.

Appendix

Reasons for non-interview

Every year some individuals did not respond to the interview for the following reasons:

- parent and/or youth refusal
- unable to locate family unit and/or youth
- deceased
- very difficult case, do not refield

References

James W. Albrecht, Per-Anders Edin, Marianne Sundström, and Susan B. Vroman. Career interruptions and subsequent earnings: A reexamination using Swedish data. *The Journal of Human Resources*, 34(2):294–311, Spring 1999.

Sigal Alon, Debra Donahoe, and Marta Tienda. The effects of early work experience on young women's labor force attachment. *Social Forces*, 79(3):1005–1034, March 2001.

Charles L. Baum. The effect of work interruptions on women's wages. *Labour*, 16(1):1–36, 2002.

Christian Belzil and Jörgen Hansen. Unobserved ability and the return to schooling. *Econometrica*, 70(5):2075–2091, September 2002.

A. S. Blinder. Wage discrimination: Reduced form and structural estimates. *The Journal of Human Resources*, 8:436–455, 1973.

G. Chamberlain. *Panel data*, volume 2, chapter in *Handbook of Econometrics*, pages 1247–1318. Amsterdam: North Holland, 1984.

Christian Dustmann and Maria Engracia Rochina-Barrachina. Selection correction in panel data models: An application to the estimation of females' wage equations. *Econometrics Journal*, 10:263–293, 2007.

Rosella Gardecki and David Neumark. Order from chaos? The effects of early labor market experiences on adult labor market outcomes. *Industrial and Labor Relations Review*, 51 (2):299–322, January 1998.

James J. Heckman. Sample selection bias as a specification error. *Econometrica*, 47(1):153–162, January 1979.

Julie L. Hotchkiss and M. Melinda Pitts. Female labour force intermittency and current earnings: switching regression model with unknown sample selection. *Applied Economics*, 37: 545–560, 2005.

Ben Jann. The Blinder-Oaxaca decomposition for linear regression models. *The Stata Journal*, 8(4):453–479, 2008.

Kristen Keith and Abagail McWilliams. The wage effects of cumulative mobility. *Industrial and labor relations review*, 49(1):121–137, October 1995.

Moon-Kak Kim and Solomon Polachek. Panel estimates of male-femaler earnings functions. *The Journal of Human Resources*, 29(2, Special Issue: Women's Work, Wages and Well-Being):406–428, Spring 1994.

Audrey Light and Manuelita Ureta. Early-career work experience and gender wage differentials. *Journal of Labor Economics*, 13(1):121–154, January 1995.

Carole F. Miller. Actual experience, potential experience or age, and labor force participation by married women. *American Economic Journal*, 21(4):60–66, December 1993.

Jacob Mincer. *Schooling, Experience, and Earnings*. New York : National Bureau of Economic Research, 1974.

Jacob Mincer. Investment in human capital and personal income distribution. *The Journal of Political Economy*, 66(4):281–302, August 1958.

Jacob Mincer and Solomon Polachek. Family investments in human capital: Earnings of women. *The Journal of Political Economy*, 82(2, Part 2: Marriage, Family Human Capital, and Fertility):S76–S108, Mar.-Apr. 1974.

Ronald Oaxaca. Male-female wage differentials in urban labor markets. *International Economic Review*, 14(3):693–709, 1973.

Wayne Simpson. Intermittent work activity and earnings. *Applied Economics*, 32:1777–1786, 2000.

Christy Spivey. Time off at what price? The effects of career interruptions on earnings. *Industrial and Labor Relations Review*, 59(1):119–140, October 2005.

Joanna K. Swaffield. Estimates of the impact of labour market attachment and attitudes on the female wage. *The Manchester School*, 75(3):349–371, June 2007.

Jeffrey M. Wooldridge. *Econometric analysis of cross section and panel data*. The MIT Press, 2002.

Jeffrey M. Wooldridge. Selection corrections for panel data models under conditional mean independence assumptions. *Journal of Econometrics*, 68:115–132, 1995.