THREE ESSAYS ON LABOR MARKET OUTCOMES

by

Anila Prakash

A Dissertation Submitted to the Faculty of the

DEPARTMENT OF ECONOMICS

In Partial Fulfillment of the Requirements
For the Degree of

DOCTOR OF PHILOSOPHY

In the Graduate College

THE UNIVERSITY OF ARIZONA

2015

THE UNIVERSITY OF ARIZONA
GRADUATE COLLEGE

As members of the Dissertation Committee, we certify that we have read the dissertation prepared by Anila Prakash titled Three Essays on Labor Market Outcomes and recommend that it be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy .

Date: 27 April 2015

Ronald L. Oaxaca

Date: 27 April 2015

Tiemen Woutersen

Date: 27 April 2015

Price V. Fishback

Date: 27 April 2015

Jessamyn Schaller

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copies of the dissertation to the Graduate College.

I hereby certify that I have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.

Date: 27 April 2015

Dissertation Director: Ronald L. Oaxaca

STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at the University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this dissertation are allowable without special permission, provided that an accurate acknowledgment of the source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the head of the major department or the Dean of the Graduate College when in his or her judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

SIGNED: Anila Prakash

ACKNOWLEDGEMENTS

This dissertation would not have been possible without the guidance and encouragement of my advisors: Ronald L. Oaxaca, Tiemen Woutersen, Price V. Fishback and Jessamyn Schaller. Professor Oaxaca for always encouraging me even when all I had were half-researched ideas. Professor Woutersen for his optimism and his unstinting belief in my abilities. Professor Fishback for the understanding and support over my trials from the very first year. Professor Schaller for the thoughtful remarks and crucial improvements. The research would not be where it is if not for their support. I would also like to thank all my classmates, especially Xing (Michelle) Liu, Hoa Nguyen, Jianjing Lin and Mariana Zerpa. Graduate life becomes a lot easier when you can share it.

Finally, I thank my parents, my brother and Aditi. They always support my decisions, even when they do not understand them. They are the best gift I have.

DEDICATION

To Rishabh for loving me as I meander through life

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

The three chapters in this dissertation look at different aspects of the labor market and its players. The first chapter estimates the impact of using the internet for job search on job match quality. Using both the semi-parametric Meyer (1990) model and the non-parametric Hausman Woutersen (2014) hazard model, the paper finds that exit rate from employment is at least 28% lower when internet is used as a job search tool. The second chapter looks at the effect of past unemployment on future wages. It is believed that employers may use past unemployment as a signal of low productivity. In this situation workers with a history of unemployment may receive lower wages. The paper uses the Machado Mata (2005) quantile decomposition technique to decompose the wage difference into differences due to characteristics and differences due to rewards. Results indicate that workers with an unemployment spell of more than three months receive at least 12% lower wages and that more than 40% of this wage difference can be attributed to the lower rewards received by the previously unemployed.. The last chapter[1] focuses on human capital formation and looks at some of the reasons behind the low levels of schooling India. Using the Indian Household Development Survey (2005), the paper finds that income continues to be an important factor behind the low level of primary school enrollment. On average, poor students have at least 3% lower enrollment rates, when compared to similar skilled non-poor students.

[1]This work is in collaboration with Rishabh Sinha (Arizona State University)

CHAPTER 1

Does Internet Job Search Result in Better Matches?

1.1 Introduction

The internet has significantly changed the way in which information is disseminated. Information is now cheaper, up-to-date and much more accessible. This diffusion of information has had some major impacts on the labor market. It allows employees to have greater flexibility in working hours and location and passively/actively be on the search for better opportunities. The benefit is not one-sided. Employers can now access much more information on prospective employees enabling a better selection and use geographic differences in factor costs to their advantage. Autor (2001) discusses three important ways in which the internet affects the labor market- how workers and firms search for one another, how labor services are delivered and how local markets shape labor demand. Arguably, important changes would be on the way search happens and employer-employee matches are made. The decrease in cost of search and increase in available information will make the matching process much more efficient in terms of both speed and quality. It will not only help workers find jobs more suited to their skills but also help employers identify employees more suited to their organization.

To understand how the internet makes job search more efficient, it is important to look at the different ways in which it has changed the market. One of the most significant changes has come with the advent of the job-posting boards. A worker can apply directly for vacancies on a company web page or through a job search website (e.g. Monster, Career Builder). Job portals like Monster also offer additional services for a fee. For someone who is actively looking for jobs, using these services ensures that they are differentiated from other users. For instance, the Resume Distribution service by Monster sends resumes directly to recruiter's who are looking to hire. This service not only accelerates the job search process but also helps employers differentiate between active and passive searchers. LinkedIn is an alternate platform for professional networking where workers maintain informal resumes and use it to find jobs, people and new business opportunities. This forum functions more like a social networking site where managers and colleagues provide recommendations and comments for workers that boosts their profile. This assessment is a source of additional information for an employer. Employers can also use MonsterTrak for institutionally targeted job postings. For example, employers can pay to have job postings sent only to graduates from Harvard or other designated pools. For local opportunities, Craigslist works like the yellow pages and provides information on jobs available. Kroft and Pope (2014) use data from Craigslist and find that the website's local expansion has to some degree crowded out newspaper advertisement. Social networking sites like Twitter and Facebook also help job searchers connect to recruiters. In addition to these job postings, there is also a huge amount of '*insider*' information available. Websites like Vault or Glassdoor allow current and former employees to anonymously discuss and provide information about their bosses and the culture at work. Salary.com provides detailed information about salaries and work environments, including salary ranges for various positions at specific companies.

Given this abundance of information it is no surprise that more and more workers and employers are now using the internet. 24.2 million people visited job-search sites in January, 2012 which was 27% more when compared to December, 2011.[1] Focusing on job seekers, in 1998 only 15% of the unemployed workers used the internet as a job search tool

[1]Numbers obtained from ComScore media release January 2012.

and approximately 22% of them had home internet access.[2] In 2008, this number has increased significantly with 74% of the unemployed job-seekers using the internet to look for jobs and 61% of them having home internet access.[3] Similarly, more and more employers are now using the internet, not only for collecting resumes but also to better scrutinize a prospective employee. According to the 2007 report of the U.S. based Society for Human Resource Management, on average, private and public sector organizations attributed 44% of their new hires the previous year to e-recruiting, around 50% of the firms used online search engines to review/collect information on a potential job candidate and roughly 20% of these firms reported eliminating a candidate based on the information discovered.

Given the enormous possibilities enabled by the internet, researchers have tried to evaluate its impact on labor market outcomes. Kuhn and Skuterud (2004), one of the first papers to measure the impact of the internet on unemployment durations, used 1998-2000 data from the Computer and Internet Supplements published by Current Population Survey (CPS).[4] Their results indicated that either internet job search (IJS) was ineffective in reducing unemployment durations or that IJS workers[5] were negatively selected on unobservables. Fountain (2005) finds that IJS workers have only a very small advantage in obtaining a job over non-internet users. Czernich (2011) investigates the effect of the spread of broadband internet on the unemployment rate and the results indicate absence of any causal link between the two. Crandall *et al.*(2007) find that employment in manufacturing, services and private nonfarm sectors is positively related to broadband penetration. Atasoy (2013) shows that broadband expansion lead to a 1.8% increase in the employment rate, with larger effects in rural and isolated areas. Dettling (2013) finds that internet increases the proba-

[2] Kuhn & Skuterud (2004) using data from the 1998 CPS Computer and Internet Supplements.

[3] Kuhn & Mansour (2014) using the 2008 NLSY97 data.

[4] CPS is a monthly labor force survey providing estimates of the economic status and activities of the population of the United States. Questionnaires on Internet and Computer Use were administered as a supplement.

[5] Workers who use the internet as a job search tool are referred to as IJS workers

bility of labor force participation for married women. Stevenson (2006) finds that internet use leads to an increase in flows from employment to employment and also greater wage growth when changing jobs. Kuhn and Mansour (2014) revisit the analysis by Kuhn and Skuterud (2004). The authors use National Longitudinal Survey of Youth 1997 (NLSY97) data to re-evaluate the claim that the internet has no impact on unemployment duration. Their results show that the internet leads to a 25% decrease in unemployment durations.

The focus of my paper is primarily on the impact of internet on the quality of employer-employee matches. While it is difficult to empirically quantify match quality, I believe that a good match can be identified from a lengthy duration. A worker who has found a job well suited to his abilities is less likely to shift to a new job. Similarly, firms are less likely to fire workers who are a better fit in the organization. Thus improved matching would cause one or both of these effects simultaneously (less quitting and less firing) and imply longer job duration for a worker. Past research also suggests the same effect. Jovanovic (1979) argues that the quality of a match is not known but must be experienced. Akerloff, Rose and Yellen (1988) provide evidence that match aspects of a job negatively affect the probability of quitting, implying that a good match would lead to lower quits and hence a higher duration. Assuming that 'good matches endure' Bowlus (1995) measures the quality of job matches across business cycles by looking at tenure. Similarly a number of papers focusing on the effect of unemployment insurance on post-unemployment outcomes use duration as a proxy for match quality (Centeno (2004), Centeno and Novo (2006)). Drawing from this vast body of related research, the quality of matches in this paper is identified using job tenure.

There has not been a lot of research on the impact of the internet on job match quality. Krueger (2000) suggests that given the low cost of posting jobs online and the speed and ease with which a worker can apply for different jobs, the internet should lead to improved match quality. However, he does not test this result empirically. Mang (2012) focuses on the impact of the internet on job match quality. The author uses German Socio-Economic Panel (SOEP) data to regress the use of the internet on a variety of matching outcomes (satisfaction, commute time, working hours and job security). His results indicate that job changers who found their new job online are better matched than their counterparts who found their new job through newspapers, friends, job agencies and other channels. Hadass (2004) investigates the impact of the spread of online recruiting on the matching of workers and firms. Using data from a US-based multinational manufacturing firm for the period 1995-2002, the paper finds that internet hires are not significantly different from print advertising hires but have lower duration when compared to hires made through employee referrals. While my paper also attempts to measure the impact of the internet on job match quality, the methodology and data used are very different.

In this paper I use the NLSY97 data to estimate the impact of the internet on job match quality, using duration as a proxy. In 2008 the survey included questions on internet usage, which help identify if a worker used the internet as a job search tool. The exit rate estimation is done using the Meyer (1990) proportional hazards model. The indicator for a worker who used the internet for job search is used as an explanatory variable and helps identify the impact of online job search on duration. Different specifications of the model are used to test for the robustness of the results. Across all models internet usage has a negative and statistically significant impact on the exit rate from employment. A conservative estimate suggests that internet search reduces the exit rate by 28%. In addition to the Meyer (1990) hazard model, I also employ the Hausman and Woutersen (2014) proportional hazard model. This model builds on the existing hazard model literature, by removing the distributional assumption on the heterogeneity component. The model conditions out the heterogeneity distribution by computing the probability of a worker surviving period t, compared to the original model which computes the probability of a worker surviving up to period t. Methodological details for both models are presented in Section 1.3.1 and

1.3.2. Even in this non-parametric formulation, the model estimates 18% lower exit rates for workers who use the internet for job search. Detailed results obtained from both models are reported in Section 1.4.

Any analysis that tries to examine a causal link between the use of the internet and the labor market must control for selection bias. To control for this, a number of robustness tests are conducted in this paper. Armed Services Vocational Aptitude Battery (ASVAB) scores are included to control for unobserved ability across workers and variables are added to control for differences in search intensity, in the hazard model. Other tests include redefining the IJS worker, checking for the effect of using newspaper ads for job search and dividing workers on the basis of their skill.[6] In addition, the paper also includes an instrumental-variable (IV) regression approach to control for endogeneity. Following from Choi (2011), variation in the adoption of the internet across industries is used to capture the exogenous variation in the adoption of the internet across workers. Across all these different estimation strategies the negative effect of the internet on exit rate persists. A more detailed discussion regarding endogeneity concerns is provided in Section 1.4.3 and 1.4.4. While it is impossible to perfectly control for endogeneity, the robustness tests indicate strong results in favor of the effect of internet usage.

1.2 Data and Summary Statistics

The data for this analysis is taken from the National Longitudinal Survey of Youth (NLSY97). The NLSY97 is a nationally representative sample of 8,984 youths who were 12 to 16 years old as of December 31, 1996. In wave 12 (2008-2009), the respondents were 24 -28 years of age and asked questions related to the usage and access of the internet. With

[6] Two groups of workers were created on the basis of their education. One group of workers had at least received college education, while the other group had not.

regard to job search, the respondents had to list which of the twelve job search activities[7] they engaged in and which of these methods involved use of the internet. These responses provide information on whether or not internet was used as a job search tool and was used to construct the indicator variable for an IJS worker.

The dataset includes only those individuals who were employed as of the 2008-2009 (Round 12), had started their current job after January 1, 2006, were not in the military and had been employed for at least 13 weeks (as of the interview date). I then obtained the end date of their current job using all surveys after round 12 until the latest survey currently available (Round 15 - 2010/2011) so that employment duration could be approximated. Missing information for exact job start and end dates prevents calculation of the exact duration. Given this restriction it was only possible to calculate job duration at a monthly level. The final dataset consisted of 2,922 respondents, of which 47.3% used the internet as a job search tool and 39.6% of the respondents are still employed at the same job leading to a right censored data set.

Table 1.1 provides the means of the variables used in the analysis (calculated separately for IJS and Non-IJS workers) for the final dataset. Some of the means suggest that IJS workers have observable characteristics that are associated with better match quality. For example, the IJS workers have slightly higher hourly pay with lower variance[8], a higher average job duration, higher representation of workers with at least a college degree, higher average ASVAB scores and are more likely to reside in urban areas (and hence are not limited by the types of jobs available). On the other hand more than 75% of the non-IJS workers have not completed college, suggesting a link between education and use of the internet. Also it appears that IJS workers, on average, search for jobs more actively. Search intensity is

[7] The job search methods included contacting employers, public/private employment agencies, friends, school/university employment center, unions, sending out resumes, placing/looking at ads, attending job training courses and other active/passive methods.

[8] The difference in the hourly wages is not statistically significant.

created as a proxy for the intensity of job search across workers and is a sum of all the methods (12 in total) a worker used to look for the current job. On average, IJS workers were using more than two search methods which is higher than that of non-IJS workers.[9] The workers are roughly equally split across race, marital status and the four regions but there is a much higher percentage of female IJS workers. Also as expected, a much higher fraction of the IJS workers have access to the internet.

Table 1.2 includes average duration across different sub-groups for the IJS and non-IJS workers. On average, job duration of workers who used the internet to search for jobs is larger by three months. This trend holds even when we compare across the gender groups. While females have a lower average duration compared to males, both male and female IJS workers have longer average duration. Looking at the difference in average duration by education, we can see that the internet contributes to a higher average job duration across all education levels, with workers in lower education categories also benefitting. This suggests that the positive effect of internet search may not be restricted to workers of a specific skill set (assuming education as a proxy for skill). Differences in duration by wage indicate an interesting result. At low wages (below \$20 per hour), IJS workers have marginally higher average job duration. However, for IJS workers earning more than \$20 per hour the average job duration is significantly larger. This result indicates that IJS workers earning the highest wages may be deriving the maximum benefit. In the context of occupation and industry, internet job search has a positive relation with duration across almost all categories. However there are some categories with lower duration and very few IJS workers, e.g., Agriculture, Utilities and Construction. This lack of IJS workers in certain industries/occupations suggests the presence of selection effects. It maybe the case that not all occupations/industries use the internet for job postings and this may bias the results. Looking at averages across race we can see that people across all races benefit

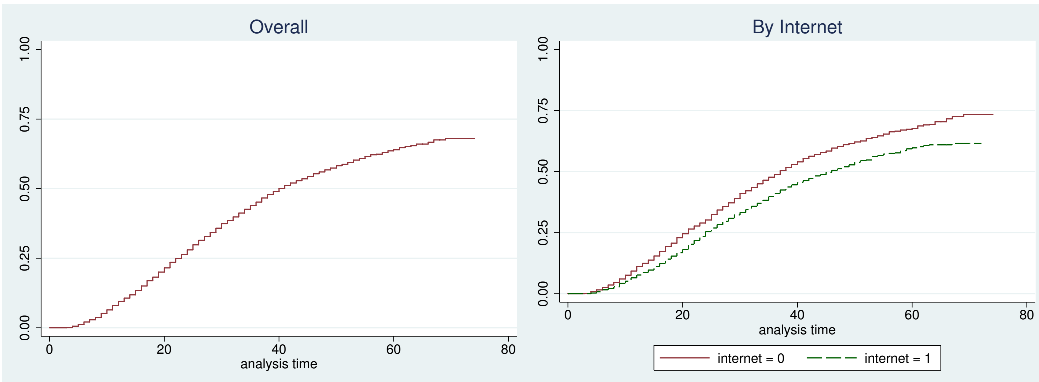
[9] The difference in means, across IJS and Non-IJS workers, for job duration, college education, ASVAB scores and search intensity are statistically significant.

from using the internet. However African-Americans seem to benefit the least. Similarly across the four geographic regions IJS workers always have longer durations.

Next I look at some details about the IJS workers (Table 1.3) from this dataset. Of the 47.3% of workers who used the internet as a search tool 93% currently have internet access (at work, home, cafe etc.), 70% use the internet several times a day and 83% have home internet access. More than 80% of them use the internet to read the news and for online banking and almost everyone uses email. In comparison, only 66% of the non-IJS workers have home internet access and less than 60% of them use the internet daily. Also, while 80% of the workers use email, very few of the workers use the internet for other purposes. The most relevant difference between the two is reflected in availability of the internet at work. Only 45% of the non-IJS workers have internet access at work, compared to more than 70% of the IJS workers. It is important to note that these numbers are not an accurate representation of the internet access/usage at the time the worker started their current job. The only conclusion we can safely draw is that compared to non-IJS workers, the IJS workers use the internet more and that a larger fraction of them have internet access.

I next plot the Kaplan-Meier (KM) hazard estimates for the entire sample and across some categorizations. The KM estimate is a nonparametric method used to estimate the survival probability (P_t) for each of the t time periods. The graphs below plot the probability of a person exiting a job at period t, i.e., the hazard rate (1-P_t). This estimate has some drawbacks primarily because it assumes all individuals to be homogeneous and can be unreliable toward the end of the analysis period where less data is available. However, it provides a better comparison than averages as it accounts for the right-censored dataset. From the graphs below, it can be seen that the hazard rate is positively linked to duration. More importantly, the hazard plot for the internet users lies below that of the non-IJS workers, suggesting that the internet helps workers find the right jobs and hence reduces the probability of job exits. At each point in time, the probability of quitting/exiting a job is much

Figure 1.1: KM Hazard Rate



(a) Overall (b) IJS vs. Non- IJS Workers

lesser for an IJS worker. Note that this positive effect of the internet persists even after accounting for the right-censored data.

The appendix includes KM Hazard graphs plotted separately for IJS and non-IJS workers across sub-groups of education, gender, race and region. Across all categories the hazard rates are lower for the IJS workers.

1.3 Model

The KM metric suggests that internet has a negative effect on the exit rate from employment. However it would be premature to draw this conclusion as the metric has several drawbacks. Most importantly, the homogeneity assumption made by the KM metric makes it an unreliable estimate. To correctly model the impact of the internet, a much more flexible modeling structure that can adjust estimates for all influential covariates, is required. The econometric methodology followed in this paper consists of estimating a proportional hazard model, where the hazard is the exit rate from employment. There are two approaches used in the paper. The first is the Prentice and Gloeckler (1978) extension proposed by

Meyer (1990) and the second is the Hausman and Woutersen (2014) model which relaxes the heterogeneity distribution assumption made by the Meyer (1990) model.

1.3.1 Model 1: Meyer (1990)

The estimation methodology used in this paper is the Prentice and Gloeckler (1978) hazard model extension as proposed by Meyer (1990). The presence of time-varying covariates and a censored dataset make duration modeling an ideal choice. This model is selected as it does not require continuous time data and can account for unobserved heterogeneity across individuals. The proportional hazard function represents the probability of losing a job, conditional on being employed until period t. If T_i is the length of a worker's employment spell, then the hazard (λ_{it}) of person i at time t takes the following form:

$$\lambda_{it} = \lambda_0(t)\exp(Z'_{it}\beta)$$

where $\lambda_0(t)$ is the baseline hazard common to all units of population, Z_{it} is the covariate vector consisting of both time dependent and time invariant variables and β is vector of unknown parameters to be estimated. The vector Z includes both job (industry, occupation, hourly pay) and individual (marital status, location, gender, race and schooling) characteristics. The time varying covariate included is the state unemployment rate.[10] In this dataset the exact job duration is not known, but the month in which a job ended is available. As the durations observed are discrete, it is assumed that time-varying covariates only vary across intervals (monthly).[11] The probability that the employment lasts until $t + 1$ given that it has lasted until t, as a function of the hazard is shown below:

$$\begin{aligned} P[T_i \geq t + 1 | T_i \geq t] &= \exp\Big[- \int_t^{t+1} \lambda_i(u) du \Big] \\ &= \exp[-\exp(z_i(t)'\beta) + \gamma(t))] \end{aligned}$$

[10] Monthly unemployment rate is obtained from the Bureau of Labor Statistics website.

[11] This is a standard assumption, essential when using discrete data.

where $\gamma(t) = \ln\left(\int_t^{t+1} \lambda_0(u)du\right)$.

The baseline hazard ($\lambda_0(t)$) can be estimated by either assuming a parametric distribution or relaxing the distributional assumption. Meyer (1990) observes that an incorrect specification of the baseline model leads to inconsistent estimates while a correct specification provides only a small efficiency increase. He suggests that in the presence of time-varying covariates, a non-parametric baseline ensures consistency of estimates. In this paper both a parametric (Weibull form) and a non-parametric baseline model are estimated. For the non-parametric baseline model, dummy variables corresponding to each duration interval are created (74 dummies as the maximum duration observed is 74 months). As the number of people whose job ended in the first 5 months and those whose job ended after 55 months was very low, a piecewise constant baseline hazard function is estimated, assuming a constant baseline hazard for people whose job ended in the first 5 months and for people who were employed beyond 55 months.[12] Different combinations of the piecewise hazard were tried and it had no effect on the results. This model gives the following log-likelihood function (below), which is then estimated in STATA using MLE maximization:

$$L(\gamma, \beta) = \sum_{i=1}^{N} \Big[\delta_i \log[1 - \exp\{-\exp[\gamma(k_i) + z_i(k_i)'\beta]\}] \\ - \sum_{t=1}^{k_i-1} \exp[\gamma(t) + z_i(t)'\beta]\Big]$$

where, $\gamma = [\gamma(0), \gamma(1)..\gamma(T-1)]'$, C_i= censoring time, δ_i=1 if $T_i \leq C_i$ and k_i=min(T_i, C_i).

Further, by introducing an additional component, the proportional hazard specification can be adjusted to control for unobserved heterogeneity across individuals. This component

[12]The non-parametric model requires observations for each duration interval. The piecewise constant hazard is assumed to prevent loss of data.

summarizes the impact of omitted variables on the hazard rate when missing regressors may be intrinsically unobservable or unobserved in the data available. Estimating a model without accounting for heterogeneity will lead to an under-estimation of the duration dependence parameter (Lancaster (1990)). After augmenting for unobserved heterogeneity, the proportional hazard model takes the following form:

$$\lambda_{it} = \nu_i \lambda_0(t) \exp(Z'_{it}\beta)$$

where ν_i is introduced to account for unobserved heterogeneity. To implement this approach, the following assumptions are made: (i) ν has a gamma distribution, (ii) it is independent of the observed covariates (Z_{it}) and (iii) that it enters the model in a multiplicative form. The distributional assumption is made so that the unobserved component can be integrated out. The gamma distribution is selected for tractability purposes.[13] It is important to note that this model is non-parametrically identified and it's identification does not rely on either the distributional assumption on the heterogeneity distribution or the functional form assumptions on the baseline hazard (Ridder and Woutersen (2003)). While Heckman and Singer (1984) demonstrate that for a given parametric baseline hazard function results can be very sensitive to the choice of the parametric form for the frailty distribution, Meyer (1990) believes otherwise. In his paper Meyer suggests that once the baseline model is non-parametric the choice of heterogeneity distribution may be unimportant. The following log-likelihood is estimated assuming a gamma distribution for ν distributed with mean 1 and variance σ^2:

[13] It is easy to derive the closed form expressions of unconditional survival, cumulative density and hazard function for the gamma distribution.

$$L(\gamma,\beta,\mu) = \sum_{i=1}^{N} \log\Big\{\Big[1+\sigma^2.\sum_{t=0}^{k_t-1}\exp\{\gamma(t)+z_i(t)'\beta\}\Big]^{-\sigma^{-2}} - \delta_i\Big[1+\sigma^2.\sum_{t=0}^{k_t}\exp\{\gamma(t)+z_i(t)'\beta\}\Big]^{-\sigma^{-2}}\Big\}$$

Another issue for which I need to account is the case where one individual has multiple employment spells. The unobserved heterogeneity component makes it difficult to rule out correlation across observations if multiple spells of the same individual are introduced. To control for this, the analysis is restricted to single spell data. If an individual re-enters with a new job, the information is excluded. Additionally, the model also assumes that true duration ($D(t_i)$) is independent of both the starting time a_i (the date on which an individual started the job) and censoring time c_i (the date after which we no longer observe the individual).

$$D(t_i^*|z_i,a_i,c_i) = D(t_i^*|z_i)$$

As the censoring time is the same for all individuals, it is easy to see that true duration is independent of the censoring time. However different starting dates make it is difficult to assume that it is independent of the starting date, as there might be some seasonal impact on duration. To ensure that the above assumption holds, dummy variables for the different starting dates are included as controls. The result section (Section 1.4.1) discusses the results obtained with and without a parametric baseline, and with and without the gamma distributed heterogeneity component.

1.3.2 Model 2: Hausman and Woutersen (2014)

While the Meyer (1990) proportional hazard model accounts for unobserved heterogeneity across individuals, it makes a few strong assumptions on its structure. Namely, v has to follow a pre-defined distribution (gamma in this paper). Hausman and Woutersen (2014) present simulations and provide a theoretical result which shows that a non-parametric estimation of the baseline hazard with gamma heterogeneity yields inconsistent estimates if the true distribution is not gamma. They further adapt the proportional hazard model such that it can still account for unobserved heterogeneity without making any parametric specification or nonparametric estimation. Horowitz (1999) was the first paper to estimate the baseline hazard function and the distribution of unobserved heterogeneity non-parametrically. However his approach requires all the regressors to be time-invariant. Moreover, the regression coefficient estimators have a slow rate of convergence and were not $N^{-1/2}$ consistent. The integrated baseline hazard and regressor parameters following the Hausman and Woutersen (2014) methodology converge at the regular rate of $N^{-1/2}$ where N is the sample size. Further this model allows for discrete measurement of durations and time-varying regressors.

As it is empirically difficult to recover the true distribution of the unobserved heterogeneity, estimators that rely on the estimation of its distribution may be unreliable. Hence, Hausman and Woutersen (2014) intuitively condition out the heterogeneity distribution and avoid any estimation of its distribution. Their new estimator is related to Han's (1987) estimator but contrary to Han (1987) their model can handle time-varying regressors. In particular, the model gives the following minimization function:

$$Q(\beta, \delta) = \frac{1}{N(N-1)} \sum_i \sum_j \sum_{l=1}^{K} \sum_{k=1}^{K} [1\{T_i \geq l\} - 1\{T_j \geq k\}] 1\{Z_i(l; \beta, \delta) < Z_j(k; \beta, \delta)\},$$

where, for each period t, $Z_i(l,\beta,\delta) = \sum_{t=1}^{l} \exp\{z_i(t)\beta + \delta_t; \delta_{0,t} = \ln\left(\int_{t-1}^{t} \lambda(t)dt\right)$ and K is the maximum observed duration.

Thus $Z_i(l,\beta,\delta)$ is the index for the l^{th} period. In particular the above function compares two different individuals by taking into account the outcome in each period through the parameters for the baseline hazard (δ). The probability that individual i survives period l is larger than the probability that individual j survives period k if and only if $Z_i(l,\beta_0,\delta_0) \leq Z_j(k,\beta_0,\delta_0)$. The outcomes of individual i and j along with the probabilities $(Z_i(l), Z_j(k))$ yield an objective function that is able to identify both β and δ. However, the function above contains a double sum and hence is computationally cumbersome. To reduce the number of computational operations a rank operator d_k is introduced; where $d_k = 1\{T \geq k\}$ for vector T of length N (number of people). Vector d can now be created by stacking d_k for all $k = 1, 2, ...K$, giving a vector $NK \times 1$. Similarly Z can be constructed by stacking Z_k for all $k = 1, 2, ...K$. Now both d and Z are of size $NK \times 1$ and $Q(\beta,\delta)$ can be re-written as below:

$$Q(\beta,\delta) = \frac{1}{N(N-1)} \sum_{k=1}^{NK} d(k)\Big[2 * Rank\{Z(k)\} - NK\Big]$$

The computational burden to calculate this above simplified function is now reduced to N ln N. Since the objective function is non-smooth, pattern-search methods available in MATLAB are used to minimize the function.[14]

The initial starting values are obtained from the Meyer (1990) model after controlling for unobserved heterogeneity. The standard errors from this estimation are used to construct the bounding box that is then used to bound the parameter space for optimization. Follow-

[14]To find details on the convergence and the consistency of the estimator refer to Hausman and Woutersen (2014). Details on pattern search methodology can be found in Kolda, Lewis and Torczon (2003) and Audet and Dennis (2000).

ing from the paper I start with a bounding box of ±3 standard errors. In each iteration the algorithm evaluates the objective function at all possible values. If an improvement is found then the size of bounding box is increased. This process is continued till convergence. In order to improve accuracy of estimates, once the parameter values stabilize the size of the bounding box is re-centered around these new estimates.

What is important in this approach is that it focuses on the probability than an individual i survives period l from time $t = 0$. This permits a convenient treatment of the heterogeneity distribution. In particular the Meyer (1990) model measures the probability than an individual i survives period l, given he has survived till $l - 1$. By focusing on survival from the beginning of the sample, the authors have eliminated the requirement to specify a heterogeneity distribution since no survival bias (dynamic sample selection) occurs in the sample comparisons.

1.4 Results & Selection Concerns

1.4.1 Results - Model 1

The effect of the internet on job match quality is measured using job tenure for workers. Results from both parametric and non-parametric baseline hazard models, with and without controls for unobserved heterogeneity (ν) are reported. Across all specifications the internet has a negative and significant impact on the exit rate from employment. While there is a difference in point estimates when controlling for ν, the difference in results is relatively small. In all cases the heterogeneity component is significant and slightly increases the effect of internet on exit rate.

To discuss the impact of the internet on job match quality, I first list the different estimates obtained from all the survival models estimated with and without the heterogeneity component. All the estimates indicate that the internet has a negative and significant effect on the exit rate from employment. Using online job search has a negative effect of -0.16 (Table 1.4 - Spec 1) on the exit rate from employment (in the non-parametric survival model without the heterogeneity component) and conditional on unobserved heterogeneity the estimate shows a larger negative effect -0.33 (Table 1.4 - Spec 5). In the parametric model without v, job search using the internet reduces the exit rate as evidenced by the -0.20 coefficient (Table 1.5 - Spec 1) and conditional on unobserved heterogeneity the coefficient effect increases to -0.23 (Table 1.5 - Spec 5). The absolute size of the estimates are larger when we control for v and this is a standard empirical result from survival literature. More importantly, all four versions of the model agree that workers who use online job search have a higher job duration that in turn implies a better match quality. Focusing on the results obtained from the non-parametric baseline model, the -0.16 coefficient implies that the baseline hazard for a worker who used the internet to look for jobs is reduced by 15%[15] after controlling for other observables. After controlling for unobserved heterogeneity, the baseline hazard of IJS workers is 28% ($exp(-0.33)$ is 0.72) lower. Both estimates are statistically significant. After controlling for all other variables the probability of a IJS worker exiting in period t, conditional on surviving till $t-1$, is 28% (or 15%) lower when compared to a non-IJS worker.

I next focus on the point estimates for all other controls (Table 1.4 and Table 1.5). The estimates reveal that in almost all cases the control variables have the same sign. Concentrating firstly on the parametric baseline model, without controls for v (Table 1.5 - Spec 1), it can be observed that the duration coefficient is negative and significant. This suggests that the probability of a worker with a higher duration to leave their job is lower when compared to a worker with lesser duration. The other covariates included in the model have the expected

[15] The hazard ratio is calculated as $exp(\beta)$. For IJS workers the hazard ratio ($exp(-0.16)$) is 0.85. This implies that the probability of exit for an IJS worker is 0.85 times the probability of exit for a non-IJS worker or that the exit rate is 15% lower for an IJS worker.

signs. Females have a higher exit rate, when compared to males. This maybe due to family or child-care concerns. Higher education leads to a decrease in the exit rate. The more a worker is educated; the lower is his exit rate. This maybe because more educated workers are able to obtain better-matched jobs or it maybe the case that workers with low education exit the labor market to complete their schooling/college. If the second scenario is true then the positive effect captured may not necessarily reflect a better match quality. However, the positive estimate does imply that workers with higher education have higher average job duration. Marriage has a significant negative effect on exit rate. I believe, that this may be partly due to personal concerns. An increase in family responsibilities (for example the presence of children) could lead to a decrease in the incidence of switching. Also, white workers have a lower exit rate when compared to workers of other races. Urban location has a positive relationship with the exit rate but the effect is not significant. Another interesting result can be seen when we observe the unemployment rate. The unemployment rate can affect exit rates in two ways. Higher unemployment rates imply a low level of economic activity and in such a situation there is a higher probability of a worker getting fired. On the other hand, higher unemployment rates would make workers more cautious and they would not quit jobs easily. Empirically however the firing effect is stronger. The unemployment rate has a positive and significant effect on the exit rate. The higher the state unemployment rate, the higher the probability of a worker exiting employment.

Focusing on the results from the parametric baseline model with controls for ν (Table 1.5 - Spec 5), the most important thing to note is that the heterogeneity component introduced in the model is significant. This suggests that controlling for unobserved heterogeneity is not only crucial but also that the estimates obtained are more robust. As mentioned before, the size of the coefficient is larger when we control for ν. With regards to the covariate coefficients, it can be clearly seen that except for the estimate on the duration, all other covariates have the same signs as was observed in the parametric baseline model without controls for ν. Duration now increases the probability of exit and this result is in line with the KM Hazard graphs plotted earlier, where across all classifications, it was observed that hazard rate and duration are positively linked. Same as before, the unemployment rate is positively linked to the exit rate, and being white, married or highly educated all lead to lower exit rates. A difference in results is also observed for the urban coefficient. While the estimate is still positive (implying workers in urban areas have lower job duration), the effect is now significant.

This following discussion is based on the non-parametric baseline model, without controls for ν (Table 1.4 - Spec 1). All of the controls have similar coefficients, though there are some differences in the significance of parameters. The urban location indicator is now positive and significant. The coefficient shows that workers in urban locations have lower average duration when compared to workers in rural areas. This may be due to the higher number of job options available to workers in urban areas, which may lead to higher switching. Since the analysis does not differentiate between voluntary and involuntary quits, it may also be the case that there is a higher incidence of firing workers in urban locations, where a larger supply of workers is available. It is however impossible to differentiate between the two effects, as data on the nature of job exit is not available. All other controls have the same sign as obtained in the parametric model. Being unmarried, female, lowly educated or not white all lead to higher exit rates and lower job duration. The unemployment rate has a positive and significant coefficient, implying that areas with large pools of unemployed workers and/or decreasing economic activity have lower average job duration.

Next I focus on the non-parametric baseline model with controls for ν (Table 1.4 - Spec 5). As was the case for the parametric baseline model, ν is significant, implying that the coefficients obtained from this model are more robust. Also, Meyer (1990) suggested that more consistent estimates might be obtained using a non-parametric baseline model in the presence of time-varying coefficients. Hence, it is safe to say that under this modeling structure, these results may be the most consistent and robust. For all controls the sign of the coefficient is the same, however the magnitude is different. Notably, the indicator for urban location is no longer significant. All other controls have the same sign and are significant across all models. Thus we can conclude, that the unemployment rate is positively linked to the exit rate. Also, education, marriage and being white reduce the exit rates and workers with any of these three characteristics generally have higher job duration.

Beyond providing consistent estimates, the non-parametric formulation provides an additional benefit. The baseline hazard rates provide the probability of exit, common to everyone in the population, at each time point. Table 1.4 (Spec 1 & 5) provide the baseline hazards for twelve duration intervals. Dur1 (month 1-5) and Dur12 (beyond 55 months) were created to prevent loss of data. The rest of the bins (Dur2-Dur11) are five month groups, each created only for ease of presentation purposes and have no impact on the point estimates.[16] The coefficient on each of these twelve variables, provide the common baseline hazard or the probability of exit in each group, common to the entire population. For example, the coefficient on Dur 9 provides the probability of exit in period 9, given a person has survived till period 8. These estimated coefficients on the duration interval dummies provide information about the shape of the baseline hazard.

To interpret these results, I focus on the change in coefficients as duration increases. In the model without the heterogeneity component, the probability of exit remains more or less similar as duration increases. On the other hand once unobserved heterogeneity is accounted for, the results are significantly different and the exit rate decreases as duration increases. This result is in line with job-match theories that suggest that a worker's probability of quitting decreases as tenure increases. Theory suggests that workers enter employment with incomplete information. This could be with respect to working condi-

[16] I have calculated hazards with different month groups and the point estimates don't change significantly. These results are not included in the paper.

tions, expectations of future or other factors. The same holds true for firms, as they are not aware if a worker will perform according to their expectations. Once the job starts, both workers and firm gain information. If unfavorable signals are received by either party, then quitting/firing is the next logical step. Discovery of these unfavorable signals is most likely to occur in the initial stages of employment, leading to higher quits in the beginning. Over time, if a worker stays in the same job then it implies that both parties received 'good' signals and the probability of quitting decreases.

As a rough check, I also analyze the robustness of the IJS estimate by including new controls. Firstly, the model is re-estimated to address the concern of unobserved ability across workers. It may be the case that workers with higher ability were much faster in adopting the internet and the use of internet then acts as a signal of productivity. In this case the increased job duration would then be due to the difference in unobserved ability and not due to the increase in match quality. To control for these unobservable differences, I use the ASVAB[17] scores provided by the NLSY97. The score included in this analysis is the ASVAB Math-Verbal Score, which is closer to the Armed Forces Qualification Test (AFQT) scores used by the Department of Defense. This score has been widely used in literature as a measure of cognitive achievement, aptitude and intelligence (Carneiro and Heckman (2002), Belley and Lochner (2007)). Following from previous literature, if we believe that ASVAB scores are a good proxy for ability across workers, then the IJS coefficient after controlling for the score should reflect the true effect of using the internet on job match quality. The results are included in Spec 2 & Spec 6 for both the parametric (Table 1.5) and non-parametric models (Table 1.4). From the results, it can be observed that the negative coefficient of the internet becomes slightly larger in the baseline parametric model and is significant. In the non-parametric model, there is a small increase in the negative effect and the baseline hazard for IJS workers is reduced by 34% (18% without ν),

[17] ASVAB provides details on the ASVAB score and how the score used in this analysis is constructed by NLSY.

after controlling for other variables. Thus even after introducing this new control, the IJS coefficient remains negative and significant.

For the second check I introduce controls for the difference in search intensity across workers. It can be argued that better matching of employee and firm may result from the more aggressive nature of job search made by one employee when compared to another. If it is believed that IJS workers search more intensively, then the negative internet coefficient may simply be due to the search effort put in by a IJS worker. Also, intensity of job search can vary among IJS workers. To correct for these biases I introduce two variables (i) search intensity and (ii) frequency of internet use.[18] The variable search intensity is a count of the number of job search methods used by a worker when looking for the current job and is introduced as a proxy for the intensity of job search. The more methods a worker uses, the higher is his intensity of job search. It is possible that a worker uses only one method extensively and in that situation while the job search intensity is high, this proxy will fail to capture that effect. There is however no measure of the time spent on job search and this proxy functions like a basic control. The second variable introduced is frequency of internet use. This is a categorical variable provided by the NLSY97 and proxies as a control for intensity of job search across IJS workers. The more frequently a worker uses the internet, the higher is his proficiency and this could translate into better job search using the internet. Looking at the results for both the parametric (Table 1.5) and non-parametric (Table 1.4) models (Spec 3 & Spec 7), the negative effect of the internet decreases slightly in both models. Conditional on unobservable heterogeneity, the hazard rate for IJS workers is 25% lower in the non-parametric model. While the proxies introduced in the model are not perfect controls for search intensity across workers, they do to a certain extent control for the volume of search done by a worker. Even after controlling for this effect, the negative effect of the internet on the baseline exit rate from employment persists.

[18] 52 observations were lost due to missing data on frequency of internet use. As almost all (50 obs.) were non-IJS workers, this might have biased the internet coefficient.

As a last test, all three variables were simultaneously introduced in the analysis (Table 1.4 & Table 1.5). The dataset is much smaller with only 2,372 workers of which 1,170 are IJS workers. Even in this smaller dataset with all controls (Spec 4 & Spec 8) internet has a negative and significant impact on the exit rate. In fact the negative effect of internet slightly increases across all specifications. In summary, these three analyses show that the negative effect of internet on exit rate is consistent. Adding more variables and reducing the dataset has no major impacts on either the value or significance of the estimate. These are however rough measures and they merely check the robustness of the coefficient across different specifications. Section 1.4.3 and 1.4.4 focus on a more detailed analysis to check for selection/endogeneity concerns.

1.4.2 Results - Model 2

This section focuses on the results obtained from the Hausman and Woutersen (2014) hazard model. This model specifies a non-parametric baseline hazard with controls for unobserved heterogeneity, without imposing the gamma distribution assumption. Table 1.6 provides a summary of the results obtained under this model. Note that the model controls for unobserved heterogeneity but no longer imposes the distributional assumption. The IJS coefficient under this model is -0.19 that implies that the baseline hazard for a worker who used the internet to look for jobs is roughly 18% lower. Compared to the original Meyer (1990) non-parametric baseline model with controls for unobserved heterogeneity (Table 1.4 - Spec 5), the negative effect is slightly smaller. However it is still negative and highly significant. Thus even under this more flexible structure, the negative effect of internet job search on the exit rate from employment persists.

Other controls have similar effects as was observed before. Female workers have higher exit rates while white workers have a higher average duration. Both, education and marriage decrease the exit rate, while workers in an urban location have much lower tenures. The unemployment rate has a positive and significant coefficient and the effect is now stronger. Table 1.7 reports the hazard estimates for each duration period for the first 52 months. Note that these numbers are not comparable to the Meyer (1990) estimates. To enable conditioning out the heterogeneity distribution requirement, the Hausman and Woutersen (2014) model calculates the exit rate for period t. In Meyer (1990), the model specifies the probability of exit in period t, given the person has survived till period $t-1$. While there are fluctuations in the exit rate over time, the overall trend for these duration estimates is similar. As suggested by job-matching theory, the probability of a person exiting employment decreases as time increases. This suggests that, the probability of quitting is higher in the beginning and decreases over time, similar to what was observed in Table 1.4.

1.4.3 Selection Concerns

As mentioned in the introduction, selection issues can significantly bias the results obtained. It can be argued that IJS workers are fundamentally different from non-IJS workers and that the internet coefficient is only capturing these unobservable differences across the two worker categories. As there is no direct way to measure the counterfactual - what would have happened to the IJS workers had they not been able to search online, a selection bias may exist. In addition to this, the modeling specification assumes independence between unobserved heterogeneity (v_i) and the observables (Z_i). Chamberlain (1985) shows that the selection bias will generally remain if unobserved and observed covariates are assumed to be independent. In the presence of selection bias, the estimates obtained may be biased upward. To deal with this issue, this analysis relies on a series of robustness checks to check for consistency of coefficients across the survival models. While the tests are not a perfect control for endogeneity, similar results across the different tests conducted indicate that IJS workers are being better matched.

Firstly, I redefine IJS workers by restricting the definition such that only the true efficiency effect of using the internet is captured. Instead of including all the twelve different job search methods in which the worker could have used the internet, a worker is now defined as an IJS worker if he/she used the internet to either- contact the employer directly or send out resumes. In comparison to surfing employment websites or emailing friends for jobs, the above two methods imply more active internet usage for job search. The worker can use the internet to sort jobs/employers he is interested in from the multitude available. The worker may use the information made available by the internet to select which employer to contact or which job to apply for. Under this new restricted definition there are now 1,025 IJS workers of which 554 (54%) contacted the employer directly and 736 (72%) sent out resumes. Table 1.8 reports the hazard model results for both the parametric and non-parametric models, after controlling for unobserved heterogeneity. For the parametric model, there is only a small change in the IJS coefficient. For the non-parametric model, there is a decrease in the negative impact of the internet on the exit rate, however the effect is still significant. Under this new definition, the probability of an IJS worker exiting the workforce is 15% lower when compared to non-IJS workers.

The second test conducted is in line with the analysis done in Dinardo & Pischke (1997). In their paper the authors question the positive effect of computer use on wages as argued in Krueger (1993). The authors find that in addition to computers other tools that require no proficiency like pencils, calculators and telephones also have a positive and significant effect on wages. They argue that since there is no skill involved in using a pencil, the positive estimate simply reflects selection effects. To test if this selection effect explains the IJS coefficient, I estimate the hazard analysis with another method of job search. In particular, I create an indicator for workers who - Placed/Answered or Looked at ads. Of 2,922 workers, 814 (27.8%) reported yes to either looking at or answering ads. Both of these job search methods don't provide any information efficiency- very little information is available and it is not updated frequently. Hence, there is no reason to expect that job match quality will improve. Also these methods do not hint at any unobservable ability on the part of workers as almost any job-seeker can look for job postings. Since placing or looking at ads does not provide any improved/increased information or reflect on any unobservable skill, it is believed that workers using this method for job search should not find any significant effect on their exit rate from employment. Table 1.9 presents the estimates from the model. The indicator has a positive coefficient implying that workers who use either of these methods have a higher probability of quitting. What is more important is that under both specifications, the coefficient is small and insignificant. This result suggests that the effect captured by the IJS coefficient may be causal and not merely due to selection effects.

The third test attempts to check the impact of internet on the exit rate across workers with different skills. To control for this, workers are divided into two separate groups based on their education (workers with and without a college degree). This division is based on previous empirical research, which classifies high skill workers by measuring the number of college graduates and low skill workers by measuring the number of high school graduates (Autor, Katz, and Kearney (2008)). The rationale behind this split stems from the belief that higher education raises productivity in general. Table 1.10 presents the results for the two groups of workers. Focusing firstly on the low skill workers or workers who have not completed college, in the parametric model the internet coefficient is similar to that in the baseline model. In the non-parametric model the internet coefficient slightly decreases (when compared to the baseline model). The baseline exit rate from employment is 14% lower for low skill IJS workers, after controlling for other variables and unobservable heterogeneity.

For the high-skilled workers or workers with at least a college degree, in the parametric model, the negative effect of the internet on the exit rate is similar to the baseline results.

In the non-parametric model the negative coefficient is much higher when compared to the baseline numbers. The exit rate from employment for high-skill IJS workers is roughly 40% lower. This large number suggests that high skill workers are deriving much larger benefits from using the internet. What is more important to note is that for both the categories of workers, the effect of the internet on the exit rate is negative and significant. Whether a worker is high skill or low skill, internet job search leads to a higher average duration. This suggests that even though the workers may vary significantly in unobservables across the two groups, the negative effect of internet on exit rate from employment is consistent.

1.4.4 Selection Concerns - IV Estimation

One important problem when dealing with the effect of internet on related outcomes is the non-random assignment of internet usage. Since the choice of using the internet is determined by the individual worker, it can be argued that the results suffer from endogeneity. Though the previous section provides some suggestive evidence, I additionally use the instrumental variable approach to further test for selection concerns. This test is conducted on the modified version of the data (ignore censoring and survival settings). In particular, I estimate the following relationship:

$$Dur_i = \beta IJS_i + \gamma X_i + \epsilon_i$$

where, Dur_i = Job duration of worker (in months) i; $IJS_i = 1$ if a worker used the internet to look for the current job and X_i is a vector of controls.

The parameter of interest in this analysis is β and can be estimated using ordinary least squares (OLS) Table 1.12.[19] From the table we can see that the coefficient for internet job search is positive and significant. The job duration for IJS workers on average is greater by 1.35 months compared to non-IJS workers. However, given the concern about endogene-

[19]The OLS estimates are mainly introduced as a comparison for the instrumental variable estimates.

ity, the OLS estimates can be biased. For instance, if the use of internet implies a person with higher ability, then β may be biased upwards. The positive correlation between the probability of using the internet and the probability of being able to search more effectively because of this higher ability may lead to a better job match and hence a higher job duration. It can also be argued that β only captures the unobservable ability of the worker and does not reflect the increase in match quality due to the increase in information efficiency introduced by the internet. Thus in order to estimate the true effect, this endogeneity issue needs to be addressed. To control for this, an instrumental variable (IV) framework is used here.

Adapting the IV used by Choi (2011), in my analysis I use the exogenous variation in the use of computers across industries as an instrument for IJS workers. I believe that more workers used online job search across industries that had higher computer usage. Industries that require higher computer usage force workers to adapt more quickly to innovations in this field. The higher the probability of using the computer in an industry, the higher is the probability of a worker being more familiar with the internet and also using it in his/her job search process. This process works two ways where industries that use computers more, shift more quickly to the online platform using it for different services, including the job application process. I therefore expect that the variation in the adoption of the internet (for job search) across workers is linked to the baseline computer usage across industries. The data for the baseline computer usage by industry comes from October 2003 Internet and Computer Use Survey Supplement to the CPS.[20] The survey asked all employed respondents "Does .... use computer at his/her main job?" This helped calculate computer usage at the industry level. Computer Usage Percentage (C_{03}) in an industry is defined as the number of employed people who responded yes when asked if they used a computer in their main job divided by the total number of employed persons in the industry.

[20]This was the last year when this question was asked.

These percentages were calculated at the 4 digit 2002 Census codes level and then attached to the final NLSY97 dataset in my analysis. Average computer usage across the main industry categories are reported in Table 1.11.

Note that this instrument thus forces all workers in the same industry to have the same speed of internet adoption. Thus if any worker was quicker in using the internet, compared to other workers across his industry, then his/her unobservable ability will not impact the instrument. While potentially this instrument is independent of job duration for a worker, it may be biased when baseline computer usage rates across industries are linked to employment outcomes. In the case of this analysis, if average job duration is higher in industries with high computer usage, then the coefficient may be upwards biased. As a rough check Table 1.11 also includes the average job duration across the different industries. Average job duration is mostly similar across industries, irrespective of the level of computer usage. For example, Finance, Insurance and Information have the highest computer usage (83%), but their average job duration is the same as in the Transportation and Warehouse Industry, which has a much lower computer usage (44%). Similarly Construction and Manufacturing industries have almost the same average duration, but the computer usage in Manufacturing is double. To prevent potential biases, the final IV analysis includes a wide variety of controls, including personal, firm, region, industry (2 digit) and occupation controls.

Using C_{03} as my instrument, I estimate the following linear relationship:

$$Dur_i = \beta_1 \widehat{IJS}_i + \beta_2 X_i + \epsilon_i$$

Where $\widehat{IJS}_i$ is derived from the predicted values of the following first stage relationship:

$$IJS_i = \theta_1 C_{03} + \theta_2 X_i + \mu_i$$

The results for this first stage of the regression are reported in Table A1.[21] The coefficient on C_{03} is positive and significant. The tests reported at the bottom of the table help examine

[21] Estimation is done using the linear probability model with robust standard errors.

the strength of the instrument. Both the Durbin (1954) and Hausman (1978) & Wu (1974) statistics are staistically significant, which implies that $\widehat{IJS_i}$ is endogenous. The F statistic reported is statistically significant and much larger than the critical value 10, rejecting the null hypothesis that the instrument is weak. The partial R^2 is 0.024 which suggests that the standard errors for $\widehat{IJS_i}$ will be inflated approximately seven times. In addition to this the Stock and Yogo (2005) test of weak instruments suggests that C_{03} does not suffer from a weak instrument problem.

Table 1.12 presents the main IV estimation results of the effect of the internet on job duration. The regression includes dummies for region, industry and occupation and robust standard errors have been reported.[22] The coefficient on $\widehat{IJS_i}$ after IV is positive and significant, and approximately 10 times the size of the OLS coefficient. The coefficient implies that the job duration for IJS workers on average is greater by approximately 13 months, when compared to non-IJS workers. The effects of the other covariates are mostly similar to the OLS estimates. Being female has a negative impact on duration while being white or married has a positive and significant impact on duration. Most importantly, this large, positive and significant component on $\widehat{IJS_i}$ after controlling for the endogeneity issue suggests that online job search has helped workers find better jobs, which has lead to a higher average job duration.

As a last test some regressions are conducted with different explanatory variables while ignoring the censoring issue on job duration. If the internet coefficient changes significantly across the different specifications, then it would indicate that internet does not affect job match quality and is merely picking up effects from other variables. To test this the following specifications were considered - Spec 1 : including all variables, Spec 2 : excluding occupation controls, Spec 3: excluding industry controls, Spec 4: excluding industrial and

[22]IV estimation excluding region, industry and occupation dummies were also carried out but there was no significant change in the size and significance of the estimates.

occupation controls, Spec 5: excluding state, industrial and occupation controls and Spec 6: excluding state, industrial and occupation controls and some other explanatory variables (Table 1.13). The internet coefficient is positive and statistically significant across all specifications. In the base specification with all variables (Spec 1), the internet coefficient is 1.21. When occupation or industry controls are removed then the coefficient increase slightly. Removing both occupation and industry together increases the internet coefficient slightly while dropping all state, industrial and occupation dummies has almost no impact. In the last specification when almost all variables are dropped, the coefficient is only slightly larger than the original. Minor changes in the internet coefficient across these specifications suggest that what we observe is the true impact of the internet on duration, unbiased by selection effects.

The above analysis shows that the results are robust to a wide range of controls and sample restrictions. In the duration analysis, there are only slight changes in the internet coefficient when controls for ability and search intensity are introduced. Re-defining the IJS variable has almost no impact. Ability controls suggest larger benefit for college graduates, but there is still a negative and significant impact for workers without a college degree. The IV results indicate a strong positive effect of the internet on job duration. Similarly in the regression analysis, there are very small changes in the coefficient value across the six specifications and internet is always positively associated with job tenure.

1.5 Conclusion

The rise of the internet has significantly impacted economies across the globe. Information is made available at lower costs and also disseminated quickly. Valuable information is a key building block of economic relations and increasing the internet penetration is causing these economic relations to evolve. In this process the labor market has also been wired, and both employers and job seekers use the internet to gather important information about each other to help them in their decision making. The central objective of the paper is to find if the internet is helping job seekers become employed in jobs suitable to their skills. The paper uses job duration as a proxy of job match quality and employing the Meyer (1990) hazard model shows that internet use is a factor determining the exit rates of workers. A conservative estimate shows that the exit rates of workers who use the internet for searching jobs are 28% lower than workers who don't. One of the criticisms of the Meyer (1990) model is the specification of the heterogeneity distribution which might give inconsistent estimates. To correct for this problem I use the Hausman and Woutersen (2014) hazard model that eliminates the requirement. Under this methodology, internet search reduces exit rate by 18%.

Among other findings, the model estimates marital status, gender and race of a worker to be important factors determining the exit rates. While a married worker has a lower exit rate than an unmarried one, the exit rate of a female worker is higher than a male worker. Also state unemployment rates have positive effects on the exit rate implying stronger firing effects.

Unobservable worker characteristics, and selection of workers and employers can systematically bias the estimates. However, I find no evidence of such effects using a series of robustness tests. Internet usage continues to be a factor impacting exit rate when controls for ability, intensity of job search and intensity of internet use are included in the model. The estimates of the hazard models also show that internet has a negative impact on the exit rate, even after controlling for differences in skill or restricting the definition of an IJS worker. In addition to this, IV estimation results also indicate a positive and significant relation between internet use and job duration.

Endogeneity concerns make it difficult to net out the effect of the internet. In such a situation a natural experiment seems to be the one of the ways forward. Nonetheless, the robust estimate of internet across the various specifications provide belief on the validity of the coefficient.

1.6 Tables

Table 1.1: Sample Means

Variable	Internet = 0 Mean	S.D.	Internet = 1 Mean	S.D.	Difference[†]
Hourly Pay	17.75	36.99	17.95	27.25	−0.20
Female	0.46	0.50	0.58	0.49	−0.12***
Duration	36.20	17.82	39.21	17.36	−3.01***
Not Completed School	0.50	0.50	0.27	0.44	0.23***
Not Completed College	0.27	0.44	0.25	0.43	0.02
Completed College	0.23	0.42	0.48	0.50	−0.25***
Married	0.28	0.45	0.29	0.46	−0.01
Age	25.78	1.41	25.84	1.46	−0.06
ASVAB Score[‡]	43.51	28.25	56.27	28.52	−12.76***
North East	0.16	0.36	0.15	0.36	0.01
North Central	0.20	0.40	0.22	0.41	−0.02
South	0.42	0.49	0.37	0.48	0.05**
West	0.21	0.41	0.25	0.44	−0.04
Internet Access	0.78	0.42	0.93	0.26	−0.15***
Internet Access (home)	0.66	0.47	0.83	0.37	−0.17***
Search Intensity	1.37	0.78	2.59	1.67	−1.22***
Urban Indicator	0.76	0.43	0.82	0.39	−0.06***
White	0.60	0.49	0.60	0.49	0.00
Black	0.25	0.43	0.25	0.43	0.00
Sample	1,540		1,382		

[†]The numbers represent the difference in mean characteristics between Non-IJS workers and IJS workers.
*** Significant at 1%, ** Significant at 5%, * Significant at 10%
[‡]ASVAB scores were available for only 2,410 observations of which 1,171 were IJS workers.

Table 1.2: Average Duration

Variable	Internet = 0 Mean	S.D.	Obs.	Internet = 1 Mean	S.D.	Obs.
Overall	36.20	17.82	1540	39.21	17.36	1382
Male	36.86	18.03	837	39.99	18.11	586
Female	35.40	17.55	703	38.63	16.78	796
Education						
Not Completed School	34.43	17.72	764	36.05	17.59	375
Not Completed College	36.39	17.64	417	38.44	18.29	340
Completed College	39.73	17.77	359	41.37	16.44	667
Hourly Pay						
0–10	31.23	17.25	612	32.29	16.96	286
10–15	37.75	17.07	476	37.80	17.52	461
15–20	39.96	17.48	197	40.98	16.84	322
>20	42.30	17.77	255	45.79	15.26	313
Occupation						
Management Related	41.69	16.86	103	43.38	16.90	183
Professional Specialty	37.79	17.77	302	41.41	16.13	468
Tech., Sales & Admin. Support	35.69	17.58	347	37.28	17.84	387
Service Occupations	34.04	17.57	386	35.37	18.07	171
Farming, Fishing, & Forestry	34.11	20.52	9	44.00	0.00	1
Precision Prod., Craft, & Repair	36.95	18.16	219	38.86	17.87	81
Setter, Operators, & Tenders	35.14	18.25	174	35.14	17.60	91
Industry						
Agri., Forestry & Fisheries	33.91	20.85	11	30.00	19.80	2
Mining	29.63	15.82	19	33.40	17.47	5
Utilities	53.44	10.09	9	47.14	13.99	7
Construction	36.63	18.60	147	35.33	16.09	43
Manufacturing	37.13	18.75	126	40.18	17.93	93
Trade	34.91	17.81	219	37.56	17.86	174
Transportation & Warehousing	39.16	16.71	49	42.06	19.89	36
Finance, Insurance, & Information	37.86	18.29	112	41.66	16.19	197
Services Industry	35.48	17.49	808	38.22	17.26	767
Public Administration	44.90	15.90	40	48.33	16.37	58

Table 1.2: Average Duration (Continued)

	Internet = 0			**Internet = 1**		
Variable	**Mean**	**S.D.**	**Obs.**	**Mean**	**S.D.**	**Obs.**
Race						
White	37.08	17.93	926	40.21	17.30	835
African-American	34.94	17.92	387	36.10	17.25	349
Other	34.73	17.05	227	40.46	17.25	198
Region						
North East	37.05	18.14	243	42.39	17.34	213
North Central	36.03	17.36	315	38.97	17.64	301
South	35.60	17.56	653	38.56	17.10	517
West	36.92	18.56	329	38.44	17.38	351

Table 1.3: Internet Use and Access Data

Categories	**IJS**		**Non-IJS**	
Workers (Count)	1540		1382	
Current Access	1199	(78%)	1283	(93%)
Home Access	1024	(66%)	1153	(83%)
Work Access	696	(45%)	1020	(74%)
Activities				
Email	1227	(80%)	1315	(95%)
Read News etc.	1048	(68%)	1207	(87%)
Play Online Games	716	(46%)	790	(57%)
Research for Work	731	(47%)	972	(70%)
Pay Online bills/Banking	895	(58%)	1127	(82%)
Frequency of Use†				
Several times a day	689	(45%)	971	(70%)
Once a day	190	(12%)	133	(10%)
3-5 days a week	147	(10%)	137	(10%)
1-2 days a week	155	(10%)	54	(4%)
Once every few weeks	152	(10%)	52	(4%)
Less often	157	(10%)	33	(2%)
Total Workers (Count)	1,540		1,382	

† 52 observations were lost due to missing data, of which 50 were Non-IJS workers

Table 1.4: Hazard Rates: Non-Parametric Baseline

Variable	Spec 1	Spec 2	Spec 3	Spec 4	Spec 5	Spec 6	Spec 7	Spec 8
IJS Worker	-0.16***	-0.20***	-0.17***	-0.21***	-0.33***	-0.41***	-0.29**	-0.38**
	(0.05)	(0.06)	(0.06)	(0.06)	(0.13)	(0.14)	(0.15)	(0.16)
ASVAB		0.03		0.06		-0.06		0.09
		(0.12)		(0.13)		(0.30)		(0.31)
Frequent Use†			0.02	0.02			0.13***	0.13**
			(0.02)	(0.02)			(0.04)	(0.06)
Search Intensity			0.02	0.02			0.01	0.03
			(0.02)	(0.02)			(0.05)	(0.05)
Female	0.16***	0.22***	0.17***	0.22***	0.31**	0.41***	0.35***	0.42***
	(0.06)	(0.06)	(0.06)	(0.06)	(0.14)	(0.15)	(0.14)	(0.15)
White	-0.13***	-0.14**	-0.13**	-0.14**	-0.25*	-0.28*	-0.18	-0.24
	(0.05)	(0.06)	(0.05)	(0.06)	(0.13)	(0.16)	(0.14)	(0.16)
Education	-0.04***	-0.05***	-0.04***	-0.04***	-0.08***	-0.07**	-0.05**	-0.04
	(0.01)	(0.01)	(0.01)	(0.01)	(0.03)	(0.03)	(0.03)	(0.03)
Married	-0.22***	-0.25***	-0.23***	-0.26***	-0.35***	-0.38***	-0.35***	-0.39***
	(0.06)	(0.06)	(0.06)	(0.06)	(0.14)	(0.15)	(0.14)	(0.15)
Urban Ind.	0.14**	0.20***	0.13**	0.19***	0.24	0.37**	0.21	0.33*
	(0.06)	(0.07)	(0.07)	(0.07)	(0.16)	(0.17)	(0.16)	(0.18)
Unempt. Rt	0.44***	0.44***	0.44***	0.45***	0.40***	0.41***	0.41***	0.42***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)
Dur1	-7.16***	-6.81***	-7.23***	-6.89***	-9.13***	-9.03***	-9.53***	-9.42***
	(1.18)	(1.22)	(1.18)	(1.23)	(1.59)	(1.66)	(1.75)	(1.80)
Dur2	-6.05***	-5.68***	-6.15***	-5.77***	-7.43***	-7.33***	-7.86***	-7.71***
	(1.17)	(1.22)	(1.18)	(1.22)	(1.57)	(1.64)	(1.74)	(1.79)
Dur3	-6.12***	-5.80***	-6.22***	-5.89***	-6.54***	-6.49***	-6.97***	-6.87***
	(1.18)	(1.22)	(1.18)	(1.23)	(1.56)	(1.62)	(1.73)	(1.77)

Hazard models provide the exit rate from employment for person i in period t (λ_{it}). The coefficient's reflect the effect of each variable on the exit rate. All specifications include controls for occupaion, industry, states, hourly pay, internet access at home and starting period.

† Frequent Use refers to the frequency of internet usage. ASVAB score was available for 2,410 workers (Spec 2 & Spec 6). 52 workers did not report their frequency of internet use (Spec 3 & Spec 7). 550 observations were lost due to missing data on frequency of internet use and ASVAB scores (Spec 4 & Spec 8).

*** Significant at 1%, ** Significant at 5%, * Significant at 10%

Table 1.4: Hazard Rates: Non-Parametric Baseline (Continued)

Variable	**Spec 1**	**Spec 2**	**Spec 3**	**Spec 4**	**Spec 5**	**Spec 6**	**Spec 7**	**Spec 8**
Dur4	-6.20***	-5.91***	-6.32***	-6.03***	-5.75***	-5.77***	-6.22***	-6.20***
	(1.18)	(1.23)	(1.18)	(1.23)	(1.55)	(1.61)	(1.72)	(1.77)
Dur5	-6.25***	-6.02***	-6.33***	-6.11***	-5.06***	-5.19***	-5.48***	-5.58***
	(1.18)	(1.23)	(1.19)	(1.24)	(1.54)	(1.60)	(1.72)	(1.76)
Dur6	-6.32***	-5.96***	-6.44***	-6.09***	-4.48***	-4.51***	-4.95***	-4.94***
	(1.19)	(1.23)	(1.19)	(1.24)	(1.54)	(1.60)	(1.72)	(1.76)
Dur7	-6.40***	-6.00***	-6.50***	-6.09***	-4.04***	-4.02***	-4.48***	-4.43***
	(1.19)	(1.23)	(1.19)	(1.24)	(1.54)	(1.60)	(1.72)	(1.75)
Dur8	-6.39***	-6.03***	-6.52***	-6.15***	-3.52**	-3.53**	-4.00**	-3.96 **
	(1.19)	(1.23)	(1.19)	(1.24)	(1.53)	(1.60)	(1.71)	(1.75)
Dur9	-6.56***	-6.24***	-6.65***	-6.33***	-3.24**	-3.30**	-3.67**	-3.70 **
	(1.19)	(1.24)	(1.19)	(1.24)	(1.53)	(1.59)	(1.71)	(1.75)
Dur10	-6.47***	-6.12***	-6.58***	-6.24***	-2.77*	-2.83*	-3.23**	-3.26 *
	(1.19)	(1.24)	(1.20)	(1.25)	(1.54)	(1.60)	(1.72)	(1.75)
Dur11	-6.45***	-6.03***	-6.54***	-6.11***	-2.40	-2.39	-2.83**	-2.78
	(1.20)	(1.24)	(1.20)	(1.25)	(1.54)	(1.60)	(1.72)	(1.76)
Dur12	-6.42***	-5.95***	-6.52***	-6.04***	-2.00	-1.91	-2.44	-2.30
	(1.20)	(1.24)	(1.20)	(1.25)	(1.53)	(1.59)	(1.71)	(1.75)
ν					4.43***	4.27***	4.43***	4.25***
					(0.43)	(0.46)	(0.44)	(0.46)
Sample Size	109,931	90,652	108,241	89,312	109,931	90,652	108,241	89,312
Log-likelihood	-8287.5	-6838.2	-8112.6	-6706.6	-8184.4	-6757.8	-8010.3	-6624.8

Hazard models provide the exit rate from employment for person i in period t (λ_{it}). The coefficient's reflect the effect of each variable on the exit rate. All specifications include controls for occupaion, industry, states, hourly pay, internet access at home and starting period.
† Frequent Use refers to the frequency of internet usage. ASVAB score was available for 2,410 workers (Spec 2 & Spec 6). 52 workers did not report their frequency of internet use (Spec 3 & Spec 7). 550 observations were lost due to missing data on frequency of internet use and ASVAB scores (Spec 4 & Spec 8).
*** Significant at 1%, ** Significant at 5%, * Significant at 10%

Table 1.5: Hazard Rates: Parametric Baseline

Variable	Spec 1	Spec 2	Spec 3	Spec 4	Spec 5	Spec 6	Spec 7	Spec 8
IJS Worker	-0.20***	-0.24***	-0.21***	-0.26***	-0.23***	-0.28***	-0.24***	-0.30***
	(0.05)	(0.06)	(0.06)	(0.06)	(0.08)	(0.08)	(0.09)	(0.09)
ASVAB		0.23*		0.17		0.10		0.12
		(0.12)		(0.13)		(0.18)		(0.18)
Frequent Use†			-0.07***	-0.07***			0.04	0.03
			(0.02)	(0.02)			(0.03)	(0.04)
Search Int.			0.01	0.01			0.02	0.02
			(0.02)	(0.02)			(0.03)	(0.03)
ln(duration)	-0.32***	-0.32***	-0.31***	-0.32***	0.16**	0.10	0.18**	0.10
	(0.04)	(0.04)	(0.04)	(0.04)	(0.08)	(0.08)	(0.08)	(0.08)
Female	0.15***	0.20***	0.17***	0.21***	0.19**	0.25***	0.21***	0.27***
	(0.05)	(0.06)	(0.06)	(0.06)	(0.08)	(0.09)	(0.09)	(0.09)
White	-0.19***	-0.24***	-0.18***	-0.23***	-0.17**	-0.18*	-0.15*	-0.17 *
	(0.05)	(0.06)	(0.05)	(0.06)	(0.08)	(0.09)	(0.08)	(0.09)
Education	-0.08***	-0.09***	-0.08***	-0.10***	-0.07***	-0.07***	-0.06***	-0.06***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)
Married	-0.26***	-0.29***	-0.27***	-0.29***	-0.32***	-0.32***	-0.32***	-0.33***
	(0.06)	(0.06)	(0.06)	(0.06)	(0.09)	(0.09)	(0.09)	(0.09)
Urban Ind.	0.08	0.13**	0.07	0.13*	0.22**	0.28***	0.20**	0.27***
	(0.06)	(0.07)	(0.06)	(0.07)	(0.10)	(0.10)	(0.10)	(0.11)
Unempt. Rt.	0.50***	0.50***	0.50***	0.51***	0.57***	0.57***	0.57***	0.58***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)	(0.03)
ν					1.34***	1.14***	1.38***	1.15***
					(0.18)	(0.17)	(0.18)	(0.17)
Sample Size	109,931	90,652	108,241	89,312	109,931	90,652	108,241	89,312
Log-likelihood	-8470.8	-6996.2	-8284.2	-6860.2	-8368.2	-6918.2	-8189.1	-6788.7

Hazard models provide the exit rate from employment for person i in period t (λ_{it}).
†Frequent Use refers to the frequency of internet usage
ASVAB scores were available for only 2,410 observations of which 1,171 were IJS workers (Spec 2 & Spec 6). 52 observations were lost due to missing data on frequency of internet use of which 50 were Non-IJS workers (Spec 3 & Spec 7). 550 observations were lost due to missing data on frequency of internet use and ASVAB scores. Of the 2,372 observations, 1,170 were IJS workers (Spec 4 & Spec 8). *** Significant at 1%, ** Significant at 5%, * Significant at 10%

Table 1.6: Hazard Rates: Hausman and Woutersen (Model 2)

Variable	**Coefficient**
IJS Worker	-0.19***
	(0.07)
Sex (Female=1)	0.18***
	(0.07)
White	-0.16**
	(0.07)
Highest Education	-0.06***
	(0.01)
Married (=1)	-0.27***
	(0.07)
Urban/Rural Ind. (Urban=1)	0.17**
	(0.08)
Unemployment Rate	0.57***
	(0.02)
Sample Size	109,931
Baseline Hazard	Non-Parametric
Heterogeneity	Yes

The Hausman-Woutersen hazard models provide the exit rate from employment for person i in period t (λ_{it}), after controlling for unobserved heterogeneity. The coefficient's reflect the effect of each variable on the exit rate. The model includes controls for each duration interval, occupaion, industry, states, hourly pay and internet access at home.
*** Significant at 1%, ** Significant at 5%, * Significant at 10%

Table 1.7: Duration Coefficients (Model 2)

Duration Interval	Coefficient	S.E.	Duration Interval	Coefficient	S.E.
D2	-0.12	(0.22)	D27	0.26	(0.19)
D5	0.77	(0.24)	D28	0.38	(0.18)
D6	0.55	(0.25)	D29	0.41	(0.18)
D7	0.52	(0.24)	D30	0.06	(0.20)
D8	0.97	(0.20)	D31	0.21	(0.19)
D9	0.70	(0.21)	D32	0.29	(0.19)
D10	0.84	(0.19)	D33	0.27	(0.19)
D11	0.77	(0.19)	D34	0.28	(0.19)
D12	0.42	(0.21)	D35	0.16	(0.20)
D13	0.35	(0.21)	D36	0.39	(0.18)
D14	0.61	(0.19)	D37	0.24	(0.20)
D15	0.54	(0.19)	D38	0.16	(0.20)
D16	0.68	(0.18)	D39	0.07	(0.21)
D17	0.27	(0.20)	D40	0.16	(0.21)
D18	0.58	(0.18)	D41	0.07	(0.22)
D19	0.37	(0.19)	D42	-0.05	(0.24)
D20	0.65	(0.17)	D43	-0.21	(0.26)
D21	0.27	(0.19)	D44	-0.04	(0.25)
D22	0.26	(0.19)	D45	0.27	(0.23)
D23	0.41	(0.18)	D46	-0.13	(0.28)
D24	0.51	(0.17)	D47	0.00	(0.27)
D25	0.42	(0.18)	D48	-0.05	(0.28)
D26	0.23	(0.19)	D49	0.19	(0.27)

Beyond the 30^{th} interval (D30), almost all coefficients are insignificant. This could be due to the very small dataset and fewer exit observation in the later intervals.

Table 1.8: Hazard Rates: New IJS Definition

Variable	Coefficient	Coefficient
New IJS	-0.28***	-0.17***
	(0.09)	(0.06)
Sex (Female=1)	0.19**	0.16***
	(0.08)	(0.06)
White	-0.17**	-0.13**
	(0.08)	(0.06)
Highest Education	-0.07***	-0.04***
	(0.02)	(0.01)
Married (=1)	-0.32***	-0.23***
	(0.09)	(0.06)
Urban Indicator	0.22**	0.16**
	(0.10)	(0.07)
Unemployment Rate	0.57***	0.54***
	(0.02)	(0.02)
Sample Size	109,931	109,931
Log-likelihood	-8366.84	-8317.84
Baseline Hazard	Parametric	Non-Parametric
ν	Yes	Yes

Hazard models provide the exit rate from employment for person i in period t (λ_{it}). The coefficient's reflect the effect of each variable on this exit rate. Both specifications include controls for occupaion, industry, states and starting period.
The model estimates are under the new restricted definition for IJS worker.
*** Significant at 1%, ** Significant at 5%, * Significant at 10%

Table 1.9: Hazard Rates: Placed/Looked at Ads

Variable	**Coefficient**	**Coefficient**
Placed/Looked at ads	0.08	0.12
	(0.09)	(0.14)
Sex (Female=1)	0.17**	0.30**
	(0.08)	(0.14)
White	-0.15*	-0.22
	(0.08)	(0.14)
Highest Education	-0.07***	-0.08***
	(0.02)	(0.03)
Married (=1)	-0.31***	-0.34**
	(0.09)	(0.14)
Urban Indicator	0.22**	0.25
	(0.10)	(0.16)
Unemployment Rate	0.57***	0.40***
	(0.02)	(0.04)
Sample Size	109,931	109,931
Log-likelihood	-8369.25	-8185.44
Baseline Hazard	Parametric	Non-Parametric
ν	Yes	Yes

Hazard models provide the exit rate from employment for person i in period t (λ_{it}). The coefficient's reflect the effect of each variable on this exit rate. Both specifications include controls for occupaion, industry, states and starting period.
The model is estimated with an indicator for workers who atleast looked/placed ads.
*** Significant at 1%, ** Significant at 5%, * Significant at 10%

Table 1.10: Hazard Rates: Low Skill and High Skill workers

Variable	**Low Educ**	**Low Educ**	**High Educ**	**High Educ**
IJS Worker	-0.22**	-0.15**	-0.23*	-0.52 **
	(0.11)	(0.07)	(0.13)	(0.26)
Sex (Female=1)	0.12	0.13*	0.26**	0.66***
	(0.12)	(0.08)	(0.13)	(0.26)
White	-0.12	-0.10	-0.17	-0.34
	(0.11)	(0.07)	(0.13)	(0.28)
Married (=1)	-0.30***	-0.21***	-0.36***	-0.88***
	(0.11)	(0.07)	(0.14)	(0.30)
Urban Indicator	0.40***	0.25***	-0.11	-0.32
	(0.13)	(0.09)	(0.16)	(0.32)
Unemployment Rate	0.59***	0.53***	0.55***	0.49***
	(0.03)	(0.03)	(0.04)	(0.07)
Sample Size	68,070	68,070	41,861	41,861
Log-likelihood	-5695.12	-5651.61	-2639.04	-2567.76
Baseline Hazard	Parametric	Non-Parametric	Parametric	Non-Parametric
ν	Yes	Yes	Yes	Yes

Hazard models calculate the exit rate from employment for person i in period t (λ_{it}). Column 1 & 2 provide results for workers who have not completed college and column 3 & 4 provide results for workers who atleast have a college degree. All specifications include controls for occupaion, industry, states & starting period. *** Significant at 1%, ** Significant at 5%, * Significant at 10%

Table 1.11: Average Computer Usage at Work in 2003 and Average Job Duration

Industry	**Computer Usage at Work (2003)**	**Average Job Duration**
Agriculture, Forestry & Fisheries	26%	33.31
Mining	49%	30.42
Utilities	71%	50.69
Construction	26%	36.33
Manufacturing	52%	38.43
Trade	51%	36.08
Transportation & Warehousing	44%	40.39
Finance, Insurance, & Information	83%	40.28
Services Industry	56%	36.81
Public Administration	77%	46.93

Table 1.12: OLS and IV Results

Variable	OLS	I.V.
IJS	1.36***	13.83***
	(0.53)	(3.59)
Sex (Female=1)	-1.39***	-2.04***
	(0.55)	(0.62)
White	1.35***	1.68***
	(0.54)	(0.59)
Highest Education	0.48***	0.01
	(0.11)	(0.18)
Married (=1)	1.85***	1.81***
	(0.56)	(0.60)
Urban Indicator	-1.43**	-1.82***
	(0.66)	(0.72)
Industries		
Agriculture, Forestry & Fisheries	-5.13	-0.17
	(7.90)	(9.06)
Mining	-9.05***	-5.45
	(3.65)	(3.82)
Utilities	5.27***	6.89***
	(1.92)	(2.74)
Construction	-8.32***	-5.24***
	(1.79)	(2.14)
Manufacturing	-7.64***	-6.17***
	(1.50)	(1.68)
Trade	-8.02***	-6.01***
	(1.38)	(1.60)
Transportation & Warehousing	-5.08***	-3.86*
	(1.91)	(2.10)
Finance, Insurance, & Information	-7.19***	-6.81***
	(1.38)	(1.50)
Services Industry	-8.13***	-6.50***
	(1.14)	(1.33)
R^2	0.47	0.36
Sample Size	2,922	2,922
Durbin Score		14.21
Wu-Hausman F(1,2859)		13.75
Robust F		65.68

The dependent variable is the job duration of a worker (in months). The first stage coefficient for the instrument $C_0 3$ is 0.46^{***}. Results include controls for occupaion (2 digit), states and starting period ***Significant at 1% **Significant at 5% *Significant at 10%

Table 1.13: Regression Results

Variable	**Spec 1**	**Spec 2**	**Spec 3**	**Spec 4**	**Spec 5**	**Spec 6**
IJS Worker	1.21*	1.62**	1.54**	1.89***	1.92***	1.66***
	(0.69)	(0.69)	(0.69)	(0.69)	(0.68)	(0.68)
Highest Education	0.60***	0.83***	0.66***	0.85***	0.88***	0.86***
	(0.14)	(0.13)	(0.14)	(0.13)	(0.13)	(0.12)
Sex (Female=1)	-1.45**	-1.38**	-1.97***	-1.97***	-1.91***	
	(0.72)	(0.70)	(0.71)	(0.66)	(0.65)	
White	1.34*	1.47**	1.14	1.29*	1.16*	
	(0.71)	(0.71)	(0.71)	(0.71)	(0.67)	
Married (=1)	2.33***	2.66***	2.59***	2.83***	2.66***	
	(0.73)	(0.73)	(0.73)	(0.73)	(0.72)	
Urban Indicator	-2.03**	-2.09***	-2.16***	-2.23***	-1.66**	
	(0.84)	(0.85)	(0.85)	(0.85)	(0.80)	
Industry Dummies	Yes	Yes	No	No	No	No
Occupation Dummies	Yes	No	Yes	No	No	No
State Dummies	Yes	Yes	Yes	Yes	No	No
All Other Variables	Yes	Yes	Yes	Yes	Yes	No

The dependent variable is the job duration of a worker (in months). All specifications include a constant term
***Significant at 1% **Significant at 5% *Significant at 10%

1.7 Appendix

KM Graphs

The graphs below plot the Kaplan Meier exit rate (calculated separately for IJS and non-IJS workers) over the entire dataset and across subgroups of education, gender, race and region. Across all categories, IJS workers have lower exit rates.

Figure 1.2: KM Hazard Rate across Education Categories

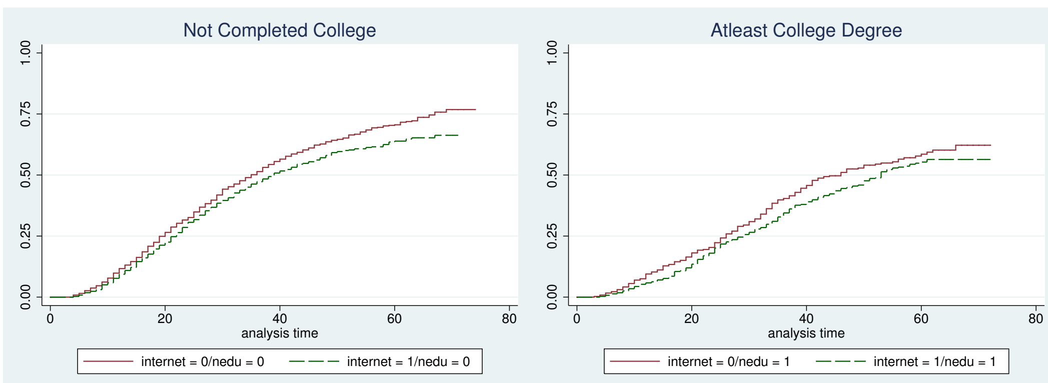


Figure 1.3: KM Hazard Rate across Gender

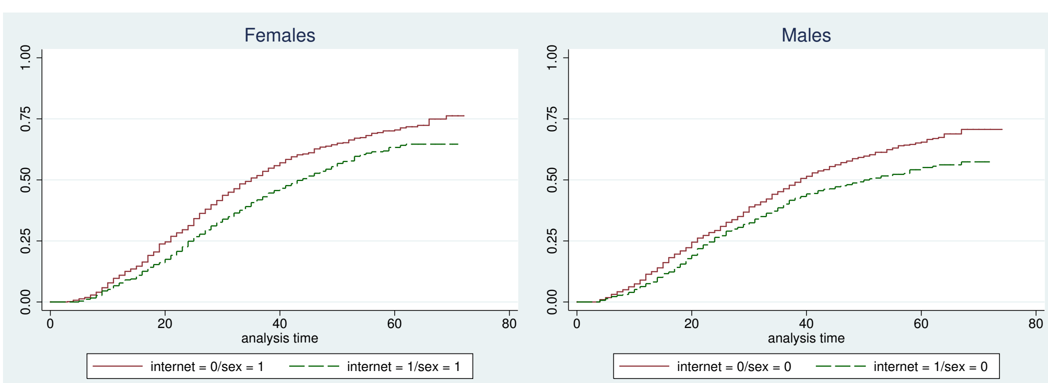


Figure 1.4: KM Hazard Rate across Race

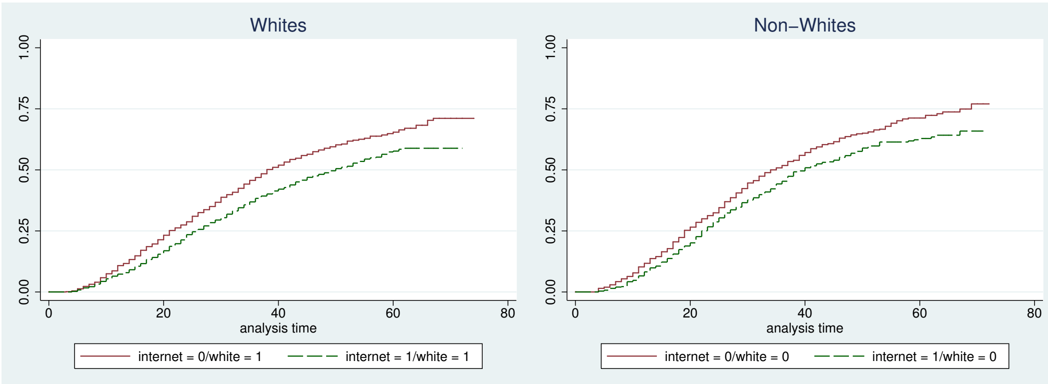
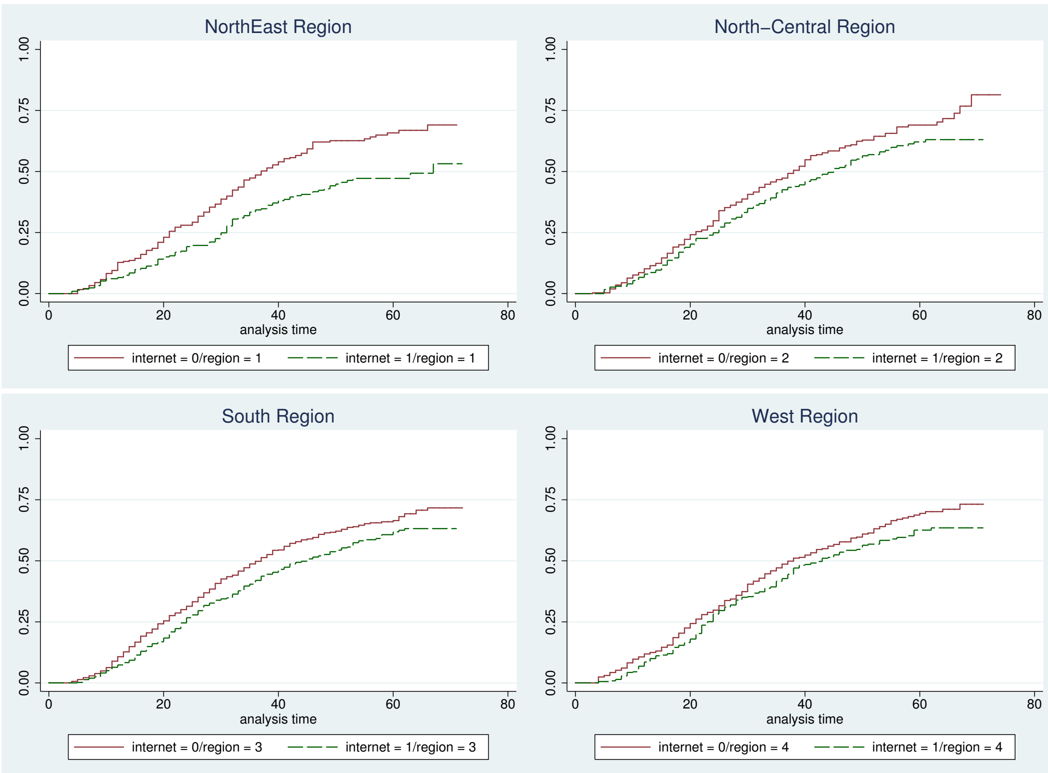


Figure 1.5: KM Hazard Rate across the 4 Census Regions



ASVAB Scores

The ASVAB is an aptitude test designed conducted by the Department of Defense, to help measure respondent abilities and predict future academic and professional success, especially in the military. The test includes ten power and two speeded subtests that measure proficiency in the following fields- Arithmetic Reasoning, Assembling Objects, Auto Information, Coding Speed, Electronics Information, General Science, Mathematics Knowledge, Mechanical Comprehension, Numerical Operations, Paragraph Comprehension, Shop Information and Word Knowledge. The scoring for the exam is based on an Item Response Theory (IRT) model. This model enables tests to be based on an examinees ability level and scores to be on the same scale irrespective of the components answered. The final ability is computed using a three-parameter logistic (3PL) model where the three parameters are - difficulty, discrimination and guessing. At the beginning of the exam, all examinees are assigned an initial score of zero (the expected mean of examinee abilities). After each response, this ability estimate is then updated, using a sequential Bayesian procedure. Once the test is completed, final ability is computed as the mode of the updated ability estimate. The final ability is then converted to a standard score on the ASVAB score scale. More information on the score can be obtained from the ASVAB website.

The ASVAB Math Verbal score provides a summary percentile score variable created by the NLS staff based of four key subtests of the twelve components mentioned above. The score is “created by first grouping respondents into three-month age groups- the oldest cohort included those born from January through March of 1980, while the youngest were born from October through December 1984. Custom sampling weights were then computed for the entire sample of respondents who had scores on all four exams and were assigned to each respondent’s scores. Within each three-month age group and using the sampling weights, NLS staff then assigned percentiles for the scores for the tests on

Mathematical Knowledge (MK), Arithmetic Reasoning (AR), Word Knowledge (WK), and Paragraph Comprehension (PC) based on the weighted number of respondents scoring below each score. Percentile scores for WK and PC were added to get an aggregate Verbal score (V) for which an aggregated intra-group, internally normed percentile was then computed. The percentile scores for MK, AR and two times the aggregated percentile for V were then summed. Finally, within each group NLS staff computed a percentile score, using the weights, on this aggregate score, yielding a final value between zero and 99. Although the formula is similar to the AFQT score generated by the Department of Defense for the NLSY79 cohort, this variable reflects work done by NLS program staff and is neither generated nor endorsed by the Department of Defense" (NLSY97 Appendix 10).

Table A1: First Stage Results

Variable	Coefficient	S.E.
C_{03}	0.46***	(0.06)
Sex (Female=1)	0.06***	(0.02)
White	-0.03*	(0.02)
Highest Education	0.03***	(0.00)
Married (=1)	0.003	(0.02)
Urban Indicator	0.04	(0.02)
NorthEast	-0.04	(0.03)
North Central	0.00	(0.03)
South	-0.04*	(0.02)
Industries		
Agriculture, Forestry & Fisheries	-0.18	(0.14)
Mining	-0.16*	(0.10)
Utilities	-0.07	(0.13)
Construction	-0.01	(0.07)
Manufacturing	0.01	(0.06)
Trade	-0.02	(0.06)
Transportation & Warehousing	0.06	(0.08)
Finance, Insurance, & Information	-0.03	(0.06)
Services Industry	-0.02	(0.05)
Occupations		
Management Related	0.11***	(0.05)
Professional Specialty	0.05	(0.04)
Technical, Sales & Admin. Support	0.07*	(0.04)
Service Occupations	-0.01	(0.04)
Farming, Fishing, & Forestry	0.03	(0.15)
Precision Production, Craft, & Repair	0.00	(0.05)
Constant	-0.18	(0.21)
Durbin Score	14.10	
Wu-Hausman F(1,2859)	13.86	
R^2	0.16	
Adj. R^2	0.15	
Partial R^2	0.024	
Robust F	65.82	
Sample Size	2,922	

Results include dummies for Job Start Period

***Significant at 1% **Significant at 5% *Significant at 10%

CHAPTER 2

Scarring Effects of Unemployment: A Quantile Approach

2.1 Introduction

The 'great recession' has highlighted several issues, one of the most important being unemployment. With unemployment rates reaching a high of 10% in October 2009 and massive layoffs,[1] a serious inquiry into the impact of unemployment is warranted. Moving beyond the negative psychological impact (Clark *et. al.* (2001) & Theodossiou (1998)) and wages foregone, there are many other aspects of future labor market experience that unemployment can significantly affect. First, it can negatively affect the offer arrival rate. Multiple papers in the search-matching literature have shown that the offer arrival rate for previously unemployed workers is lower when compared to workers with no past unemployment experience. Second, it can affect the re-employment wages. Unemployed workers, especially workers who have been unemployed for long, have much lower reservation wages when compared to similarly skilled employed workers. The lower reservation wages along with the 'stigma' associated with unemployment leads them to accept jobs without commensurate pay. Third, this low re-employment wage may further negatively impact their future wage growth potential. This is in some respects more harmful as it implies a scarring[2]

[1] The Bureau of Labor Statistics reported 3,059 mass layoff actions in February 2009, resulting in more than 0.32 million jobless workers in one month.

[2] Based on David T. Ellwood (1982) terminology where 'scars' represent persistent effects and 'blemishes' represent transitory effects.

effect.

Multiple papers have investigated the effect of the unemployment rate on re-employment wages and future wage growth. Arulampalam (2001) using the British Household Panel Survey (BHPS) data finds that a spell of unemployment leads to 6% lower re-employment wages and 14% lower wages three years after the unemployment experience. Gregg and Tominey (2005) find that early unemployment (between the ages of 16 and 23) leads to a wage scar of 13%-21% at age 42. This negative effect reduces to a 9%-11% range, if there is no repeat incidence of unemployment. Ruhm (1991) looks at displaced workers and finds that compared to non-displaced workers, they are twice as likely to experience wage reductions exceeding 25%. Couch and Placzek (2010) find that earnings losses of displaced workers are initially more than 30% and remain above 15% even six years later. Jacobson *et. al.* (1993) find that high tenure workers separating from distressed firms suffer long-term wage losses averaging 25% per year. On the other hand, Gregory and Jukes (2001) using a representative sample of British men find that there is only a 10% initial wage re-engagement wage difference which disappears within two years.

Studies have also looked at the effect of unemployment on the job-offer rate. Blau and Robins (1990) find that the offer rate per contact is greater for employed searchers than for unemployed searchers. This difference in the offer rate may be due to differences in the effectiveness of search while employed or it could be attributed to unobserved differences in search effort. Burgess (1993) shows that employed job searchers provide competition for the unemployed. In his model, when hiring rises, the employed searchers crowd out the unemployed from the new jobs. On the other hand, Eriksson and Rooth using unique data from a field experiment in the Swedish labor market, find no evidence of recruiting employers using information about past unemployment to sort workers. They however find some evidence of recruiters using contemporary unemployment as a sorting criterion.

In this paper I attempt to evaluate whether past unemployment has a negative effect on wages, beyond the initial re-employment wage. Using linked Current Population Survey (CPS)[3] data I estimate if there exists a wage difference between the previously unemployed (henceforth referred to as PU) workers and workers who do not have a recent history of unemployment, one year after the unemployment spell. Using quantile regression and quantile decomposition methods developed by Machado and Mata (2005), I estimate and decompose this wage difference at the quantile level. Decomposition is done at the quantile level which helps estimate the marginal effect of a covariate on log wages at various points in the distribution, and not just at the mean. Details on the methodology are presented in Section 2.3. To my knowledge the decomposition approach at the quantile level using the panel nature of CPS has not been done before. Thus this paper intends to contribute to the growing body of literature related to the wage effects of unemployment, by providing a more detailed distribution of the wage difference using a different panel dataset.

I find that in 2010-11, PU workers get anywhere between 9% to 25% [4] lower wages and that at least 47% of this wage difference against the PU workers is due to the difference in the rewards to characteristics. One argument made in favor of the lower wages stems from the potential skill loss that employers might perceive to occur during this period of unemployment (Pissarides (1992)). To control for this, workers are divided between short term PU and long-term PU workers, where the short-term PU workers have an unemployment spell of less than three months. It is hard to hypothesize any major skill losses in a three-month period. Moreover, the analysis evaluates the wage difference at least one year after any unemployment experience. It is even more difficult to imagine a skill loss severe enough to lead to lesser wages one year after re-employment. For the long-term PU workers, I restrict the unemployment duration to be less than a year. While there could be some potential skill loss in this duration, it would be difficult to explain very large wage

[3] The Current Population Survey (CPS) is administered by the Census Bureau using a probability selected sample of about 60,000 occupied households from all 50 states.

[4] The range comes from the differences in the duration of unemployment and the quantile level of the wage distribution.

differences one year after any such experience.

Secondly it may be argued that this analysis focuses on the very recent period post recession, and hence the wage difference could be due to other macroeconomic factors, unrelated to unemployment. Thus in addition to 2010-11, my analysis considers two additional periods (1998-99 and 2005-06). Both these periods are before the current recession and the last period (1998-99) is sufficiently far from the other two periods to have any common economic characteristics. I find that in 2005-06, the unexplained component of the wage difference ranges between 53% - 73% and in 1998-99 the unexplained component accounts from anywhere between 35%- 58% of the estimated wage difference. Thus, not only is a significant component of the wage difference unexplained, but it also exists across all the three periods in the analysis. Detailed results are discussed in Section 2.4.

It is important to note that these numbers may be upward biased and that the entire difference cannot be attributed to the scarring effect of unemployment. Unemployed workers can be both good and bad in terms of productivity and the analysis will include workers with low productivity. In this situation, unemployment is a good signal of their lower skills and the wages they receive are commensurate to their skill level. It can be said that employers would have discovered the true productivity of these workers and they would have received the lower wages irrespective of the unemployment spell. Including such workers in the sample causes an upward bias in the effect of unemployment. Unfortunately there is no way to separate good workers from bad workers and it is impossible to separate out that portion of the difference in wages due to scarring.

Another drawback of the current study is the absence of accounting for endogeneity/self-selection. Given that unemployment is not a random occurrence, sample selection is a valid concern. A few recent studies have attempted to account for sample selection when implementing quantile decomposition techniques (Albrecht, Vuuren and Vroman (2009), Chzhen *et. al.* (2012)). These studies apply a semi-parametric adaptation of the Heckman parametric procedure for quantile wage regressions, as proposed by Buchinsky (1998). However, given the lack of viable instruments and more importantly, the problem of the identification of the intercept of the wage equation,[5] the current analysis restricts its scope to the estimation of the uncorrected wage differential. Despite these concerns, the large wage differentials suggest the presence of some scarring effect.

2.2 Data & Summary Statistics

The data for the current analysis are drawn from the CPS for three distinct periods, each one lasting two years. In essence, I follow a worker for two years and try and estimate if his past unemployment has any effect on wages at least one year after the spell. The panel nature of the dataset is enabled by the survey methodology employed by CPS. Every household that enters the CPS is interviewed each month for 4 months, then not interviewed for the next 8 months and interviewed again for 4 more months. Every month the survey asks workers about their employment details for the past 52 weeks and the same person is then again asked the same questions, after 8 months. Linking the workers across the surveys provides us with the worker's employment history for the last two years. For example if a worker is first interviewed in the month of March 2009, then the survey will record his employment and other details for the last year (from March 2008- March 2009). This worker is then again interviewed in March 2010 and the survey records his employment details from March 2009- March 2010. By linking across time periods I can then reconstruct the worker's employment history from March 2008- March 2010. To create a match, the Unique Person Identifier and the Household Identification Number provided by CPS are used. However, a household may move and sometimes the same identifier is attached to the new household. To control for incorrect matches, additional controls on age, race,

[5]This is due to the conflation of the intercept in the wage equation with the constant term associated with the power series approximation of the selection term (Buchinsky (1998))

marital status, and educational qualification are included. The earliest data for this analysis cover the period from March 1997- March 1999, the second dataset ranges from March 2004- March 2006 and the latest dataset ranges from March 2009 - March 2011. March is selected to capture the additional wage information collected in March and the three distinct periods are selected to investigate if the wage difference against the PU has been present over time.

One of the most important aspects of this matched dataset is that it includes information on all weeks worked in the last two years. Using this data, I can calculate if a worker had a spell of unemployment and the exact duration of the spell. The analysis differentiates between short and long unemployment spells. All workers who had an unemployment spell of less than 3 months in the last year are categorized as short-term unemployed (PU-S). Workers who were unemployed for more than 3 months but less than 11 months[6] are included in the long-term unemployed category (PU-L). Thus, the PU-S category includes workers who were unemployed for less than 3 months last year and employed for a full 52 weeks in the current year. The PU-L category includes workers who were unemployed for any period ranging from 3-10 months in the previous year and employed for 52 weeks in the current year. The reference group (or the Not PU category) includes workers who worked a full 52 weeks in both years. The data only include male workers[7] who are currently employed as of the interview date and were between the ages of 22 & 60. Workers who had not worked for profit or pay, unpaid family workers, self-employed workers and armed force personnel are excluded. For the period 2009-2011, the dataset includes 5,952 workers who were employed for 52 weeks in both years, 380 PU-S workers and 280 PU-L workers. For 2004-2006, there are 6,269 Not PU workers, 346 PU-S workers and 190 PU-L workers. In 1998-1999, the dataset includes 5,870 Not PU workers, 372 PU-S workers and 198 PU-L

[6]The unemployment duration is restricted to less than 11 months to ensure workers were employed for some period last year and have not been employed for durations which cannot be accounted for.

[7]Female workers were excluded to simplify the analysis as their work history may include interruptions related to family and other household concerns.

workers.

Table 2.1 provides the means of some of the variables included in the analysis, calculated separately for the Not PU, PU-S and PU-L workers, for the three periods. Across all three periods, the hourly log wages of the previously unemployed workers are much lower than those of the Not PU workers. The PU-L wages are even lower than the PU-S wages, suggesting a link between unemployment duration and wages. The means on education suggest that the previously unemployed workers have observable characteristics that are associated with lower skills/wages. Across all three years, a higher percentage of the Not PU workers have completed college and the PU-L workers have lower college completion rates, even when compared to the PU-S workers. This low education could also be a reason behind the longer spells of unemployment, as it would make the PU-L workers less attractive to employers. Another reason behind the low earnings for the PU workers could be the higher incidence of switching in both industry and occupation. Past research suggests that a worker may face wage losses when switching due to the loss of value of skills obtained earlier but not required in the new field. In 2010-2011, around 30% of the PU workers switched occupations and more than 20% of the PU workers switched industries. This incidence of switching is even higher in the other two periods for the PU-L workers. Across all three periods the mean potential experience of the Not PU workers is approximately 3 years more than that for PU-L workers, which could again lead to the higher wages earned by them. Also a significantly higher percentage of the Not PU workers are married. The percentage of white and black workers is roughly the same across the three categories in all years except 1998-1999.

I calculate mean wages across education, marriage status, race, potential experience, industry and occupation for the Not PU and PU workers, for the three periods (Table 2.2, 2.3 and 2.4). Almost across all categories and periods PU workers have lower wages. For 2010-2011 (Table 2.2), the average wages is low even for PU-L workers who have completed college. Married workers earn more on average, but again the married PU-L workers earn less than the unmarried Not PU workers. The wage difference holds across race, with white Not PU workers earning the highest on average. Surprisingly, black PU-S workers earn less than black PU-L workers. The wage difference also holds across the different groups of potential experience. Across all industry categories Not PU workers earn the most with some of the largest difference in wages relative to the PU-L workers in the Finance and Public Administration sectors. Except for Hospitality and Other Services, PU-S workers earn more than the PU-L workers. Not PU workers have higher wages in all occupations except Farming. There is only one worker in the PU-S category, which might be distorting this number. Also except for Production, PU-S workers earn consistently higher wages than the PU-L workers.

Similar to the setting in 2010-2011, in 2005-2006 (Table 2.3), PU-L workers earn significantly lesser than the Not PU workers. For example, on average PU-L workers with a college degree earn less than Not PU workers without a college degree, on average. Additionally, white PU-L workers earn less on average than black Not PU workers. Similar to 2010-2011, Not PU workers earn the most across all industry categories with some of the largest difference in wages with the PU-L workers in the Finance sector. Similar trends are found in the 1998-1999 data (Table 2.4), with Not PU workers earning the most on average. The industry and occupation categories are not included as they are not similar to the classification above.[8]

From the three tables, we can observe that on average, unemployment is linked to lower wages, and that this wage difference holds across multiple categories and different periods. Also duration of unemployment has a strong effect on wages, with PU-L workers earning consistently lower amounts than both Not PU and PU-S workers. I next use measures beyond the mean to obtain a more detailed description of the wage distribution. Table 2.5

[8]The Industry and Occupation categories follow the 1990 classification

provides the wages at each of the quantiles for the three categories of workers across three periods. It can be seen that the wages of the Not PU workers are higher across all quantiles and years. This difference increases as the duration of unemployment increases. Except for the 10^{th} and 20^{th} percentile in 1998-99, the wages of PU-L workers are always lower than that of the PU-S workers. Additionally, there are different patterns of wage difference as we move from the lower to the upper quantiles, across the three periods. In 2010-11, for both the PU-S and PU-L workers, the wage difference in the lower quantiles is much larger suggesting that those in the left tail of the wage distribution suffer a much higher impact. For PU-S workers this trend holds also in the 1998-99 period. However, for the PU-L workers, in both the 2005-06 and the 1998-99 period, the higher quantiles experience a much larger wage difference. PU-L workers beyond the 50^{th} percentile, have much lower wages when compared to the Not PU workers. This suggests a larger wage impact for the right-tail PU-L workers. Figure 2.1 plots the wage difference between the PU and Not PU workers, for the three periods. At a glance, it is evident that PU-S workers have larger differences at the left tail while the PU-L workers beyond the 50^{th} percentile suffer more. This rough measure suggests that an analysis at the means (to identify the effect of previous unemployment spells) is insufficient. The quantile split suggests that workers at different ends of the wage distribution are affected differently. Figure 2.2, 2.3 & 2.4 plot the kernel wage density for the PU-S and PU-L workers separately. Similar, to what was observed from Table 2.5, these three figures clearly indcate that the difference in the wage density between the NPU and PU-L workers is much larger than the difference between the NPU and PU-S workers.

The next section outlines the modeling technique to be used in the paper to get a robust and detailed measure of the impact of a period of unemployment.

2.3 Methodology

The significant wage differences across time periods, between the PU and Not PU workers suggests the need for sophisticated modeling. To correctly estimate the impact of past unemployment on wages a flexible modeling structure that can control for all influential covariates is required. The econometric methodology followed in this paper consists of evaluating this difference at the quantile level. Following from the results reported in Table 2.5, previous unemployment differently affects the tails of the distribution. A quantile analysis will help analyze if the difference in the effect of previous employment at the tails is significant. To estimate the wage difference, I first estimate the coefficients for the PU dummies at different quantiles. This will help us answer two questions. First, I can estimate if the effect of previous unemployment is significant across quantiles. Second, I can analyze if previous unemployment has different effects at the high/low wage levels. The second step of the analysis involves decomposing this wage difference. If however, the PU-S and PU-L variables have no impact in quantile regression, then it is safe to conclude that further decomposition would yield no significant results.

2.3.1 Quantile Regression

Quantile regression as introduced by Koenker and Bassett (1978) are models in which quantiles of the conditional distribution of the response variables are expressed as functions of observed covariates. It is a technique for estimating the θ_{th} quantile of a log wage conditional on covariates. The quantile regression model assumes that q_θ is linear in x; that is, $q_{\theta(y|x)} = x'\beta(\theta)$, where q_θ is the θ_{th} quantile of the random variable y (Ln(Wages)) and provides a full characterization of the conditional distribution of wages. The coefficient vector $\beta(\theta)$ tells us the effect of any particular factor on the dependent variable in the θ_{th} quantile. It is estimated as the solution to:

$$\min \sum q_\theta(y_i - f(x_i\beta_{(\theta)})$$

Assuming a linear function, the quantile regression estimator for q_θ minimizes the following objective function:

$$q_{\theta(y|x)} = \sum_{y_i < x_i'\beta}^{N} q|y_i - x_i'\beta_\theta| + \sum_{y_i > x_i'\beta}^{N} (1-q)|y_i - x_i'\beta_\theta|$$

This non-differentiable function is minimized via the simplex method. In contrast to ordinary least squares, quantile regression minimizes $\sum|e_i|$ and is also referred to as least absolute deviations regression. In addition to providing a richer description of the data, quantile regression estimates are more robust against outliers relative to the ordinary least squares regression. Also quantile regression is semi-parametric, as it makes no assumptions about the distribution of the error process. The simplex estimate is equivalent to maximum likelihood estimate when the error term follows a Laplace distribution.

2.3.2 Decomposition

Based on Machado and Mata (2005), quantile regression techniques can be used to decompose the difference between the PU and Not PU workers log wage distributions into a component that is due to differences in labor market characteristics and a component that is due to differences in the rewards that the two groups receive for their labor market characteristics. This decomposition is in the spirit of the Oaxaca (1973) and Blinder (1973) technique, except that, rather than identifying the sources of the differences between the means of two distributions, the sources are identified at each quantile. The idea is to generate two counterfactual densities $f(y_s|x_s)$ where s represents the state of the worker. The two densities are:

- The log wage density that would arise if the PU workers were paid according to the Not PU characteristics but with PU returns $[f(y_{PU}|x_{NPU})]$ and
- The log wage density if PU workers retained their own labor market characteristics but were paid like the Not PU workers $[f(y_{NPU}|x_{PU})]$.

$f(y_{PU})|x_{NPU})$ along with the Not PU wage density $[f(y_{NPU}|x_{NPU})]$ would provide the difference in wages due the differences in characteristics.[9] $f(y_{NPU}|x_{PU})$ along with the actual PU wage distribution $[f(y(_{PU}|x_{PU})]$ will compute the difference in wages due to differences due to the differences in rewards. The Machado-Mata approach to estimating the first density is as follows:[10]

- Draw m numbers at random from (0,1), $(\theta_1, \theta_2.....\theta_m)$
- From the PU dataset, estimate the quantile coefficient vectors, $\beta_{PU}(\theta_i)$, for i=1,...,m.
- Make m draws at random with replacement from the Not PU dataset, x_i^{NPU}
- The counterfactual density $f(y_{PU}|x_{NPU})$ is then: $x_i^{NPU} * \beta_{PU}(\theta_i)$ for i=1,...,m

To generate the second density I follow similar steps as above, but now the quantile regression coefficient vectors are drawn from the Not PU dataset $(\beta_{NPU}(\theta_i))$. Also, the m[11] random draws with replacement are made from the PU dataset (x_i^{PU}). Thus,

[9] All variables included in the quantile regressions are included as controls in the decomposition exercise

[10] Machado and Mata (2005) prove that the four-step sampling procedure described below gives a consistent estimator of the counterfactual distribution

[11] Following from the Machado and Mata (2005) and Albrecht, Vuuren and Vroman (2009) paper, I use 1000 replications

the counterfactual density $f(y_{NPU}|x_{PU})$[12] is then: $x_i^{PU} * \beta_{NPU}(\theta_i)$. From the linearity assumption of quantile regression, the following result then holds true:

$$x_{NPU}\beta_{NPU}(\theta) - x_{PU}\beta_{PU}(\theta) = (x_{NPU} - x_{PU})\beta_{NPU}(\theta) + (\beta_{NPU}(\theta) - \beta_{PU}(\theta))x_{PU}$$

where;

$(x_{NPU} - x_{PU})\beta_{NPU}(\theta)$ is the explained component or the component of the difference that can be attributed to the difference in characteristics of the workers, and,

$(\beta_{NPU}(\theta) - \beta_{PU}(\theta))x_{PU}$ is the unexplained component or the component of the difference that can be attributed to the difference in the rewards that the two groups receive for their labor market characteristics.

2.4 Results

2.4.1 Quantile Regressions Results

This section discusses the results obtained for the pooled quantile regressions with dummies for the PU-S and PU-L categories (Table 2.6 and Table 2.7). These pooled regressions impose the restriction that the returns to included labor market characteristics are the same for the two groups. The estimated PU dummy coefficients in these regressions then indicate the extent to which the wage gap is explained through previous unemployment at the various quantiles after controlling for differences in characteristics. The control variables include gender, race, education, potential experience, firmsize, marital status, state, industry and occupational dummies as controls. The coefficients have been obtained for the 10^{th}, 25^{th}, 50^{th}, 75^{th} and 90^{th} percentiles. For comparison purposes the OLS estimate is also calculated.

[12]Note that the counterfactual densities have been calculated separately for the PU-S and the PU-L workers. The text uses PU to simplify notation.

Focusing first on the PU-S coefficient (Table 2.6), it is clear that across years and quantiles, even after controlling for other factors, the effect is negative and statistically significant. In 2010-11, male PU-S workers in the 25^{th} percentile had 16% lower wages when compared to similar workers without a previous spell of unemployment. This percentage decreases to 11% in the 50^{th} percentile and then again rises in the 90^{th} percentile to 18%. In 2005-06, up to the 25^{th} percentile, PU-S workers have 15% lower wages while PU-S workers beyond the 90^{th} percentile have 17% lower wages. In 1998-99 PU-S workers in the 10^{th} percentile have 15% lower wages, but beyond the 50^{th} percentile, the wage difference is around 9%. Hence, it seems that the tail ends of the distribution suffer the most and except in 1998-99, this negative effect is largest in the 90^{th} percentile. It is also relevant to note that this negative effect of previous unemployment is present even in the oldest data and that the coefficients are similar across periods. Thus it is probably safe to assume that the scarring effects of previous unemployment were present even before the current recession. The OLS coefficient helps convey the relevance of a quantile analysis. There is significant variation in the negative effects of previous unemployment across quantiles, and an analysis at the means might have failed to reveal the workers who suffer the most.

Moving to the PU-L coefficients (Table 2.7), the coefficients are again negative and statistically significant across all years and quantiles. The negative effects are much larger in the lower quantiles in most cases. The negative difference at the 90^{th} percentile in 2010-11 was statistically not significant. In the other two years the coefficients are statistically significant at all quantiles, but the negative effect of previous unemployment seems much stronger for workers with lower wages. In 2010-11, workers up to the 50^{th} percentile receive roughly 18% lower wages when compared to similar workers without a previous spell of unemployment. This wage difference is larger in 2005-06, where PU-L workers up to the 25^{th} percentile receive more than 30% lower wages. In 1998-99 the wage difference against the PU-L workers up to the 25^{th} percentile is 22%. Again, the OLS coefficient misses this crucial difference in the impact of previous unemployment across quantiles. For example the OLS coefficient is mostly being driven by the low-wage workers and misses the lower negative effect on the high wage workers.

Comparing the PU-S and PU-L coefficients, it is clear that there is one distinct difference. While the PU-S workers seem to be suffering at both tails of the distribution, the PU-L low-wage workers suffer the major brunt of the negative effect. In addition to this, the negative coefficients are much larger for the PU-L workers, implying a much larger scarring effect. While this analysis does not decompose the wage difference into explained and unexplained components, it can be observed that the negative effect of previous unemployment on wages is present and significant even one year after the experience.

Looking at the other variables, the signs of the coefficients are roughly the same for the two categories of workers in all three periods. One of the reasons behind the similar effects is that the comparison group or the Not PU workers are the same for both the PU-S and PU-L workers. Given that only a small portion of the workers are categorized as previously unemployed, the major effects across variables is being driven by the Not PU workers, resulting in similar coefficients. Education has a positive effect on wages, with one additional year increasing wages by 4% for workers up to the 25^{th} percentile but the effect is not statistically significant in the higher quantiles.[13] Potential Experience[14], has a small positive, statistically significant and constant effect. A one-year increase in experience leads to a 3% increase in wages. Married white male workers earn more. The coefficient on white is not always statistically significant but specifically has a strong positive impact for both PU-S and PU-L workers in the 10^{th} percentile. Married workers on average receive 15% higher wages. Firm size has a positive effect on wages, with larger firms paying 5% higher in wages, while the number of employers a worker has had in the past has a negative effect,

[13] In line with the *Mincerian* wage equation Mincer (1974), the analysis includes both education squared and potential experience squared terms

[14] Potential Experience is calculated as Age- 6 - Years of Education based on Mincer (1974)

especially for the low-wage workers. For PU-S workers in the 10^{th} percentile one additional employer can lead to wages being reduced by 10% but the negative effect disappears in the higher quantiles. Similarly, both Industry Switch and Occupation Switch indicators have negative coefficients in the lower quantiles, but the effect is not statistically significant in the higher quantiles.

2.4.2 Quantile Decomposition Results

I now explore the results of the quantile regression decomposition of the wage gaps, following Machado and Mata (2005). Tables 2.8 and 2.9 report the results of the quantile decomposition exercise for the PU-S and PU-L workers respectively, for the three periods. For each period, four columns are reported. The 1^{st} column reports the raw wage difference at a particular quantile. The 2^{nd} column reports the predicted difference. This predicted difference is then divided into the explained component or the difference attributable to differences in characteristics (3^{rd} column) and the unexplained component or the difference due to differences in rewards (4^{th} column). Figure 2.5 provides a graphical representation of the absolute values for the 2^{nd} (Predicted wage difference) and 4^{th} column (Unexplained component) for the PU-S workers, while Figure 2.6 provides a similar graphical representation for the PU-L workers.

Focusing first on the predicted wage difference for the PU-S Workers, in 2010-11 the wage difference falls as we move along the quantiles (from the 5^{th} to the 20^{th} percentile) and then stabilizes around that value until about the 70^{th} percentile, after which the wage difference steadily increases. The variation is large and PU-S workers receive anywhere from 9% to 12% lower wages. In 2005-06, the wage difference follows a similar trend, except that it stabilizes at a value of -0.25, implying higher differences across quantiles. PU-S workers receive at least 11% lower wages when compared to NPU workers. In 1998-99, the wage difference initially rises up to the 35^{th} percentile, after which it steadily drops. This trend can easily be spotted using Figure 2.5. More importantly, when we look at the differences-in-reward graph of Figure 2.5, it is evident that a large portion of the difference is due to the differences in reward. Using the quantile wages, the decomposition results indicate that the anywhere between 4% to 10% of the wage difference can attributed to the differences in rewards between the PU-S and NPU workers. The unexplained component decreases as we move upward across quantiles, suggesting that low-wage workers, especially workers below the 40^{th} percentile, suffer more due to a previous spell of unemployment. An important characteristic is visible when we focus on the 3^{rd} column for each period, in Table 2.8. For the three periods, the contribution of differences in characteristics to the wage difference remains mostly constant across quantiles. This implies that almost all the variation in the wage difference across quantiles is due to the difference in rewards. This is also made evident when looking at Figure 2.5, the predicted wage difference and difference in reward plots follow a similar trend.[15]

Next I look at the PU-L coefficients (Table 2.9 and Figure 2.6). First, it is clear that the wage difference across quantiles is now much larger when compared to the PU-S wages. Across the three periods, PU-L workers receive at least 12% lower wages and this difference can go up to 26% (2005-06). Similar to the PU-S workers, the wage difference is much larger in the lower quantiles and starts dropping as we go higher. In fact in 1998-99, there is a steady drop in the wage difference once we move beyond the 30^{th} percentile. Again, the largest wage difference is observed in the 2005-06 period. Moving to the differences in reward plot, it is clear that again the differences in rewards account not only for a significant portion of the wage difference, but also accounts for almost all of the variation across quantiles. In 2010-11, workers below the 30^{th} percentile receive at least 16% lower wages and at least 52% of this difference is due to the difference in rewards. Even beyond the 30^{th} percentile

[15]The y-axis is the same on both graphs and the contribution of the unexplained component to the overall wage difference can be easily observed

differences in reward lead to 7% lower wages for the PU-L workers. In 2005-06, the wage difference is highest, with PU-L workers below the 30^{th} percentile getting less than 80% of the NPU wages. More than 55% of this lower wage can be explained due to the difference in rewards. In 1998-99, PU-L workers get at least 12% lower wages and more than 40% of this difference can be attributed to the difference in rewards. What is most important across the three periods is that the unexplained component is negative and statistically significant across all quantiles and explains a large portion of the wage difference.

Comparing the decomposition results to the quantile regression results obtained earlier, it is clear that the unexplained component leads to a lower wage difference. Also the higher difference at the tails for the PU-S workers disappears in the decomposition analysis. Now for both PU-S and PU-L workers, the left-tail workers suffer a much larger impact and the negative effect diminishes as we move to the higher quantiles. While the absolute size of the unexplained component is larger in the higher percentiles, the much higher wages lead to relatively lower levels of explanatory power. What is more important to note is that the decomposition results are in line with some of the previous studies. In my analysis the unexplained component accounts for at least 4% lower wages for the PU-S workers and at least 6% lower wages for the PU-L workers. As noted before, Arulampalam (2001) found re-employment wages to be 6% lower. This suggests that a large portion of the re-employment wage difference persists even after one year. This persistence effect is more troubling when we focus on the low-wage workers, where, workers below the 30^{th} percentile are getting at least 9% lower wages. This result also lies in the 9% - 11% range provided by Gregg and Tominey (2005) as the wage difference against the PU workers lies in the range of 4% to 16%.[16] However, compared to the 25% wage difference suggested by Ruhm (1991) and the 30% difference suggested by Couch and Placzek (2010), the numbers in this analysis are noticeably lower.

[16] The larger ranges can be attributed to quantile analysis.

The most significant result to note from the decomposition results is that the unexplained component or the difference in rewards for similar characteristics contributes to a large portion of the wage difference. The wages I decompose in this analysis are wages that the workers receive, at least one year post any unemployment experience. Moreover, the difference is larger for workers who faced a longer spell of unemployment. While it is difficult to account for unobservable ability, this large difference in wages one-year post the unemployment experience does suggest a scarring effect.

2.5 Conclusion

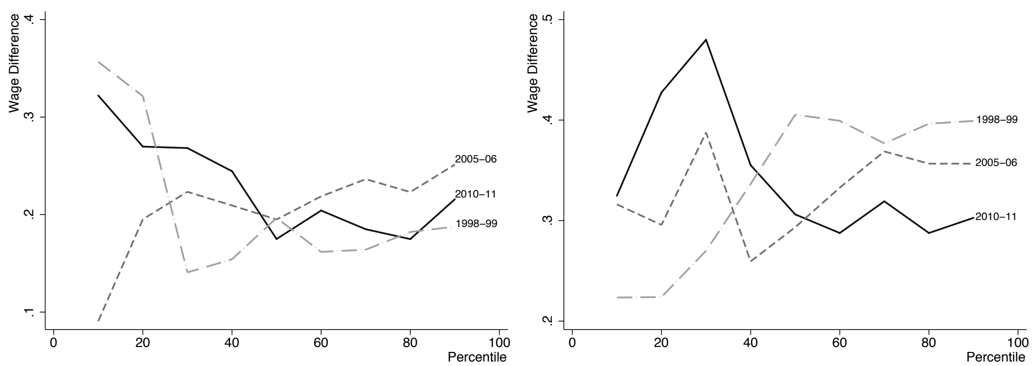
Unemployment spells can have large negative effects on workers. Not only does it lead to a wage loss in the current period but also it can negatively impact their future labor market experience. Employers perceive unemployment as a signal of low productivity and this may lead to a low offer rate for jobs, low wages at the point of re-employment and a depressed wage growth over time. Exploiting the panel nature of CPS, I follow a worker for two consecutive years. Workers are classified as PU if they had a spell of unemployment in the previous year and have been employed for more than 50 weeks in the current year. The analysis then attempts to estimate and decompose the wage difference between the PU and Not PU workers, where the Not PU workers have not had any spell of unemployment in the last two years. To capture the effect of the length of unemployment, previously unemployed workers are divided into short and long term unemployed. Additionally, three different sample periods are also considered. Applying the Machado Mata (2005) quantile decomposition methodology, I find that for 2010-11, difference in rewards explains anywhere from 57% to 73% of the wage difference against the PU-S workers. For the PU-L workers, the workers face larger wage differences. In 2010-11, 47% to 63% of the wage difference can be attributed to a difference in rewards.

Among other findings, the quantile regression results show that education and experience have a positive effect on wages. A one-year increase in potential experience leads to a 3% increase in wages. There is some variation in the effect across quantiles. Firm size has a statistically significant positive effect in the lower quantiles, but the effect decreases as a worker moves up the wage distribution. The positive effect may not be as significant for the high-income workers. An increase in the number of employers has larger negative effects in the lower quantiles. Being white leads to higher wages, but the effect is not always statistically significant. Lastly, both metro and marriage lead to a large positive and statistically significant effect across quantiles.

One of the main concerns for this analysis is the impossibility of disentangling the true productivity effect from the scarring effect. It is possible that unemployment truly signals low productivity. However as it is very difficult to estimate true productivity, any estimate of the negative/scarring effect of unemployment for a worker will also include workers who would have faced the negative experience, irrespective of his/her unemployment. The second major drawback of this paper is the absence of any accounting for endogeneity/self-selection. Due to the lack of viable instruments and the problem of the identification of the intercept term in the wage equation, the current analysis only estimates the uncorrected wage differential. Having said that, the large difference in rewards especially against the left-tail PU-S workers, suggests that some of the difference maybe attributable to scarring.

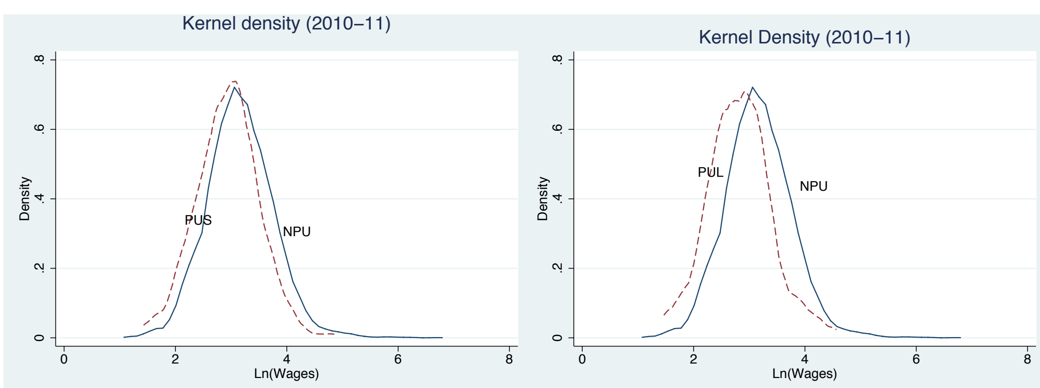
2.6 Figures

Figure 2.1: Wage Difference



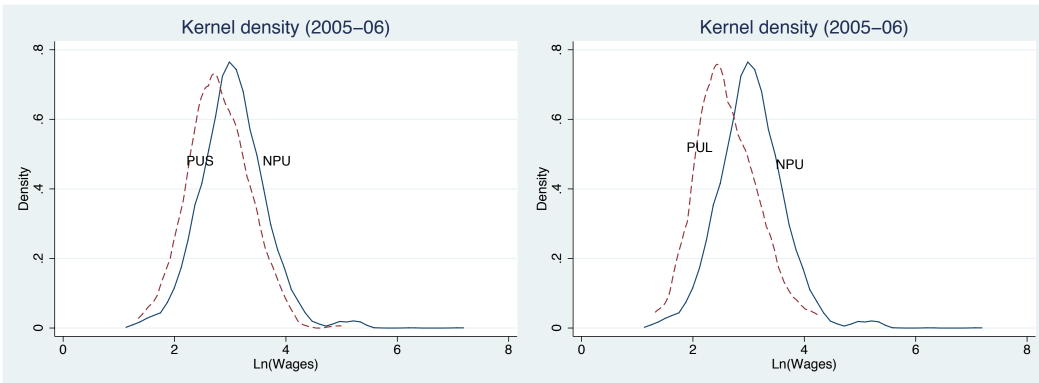
(a) PUS Workers (b) PUL Workers

Figure 2.2: Wage Density (2010-11)



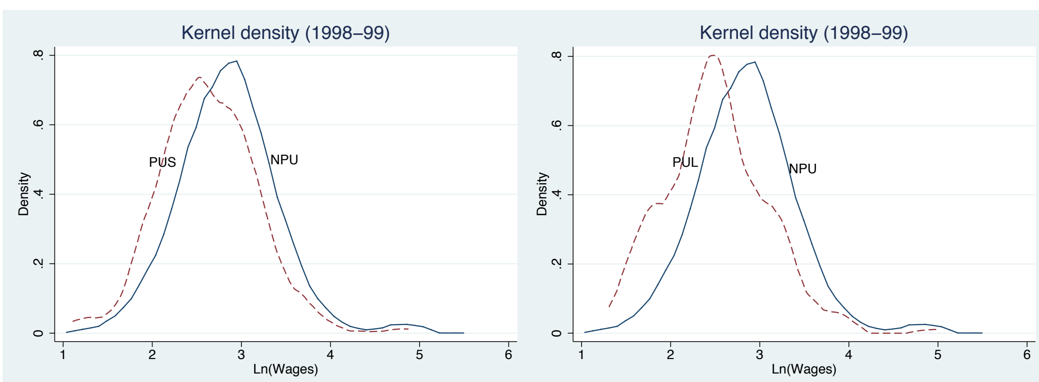
(a) PUS Workers (b) PUL Workers

Figure 2.3: Wage Density (2005-06)



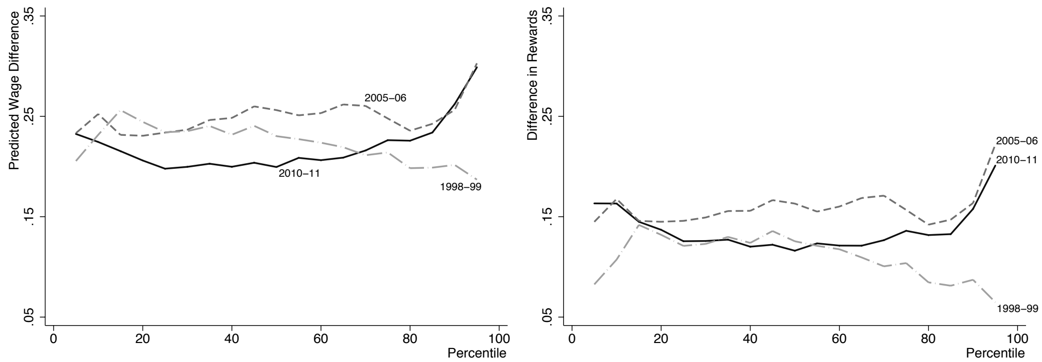
(a) PUS Workers (b) PUL Workers

Figure 2.4: Wage Density (1998-99)



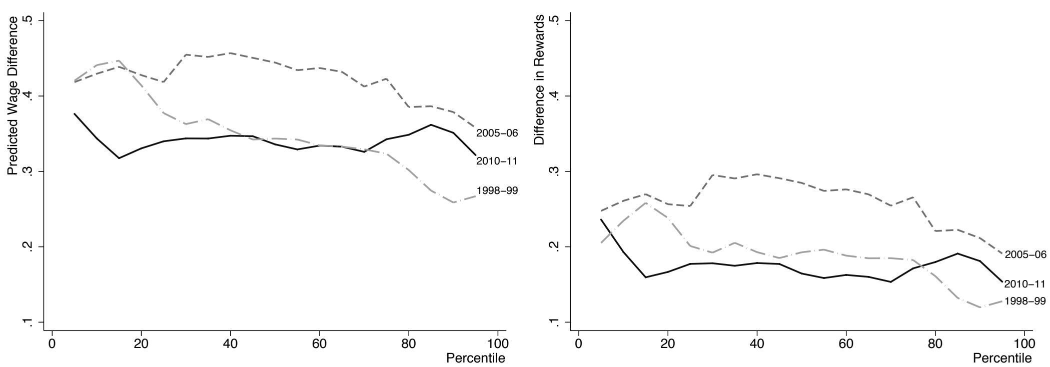
(a) PUS Workers (b) PUL Workers

Figure 2.5: Decomposition Results: PU-S



(a) Predicted Wage Difference (b) Unexplained Component

Figure 2.6: Decomposition Results: PU-L



(a) Predicted Wage Difference (b) Unexplained Component

2.7 Tables

Table 2.1: Sample Means

Variable	Not PU Mean	Not PU S.D.	PU-S Mean	PU-S S.D.	PU-S Diff†	PU-L Mean	PU-L S.D.	PU-L Diff††
Year: 2010-2011								
Ln(Wages)	3.16	0.60	2.93	0.56	0.23**	2.82	0.57	0.34**
Not Completed Coll.	0.49	0.50	0.58	0.49	-0.08**	0.61	0.49	-0.11**
Completed Coll.	0.51	0.50	0.42	0.49	0.08**	0.39	0.49	0.11**
Potential Experience	24.14	9.88	23.36	10.72	0.78	21.79	11.27	2.35**
Age	43.20	9.52	42.00	10.27	1.19*	40.15	10.75	3.05**
Married	0.71	0.45	0.65	0.48	0.06**	0.56	0.50	0.15**
Industry Switch	0.19	0.39	0.21	0.41	-0.02	0.26	0.44	-0.06**
Occupation Switch	0.27	0.44	0.30	0.46	-0.04	0.32	0.47	-0.06*
White	0.86	0.35	0.85	0.36	0.01	0.82	0.38	0.04†
Black	0.07	0.26	0.05	0.22	0.02	0.09	0.28	-0.01
Sample	5,952		380			280		
Year: 2005-2006								
Ln(Wages)	3.05	0.59	2.79	0.55	0.26**	2.62	0.56	0.43**
Not Completed Coll.	0.55	0.50	0.59	0.49	-0.04	0.64	0.48	-0.09*
Completed Coll.	0.45	0.50	0.41	0.49	0.04	0.36	0.48	0.09*
Potential Experience	24.11	9.57	22.46	10.19	1.66**	20.95	10.44	3.17**
Age	42.88	9.34	41.10	9.90	1.79**	39.42	10.22	3.47**
Married	0.75	0.43	0.65	0.48	0.10**	0.51	0.50	0.24**
Industry Switch	0.21	0.41	0.25	0.43	-0.04	0.37	0.49	-0.17**
Occupation Switch	0.27	0.45	0.28	0.45	-0.01	0.37	0.49	-0.10**
White	0.87	0.34	0.90	0.30	-0.03	0.84	0.37	0.03
Black	0.07	0.25	0.06	0.23	0.01	0.08	0.27	-0.01
Sample	6,269		346			190		

Table 2.1:Sample Means (Continued)

	Not PU		PU-S			PU-L		
Variable	**Mean**	**S.D.**	**Mean**	**S.D.**	**Diff**†	**Mean**	**S.D.**	**Diff**††
Year: 1998-1999								
Ln(Wages)	2.85	0.56	2.63	0.55	0.22**	2.51	0.57	0.34**
Not Completed Coll.	0.57	0.49	0.63	0.48	-0.06*	0.63	0.49	-0.05
Completed Coll.	0.43	0.49	0.37	0.48	0.06*	0.37	0.49	0.05
Potential Experience	22.86	9.20	21.88	10.10	0.99*	20.80	11.40	2.06**
Age	41.57	8.93	40.05	9.63	1.53**	39.12	10.49	2.45**
Married	0.75	0.44	0.68	0.47	0.07**	0.54	0.50	0.21**
Industry Switch	0.20	0.40	0.28	0.45	-0.07**	0.37	0.48	-0.16**
Occupation Switch	0.28	0.45	0.33	0.47	-0.05*	0.37	0.48	-0.09**
White	0.88	0.32	0.87	0.33	0.01	0.83	0.38	0.06*
Black	0.07	0.26	0.08	0.26	-0.001	0.11	0.31	-0.03†
Sample	5,870		372			198		

Diff† calculates the difference in mean characteristics of the NPU and PU-S and Diff†† calculates the difference in mean characteristics of the NPU and PU-L workers. **Significant at 1% *Significant at 5% †Significant at 10%

Table 2.2: Average of Ln(Wages) over Categories : 2010-2011

Categories	Not PU			PU-S			PU-L		
	Mean	**S.D.**	**Obs.**	**Mean**	**S.D.**	**Obs.**	**Mean**	**S.D.**	**Obs.**
Education									
Not Comp. Coll.	2.94	0.51	2978	2.79	0.50	222	2.71	0.55	171
Completed Coll.	3.39	0.59	2974	3.13	0.58	158	3.00	0.56	109
Marriage									
Not Married	2.98	0.57	1719	2.78	0.58	134	2.70	0.57	123
Married	3.24	0.59	4233	3.01	0.53	246	2.91	0.55	157
Race									
White	3.18	0.59	5101	2.94	0.55	322	2.84	0.58	230
Black	2.94	0.55	437	2.69	0.50	20	2.82	0.47	24
Other	3.20	0.64	414	2.96	0.67	38	2.63	0.57	26
Pot. Exp.									
≤ 15 yrs.	3.04	0.56	1210	2.87	0.57	96	2.73	0.57	88
15 - 25 yrs.	3.20	0.61	1743	2.91	0.59	99	2.72	0.55	65
25 - 35 yrs	3.24	0.61	1955	3.03	0.54	118	2.92	0.50	84
≥ 35 yrs.	3.09	0.55	1044	2.88	0.50	67	2.96	0.66	43
Industry									
Agriculture	2.58	0.50	59	2.15	0.16	2	2.05	0.68	2
Mining	3.33	0.52	76	2.88	0.43	7	2.86	0.42	7
Construction	3.12	0.54	433	3.00	0.40	54	2.89	0.54	37
Manufacturing	3.18	0.57	1020	2.96	0.52	91	2.92	0.46	40
Wh.& Rt. Trade	2.99	0.58	844	2.75	0.58	39	2.76	0.59	45
Transport & Utili.	3.17	0.53	514	3.05	0.41	22	2.80	0.51	22
Information	3.32	0.63	173	3.15	0.39	10	2.95	0.75	6
Finance	3.45	0.63	378	3.31	0.67	14	2.79	0.38	10
Prof. Services	3.39	0.64	672	2.97	0.77	45	2.86	0.72	41
Education	3.15	0.57	735	2.95	0.48	48	2.79	0.52	32
Hospitality	2.68	0.56	295	2.43	0.38	21	2.72	0.61	25
Other Services	2.90	0.50	205	2.57	0.46	9	2.87	0.32	7
Public Admin.	3.32	0.46	548	3.15	0.60	18	2.58	0.77	6

Table 2.2: Average of Ln(Wages) over Categories : 2010-2011 (Continued)

	Not PU			PU-S			PU-L		
Categories	**Mean**	**S.D.**	**Obs.**	**Mean**	**S.D.**	**Obs.**	**Mean**	**S.D.**	**Obs.**
Occupations									
Management	3.51	0.58	1118	3.33	0.51	46	3.25	0.50	30
Professional	3.38	0.54	1286	3.13	0.56	75	3.10	0.62	45
Service	2.86	0.57	711	2.49	0.47	52	2.50	0.50	48
Sales	3.16	0.64	538	3.01	0.76	23	2.63	0.53	25
Admin. Support	2.99	0.52	405	2.85	0.47	17	2.56	0.40	16
Farming	2.46	0.47	39	2.82	0.00	1	1.57	0.00	1
Construction	3.06	0.49	394	3.00	0.41	52	2.88	0.46	43
Instal. & Repair	3.04	0.44	493	3.01	0.44	25	2.97	0.51	10
Production	2.93	0.47	501	2.80	0.47	56	2.84	0.50	30
Transportion	2.87	0.51	467	2.68	0.43	33	2.66	0.52	32

Table 2.3: Average of Ln(Wages) over Categories : 2005-2006

	Not PU			PU-S			PU-L		
Categories	**Mean**	**S.D.**	**Obs.**	**Mean**	**S.D.**	**Obs.**	**Mean**	**S.D.**	**Obs.**
Education									
Not Comp. Coll.	2.84	0.50	3441	2.62	0.49	203	2.51	0.50	121
Completed Coll.	3.31	0.59	2828	3.04	0.53	143	2.80	0.61	69
Marriage									
Not Married	2.83	0.57	1542	2.69	0.54	120	2.47	0.48	93
Married	3.12	0.58	4727	2.85	0.54	226	2.76	0.58	97
Race									
White	3.07	0.59	5457	2.78	0.55	311	2.63	0.57	159
Black	2.81	0.56	411	2.95	0.54	20	2.53	0.35	15
Other	3.07	0.64	401	2.73	0.47	15	2.54	0.61	16
Pot. Exp.									
$\leq$ 15 yrs.	2.92	0.58	1189	2.72	0.53	96	2.60	0.58	63
15 - 25 yrs.	3.10	0.60	1930	2.82	0.54	99	2.58	0.54	56
25 - 35 yrs	3.10	0.59	2173	2.85	0.59	105	2.70	0.53	50
$\geq$ 35 yrs.	2.99	0.56	977	2.76	0.49	46	2.57	0.59	21
Industry									
Agriculture	2.47	0.61	71	2.23	0.37	6	2.24	0.30	4
Mining	3.27	0.58	82	2.67	0.76	5	2.62	0.96	3
Construction	2.91	0.54	569	2.69	0.57	61	2.62	0.52	28
Manufacturing	3.07	0.54	1244	2.82	0.42	61	2.80	0.70	29
Wh.& Rt. Trade	2.91	0.57	916	2.62	0.49	43	2.39	0.45	32
Transport & Utili.	3.03	0.50	554	2.83	0.41	20	2.56	0.23	9
Information	3.25	0.51	174	3.05	0.45	10	2.91	0.45	5
Finance	3.36	0.67	380	2.96	0.71	17	2.73	0.42	15
Prof. Services	3.27	0.63	577	2.96	0.71	28	2.64	0.59	20
Education	3.09	0.64	660	2.95	0.52	59	2.80	0.67	15
Hospitality	2.63	0.60	280	2.51	0.50	20	2.48	0.56	15
Other Services	2.85	0.51	228	2.91	0.50	10	2.31	0.35	7
Public Admin.	3.18	0.45	534	2.83	0.49	6	2.78	0.51	8

Table 2.3: Average of Ln(Wages) over Categories : 2005-2006 (Continued)

	Not PU			PU-S			PU-L		
Categories	**Mean**	**S.D.**	**Obs.**	**Mean**	**S.D.**	**Obs.**	**Mean**	**S.D.**	**Obs.**
Occupations									
Management	3.42	0.59	1097	3.02	0.45	36	2.84	0.57	14
Professional	3.32	0.55	1189	3.14	0.54	71	3.06	0.60	31
Service	2.71	0.56	656	2.29	0.36	29	2.33	0.33	27
Sales	3.05	0.62	610	2.89	0.58	30	2.47	0.71	17
Admin. Support	2.92	0.44	392	2.75	0.45	22	2.47	0.25	11
Farming	2.38	0.50	42	2.23	0.39	4	2.24	0.30	4
Construction	2.86	0.49	533	2.66	0.59	60	2.64	0.53	31
Instal. & Repair	2.96	0.45	547	2.82	0.51	17	2.60	0.38	14
Production	2.83	0.45	632	2.62	0.39	39	2.60	0.57	20
Transportion	2.81	0.49	571	2.68	0.34	38	2.44	0.44	21

Table 2.4: Average of Ln(Wages) over Categories : 1998-1999

	Not PU			PU-S			PU-L		
Categories	**Mean**	**S.D.**	**Obs.**	**Mean**	**S.D.**	**Obs.**	**Mean**	**S.D.**	**Obs.**
Education									
Not Comp. Coll.	2.68	0.48	3369	2.49	0.49	235	2.33	0.49	124
Completed Coll.	3.09	0.57	2501	2.86	0.57	137	2.80	0.58	74
Marriage									
Not Married	2.66	0.52	1489	2.55	0.54	119	2.45	0.55	91
Married	2.92	0.56	4381	2.66	0.55	253	2.56	0.59	107
Race									
White	2.87	0.56	5190	2.65	0.55	325	2.52	0.57	164
Black	2.66	0.49	435	2.41	0.50	28	2.16	0.42	21
Other	2.79	0.56	245	2.66	0.59	19	2.90	0.61	13
Pot. Exp.									
$\leq$ 15 yrs.	2.75	0.56	1258	2.52	0.51	99	2.46	0.54	71
15 - 25 yrs.	2.87	0.56	2048	2.73	0.57	124	2.56	0.53	51
25 - 35 yrs	2.92	0.56	1883	2.67	0.56	105	2.60	0.67	46
$\geq$ 35 yrs.	2.79	0.55	681	2.52	0.49	44	2.40	0.54	30

Table 2.5: Mean Hourly Log Wages by Percentile

	2010-11			2005-06			1998-99		
Percentile	**NPU**	**PU-S**	**PU-L**	**NPU**	**PU-S**	**PU-L**	**NPU**	**PU-S**	**PU-L**
10^{th}	1.79	1.47	1.46	1.66	1.57	1.34	1.57	1.21	1.35
20^{th}	2.00	1.73	1.57	1.86	1.66	1.56	1.75	1.43	1.53
30^{th}	2.10	1.83	1.62	1.98	1.75	1.59	1.83	1.69	1.56
40^{th}	2.16	1.91	1.80	2.04	1.83	1.78	1.91	1.75	1.57
50^{th}	2.22	2.04	1.91	2.13	1.93	1.84	1.98	1.78	1.57
60^{th}	2.26	2.06	1.98	2.19	1.98	1.86	2.01	1.84	1.61
70^{th}	2.29	2.11	1.98	2.25	2.01	1.88	2.04	1.88	1.66
80^{th}	2.33	2.15	2.04	2.26	2.04	1.91	2.10	1.92	1.70
90^{th}	2.38	2.16	2.07	2.29	2.04	1.93	2.14	1.95	1.74

Table 2.6: Quantile Regressions: PU-S

Variable	10^{th}	25^{th}	50^{th}	75^{th}	90^{th}	OLS
Year: 2010-2011						
PU-S	-0.11**	-0.16**	-0.11**	-0.10**	-0.18**	-0.14**
	(0.04)	(0.03)	(0.03)	(0.03)	(0.04)	(0.02)
Education	0.04*	0.05**	0.06**	0.05*	0.03	0.05**
	(0.02)	(0.01)	(0.02)	(0.02)	(0.02)	(0.01)
Potential Exp	0.03**	0.03**	0.03**	0.03**	0.03**	0.03**
	(0.01)	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)
Married	0.14**	0.12**	0.13**	0.12**	0.12**	0.14**
	(0.03)	(0.02)	(0.01)	(0.02)	(0.02)	(0.01)
White	0.14*	0.05	0.03	0.03	0.01	0.05**
	(0.06)	(0.03)	(0.02)	(0.03)	(0.05)	(0.03)
Metro	0.10**	0.10**	0.11**	0.11**	0.12**	0.11**
	(0.03)	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)
Industry Switch	-0.03	-0.02	-0.01	0.01	0.03	0.01
	(0.03)	(0.02)	(0.02)	(0.03)	(0.03)	(0.02)
Occupation Switch	-0.01	-0.01	0.001	0.01	0.01	0.01
	(0.03)	(0.02)	(0.02)	(0.02)	(0.03)	(0.01)
Firm Size	0.06**	0.05**	0.05**	0.04**	0.03**	0.04**
	(0.01)	(0.00)	(0.00)	(0.01)	(0.01)	(0.00)
No. Of Employers	-0.10*	-0.07*	-0.04	0.01	0.01	-0.05†
	(0.05)	(0.03)	(0.02)	(0.04)	(0.04)	(0.03)
Constant	0.89**	1.40**	1.53**	1.63**	2.00**	1.47**
	(0.28)	(0.18)	(0.17)	(0.18)	(0.20)	(0.13)
Year: 2005-2006						
PU-S	-0.15**	-0.15**	-0.12**	-0.14**	-0.19**	-0.17**
	(0.04)	(0.03)	(0.03)	(0.03)	(0.04)	(0.02)
Education	0.03†	0.03†	0.03**	0.02	0.01	0.02†
	(0.02)	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)
Potential Exp	0.03**	0.03**	0.03**	0.03**	0.03**	0.03**
	(0.01)	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)
Married	0.18**	0.15**	0.14**	0.13**	0.14**	0.15**
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)
White	0.06	0.07†	0.04	0.02	-0.02	0.02
	(0.05)	(0.04)	(0.04)	(0.04)	(0.07)	(0.03)
Metro	0.09**	0.08**	0.09**	0.10**	0.13**	0.10**
	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.01)

Table 2.6: Quantile Regressions: PU-S *(Continued)*

Variable	**10th**	**25th**	**50th**	**75th**	**90th**	**OLS**
Industry Switch	-0.04	-0.01	-0.01	0.03	0.01	0.01
	(0.02)	(0.02)	(0.01)	(0.02)	(0.03)	(0.02)
Occupation Switch	-0.04†	0.01	0.01	0.01	0.03	-0.01
	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.01)
Firm Size	0.05**	0.05**	0.04**	0.04**	0.03**	0.04**
	(0.01)	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)
No. Of Employers	-0.09**	-0.05*	-0.07**	-0.06*	-0.03	-0.05*
	(0.03)	(0.02)	(0.02)	(0.03)	(0.03)	(0.02)
Constant	0.87**	1.34**	1.56**	1.85**	2.24**	1.64**
	(0.24)	(0.21)	(0.14)	(0.18)	(0.27)	(0.14)
Year: 1998-1999						
PU-S	-0.15**	-0.11**	-0.09**	-0.08*	-0.09*	-0.10**
	(0.05)	(0.03)	(0.03)	(0.04)	(0.04)	(0.02)
Education	0.07**	0.05**	0.03*	0.01	-0.03	0.02†
	(0.02)	(0.02)	(0.01)	(0.01)	(0.03)	(0.01)
Potential Exp	0.03**	0.03**	0.03**	0.03**	0.02**	0.03**
	(0.01)	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)
Married	0.13**	0.14**	0.12**	0.13**	0.15**	0.15**
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)
White	0.09	0.07†	0.08*	0.10*	0.07	0.09**
	(0.06)	(0.04)	(0.04)	(0.05)	(0.05)	(0.03)
Industry Switch	-0.07**	-0.03†	-0.02	-0.02	-0.02	-0.03*
	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)
Occupation Switch	-0.05*	-0.08**	-0.05**	-0.04**	-0.04	-0.06**
	(0.03)	(0.02)	(0.02)	(0.02)	(0.03)	(0.01)
Firm Size	0.06**	0.06**	0.06**	0.05**	0.04**	0.05**
	(0.01)	(0.00)	(0.00)	(0.01)	(0.01)	(0.00)
No. Of Employers	-0.09**	-0.07**	-0.06**	-0.04	0.07†	-0.03
	(0.03)	(0.03)	(0.02)	(0.03)	(0.04)	(0.02)
Constant	0.87**	1.14**	1.55**	1.82**	2.20**	1.59**
	(0.28)	(0.19)	(0.15)	(0.16)	(0.31)	(0.13)
Industry Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Occupation Dummies	Yes	Yes	Yes	Yes	Yes	Yes
State Dummies	Yes	Yes	Yes	Yes	Yes	Yes

The dependent variable is the log of hourly wages. The controls include a dummy for african-americans, potential experience and education squared.

**Significant at 1% *Significant at 5% †Significant at 10%

Table 2.7: Quantile Regressions: PU-L

Variable	10^{th}	25^{th}	50^{th}	75^{th}	90^{th}	OLS
Year: 2010-2011						
PU-L	-0.20**	-0.17**	-0.19**	-0.12*	-0.10	-0.17**
	(0.06)	(0.04)	(0.03)	(0.05)	(0.07)	(0.03)
Education	0.04†	0.04**	0.06**	0.04†	0.02	0.05**
	(0.02)	(0.01)	(0.01)	(0.02)	(0.02)	(0.01)
Potential Exp	0.03**	0.03**	0.03**	0.03**	0.03**	0.03**
	(0.01)	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)
Married	0.14**	0.13**	0.13**	0.12**	0.12**	0.14**
	(0.03)	(0.02)	(0.01)	(0.02)	(0.02)	(0.01)
White	0.16**	0.06†	0.03	0.04	0.04	0.06*
	(0.06)	(0.04)	(0.03)	(0.04)	(0.05)	(0.03)
Metro	0.12**	0.09**	0.10**	0.09**	0.09**	0.10**
	(0.03)	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)
Industry Switch	-0.03	-0.02	0.01	0.02	0.06†	0.01
	(0.03)	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)
Occupation Switch	-0.02	-0.02	-0.01	-0.01	-0.01	-0.01
	(0.02)	(0.02)	(0.01)	(0.02)	(0.03)	(0.01)
Firm Size	0.05**	0.04**	0.04**	0.04**	0.03**	0.04**
	(0.01)	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)
No. Of Employers	-0.12**	-0.08*	-0.03	0.01	0.03	-0.04
	(0.04)	(0.04)	(0.03)	(0.04)	(0.04)	(0.03)
Constant	0.91**	1.41**	1.53**	1.69**	1.97**	1.45**
	(0.29)	(0.16)	(0.16)	(0.22)	(0.22)	(0.13)
Year: 2005-2006						
PU-L	-0.34**	-0.31**	-0.22**	-0.17**	-0.18**	-0.27**
	(0.10)	(0.05)	(0.06)	(0.06)	(0.06)	(0.04)
Education	0.04*	0.03†	0.03**	0.01	-0.02	0.02†
	(0.02)	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)
Potential Exp	0.03**	0.03**	0.03**	0.03**	0.03**	0.03**
	(0.01)	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)
Married	0.18**	0.16**	0.14**	0.14**	0.14**	0.16**
	(0.02)	(0.02)	(0.01)	(0.02)	(0.02)	(0.01)
White	0.05	0.07*	0.06	0.02	-0.07	0.02
	(0.04)	(0.03)	(0.04)	(0.04)	(0.06)	(0.03)
Metro	0.10**	0.07**	0.09**	0.09**	0.12**	0.10**
	(0.03)	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)

Table 2.7: Quantile Regressions: PU-L*(Continued)*

Variable	**10^{th}**	**25^{th}**	**50^{th}**	**75^{th}**	**90^{th}**	**OLS**
Industry Switch	-0.02	-0.01	-0.01	0.03	0.02	0.01
	(0.03)	(0.02)	(0.01)	(0.02)	(0.02)	(0.02)
Occupation Switch	-0.05^{*}	0.01	0.01	0.02	$0.04^{\dagger}$	0.01
	(0.02)	(0.02)	(0.01)	(0.02)	(0.02)	(0.01)
Firm Size	0.05^{**}	0.05^{**}	0.04^{**}	0.04^{**}	0.03^{**}	0.04^{**}
	(0.01)	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)
No. Of Employers	-0.08^{*}	-0.06^{*}	-0.07^{**}	-0.05	-0.01	-0.05^{*}
	(0.04)	(0.02)	(0.02)	(0.03)	(0.04)	(0.02)
Constant	0.88^{**}	1.35^{**}	1.57^{**}	1.89^{**}	2.38^{**}	1.66^{**}
	(0.22)	(0.20)	(0.14)	(0.20)	(0.22)	(0.15)
Year: 1998-1999						
PU-L	-0.22^{**}	-0.22^{**}	-0.23^{**}	-0.07	-0.13^{*}	-0.18^{**}
	(0.08)	(0.03)	(0.05)	(0.06)	(0.06)	(0.04)
Education	0.08^{**}	0.06^{**}	0.03^{*}	0.01	-0.05	0.02
	(0.03)	(0.02)	(0.01)	(0.01)	(0.04)	(0.01)
Potential Exp	0.02^{**}	0.02^{**}	0.03^{**}	0.03^{**}	0.02^{**}	0.02^{**}
	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)
Married	0.14^{**}	0.14^{**}	0.12^{**}	0.13^{**}	0.16^{**}	0.15^{**}
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)
White	0.09	$0.07^{\dagger}$	$0.07^{\dagger}$	0.07	0.05	0.07^{*}
	(0.05)	(0.04)	(0.04)	(0.06)	(0.06)	(0.03)
Industry Switch	-0.06^{*}	-0.03	-0.04^{*}	-0.02	-0.01	-0.03^{*}
	(0.03)	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)
Occupation Switch	-0.08^{**}	-0.09^{**}	-0.05^{**}	-0.04^{*}	$-0.04^{\dagger}$	-0.07^{**}
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)
Firm Size	0.06^{**}	0.06^{**}	0.05^{**}	0.05^{**}	0.04^{**}	0.05^{**}
	(0.01)	(0.00)	(0.00)	(0.01)	(0.01)	(0.00)
No. Of Employers	$-0.07^{\dagger}$	-0.06^{*}	-0.07^{**}	-0.03	0.08^{*}	-0.03
	(0.04)	(0.02)	(0.02)	(0.04)	(0.04)	(0.02)
Constant	$0.62^{\dagger}$	1.08^{**}	1.60^{**}	1.90^{**}	2.43^{**}	1.62^{**}
	(0.33)	(0.17)	(0.14)	(0.17)	(0.38)	(0.14)
Industry Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Occupation Dummies	Yes	Yes	Yes	Yes	Yes	Yes
State Dummies	Yes	Yes	Yes	Yes	Yes	Yes

The dependent variable is the log of hourly wages. The controls include a dummy for african-americans, potential experience and education squared.
**Significant at 1% *Significant at 5% †Significant at 10%

Table 2.8: Quantile Decomposition: PU-S

	2010-11				2005-06				1998-99			
θ_{th}	**R.D.**†	**P.D.**‡	**Char.***	**Coef.**§	**R.D.**†	**P.D.**‡	**Char.***	**Coef.**§	**R.D.**†	**P.D.**‡	**Char.***	**Coef.**§
5	-0.18	-0.23	-0.07	-0.16	-0.19	-0.23	-0.09	-0.14	-0.18	-0.21	-0.12	-0.08
10	-0.23	-0.22	-0.06	-0.16	-0.21	-0.25	-0.08	-0.17	-0.18	-0.23	-0.12	-0.11
15	-0.23	-0.22	-0.07	-0.14	-0.22	-0.23	-0.09	-0.15	-0.26	-0.26	-0.11	-0.14
20	-0.18	-0.21	-0.07	-0.14	-0.23	-0.23	-0.09	-0.14	-0.25	-0.24	-0.11	-0.13
25	-0.19	-0.20	-0.07	-0.13	-0.22	-0.23	-0.09	-0.15	-0.20	-0.23	-0.11	-0.12
30	-0.18	-0.20	-0.07	-0.13	-0.27	-0.24	-0.09	-0.15	-0.20	-0.24	-0.11	-0.12
35	-0.22	-0.20	-0.08	-0.13	-0.28	-0.25	-0.09	-0.16	-0.25	-0.24	-0.11	-0.13
40	-0.18	-0.20	-0.08	-0.12	-0.29	-0.25	-0.09	-0.16	-0.25	-0.23	-0.11	-0.12
45	-0.19	-0.20	-0.08	-0.12	-0.27	-0.26	-0.09	-0.17	-0.26	-0.24	-0.10	-0.14
50	-0.20	-0.20	-0.08	-0.12	-0.26	-0.26	-0.09	-0.16	-0.23	-0.23	-0.10	-0.13
55	-0.19	-0.21	-0.09	-0.12	-0.25	-0.25	-0.10	-0.16	-0.24	-0.23	-0.11	-0.12
60	-0.19	-0.21	-0.09	-0.12	-0.25	-0.25	-0.09	-0.16	-0.19	-0.22	-0.11	-0.12
65	-0.20	-0.21	-0.09	-0.12	-0.27	-0.26	-0.09	-0.17	-0.21	-0.22	-0.11	-0.11
70	-0.22	-0.22	-0.09	-0.13	-0.26	-0.26	-0.09	-0.17	-0.20	-0.21	-0.11	-0.10
75	-0.26	-0.23	-0.09	-0.14	-0.22	-0.25	-0.09	-0.16	-0.20	-0.21	-0.11	-0.10
80	-0.29	-0.23	-0.09	-0.13	-0.23	-0.24	-0.09	-0.14	-0.20	-0.20	-0.11	-0.08
85	-0.25	-0.23	-0.10	-0.13	-0.25	-0.24	-0.10	-0.15	-0.20	-0.20	-0.12	-0.08
90	-0.25	-0.26	-0.10	-0.16	-0.25	-0.26	-0.09	-0.16	-0.24	-0.20	-0.11	-0.09
95	-0.26	-0.30	-0.10	-0.20	-0.28	-0.30	-0.08	-0.22	-0.22	-0.19	-0.12	-0.06

R.D † is the raw difference in wages at each quantile level
P.D.‡ is the predicted difference in wages using quantile regression
Char.* is the difference in wages explained due to differences in the worker characteristics
Coef.§ is the unexplained difference in the wages or the component attributable to the differences in rewards

Table 2.9: Quantile Decomposition: PU-L

	2010-11				2005-06				1998-99			
θ_{th}	**R.D.**†	**P.D.**‡	**Char.***	**Coef.**§	**R.D.**†	**P.D.**‡	**Char.***	**Coef.**§	**R.D.**†	**P.D.**‡	**Char.***	**Coef.**§
5	-0.26	-0.38	-0.14	-0.24	-0.29	-0.42	-0.17	-0.25	-0.41	-0.42	-0.22	-0.21
10	-0.34	-0.34	-0.15	-0.19	-0.37	-0.43	-0.17	-0.26	-0.41	-0.44	-0.21	-0.23
15	-0.32	-0.32	-0.16	-0.16	-0.45	-0.44	-0.17	-0.27	-0.43	-0.45	-0.19	-0.26
20	-0.31	-0.33	-0.16	-0.17	-0.37	-0.43	-0.17	-0.26	-0.43	-0.41	-0.18	-0.24
25	-0.33	-0.34	-0.16	-0.18	-0.41	-0.42	-0.16	-0.25	-0.34	-0.38	-0.18	-0.20
30	-0.33	-0.34	-0.17	-0.18	-0.50	-0.45	-0.16	-0.30	-0.30	-0.36	-0.17	-0.19
35	-0.36	-0.34	-0.17	-0.17	-0.48	-0.45	-0.16	-0.29	-0.34	-0.37	-0.16	-0.21
40	-0.33	-0.35	-0.17	-0.18	-0.48	-0.46	-0.16	-0.30	-0.33	-0.35	-0.16	-0.19
45	-0.34	-0.35	-0.17	-0.18	-0.47	-0.45	-0.16	-0.29	-0.34	-0.34	-0.16	-0.19
50	-0.33	-0.34	-0.17	-0.16	-0.48	-0.44	-0.16	-0.28	-0.36	-0.34	-0.15	-0.19
55	-0.29	-0.33	-0.17	-0.16	-0.46	-0.43	-0.16	-0.27	-0.38	-0.34	-0.15	-0.20
60	-0.32	-0.33	-0.17	-0.16	-0.50	-0.44	-0.16	-0.28	-0.37	-0.33	-0.15	-0.19
65	-0.34	-0.33	-0.17	-0.16	-0.43	-0.43	-0.16	-0.27	-0.37	-0.33	-0.15	-0.18
70	-0.32	-0.33	-0.17	-0.15	-0.45	-0.41	-0.16	-0.25	-0.37	-0.33	-0.14	-0.18
75	-0.35	-0.34	-0.17	-0.17	-0.44	-0.42	-0.16	-0.27	-0.36	-0.32	-0.14	-0.18
80	-0.43	-0.35	-0.17	-0.18	-0.42	-0.39	-0.16	-0.22	-0.23	-0.30	-0.14	-0.16
85	-0.39	-0.36	-0.17	-0.19	-0.43	-0.39	-0.16	-0.22	-0.25	-0.27	-0.14	-0.13
90	-0.43	-0.35	-0.17	-0.18	-0.45	-0.38	-0.17	-0.21	-0.30	-0.26	-0.14	-0.12
95	-0.33	-0.32	-0.17	-0.15	-0.34	-0.36	-0.17	-0.19	-0.36	-0.27	-0.14	-0.13

R.D † is the raw difference in wages at each quantile level
P.D.‡ is the predicted difference in wages using quantile regression
Char.* is the difference in wages explained due to differences in the worker characteristics
Coef.§ is the unexplained difference in the wages or the component attributable to the differences in rewards

CHAPTER 3

Poverty, Ability and Primary School Enrollment in India

3.1 Introduction

Universalization of primary schooling by 2015 was one of the eight Millennium Development Goals adopted at the United Nations (UN) Summit in 2000. This goal ensures that children receive a basic foundation of knowledge and an increased accessibility to a living wage later in life. Assessment of the progress made towards this goal shows that South Asia along with sub-Saharan Africa are lagging behind in attaining this objective (Glewwe and Zhao (2007)). With almost 70% of the primary school age population in the region coming from India, the country's poor performance has been a primary driver of the slow progress.

Low levels of schooling can be driven from both the supply and demand side. Children may not be attending schools either due to their low ability or the high costs associated with education. In addition, the quality of schools is also be an important factor. In India, all three factors play a significant role. In terms of ability, according to Pratham's[1] Annual Status of Education 2013 report, about 50 percent of children in class five cannot read class two texts and only 26 percent of students in the fifth class can do a division problem. In the 2009 Programme for International Student Assessment (PISA),[2] India ranked at the

[1] Pratham is a non-governmental organization (NGO) with the mission to improve the quality of education in India.

[2] The PISA is a worldwide study by the OECD in member and non-member nations of 15 year old school

bottom and scored two standard deviations below the average score for Organisation for Economic Co-operation and Development (OECD) countries. In addition, schools in India face problems with both teacher quality and lack of standard facilities. Teacher absenteeism in India is exorbitant, with 25 percent never showing up for work (Kremer, et al. (2006)) . In another study of 188 government-run primary schools, Basu (2004) finds that 59% of the schools had no drinking water and 89% had no toilets. Finally, as is the case in most developing countries, the opportunity cost of education is a significant reason behind the low levels of schooling. While tuition in public-primary schools is negligible and almost completely subsidized, the overhead costs of books and uniforms can be quite high relative to income, dissuading poor families from sending their children to school (Public Report on Basic Education for India Team (1999)). In addition, children are also an economic resource. Jacoby and Skoufias (1997) find that child labor helps smooth the incomes of rural Indian families. Given the multitude of problems faced by the government in addition to the ever expanding population, it is no surprise that India lags behind in the field of education.

Past research in this field suggests multiple avenues through which the government could impact school enrollment and completion rates. Schultz (2004) evaluated the Progresa Program, which provides poor mothers in rural Mexico with education grants, and found that on average there was a 3.5% increase in enrollment. Morley and Coady (2003) evaluate the cash-for-education programs and find that these transfers not only reduce poverty in the short run, but the additional education of the children from poor families breaks the long-run cycle of poverty by increasing their earning potential. Using data from Mexico, Brazil, Bangladesh, Nicaragua, Honduras, and Chile the authors find significant positive impact on education and poverty outcomes. Kremer, et al. (2002) evaluate a program in which an NGO, Internationaal Christelijk Steunfonds Africa (ICS), provided uniforms,

childrens' scholastic performance on mathematics, science, and reading. It was first performed in 2000 and is repeated every three years. India did not participate in the future assessments.

textbooks, and classroom construction to seven randomly selected schools in Kenya. The authors find not only a decrease in dropout rates but also a 50% increase in class sizes. Ahmed (2004) found that the School Feeding Program instituted by the Government of Bangladesh and the U.N. World Food Programme, raised school enrollment by 14.2% and reduced the probability of dropping out of school by 7.5%. Using twenty years of panel data from Bangladesh, Woutersen and Khandker (2014) find that access to micro credit not only stabilizes household income but also leads to an increase in school enrollment rates. In relation to raising the quality of schools, Duflo, Hanna and Ryan (2012) estimate the effect of monitoring and financial incentives on absenteeism. Using data from the NGO Seva Mandir, the authors find a 21% drop in instructor absenteeism, once pay was linked to attendance. Similarly, an increase in facilities can have significant effects on school enrollment rates. Using data from the innovative bicycle program in Bihar (a state in India), Muralidharan and Prakash (2013) find that the provision of bicycles to girls increased enrollment in secondary school by 30%. Harounan , et al. (2013) find that the girl-friendly primary school program in Burkina Faso led to a 5% increase in girls' enrollment rates.

As suggested above, India faces multiple problems in the field of education and there are many ways in which the government could affect the schooling decision. It is therefore important to determine which factor is most relevant and will have the largest impact. The high correlation between family income, skill and school quality make it difficult to disentangle the three effects. The inability to distinguish these explanations creates a quandary because each factor yields a different policy prescription. If families are financially constrained, then an argument can be made in favor of transfer programs that reduce opportunity cost. On the other hand, low skill or low school quality involve arguments in favor of improving the quality of education, and financial aid will not have much effect. The current analysis attempts to estimate the importance of the three different effects separately. Related research in the United States has mostly been through indirect measures of credit constraints. In this analysis, we utilize the highly detailed Indian Household Development Survey. In addition to low income indicators, this nationally representative survey includes scores from three tests (reading, writing and math) conducted for all children between the ages 8-11.[3] This test provides the proxy for ability across students, and school enrollment rates are used as the proxy for the schooling-choice decision. To control for school quality, we use the smallest geographical unit included. The analysis thus includes around 650 fixed effects to control for school quality at the village/neighborhood level. Given the high variation in school quality, the analysis makes a more convincing argument by comparing enrollment rates across students within the same school district.

The modeling methodology employed in this paper is the linear probability model. While a probit/logit model might have provided better estimates, the presence of a large number of fixed effects place restrictions on the modeling technology. By including controls for income, ability and school quality, we use the linear probability model to separately estimate the effect of each of the factors on school enrollment rates. The model calculates the effect of each of the three factors separately on the probability of enrollment for a child. Three separate regressions are conducted for the three tests due to the variance in the availability of the three scores. Other controls include household features, caste and additional individual controls. After controlling for school quality, results indicate that both income and ability have significant effects on enrollment rates. An increase of a point in the test score[4] could lead from anywhere between a 2 percent to 9 percent increase in the school enrollment rates, depending on the test. This positive effect is statistically significant across the three tests. Focusing on the 13% lower enrollment rates (given a raw enrollment rate of 86.15%), the result suggests that ability is one of the important reasons behind the low enrollment rate. More importantly, we also find that financial constraints still have a significant negative effect on enrollment rates. There is a 6% raw enrollment difference between

[3]Details on the score are included in the Data section.

[4]Details on test score are provided in Section 3.2

the poor and the non-poor similarly skilled children. The results show an enrollment difference anywhere in the range of 3% to 9% due to financial constraints. This result implies that if not all, possibly a large portion of the raw enrollment differential between the rich and poor children can be due to financial constraints. Details on the equation and the results are included in section 3.3.

An important concern with the analysis is the selection/omitted variables problem. First, the proxy for ability may be biased against students without schooling. The proxy uses scores on math, reading and writing tests and hence could be biased against students without formal schooling. Additionally, a higher score on the tests may lead parents to believe that the child does not need schooling right now, and that may affect the enrollment decision negatively. Given that ability and enrollment are positively related, the results imply that this negative effect on enrollment is not very strong. The more important concern is that the test is not a measure of the true cognitive ability and the current analysis is not able to control for this effect. Also we argue that children from poor income households who are not enrolled, are financially constrained. Again, the analysis cannot identify the causality. At best we present some evidence that suggests that children are constrained. If children from poor families are choosing to not go to school then the analysis cannot identify the same. Having said that, it is also important to note that these estimates are most probably underestimating the true effects for several reasons. First, enrollment rates are much higher when compared to attendance or graduation rates. Assuming that the poor students have significantly lower graduation or attendance rates, these negative effects could be much stronger. Moreover, the financial constraint effect could be picked up by other variables. In India, high caste is associated with relatively higher incomes. The high caste and Brahmin families have higher average incomes compared to Adivasis, Muslims and other lower castes. The positive relationship between income and school enrollment rate implies that some of the income effects on enrollment would be captured by the caste coefficients. In addition to this, poorer children could be performing worse on the test due to lower school quality and lower levels of education. This means that the test score might underestimate the true ability of the poorer children. Again, given the positive relation between ability and enrollment, the true effect of financial constraints will be underestimated.

From the point of policy implementation, this is a disturbing result. India has launched multiple transfer programs over the years to reduce the opportunity cost of attending school. For example, the Mid-Day Meal Program launched by the Indian government in 1995 aimed at boosting enrollment, retention, and attendance rates for children, while also improving nutrition and health outcomes. However, almost two decades hence, the low schooling rates in India suggest that the government actions thus far have not fully offset the problem.

3.2 Data

The data for the analysis are taken from the India Human Development Survey (IHDS) 2005 which is publicly available. The survey is nationally representative and consists of over 41,000 households. The survey contains information on a multitude of topics including the education, health and economic characteristics of the members of the household. However, most importantly the survey conducted learning tests on all children aged 8–11 which evaluated children on their ability to read, write and solve math problems. Critical to our methodology, we use the children's performance on these tests as a measure of their learning ability.

3.2.1 Variables and Sample Selection

We are interested in finding the role played by financial constraints, ability and school quality in determining primary school enrollment and accordingly a child's enrollment status serves as the dependent variable. In order to control for the financial condition of the household, we categorize households into two groups depending on whether they were classified as being below poverty line or not.[5] We also control for the number of adults, teens and children in a household to account for differences in household size and composition. Among other controls, we use age, gender, caste/religion, rural-urban area and the schooling of the most educated person in the household.

In order to control for ability, we use the performance of children on the learning tests conducted during the survey. These tests were applicable to all children between the ages of 8–11 irrespective of whether they were enrolled at the time or not. The evaluation consisted of three tests: reading, solving math problems and writing. The reading tests were available in 13 languages, including English, and a child was free to choose any one of them. The children were first asked to identify alphabets and if able to do so were asked to read simple words. The next two levels checked a child's ability to read a paragraph followed by the ability to read a story. The math test evaluated children on their ability to identify two-digit numbers, solving subtraction problems, and solving division exercises. The writing test classified students on the basis of whether they were able to write a paragraph with two or fewer mistakes or not. Due to the nature of the scoring scale, we do not aggregate scores and instead report results separately for each test. Finally, we use geography fixed effects in order to account for differences in school quality.

Observations amounting to approximately 28% of the sample (children aged 8–11) are dropped from our regressions because of missing data on one or more variables used in

[5] Household income could be used as an alternative. However, income is not populated for a large number of children and a large chunk of the data has negative values.

our regression equations. This leaves us with information on roughly 12,000 children. The base sample includes all children for whom all variables are observed. There is a slight variation in sample size across the three tests, depending on availibility of scores. Table 3.1 offers a comparison of the summary statistics of the sub-samples with respect to the base sample. The enrollment rates in the sub-samples are approximately 1.1 percentage-points higher than in the base sample. In line with this, we see higher years of schooling of the most educated household adult and more representation of boys, children from urban area and children from Brahmin and high-caste households. Also according to expectations, there is lower size of household with fewer children in the household. However, none of the differences in summary statistics across samples point to a potent selection problem.

3.2.2 Facts on Poverty, Ability and Enrollment

The goal of this paper is to disentangle the effects of a household's financial status from a child's learning ability on the enrollment decisions made by the households. We begin by documenting how enrollment rates differ for children from low-income households as well as for children with low- ability. We also show that the household financial status is correlated with the learning ability.

Table 3.2 compares characteristics of below poverty households to ones above the poverty line. On average, below poverty households have 0.84 more members than above poverty households and this difference also translates in to more teens and children per household. The mean years of schooling of the most educated adult of the households vary significantly across groups. The mean for below poverty households stands at 4.4 years which is 3.3 years less than the mean observed in above poverty households. Brahmin and other high-caste households are factors more likely to be above poverty while the Dalit (Scheduled Castes), the Adivasi (Scheduled Tribes) and the Muslim households are more likely to come from the below poverty group. Most importantly enrollment rates of children aged 8-11 vary significantly across the two income groups. In particular, children from below poverty households are 6.5% less likely to be enrolled relative to children from above poverty households.

Figure 3.1 shows how enrollment rates vary across different levels of learning ability in reading, math solving and writing. The main observation drawn from looking at the figures is that children with higher ability are more likely to be enrolled relative to children with lower ability. For example, the enrollment rate rises by 12.7, 8.9 and 4.4 percentage points moving from lowest ability to highest ability level in reading, math solving and writing respectively. It is also noteworthy that most of the gains in enrollment are realized at the first step when children move from demonstrating no measurable learning ability to the marginal level of ability.[6]

While enrollment rates respond to both ability and financial status of the household, the two variables are not independent and share a systematic relationship. Children from households above the poverty line on average have higher learning ability as measured by the test scores. Figure 3.2 characterizes this association between ability and household income. Less than 10% of the children from above poverty households were not able to read at least a letter in the reading test, while more than 40% were proficient in reading a story given to them. In sharp contrast, the distribution of children from below poverty households was uniform across reading ability levels. In the same vein, only 12% of children from above poverty households were unable to identify numbers compared to an alarming 31% for children from below poverty households. In writing tests children from above poverty households performed much better than children from below poverty households in which one out of every two children was not able to write a sentence without making

[6]This marginal learning ability relates to the ability to read letters and identify numbers for reading and math solving respectively.

two or less mistakes.

3.2.3 School Quality

School quality varies a great deal across India and is an important consideration while making enrollment decisions. On one hand there are private schools that have well-trained teachers together with all modern facilities, while on the other many public schools lack basic facilities like toilets and roofed classrooms. Returns to schooling are likely to vary significantly across schools due to large differences in school quality and this in turn may cause low enrollment in areas that have low quality schools. As such it becomes crucial to control for school quality. In the absence of any direct measures of school quality, we use area level fixed effects in order to control for such variations. The smallest unit of area that we have corresponds to a village or a neighborhood. There are around 650 villages/neighborhoods and on average each area unit has 19 children within the ages of 8–11. It is highly probable that many, if not most, of these area units have a single school and our area fixed effects are able to capture a considerable amount of heterogeneity in school quality.

Figure 3.3 shows how the relationship between ability, household income and enrollment vary across geographical area. Each dot here represents a geographical area unit.[7] We find that geographical areas with a higher proportion of above poverty households have higher enrollment rates (Figure 3.3a). However, the crucial point here is that there is a significant deviation of many area units from the linear fit line. The correlation between enrollment rate and reading ability is somewhat less compared to the previous case but is still statistically significant at the 5% level. More importantly, as seen in Figure 3.3b reading ability

[7]The mean score in reading is calculated by assigning consecutive integer scores for increasing level of ability starting with 0 for the lowest ability specification. We present reading results only for the sake of brevity. The relationship between ability and poverty, and ability and enrollment is robust to using performance on maths and writing tests.

has limited explanatory power in explaining variation in enrollment rates across area units. More than 40% of the area units that have a mean reading score between 3 and 4 have actual enrollment rates that are 10 percentage points below the predicted rates. Similarly, of all the area units with a mean reading score in the range of 0–2, roughly 28% have actual enrollment rates that surpass the predicted values by more than 10 percentage points. Figure 3.3c shows that although mean reading scores is strongly correlated with proportion of households above the poverty line, there is still a lot of variation across the area units. Given this evidence it becomes necessary to account for differences in enrollment rates across area units that are symptomatic of differences in school quality. Finally, in Figure 3.4 we show how enrollment rate varies across some of the other controls that we use in our regressions. Enrollment rates are lower for girls and children from disadvantageous caste and religion groups like Adivasis (ST), Muslims and Dalits (ST). Enrollment rate decreases dramatically with a decline in the years of schooling of the most educated household adult.

3.3 Regression Results

Three main reasons could lead to low levels of enrollment (i) low ability, (ii) low school quality and/or, (iii) low income constraints. Given the rich dataset, we attempt to disentangle the three effects and estimate the impact of each of the three effects on the low levels of primary school enrollment in India. Most importantly, this analysis attempts to investigate if income constraints lead to the low levels of primary school enrollment. To estimate the school enrollment gaps between low income and higher income similarly skilled students, we use a linear probability model. While a probit/logit model would give better results, large number of fixed effects[8] cannot typically be added to a traditional probit model without inducing bias in the coefficients and standard errors (Heckman (1981)). We thus

[8]There are more than 600 districts included in the data

use linear probability models for consistency.[9] The linear probability model is of the form-

$$\text{Enroll}_{ij} = \sum_{k=0}^{N} [\alpha_k \text{Score}_{ij}^k + \beta_k (\text{Poor}_{ij} * \text{Score}_{ij}^k)] + \gamma X_{ij} + \text{Dist}_j + \epsilon_{ij}$$

In the above equation, Enroll_{ij} indicates if child i in district j is currently enrolled in school, Score_{ij}^k represents score k of child i in district j and α_k calculates the effect of Score^k on the enrollment decision. k ranges from 0-N, where N is the highest score on each test.[10] Poor_{ij} is the low income indicator created by the government based on the poverty line, X_{ij} is the vector of individual controls including household characteristics, caste, gender and parents education levels, and Dist_j represents district fixed effects.

In this specification, β_k thus compares the enrollment decision of poor students relative to higher income students, with the same score and within the same school district. Poor is not included as a separate control to ensure that the comparison group for each poor child is a similarly skilled non-poor child. The β's can be thought of as imperfect measures of income constraint, if we assume that similar scores imply similar levels of skills and that district controls function as control for school quality. A negative and significant β thus implies that similar skilled poor students are constrained from enrolling due to income constraints.

Table 3.3 shows the results for the linear probability model for the three ability tests. The ability level coefficients represent the effect of ability on enrollment decision and the Ability*Poor coefficients compare the enrollment decision between poor and rich children. Firstly, we can clearly see that all ability level coefficients are positive and significant. In all tests, the lowest score (Ability 1/Score 0) is the excluded category. This suggests that an

[9] Robustness checks include probit results without district fixed effects.

[10] N=3 in Math's, 4 in Reading and 2 in Writing. Separate regressions are run for the three scores due to differences in sample size.

increase in the score leads to significantly higher enrollment rates. If we believe the scores to be a proxy for ability, then the coefficients imply that there is a positive relation between ability and enrollment. For the math tests, there is a 5.8% increase in enrollment rates when a child can identify numbers, compared to a child who cannot. Across the different math scores, enrollment increases by about 5.5% when a child moves from being unable to identify numbers (Ability 1/Score 0) to being able to divide (Ability 4/Score 3). The reading scores suggest larger enrollment effects. Compared to a child who cannot read, there is on average more than 8% higher enrollment rates, as the child scores higher. Enrollment rates increase by 9.7% when a child is able to read words (Ability 3/Score 2) compared to a child who cannot read (Ability 1/Score 0). For the writing test there were only two scores. Ability 1/Score 0 if the child makes more than 2 mistakes while writing and Ability 2/Score 1 if there are less than two mistakes. Enrollment rates are 2.4% higher for children who score one. Thus, we can see that even after including school quality controls (district fixed effects), ability has a positive effect on enrollment rates. Across all three tests, enrollment rates increase by at least 2%, when a child scores more than the least score in any test.

The second set of coefficients is the most important from the point of view of the analysis. The Ability * Poor coefficients indicate the difference in enrollment decisions between the rich and poor children. We can clearly observe that almost all numbers are negative and statistically significant. The negative β's show that, on average, poor students have lower enrollment rates when compared to richer students and that there may be some financial constraint binding the enrollment decision of the lower income children. This holds true across the three separate regressions. In the math test, the lowest ability (Ability 1/Score 0) poor children have 5.8% lower enrollment rates, compared to the lowest ability not-poor children. For the reading test, the average difference across the first three scores is much higher. For example, poor children with the lowest reading ability (cannot read) have on average 9.1% lower enrollment rates compared to the lowest ability not-poor children.

Except for the Ability 4/Score 3 (can read paragraphs), the poor children always have significantly lower enrollment rates when compared to the non-poor children. In the writing test for both scores, poor children have lower enrollment rates. Even for the lowest writing ability poor children (makes more than two mistakes), there is an enrollment difference of 6.6%. Thus across almost all ability levels, poor children have lower enrollment rates. Also these gaps are largest in the low ability score categories. These negative and significant coefficients imply the presence of some sort of liquidity or income constraint. From the point of policy, providing government aid in the form of fees reduction, meals *etcetera.* could lead to significant improvement in the school enrollment rates.

There are some common factors across the three tests. The value of the coefficients is similar across the regressions, implying robust results. Female children have approximately 1 percent lower enrollment rates, though the effect is not always statistically significant. Age again has a negative and significant effect on enrollment rates, implying that older children have a higher tendency to leave school. This may be due to the increase in opportunity cost in terms of a job opportunity. Household characteristics like size of households and number of children in household have almost no effect on enrollment rates. In line with related literature the education of parent/adult has a positive and significant effect on enrollment rates. One year of additional schooling for parent leads to a 0.2 percent increase in enrollment rates. Urban location could have two contradictory effects on enrollment rate. Better schools in an urban location would imply an increase in enrollment rate while the increase in job opportunities would imply a decrease. The significant negative coefficient implies that for children between the ages of 8-11, the opportunity cost of schooling is higher, leading to about a 5% lower enrollment rate. Caste system has a significant relevance for enrollment rates. Caste system effects on enrollment and opportunity depend on the rank of the caste in society. In this analysis, the excluded caste is Brahmin, which is the highest ranked caste in India. Compared to Brahmins, all other castes have lower enrollment rates.[11] On average, other castes have roughly 3% lower enrollment rates. The lower enrollment rates reflect both the negative school experience and the lack of visible future growth potential for the lower caste children. These results suggest that ability and income have the strongest effect on enrollment rates. While other controls (barring urban indicator) have consistent coefficients, the effects are small and not always significant.

As mentioned earlier, probit regression would provide better estimates as the results are not biased. Horrace and Oaxaca (2006) show that, as the relative proportion of linear probability model predicted probabilities that fall outside the unit interval increases, the potential bias of the linear probability model increases.[12] However, the presence of about 650 district fixed effects as a control for school quality, preclude the use of a probit or logit specification. As a robustness check, Tables 4 to 6 reports both the linear probability results and the marginal effects from the probit regression, without district fixed effects. This helps not only get comparative probit estimates but also we can check the robustness of the linear probability model estimates, when we exclude district effects. Table 3.4 presents the results for the Math test, Table 3.5 includes the reading test estimates and Table 3.6 includes the writing test results. The baseline results are included in the tables for ease of comparison.

In the math's test results (Table 3.4), we can see that ability has a positive and significant effect on enrollment rate. Compared to baseline results, exclusion of district controls has not had significant effect on the ability controls. In the probit specification, the results are positive, statistically significant and slightly larger in effect. An increase in ability by one point leads to at least a 7% increase in enrollment rates. For the Ability * Poor controls, the coefficient is always negative in both the linear probability model and the probit regression, however the effect is not always statistically significant. On average the linear probabil-

[11] High caste includes other high caste categories, and while the coefficient is negative it is small and insignificant

[12] In the LPM model with district effects, approximately 86% of the predicted Y lies within 0 and 1. In the LPM model without district effects, predicted Y always lies between 0 and 1.

ity model suggests that enrollment rates are 5% lower for poor children but the effect is insignificant for children with mid-level (score of 2 or 3) mathematical scores. The probit estimate is only statistically significant for children with Ability 2/Score 1 but negative and insignificant in all score categories. In the reading test results (Table 3.5), again the ability coefficients are always positive and significant. Excluding district effects increases the effect of ability. Compared to the lowest reading score students, children who can read have 11% higher enrollment rates. In the probit estimation also, the coefficients are positive and the marginals indicate similar effects. More importantly, the probit and LPM estimates are pretty similar, suggesting robust results. Moving to the Ability * Poor dummies, the exclusion of district dummies in the LPM model leaves the coefficients negative, but not always significant. In the probit results, the effect is almost always negative and significant. The largest effect is found in children who can read letters. Poor children have 4.6% lower enrollment rates compared to the non-poor children. In the last table, Table 3.6, the results for the writing test are included. Compared to the base results, the removal of district fixed effects slightly increases the ability effect. Compared to the 2.4% of the baseline specification, now an improvement in writing scores leads to a 3.2% increase in enrollment rates. For the probit model also the numbers are positive and significant and the marginal suggests a 2.9% increase in enrollment rates. For the Ability * Poor dummies, the exclusion of district dummies has a small effect. The coefficients in both models are negative and significant. In the probit model, low ability poor students have 3.7% lower enrollment rates and high ability poor students have 1.7% lower enrollment rates compared to low and high skill non- poor children, respectively.

The above tables suggest that the numbers are significant and robust to major changes. Focusing on the baseline results (Table 3.3), it is evident that both skill and income have significant effects on enrollment rates. It is also important to note that the Ability * Poor coefficients are possibly underestimating the true liquidity constraints. From the ability coefficients we know that enrollment rates increase as ability increases. However, it could be the case that poor children score lower on the tests due to low quality/levels of schooling. In this case, the scores would underestimate the ability categories of the poor children. This underestimation would imply that we are comparing higher ability poor students with lower ability non-poor children and hence underestimating the true liquidity constraints on enrollment rates. This is an important result for the purposes of financial aid policy. While there have been some policies in the form of mid-day meals and bicycle loan programs, there is still clearly scope for policymakers to increase aid leading to higher primary school enrollment rates.

3.4 Conclusion

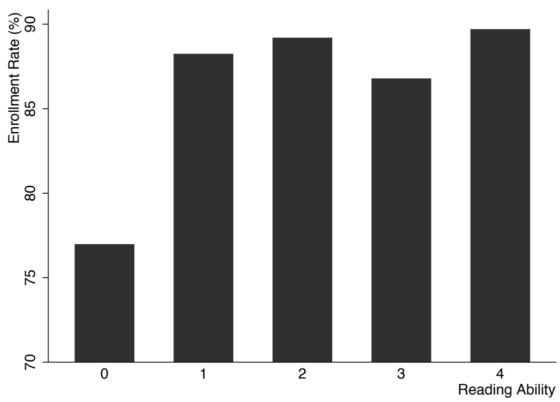
Universal access to education is a huge concern for India. The situation is worse when we look at primary school education, where despite removing fees and instituting various programs, the government has not been able institute 100% enrollment. Large opportunity costs and the use of children for income smoothing suggest that income constraints might still be a large constraint behind the low levels of enrollment and even lower rates of attendance and passing. Past research suggest three major causes behind the low levels of schooling. Using the highly detailed IHDS data, this analysis attempts to disentangle and estimate the effects of the three causes independently. A linear probability model is used and proxies for ability, school quality and income are included in the analysis. We find that all three factors contribute to the low enrollment rates. More importantly, income is still a binding constraint and accounts for almost 3 percent of the enrollment difference between similarly skilled poor and non-poor children, between the age of 8 and 11.

Among other findings, education of family member, gender, caste and location have an effect on enrollment rates. An increase in year of education of a family member leads to 0.2 percent increase in enrollment rates and unsurprisingly female children have lower enrollment rates. Lower/Minority castes like Muslims, Adivasi's and others have roughly 3 percent lower enrollment rates when compared to the high caste Brahmins. Household size has a negative effect on enrollment but the effects are not significant.

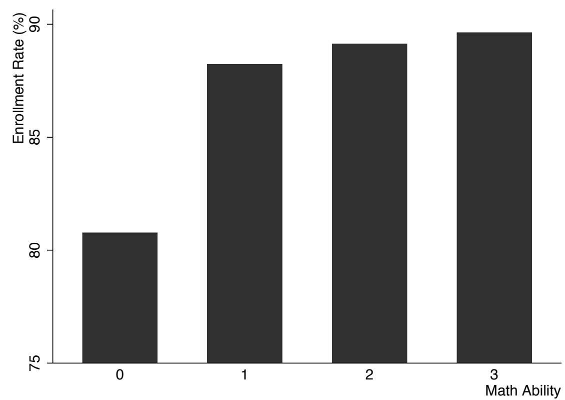
An important concern for this analysis is the viability of the ability scores and selection concerns. While the scores for math, reading and writing reveal ability to some extent, they are not a true measure of the cognitive ability of the children. The results could be biased if low ability kids with better schooling perform better than high ability children. While there is no way to correct for this bias, we argue that there is a higher probability of high-income kids performing better on the tests. In this situation, the results are underestimating the true income constraint. However, we have no way to test this hypothesis.

In addition to this, passing rates from a class and class attendance rates can also be used as indicators for the schooling choice decision. Both reflect on the constant work done by the child throughout the year. However, at this point the data does not include this information. By the end of this year, IHDS plans to release a new panel of the data for the year 2011-12. Revisiting this analysis with new data on the children will provide a better estimate of the effects of ability and income on schooling. Especially, we also hope to track the effect of change in poverty status of a family on the schooling decision. Nevertheless, for now the large significant effects across regressions suggest that income constraints are still far from resolved.

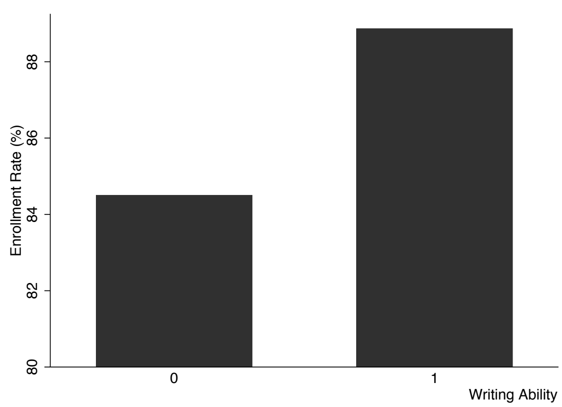
Figure 3.1: Enrollment Rates Across Learning Ability



(a) Reading



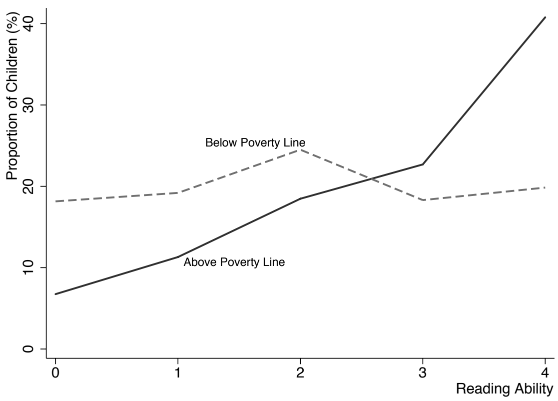
(b) Math



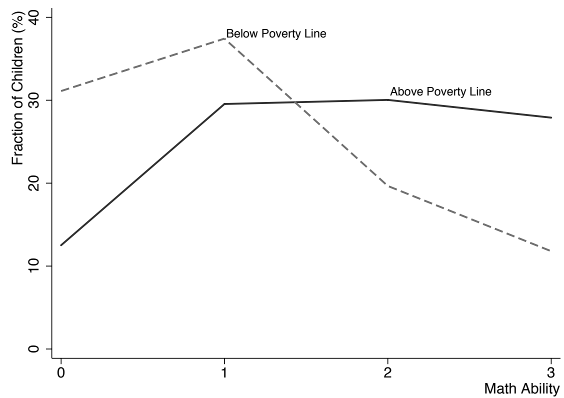
(c) Writing

The above figures shows how enrollment rates vary with ability. The ability level corresponding to various test are as follows. Reading, 0: Can not read, 1: Able to read letter, 2: Able to read word, 3: Able to read paragraph and 4: Able to read story. Math, 0: Can not identify numbers, 1: Able to identify numbers, 2: Able to solve subtraction problems and 3: Able to solve division problems. Writing, 0: Can not write with less than 2 mistakes and 1: Can write with less than 2 mistakes.

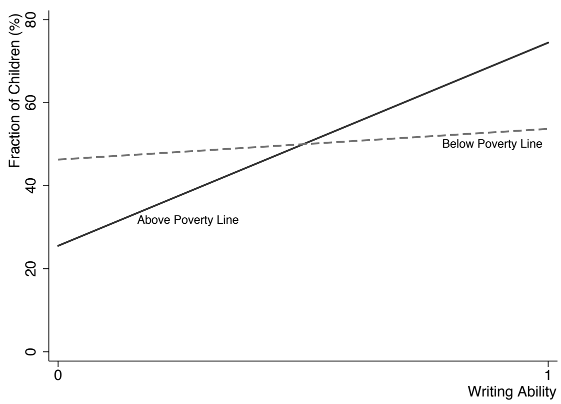
Figure 3.2: Distribution of Children Across Ability Levels by Income Type



(a) Reading



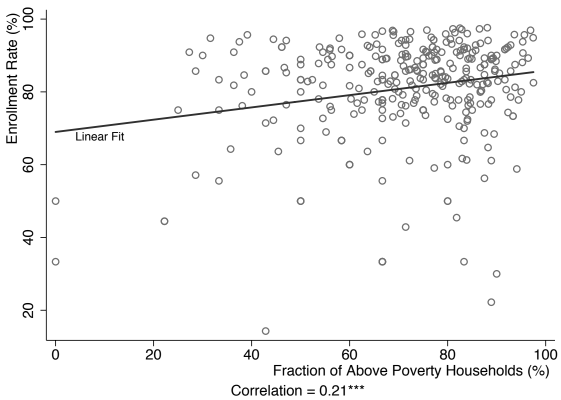
(b) Math



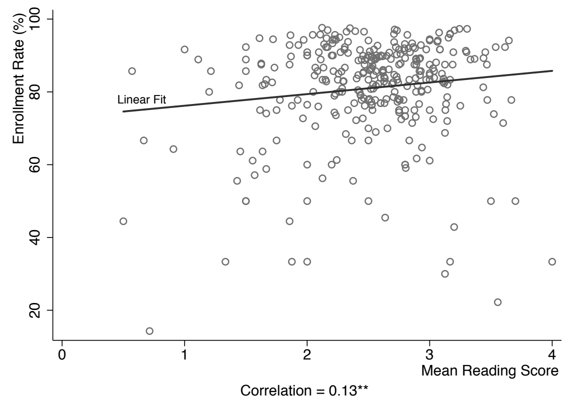
(c) Writing

The above figures demonstrate how children from different income groups are distributed across the various ability levels. The ability level corresponding to various test are as follows. Reading, 0: Can not read, 1: Able to read letter, 2: Able to read word, 3: Able to read paragraph and 4: Able to read story. Math, 0: Can not identify numbers, 1: Able to identify numbers, 2: Able to solve subtraction problems and 3: Able to solve division problems. Writing, 0: Can not write with less than 2 mistakes and 1: Can write with less than 2 mistakes.

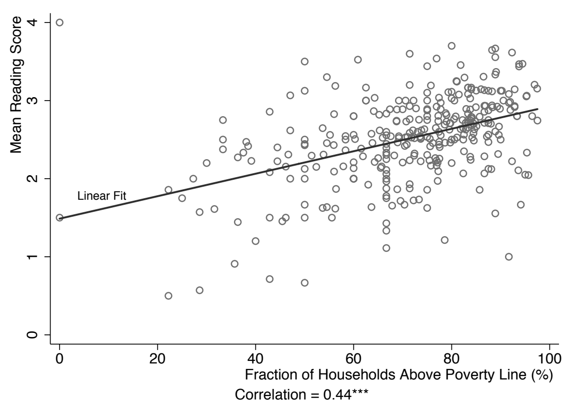
Figure 3.3: Ability, Poverty and Enrollment Across Geographical Area Units



(a) Enrollment vs Poverty



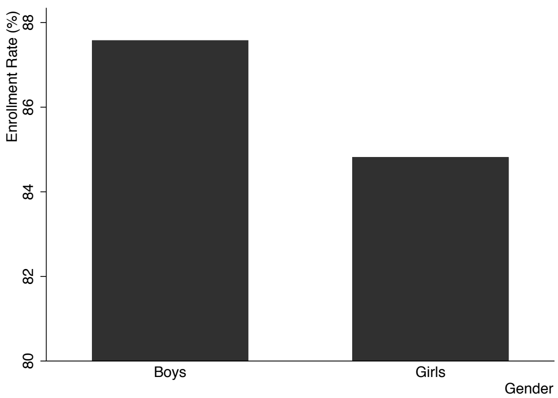
(b) Enrollment vs Ability



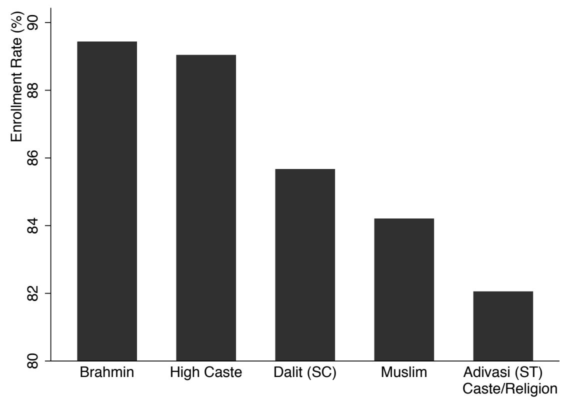
(c) Ability vs Poverty

The above figures demonstrate show the variation in enrollment, poverty and ability across primary sampling units (village/neighborhood). Each dot represents a village or a neighborhood. The mean score in reading is calculated by assigning consecutive integer scores for increasing level of ability starting with 0 for the lowest ability specification. We present reading results only for the sake of brevity. The relationship between ability and poverty, and ability and enrollment is robust to using performance on maths and writing tests.

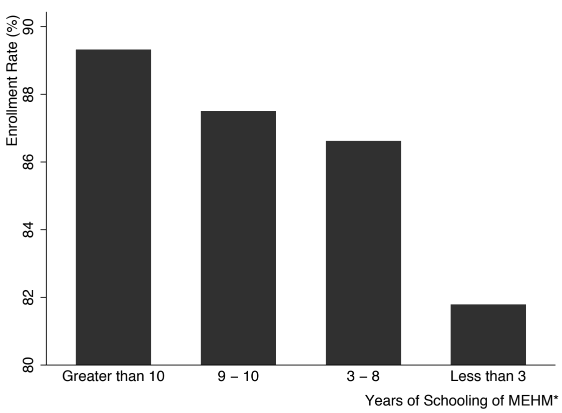
Figure 3.4: Variation in Enrollment Rates Across Other Controls



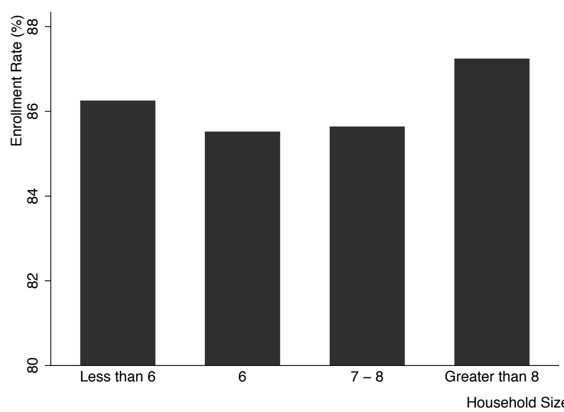
(a) Gender



(b) Caste/Religion



(c) Years of Schooling of MEHA*



(d) Size of the Household

The above figures show how enrollment rate varies across some of the controls used in our regressions. In figures 3.4c and 3.4d, the four groups represent the four quarlites of years of schooling of the most educated household adult and household size respectively.

*MEHA: Most educated household adult.

3.5 Tables

Table 3.1: Summary Statistics

Variable	Base Sample	Test of Ability Reading	Maths	Writing
Enrollment Rate (%)	86.15	87.25	87.26	87.24
Fraction Below Poverty Line (%)	28.81	27.38	27.41	27.41
Mean Score		2.62	1.59	0.69
Age	9.47	9.47	9.47	9.47
Fraction of Boys (%)	53.05	53.20	53.23	53.09
Size of the Household	6.75	6.74	6.74	6.74
No. of Teens in Household	0.62	0.62	0.62	0.62
No. of Children in Household	3.22	3.21	3.21	3.21
Years of Schooling of MEHA†	6.88	6.94	6.94	6.94
Fraction of (%)				
Brahmins	5.66	5.72	5.73	5.74
High-Caste	15.40	15.49	15.49	15.51
Other Backward Caste	33.73	33.88	33.84	33.85
Dalit	21.40	21.58	21.55	21.60
Adivasi	6.81	6.62	6.61	6.58
Muslim	14.32	14.02	14.03	14.01
Sikh/Jain	1.63	1.66	1.66	1.67
Christian	1.06	1.05	1.05	1.05
Fraction in Urban Area (%)	29.76	29.80	29.79	29.84
Number of Observations	12,028	11,568	11,522	11,470

Note that the number of observations in the base sample change with a change in the metric being considered.

†Most educated household adult

Table 3.2: Poor Vs. Non-Poor Households

Variable	Household Status Above Poverty Line	Below Poverty Line
Enrollment Rate (%)	87.66	81.13
Age	9.49	9.42
Fraction of Boys (%)	53.79	50.59
Size of the Household	6.55	7.39
No. of Teens in Household	0.59	0.73
No. of Children in Household	3.03	3.86
Years of Schooling of MEHA†	7.64	4.37
Fraction of (%)		
Brahmins	6.53	2.77
High-Caste	18.47	5.15
Other Backward Caste	34.61	30.79
Dalit	19.82	26.65
Adivasi	4.59	14.19
Muslim	12.73	19.59
Sikh/Jain	2.04	0.25
Christian	1.19	0.61
Fraction in Urban Area (%)	30.62	26.90

†Most educated household adult

Table 3.3: Linear Probability Model: Effect on Enrollment

Variable	Math	Reading	Writing
Ability Levels			
Ability 2	0.058^{**}	0.095^{**}	0.024^{*}
	(0.02)	(0.02)	(0.01)
Ability 3	0.057^{**}	0.097^{**}	
	(0.02)	(0.02)	
Ability 4	0.066^{**}	0.057^{*}	
	(0.02)	(0.02)	
Ability 5		0.093^{**}	
		(0.02)	
Ability Level & Poor			
Ability 1 * Poor	-0.058^{*}	-0.091^{*}	-0.066^{**}
	(0.03)	(0.04)	(0.02)
Ability 2 * Poor	-0.044^{**}	-0.060^{**}	-0.032^{**}
	(0.01)	(0.02)	(0.01)
Ability 3 * Poor	-0.019	-0.044^{**}	
	(0.02)	(0.02)	
Ability 4 * Poor	$-0.058^{\dagger}$	0.007	
	(0.03)	(0.02)	
Ability 5 * Poor		-0.041^{*}	
		(0.02)	

**Significant at 1% *Significant at 5% †Significant at 10%

The ability level corresponding to various test are as follows. For reading, Ability 1: Can not read, Ability 2: Able to read letter, Ability 3: Able to read word, Ability 4: Able to read paragraph and Ability 5: Able to read story. For math, Ability 1: Can not identify numbers, Ability 2: Able to identify numbers, Ability 3: Able to solve subtraction problems and Ability 4: Able to solve division problems. For writing, Ability 1: Can not write with less than 2 mistakes and Ability 2: Can write with less than 2 mistakes.

Table 3: Linear Probability Model: Effect on Enrollment *(Continued)*

Variable	**Math**	**Reading**	**Writing**
Female	-0.011	-0.012	-0.013†
	(0.01)	(0.01)	(0.01)
Caste			
High-Caste	-0.009	-0.008	-0.010
	(0.02)	(0.02)	(0.02)
Other Backward Caste	-0.028†	-0.030†	-0.030†
	(0.02)	(0.02)	(0.02)
Dalit	-0.026	-0.028	-0.030†
	(0.02)	(0.02)	(0.02)
Adivasi	-0.026	-0.024	-0.028
	(0.02)	(0.02)	(0.02)
Muslim	-0.012	-0.012	-0.014
	(0.02)	(0.02)	(0.02)
Sikh/Jain	-0.015	-0.016	-0.017
	(0.03)	(0.03)	(0.03)
Christian	-0.031	-0.033	-0.037
	(0.04)	(0.04)	(0.04)
Age	-0.007†	-0.007*	-0.006
	(0.00)	(0.00)	(0.00)
Size of Hh	0.001	0.001	0.001
	(0.00)	(0.00)	(0.00)
No. of Children in Hh	-0.002	-0.002	-0.002
	(0.01)	(0.01)	(0.01)
No. of Teens in Hh	-0.002	-0.002	-0.001
	(0.01)	(0.01)	(0.01)
Urban	-0.054**	-0.053**	-0.055**
	(0.01)	(0.01)	(0.01)
Years of Schooling of MEHA†	0.002†	0.002†	0.002*
	(0.00)	(0.00)	(0.00)
Constant	0.984**	0.969**	1.006**
	(0.06)	(0.06)	(0.07)
R- Squared	0.17	0.18	0.17
District Controls	Yes	Yes	Yes
Number of Observations	11,522	11,568	11,470

**Significant at 1% *Significant at 5% †Significant at 10%
†Most educated household adult

Table 3.4: Linear Probability & Probit Model: Effect on Enrollment (Math)

Variable	**LPM**††	**LPM (N.D.)**††	**Probit (N.D.)**††
Ability Levels			
Ability 2	0.058**	0.058**	0.074**
	(0.02)	(0.02)	(0.01)
Ability 3	0.057**	0.069**	0.079**
	(0.02)	(0.02)	(0.01)
Ability 4	0.066**	0.073**	0.086**
	(0.02)	(0.02)	(0.01)
Ability Level & Poor			
Ability 1 * Poor	-0.058*	-0.051†	-0.015
	(0.03)	(0.03)	(0.01)
Ability 2 * Poor	-0.044**	-0.015	-0.028*
	(0.01)	(0.02)	(0.01)
Ability 3 * Poor	-0.019	-0.008	-0.010
	(0.02)	(0.02)	(0.02)
Ability 4 * Poor	-0.058†	-0.077†	-0.029
	(0.03)	(0.04)	(0.02)
District	Yes	No	No

**Significant at 1% *Significant at 5% †Significant at 10%

††LPM is the base linear probability model with district fixed effects. LPM (N.D.) is the linear probability model without district controls. Probit (N.D) reports the marginal effects from the probit model without district controls

The ability level corresponding to math- Ability 1: Can not identify numbers, Ability 2: Able to identify numbers, Ability 3: Able to solve subtraction problems and Ability 4: Able to solve division problems.

Table 3.5: Linear Probability & Probit Model: Effect on Enrollment (Read)

Variable	**LPM**[††]	**LPM (N.D.)**[††]	**Probit (N.D.)**[††]
Ability Levels			
Ability 2	0.095**	0.125**	0.113**
	(0.02)	(0.03)	(0.02)
Ability 3	0.097**	0.116**	0.113**
	(0.02)	(0.03)	(0.02)
Ability 4	0.057*	0.081**	0.076**
	(0.02)	(0.03)	(0.02)
Ability 5	0.093**	0.117**	0.115**
	(0.02)	(0.03)	(0.02)
Ability Level & Poor			
Ability 1 * Poor	-0.091*	-0.060	-0.025
	(0.04)	(0.04)	(0.02)
Ability 2 * Poor	-0.060**	-0.049	-0.046**
	(0.02)	(0.03)	(0.02)
Ability 3 * Poor	-0.044**	-0.031†	-0.028†
	(0.02)	(0.02)	(0.02)
Ability 4 * Poor	0.007	0.023	0.019
	(0.02)	(0.02)	(0.02)
Ability 5 * Poor	-0.041*	-0.048*	-0.032*
	(0.02)	(0.02)	(0.02)
District	Yes	No	No

**Significant at 1% *Significant at 5% †Significant at 10%

††LPM is the base linear probability model with district fixed effects. LPM (N.D.) is the linear probability model without district controls. Probit (N.D) reports the marginal effects from the probit model without district controls

The ability level corresponding to reading- Ability 1: Can not read, Ability 2: Able to read letter, Ability 3: Able to read word, Ability 4: Able to read paragraph and Ability 5: Able to read story.

Table 3.6: Linear Probability & Probit Model: Effect on Enrollment (Write)

Variable	**LPM**††	**LPM (N.D.)**††	**Probit (N.D.)**††
Ability Levels			
Ability 2	0.024*	0.032**	0.029**
	(0.01)	(0.01)	(0.01)
Ability Level & Poor			
Ability 1 * Poor	-0.066**	-0.053*	-0.037**
	(0.02)	(0.02)	(0.01)
Ability 2 * Poor	-0.032**	-0.020†	-0.017†
	(0.01)	(0.01)	(0.01)
District	Yes	No	No

**Significant at 1% *Significant at 5% †Significant at 10%

††LPM is the base linear probability model with district fixed effects. LPM (N.D.) is the linear probability model without district controls. Probit (N.D) reports the marginal effects from the probit model without district controls

The ability level corresponding to writing- Ability 1: Can not write with less than 2 mistakes and Ability 2: Can write with less than 2 mistakes.

REFERENCES

[1] Ahmed, A. U. (2004). Impact of Feeding Children in School: Evidence from Bangladesh. *International Food Policy Research Institute*, Washington, DC.

[2] Akerlof, George, Andrew Rose and Janet Yellen (1988). Job Switching and Job Satisfaction in the U.S. Labor Market. *Brooking Papers on Economic Activity*, 2, 495-594.

[3] Albrecht, J., A. van Vuuren, and S. Vroman (2009). Counterfactual Distributions with Sample Selection Adjustments: Econometric Theory and an Application to the Netherlands. *Labour Economics*, 16, 4, 383-396.

[4] Arulampalam, W (2001). Is Unemployment Really Scarring? Effects of Unemployment Experiences on Wages . *The Economic Journal*, 111, 475, 585-606.

[5] Atasoy, Hilal (2013). The Effects of Broadband Internet Expansion on Labor Market Outcomes, *Industrial and Labor Relations Review*, 66, 2, 315-345.

[6] Audet, C. and J. E. Dennis, Jr. (2000). Pattern Search Algorithms for Mixed Variable Programming. *SIAM Journal on Optimization*, 11, 573-594.

[7] Autor, D. H. (2001). Wiring the Labor Market. *Journal of Economic Perspectives*, 15, 1, 25-40.

[8] Autor, D. H., Lawrence F. Katz and Melissa S. Kearney (2008). Trends in U.S. Wage Inequality: Revising the Revisionists. *The Review of Economic and Statistics*, 90, 2, 300-323.

[9] Basu, Kaushik (2004). Combating India's Truant Teachers. *BBC News*

[10] Belley, Philippe and Lance Lochner (2007). The Changing Role of Family Income and Ability in Determining Educational Achievement. *Journal of Human Capital*, 1, 1, 37-89.

[11] Blau, D. M. and Robins, P. K. (I990). Job Search Outcomes for the Employed and Unemployed. *Journal of Political Economy*, 98, 637-655.

[12] Blinder, A. S. (1973). Wage Discrimination: Reduced Form and Structural Estimates. *Journal of Human Resources*, 8, 4, 436-55.

[13] Bowlus, A. J. (1995). Matching Workers and Jobs: Cyclical Fluctuations in Match Quality. *Journal of Labor Economics*, 13, 2, 335-350.

[14] Buchinsky, M. (1998). The Dynamics of Changes in the Female Wage Distribution in the USA: A Quantile Regression Approach. *Journal of Applied Econometrics*, 13, 1, 1-30.

[15] Burgess, Simon (1993). A Model of Competition Between Unemployed and Employed Job Searchers: An Application to the Unemployment Outflow Rate in Britain. *The Economic Journal*, 103, 420, 1190-1204.

[16] Carneiro, Pedro and James J. Heckman (2002). The Evidence on Credit Constraints in Post- Secondary Schooling. *The Economic Journal*, 112, 482, 705-734.

[17] Centeno, Mario (2004). The Match Quality Gains from Unemployment Insurance. *The Journal of Human Resources*, 39, 3, 839-863.

[18] Centeno, Mario and Alvaro A. Novo (2006). The Impact of Unemployment Insurance Generosity on Match Quality Distribution. *Economics Letters, Elsevier*, 93, 2, 235-241.

[19] Chamberlain, G. (1985). Heterogeneity, Omitted Variable Bias, Duration Dependence. *Longitudinal analysis of labor market data*, Cambridge University Press.

[20] Chaudhury, Nazmul, Jeffrey Hammer, Michael Kremer, Karthik Muralidharan, and F. Halsey Rogers (2006). Missing in Action: Teacher and Health Worker Absence in Developing Countries. *Journal of Economic Perspectives*, 20, 1, 91-116.

[21] Choi, Eleanor J. (2011). Does the Internet Help the Unemployed Find Jobs?. *Available at SSRN 2152727*.

[22] Chzhen, Y. and K. Mumford and C. Nicodemo (2012). The Gender Pay Gap in the Australian Private Sector: Is Selection Relevant across the Wage Distribution? *Institute for the Study of Labor (IZA)*.

[23] Clark, Andrew, Yannis Georgellis and Peter Sanfey (2001). Scarring: The Psychological Impact of Past Unemployment. *Economica*, 68, 270, 221-241.

[24] Crandall, R. , Wlliam Lehr and Robert Litan (2007). The Effects of Broadband Deployment on Output and Employment: A Cross-sectional Analysis of U.S. Data. *Issues in Economic Policy*, 6.

[25] DiNardo, John E. and Pischke, Jorn-Steffen (1997). The Returns to Computer Use Revisited: Have Pencils Changed the Wage Structure Too? *The Quarterly Journal of Economics*, 112, 1, 291-303.

[26] Duflo, Esther, Rema Hanna, and Stephen P. Ryan (2012). Incentives Work: Getting Teachers to Come to School. *American Economic Review*, 102, 4, 1241-78.

[27] Durbin, J. (1954). Errors in Variables. *Review of the International Statistical Institute*, 22, 23-32.

[28] Ellwood, David T. (1982). Teenage Unemployment: Permanent Scars or Temporary Blemishes? *The Youth Labour Market Problem: its Nature, Causes and Consequences*, University of Chicago Press, 349-385.

[29] Eriksson, Stefan, and Dan-Olof Rooth (2014). Do Employers Use Unemployment as a Sorting Criterion When Hiring? Evidence from a Field Experiment. *American Economic Review*, 104, 3, 1014-39.

[30] Fountain, Christine (2005). Finding a Job in the Internet Age. *Social Forces*, 83, 3, 1235-1262.

[31] Glewwe, P. and Zhao, M. (2007). Attaining Universal Primary Completion by 2015: An Evaluation of Cost Estimates? *In Educating All Children: A Global Agenda*, MIT Press, 415-454.

[32] Government of India (2003). Selected Educational Statistics, 2002-03. *Department of Education*, Ministry of Human Resource Development.

[33] Gregg, P. and Tominey, E. (2005). The Wage Scar from Youth Unemployment. *Labour Economics*, 12, 4, 487-509.

[34] Gregory, Mary and Robert Jukes (2001). Unemployment and Subsequent Earnings: Estimating Scarring among British Men 1984-94. *The Economic Journal*, 111, 475, 607-625.

[35] Hadass, Y.S. (2004). The Effect of Internet Recruiting on the Matching of Workers and Employers, mimeo.

[36] Hausman, Jerry A. (1978). Specification Tests in Econometrics. *Econometrica*, 46, 6, 1251-1271.

[37] Hausman, Jerry A. and Tiemen Woutersen (2014). Estimating a Semi-Parametric Duration Model without Specifying Heterogeneity. *Journal of Econometrics*, 178, 114-131.

[38] Heckman, J. (1981). The Incidental Parameter Problem and the Problem of Initial Conditions in Estimating a Discrete Time-Discrete Dtata Stochastic Process. *Structural*

Analysis of Discrete Data with Econometric Applications, MIT Press, Cambridge, 179-195.

[39] Heckman, J. and B. Singer (1984). A Method for Minimizing the Distributional Assumptions in Econometric Models for Duration Data. *Econometrica*, 52, 2, 271-320.

[40] Horowitz, J. L. (1999). Semiparametric Estimation of a Proportional Hazard Model with Unobserved Heterogeneity. *Econometrica*, 67, 1001-1028.

[41] Horrace, William C. and Ronald L. Oaxaca (2006). Results on the Bias and Inconsistency of Ordinary Least Squares for the Linear Probability Model. *Economics Letters*, 90, 3, 321-327.

[42] Jacobson, Louis S., Robert J. LaLonde, and Daniel G. Sullivan (1993). Earnings Losses of Displaced Workers. *American Economic Review*, 83, 4, 685-709.

[43] Jacoby, Hanan G., and Emmanuel Skoufias (1997). Risk, Financial Markets, and Human Capital in a Developing Country. *The Review of Economic Studies*, 64, 3, 311-335.

[44] Jovanovic, Boyan (1979). Job Matching and the Theory of Turnover. *Journal of Political Economy*, 87, 5, 1, 972-990.

[45] Kazianga, Harounan, Dan Levy, Leigh L. Linden, and Matt Sloan (2013). The Effects of 'Girl-Friendly' Schools: Evidence from the BRIGHT School Construction Program in Burkina Faso. *American Economic Journal: Applied Economics*, 5, 3, 41-62.

[46] Koenker, R. W. and G. J. Bassett, (1978). Regression Quantiles. *Econometrica*, 46, 1, 33-50.

[47] Kolda, T. G., Lewis, R. M., and Torczon, V. (2003). Optimization by Direct Search: New Perspectives on Some Classical and Modern Methods. *SIAM Review*, 45, 383-482.

[48] Kremer, Michael, Sylvie Moulin, and Robert Namunyu (2002). Unbalanced Decentralization. *Harvard University. Cambridge, MA*.

[49] Kroft, K. and D. Pope (2014). Does Online Search Crowd Out Traditional Search and Improve Matching Efficiency? Evidence from Craigslist, *Journal of Labor Economics*, 32, 2, 259-303.

[50] Krueger, Alan B (1993). How Computers Have Changed the Wage Structure: Evidence from Microdata, 1984-1989. *The Quarterly Journal of Economics*, 108, 1, 33-60.

[51] Krueger, Alan B. (2000). The Internet is Lowering the Cost of Advertising and Searching for Jobs. *New York Times*.

[52] Kuhn, Peter and Mikal Skuterud (2004). Internet Job Search and Unemployment Durations. *American Economic Review*, 94, 1, 218-232.

[53] Kuhn, Peter, and Hani Mansour (2014). Is Internet Job Search Still Ineffective?. *The Economic Journal*, 124, 581, 1213-1233.

[54] Lancaster, Tony (1990). The Econometric Analysis of Transition Data. *Cambridge University Press*.

[55] Machado, J. A. and J. Mata (2005). Counterfactual Decomposition of Changes in Wage Distribution Using Quantile Regression. *Journal of Applied Econometrics*, 20, 445-465.

[56] Mang, Constantin (2012). Online Job Search and Matching Quality, Ifo Working Paper, 147.

[57] Meyer, Bruce D. (1990). Unemployment Insurance and Unemployment Spells. *Econometrica*, 58, 4, 757-782.

[58] Morley, Samuel, and David Coady (2003). From Social Assistance to Social Development: Targeted Education Subsidies in Developing Countries. *Peterson Institute Press*, All Books.

[59] Muralidharan, Karthik, and Nishith Prakash (2013). Cycling to School: Increasing Secondary School Enrollment for Girls in India. *National Bureau of Economic Research*, No: 19305.

[60] Oaxaca, Ronald L. (1973). Male-Female Wage Differentials in Urban Labor Markets. *International Economic Review*, 14, 3, 693-709.

[61] 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.

[62] PROBE Team (1999). Public Report on Basic Education (PROBE) in India. *Oxford University Press*, Delhi.

[63] Ridder, Geert and Tiemen Woutersen (2003). The Singularity of the Information Matrix of the Mixed Proportional Hazard Model. *Econometrica*, 71, 1579-1589.

[64] Ruhm, C. (1991). Are Workers Permanently Scarred by Job Displacements? *American Economic Review*, 81, 1, 319-324.

[65] Schultz, T. Paul (2004)."School Subsidies for the Poor: Evaluating the Mexican Progresa Poverty Program. *Journal of development Economics*, 74, 1, 199-250.

[66] Stevenson, Betsey (2006). The Impact of the Internet on Worker Flows, mimeo, *The Wharton School*, University of Pennsylvania.

[67] Stock J, Yogo M. (2005). Testing for Weak Instruments in Linear IV Regression. In Andrews DWK Identification and Inference for Econometric Models New York, *Cambridge University Press*, 80-108.

[68] Theodossiou, I (1998). The Effects of Low-Pay and Unemployment on Psychological Well-Being: A Logistic Regression Approach. *Journal of Health Economics*, 17, 1, 85-104.

[69] Woutersen, Tiemen, and S. Khandker (2014). Estimating the Long-Run Impact of Microcredit Programs on Educational Outcomes. *Unpublished Manuscript*.

[70] Wu, De-Min (1974). Alternative Tests of Independence between Stochastic Regressors and Disturbances: Finite Sample Results. *Econometrica*, 42, 3, 529-546.