Does Internet Job Search Result in Better Matches?

Anila Prakash*
University of Arizona
October 2014

Abstract

The internet enables both employers and job seekers to gather valuable information about each other in a cheap and timely manner. This increased information-efficiency hints that quality of job matches may improve as the internet penetrates the labor market. The paper documents evidence of improved match quality in the presence of online job search. Drawing from previous research, tenure is used as a proxy for match quality. Impact of the internet on exit rate from employment is estimated using both the Meyer (1990) and the Hausman and Woutersen (2014) proportional hazards model. A conservative estimate shows that exit rates are lowered by at least 28% when the internet is used as a job search tool. Multiple robustness tests indicate consistency of estimates across different specifications.

Keywords: Internet job search, Proportional hazards model, Matching

JEL classification: J64, C41

*Email: anilap@email.arizona.edu. I thank Price V. Fishback, Ronald L. Oaxaca, Jessamyn Schaller, Tiemen Woutersen and Mo Xiao for their valuable comments and suggestions. I acknowledge helpful comments on this paper by conference participants at the North American Summer Meeting of the Econometric Society (2013), Southern Economic Association Annual Meeting (2013) and University of Arizona.

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 utilize 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 (2012) 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 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 probability 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 (2011) 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.

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

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

[3] Kuhn & Mansour (2011) 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

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 3.1 and 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 4.1 and 4.2.

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 4.3 and 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.

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 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

[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.

[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.

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 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 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 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

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

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

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

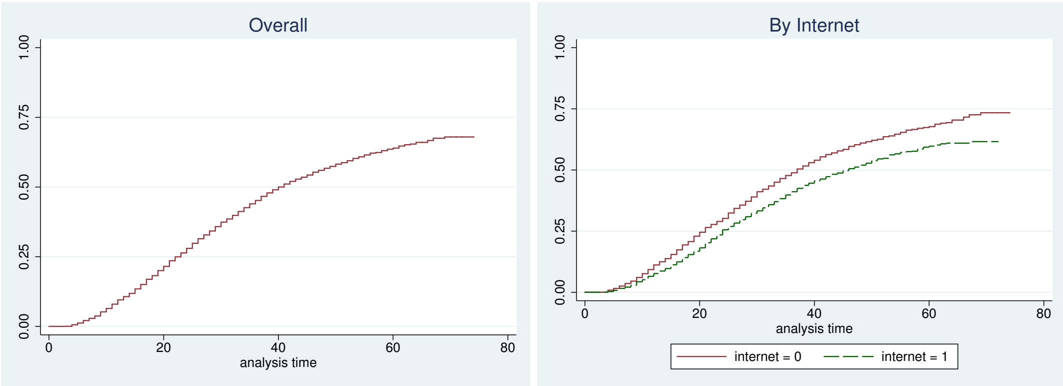


Figure 1: KM Hazard Rate plotted for all workers vs. the KM Hazard Rate plotted separately for IJS and Non-IJS workers

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.

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.

3.1 Model 1: Meyer (1990)

The primary approach for estimation used in this paper is the Prentice and Gloeckler (1978) extension 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:

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

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

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:

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

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

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

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

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 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:

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

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 indepen-

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

dent 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 4.1) discusses the results obtained with and without a parametric baseline, and with and without the gamma distributed heterogeneity component.

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 predefined 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 required time-invariant regressors, the regression coefficients had at 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. Following 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.

[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).

4 Results & Selection Concerns

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 (v) 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 v, 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 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 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 5 - Spec 1) and conditional on unobserved heterogeneity the coefficient effect increases to -0.23 (Table 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 4 and Table 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 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

[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.

duration. The other covariates included in the model have the expected 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 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 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 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 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

[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.

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 conditions, 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 5) and non-parametric models (Table 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), 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

[17] ASVAB provides details on the ASVAB score and how the score used in this analysis is constructed by NLSY.
[18] 52 observations were lost due to missing data on frequency of internet use. As almost all (50 obs.) were non-IJS

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 5) and non-parametric (Table 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.

As a last test, all three variables were simultaneously introduced in the analysis (Table 4 & Table 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 4.3 ad 4.4 focus on a more detailed analysis to check for selection/endogeneity concerns.

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 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 4 - Spec 5), the negative effect is slightly smaller.

workers, this might have biased the internet coefficient.

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 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 4.

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 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 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 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.

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 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 endogeneity, 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

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

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. 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 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 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.

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

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 the strength of the instrument. Both the Durbin (1954) and Hausman (1978) & Wu (1974) statistics are significant, which implies that $\widehat{IJS}_i$ is endogenous. The F statistic reported is significant and much larger than the critical value 10, rejecting the null hypothesis that the instrument is weak. 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 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 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 13). The internet coefficient is positive and 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 in-

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

[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.

ternet 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.

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.

References

[1] 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.

[2] Atasoy, Hilal (2013). The Effects of Broadband Internet Expansion on Labor Market Outcomes, *Industrial and Labor Relations Review*, Vol. 66, No. 2, 315-345.

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

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

[5] 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 (May), 300âĂŞ323.

[6] 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.

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

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

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

[10] Centeno, Mario and Alvaro A. Novo (2006). The impact of unemployment insurance generosity on match quality distribution. *Economics Letters, Elsevier*, 93, 2, 235-241.

[11] Chamberlain, G. (1985). Heterogeneity, omitted variable bias, duration dependence, In: Heckman J J, Singer B (eds.), *Longitudinal analysis of labor market data*, Cambridge University Press.

[12] Choi, Eleanor J. (2011). Does the Internet Help the Unemployed Find Jobs?, mimeo , U.S. Bureau of Labor Statistics.

[13] 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.

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

[15] 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.

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

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

[18] Hausman, Jerry A. (1978). Specification tests in econometrics. *Econometrica*, 46(6), 1251-1271.

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

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

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

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

[23] 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.

[24] Kroft, K. and D. Pope (2012). Does online search crowd out traditional search and improve matching efficiency? Evidence from Craigslist, *forthcoming Journal of Labor Economics*.

[25] 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.

[26] Krueger, A. B. (2000). The Internet is lowering the cost of advertising and searching for jobs. *New York Times*.

[27] Kuhn, Peter and H. Mansour (2011). Is Internet Job Search Still Ineffective?. *IZA Discussion Paper*.

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

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

[30] Mang, Constantin (2012). Online Job Search and Matching Quality, mimeo, Ifo Institute, Poschingerstr.

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

[32] 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.

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

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

[35] 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.

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

6 Tables

Table 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 of non-IJS workers with respect to mean characteristics of 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 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
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 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 4: Baseline Non-Parametric

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)
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 5: Baseline Parametric

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
Ability	No	Yes	No	Yes	No	Yes	No	Yes
Search Int.	No	No	Yes	Yes	No	No	Yes	Yes
Heterogeneity	No	No	No	No	Yes	Yes	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 the exit rate. All specifications include controls for occupaion, industry, states, hourly pay and internet access at home.

†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 6: Hausman and Woutersen (Model 2) Results

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 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 8: 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 9: 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 10: 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 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. 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 and starting period
*** Significant at 1%, ** Significant at 5%, * Significant at 10%

Table 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 12: OLS and IV Results

Variable	Coefficient- OLS	Coefficient-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)
Constant	19.10**	11.64
	(8.50)	(9.25)
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_03 is 0.46^{***}. Results include controls for occupaion (2 digit), states and starting period

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

Table 13: Regression Results

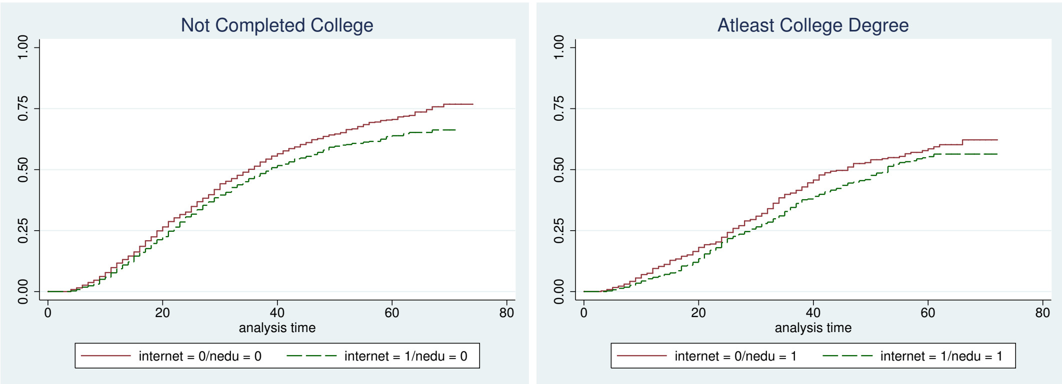
Variable	Spec 1	Spec 2	Spec 3	Spec 4	Spec 5	Spec 6
IJS Worker	1.21^{*}	1.62^{**}	1.54^{**}	1.89^{***}	1.92^{***}	1.66^{***}
	(0.69)	(0.69)	(0.69)	(0.69)	(0.68)	(0.68)
Highest Education	0.60^{***}	0.83^{***}	0.66^{***}	0.85^{***}	0.88^{***}	0.86^{***}
	(0.14)	(0.13)	(0.14)	(0.13)	(0.13)	(0.12)
Sex (Female=1)	-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%

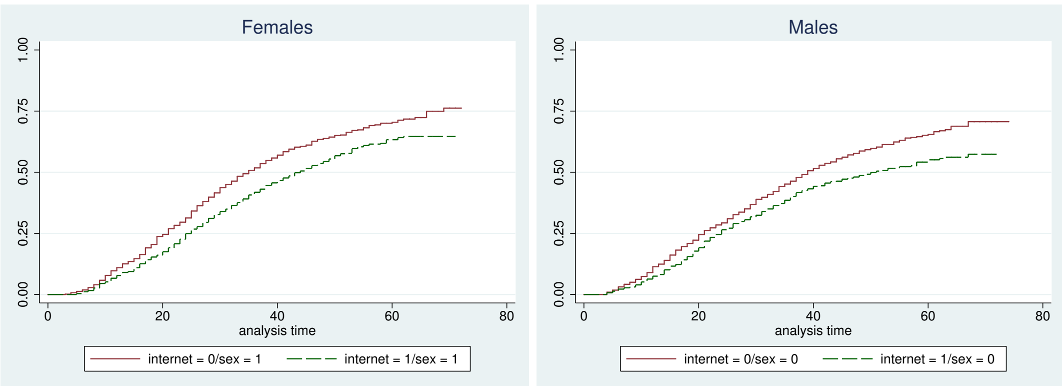
7 Appendix

KM Graphs

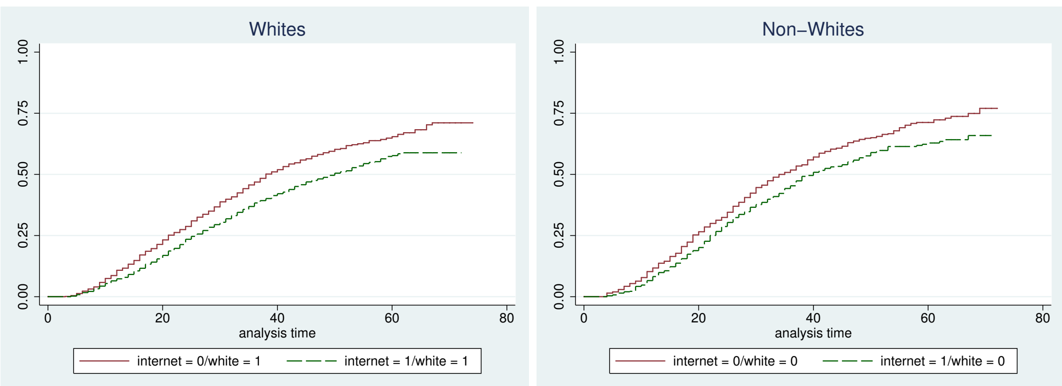
The graphs below plot the Kaplan Meier exit rate (calculated separately for IJS and non-IJS workers) over the entire dataset and across subgroups of education, gender, race and region. Across all categories, IJS workers have lower exit rates.



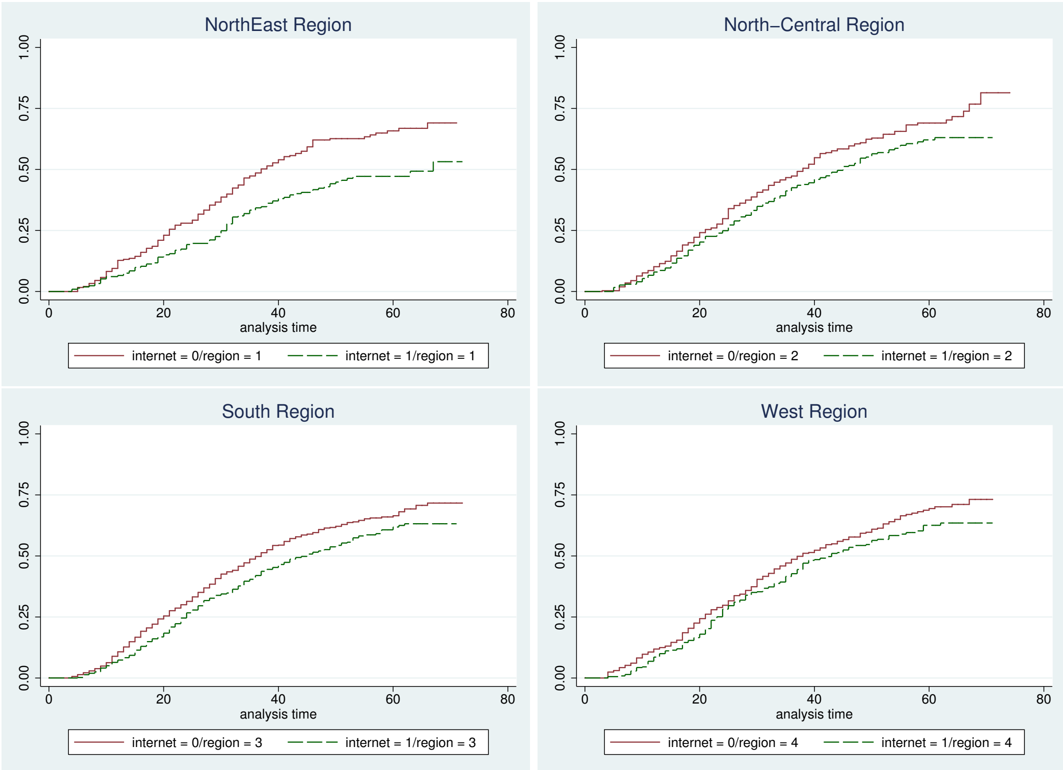
KM Hazard Rate plotted separately for IJS and Non-IJS Workers across Education Categories



KM Hazard Rate plotted separately for IJS and Non-IJS Workers across Gender



KM Hazard Rate plotted separately for IJS and Non-IJS Workers across Race



KM Hazard Rate plotted separately for IJS and Non-IJS Workers 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%