PERSPECTIVES

PHOENIX CENTER FOR ADVANCED LEGAL & ECONOMIC PUBLIC POLICY STUDIES

Internet Use and Job Search: More Evidence

George S. Ford, PhD*

January 26, 2010

Introduction

In POLICY PAPER NO. 39, my colleagues and I presented evidence indicating that Internet use, whether dialup or broadband, reduces discouragement in the labor market.[1] Discouragement is a formal classification of jobless persons and implies the person wants a job but has stopped searching for work due to unfavorable beliefs about employment prospects. Broadband Internet use, whether use home or public connections, was found to reduce the probability of becoming "discouraged" by about 50%. Dialup Internet use also had a large effect, reducing the probability of becoming discouraged with the labor market and by about one-third. These results were based on an analysis of data from the *Current Population Survey* published by the U.S. Census Bureau.

In an effort to measure causal effects, the empirical methods adopted in that PAPER focused on the conditions of unconfoundedness and covariate overlap, resulting in a multivariate, multinomial logit model with Propensity Score trimming. This empirical model was based largely on the work of Lechner (2002) and Crump et al. (2009), as detailed in Imbens and Wooldridge (2009) and Angrist and Pischke (2009).[2]

While we used the multinomial logit model in POLICY PAPER NO. 39, there are other methods by which to estimate the treatment effect of Internet use on discouragement. As mentioned in the PAPER, Lechner (2002) proposes the use of pair-wise estimation of the treatment effects using a Propensity Score matching technique, a statistical procedure frequently seen in the epidemiology literature.[3] The matching approach divides the sample in to treatment pairs (e.g., no dose, low dose; no dose, high dose; low dose, high dose), and then creates a sample of treated and untreated observations by matching them based on similarities in characteristics (as summarized by the Propensity Score). Using the matched sample, the treatment effect for each pair can be computed. For L treatments, there are L(L - 1) pair-wise combinations.

> ***As in POLICY PAPER NO. 39, the treatment effects are large—Broadband use at home and at public sites reduces discouragement by about 60%. The effect of Dialup is smaller, but still large, at about 30 to 40%, depending on the definition of discouragement. Overall, the results are highly comparable to those in POLICY PAPER NO. 39…***

Rather than use the pair-wise approach, our multinomial logit approach computed the treatment effect "structurally," that is for all pairs simultaneously. This joint estimation using the "structural" approach had obvious appeal. Moreover, we faced a trichotomous outcome (unemployment; discouragement; marginally attached but not discouraged), so not only was a pair-wise division of the treatments necessary but also a pair-wise division of the outcomes. The means differences of an order statistic has no meaning. Thus, implementation of pair-wise matching methods requires even more sample divisions than in the prototypical case of a continuous outcome (e.g., wages or income).

We believe POLICY PAPER NO. 39 includes the best estimation approach of those available. Nevertheless, computing the effects of Internet use on discouragement using different methodologies may corroborate the results from our chosen methodology presented in POLICY PAPER NO. 39. In this PERSPECTIVE, I detail the results of the pair-wise matching approach, as proposed by Lechner (2002), to the Internet use and job search problem.[4]

Creating Pairs

As described in POLICY PAPER NO. 39, we have three treatments, broadband Internet use (B), Dialup Internet use (D), and Public site Internet use (P). No Internet use is labeled "0." Likewise, we have three outcomes: unemployment (u); marginally attached and discouraged (d); and just marginally attached (m). For pair-wise comparisons, we must convert all these into dichotomous treatment/outcome sets. That is, we compute the effect of Dialup on being discouraged, ignoring other treatments and outcomes (i.e., we consider the outcome-treatment pair [uD, dD] and ignore [uP, dP] or [uD, mD] and so forth). The effect of Dialup on the "just marginal" outcome must be estimated separately.

Lechner (2002) proposes estimating the effects for all possible treatment pairs including between-treatment comparisons (i.e., 0D, 0B, 0P, DB, DP, and BP).[5] With a trichotomous outcome, we must also consider the outcome u-d independently of outcome u-m. And, with two definitions of discouragement, a complete pair-wise analysis would render a very large number of tests. As such, we limit our analysis and focus on the key findings of our earlier paper. Specifically, we estimate the treatment effects only for discouragement (the u-d outcome), ignoring the effects on the "just marginal" classification. Furthermore, we limit our attention to Internet use versus non-use, rather than consider the between-treatment pairs (DB, DP, and BP). This limited focus is based purely on expositional convenience, not on theoretical grounds. It is limiting in that we cannot empirically test for equal treatment effects across the treatments.[6]

The Propensity Score

The first step in the pair-wise matching approach is to estimate the Propensity Scores (labeled $\hat{s}_t$, where t = D, B, or P). There are three Propensity Scores to estimate: $\hat{s}_{\mathrm{D}}$, $\hat{s}_{\mathrm{B}}$ and $\hat{s}_{\mathrm{P}}$. Once the scores are estimated, the observations in the sample are matched on the score using radius and kernel matching algorithms. Matching occurs only within the relevant sample. For example, one sample includes only those with Dialup or no Internet use and only those either unemployed or discouraged (e.g., the treatment-outcome pair is [uD, dD]). With these matched samples, the treatment effect (on the treated, or "ATT") can be computed. All calculations are based on the *psmatch2* command in STATA 11. We match control observations to the treated observations using radius (r = 0.001) and kernel (Gaussian; bw = 0.04) matching algorithms.[7] The common support option is invoked.

Covariates in the Propensity Score model include: a dummy variable equal to 1 if there are children 18 or younger in the home; a dummy variable equal to 1 if the respondent is male; a dummy variable equal to 1 if the respondent has a college education; a dummy variable equal to 1 if the respondent does not have a high school degree; a dummy variable equal to 1 if the respondent is Caucasian; a dummy variable equal to 1 if the respondent is an immigrant; a dummy variable equal to 1 if the respondent lives in a metro area; a dummy variable equal to 1 if the respondent is a veteran; a dummy variable equal to 1 if the respondent is currently in school; a set of income dummy variables indicating incomes ≤ \$20,000, \$20,000 to \$40,000, \$40,000 to \$60,000, and \$60,000 to \$100,000 (with a "> \$100,000" dummy omitted); set of dummy age variables indicating persons 20 years or younger, between 20 and 40 years, and between 40 and 60 years (with a "> 60 years" dummy left out to avoid the dummy trap); a dummy variable equal to 1 if the respondent is married; a variable measuring household size, and three regional dummy variables (with the fourth excluded to avoid the dummy trap). In all, there are 20 covariates and a constant term.

The Propensity Score models are all estimated as logit and they all perform reasonably well.[8] The ROC statistics for the Dialup, Broadband and Public use equations are 0.78, 0.68, and 0.83 respectively. The statistics are very good for the Dialup and Broadband equations, and marginally acceptable for the Public use equation.[9] Public use is the most difficult to explain, which is not very surprising. The Propensity Score regressions perform well in terms of the pseudo-R^2 test proposed by Sianesi (2004).[10]

Put simply, matching algorithms generate weights for each observation based on the quality of match to treated observations. These weights are then used to compute a weighted average means differences between the treated and untreated observations. An alternative approach is to use these weights in a weighted logit regression of the treatment effect using the same X covariates from POLICY PAPER NO. 29.[11] This alternative should produce more efficient estimates relative to pure matching. I applied this alternative technique and the results were almost identical to those of the standard matching algorithm. As such, I do not report them here. I do note that the logit estimates were slightly more efficient (i.e., larger t-statistics) in most cases, but statistical significance is high and treatment effects large in both approaches.

Results

As is known, matching often reduces the sample sizes by discarding observations, thereby trading off efficiency for a reduction in bias. Sample sizes for control and treated observations are provided in Table 1. The sample sizes remain large with kernel matching, with a loss in some cases of about 30% of available observations. Kernel matching uses essentially all observations, assigning low weights (perhaps zero) to those observations that are poor matches.

Table 1. Matched Samples

	ATT, Radius	ATT, Kernel	Available
BLS Discouraged			
Untreated	788	788	788
Dialup	241	323	334
Broadband	1,022	1,404	1,478
Public	364	432	434
Information-Related Discouragement			
Control	888	888	888
Dialup	258	348	362
Broadband	1,045	1,493	1,557
Public	387	451	452

Significance (* 5%, ** 10%)

The estimated average treatment effects on the treated ("ATT") are summarized in Table 2. They are reported as percentage differences between the control and treated mean outcomes for comparability to POLICY PAPER NO. 39. All but one treatment effect is significant at the 5% level or better, and all are significant at the 10% level.

Table 2. Summary of Results

	ATT, Radius	ATT, Kernel
BLS Discouraged		
Dialup	-54%*	-41%*
Broadband	-59%*	-65%*
Public	-62%*	-64%*
Information-Related Discouragement		
Dialup	-26%**	-30%*
Broadband	-60%*	-57%*
Public	-60%*	-62%*

Significance (* 5%, ** 10%)

As in POLICY PAPER NO. 39, the treatment effects are large—Broadband use at home and at public sites reduces discouragement by about 60%. The effect of Dialup is smaller, but still large, at about 30 to 40%, depending on the definition of discouragement. Overall, the results are highly comparable to those in POLICY PAPER NO. 39, though they resemble the unconditional treatment effects in Table 3 of that PAPER more than the others.

In all, the pair-wise matching approach to measuring the treatment effects of Internet use on labor market discouragement gives us no reason to qualify the results from POLICY PAPER NO. 39. The treatment effects are large and statistically significant irrespective of the estimation approach. Such corroboration is encouraging.

Conclusion

In POLICY PAPER NO. 39 we presented evidence showing that Internet use significantly and sizably reduced labor market discouragement. In that PAPER, multinomial logit regression was combined with Propensity Score methods to estimate causal effects. In this PERSPECTIVE, treatment effects were estimated using pair-wise matching algorithms. The estimated treatment effects are similar in size and remain statistically significant.

We conclude, once more, that the evidence supports the hypothesis that Internet use reduces discouragement in labor markets, and the hypothesis that the effect of Public use is similar to home Broadband use, but the effect of Dialup is large but smaller than these other forms of use.

NOTES:

* **Dr. George Ford** ***is Chief Economist of the Phoenix Center for Advanced Legal and Economic Public Policy Studies. The views expressed in this*** **PERSPECTIVE** ***do not represent the views of the Phoenix Center, its Adjunct Follows, or any if its individual Editorial Advisory Board Members.***

1 T. Randolph Beard, George S. Ford and Richard P. Saba, *Internet Use and Job Search*, PHOENIX CENTER POLICY PAPER NO. 39 (January 2010)(available at: http://www.phoenix-center.org/pcpp/PCPP39Final.pdf).

2 M. Lechner, *Program Heterogeneity and Propensity Score Matching: An Application to the Evaluation of Active Labor Market Policies*, 84 REVIEW OF ECONOMICS AND STATISTICS 205-220 (2002); R. Crump, V. Hotz, G. Imbens and O. Mitnick, *Dealing with Limited Overlap in Estimation of Average Treatment Effects*, 96 BIOMETRIKA 187-199 (2009); *see also* M. Lechner, *Identification and Estimation of Causal Effects of Multiple Treatments under the Conditional Independence Assumption*, in ECONOMETRIC EVALUATION OF LABOR MARKET POLICIES (M. Lechner and F. Pfeiffer eds. 2001) at 43-58; G. Imbens and J. Wooldridge, *Recent Developments in the Econometrics of Program Evaluation*, 47 JOURNAL OF ECONOMIC LITERATURE 5-86 (2009); Angrist and J. Pischke, MOSTLY HARMLESS ECONOMETRICS (2009).

3 *Supra* n. 2.

4 *Supra* n. 2.

5 *Supra* n. 2.

6 A problem with between treatment matching is that our treatments are not ordered. Thus, it is not clear which to treat as the treatment and which the control. It matters, since matching finds a control group for the treated, and the matches are not symmetric. We found evidence of significant differences.

7 The bandwidth is selected by the rule-of-thumb in M. Caliendo, MICROECONOMETRIC EVALUATION OF LABOUR MARKET POLICIES (2006), at 52.

8 O. Baser, *Too Much Ado About Propensity Score Models? Comparing Methods of Propensity Score Matching*, 9 VALUE IN HEALTH 377-385 (2006).

9 D. Hosmer and S. Lemeshow, APPLIED LOGISTIC REGRESSION (2000) at 162.

10 B. Sianesi, *An Evaluation of the Active Labour Market Programmes in Sweden*, 86 THE REVIEW OF ECONOMICS AND STATISTICS 133-155 (2004).

11 Baser, *supra* n. 8.

PHOENIX CENTER FOR ADVANCED LEGAL & ECONOMIC PUBLIC POLICY STUDIES
5335 Wisconsin Avenue, NW, Suite 440
Washington, D.C. 20015
Tel: (+1) (202) 274-0235 • (+1) (202) 318-4909
www.phoenix-center.org