УДК 330.47

Vakhitova H. O.

CERTIFICATION OF IT PROFESSIONALS

This paper examines the rate of return to Microsoft certification using a data set colleted by Microsoft Certified Professional Magazine and a control group from the Current Population Survey. The analysis indicates that this return is positive and significant. Whenever we control for occupation it drops from 33 % to 2 %-10 %, depend on specification, comparable to return form other forms of human capital. A measurement bias is also discussed

Introduction

Incredible development of information technologies beginning mid 1990s created a huge demand for IT specialist in developed countries. Industry-commissioned studies in US revealed the shortage of IT specialists in the late 1990 s. In a year 2000 Germany opened its border for foreign IT specialists to fulfill an estimated 75,000 jobs shortfall in the area of information technologies. Even non-IT-majors all around the world were taking a great number of IT courses, and courses that were not explicitly about IT often contained a fair amount of IT-related content. Acknowledging this situation in the IT labor market, several leading IT companies proposed various certification programs for IT specialists, aimed to assure a certain qualification of their holders.

Despite the huge interest in business community, certification of IT specialists has not received a lot of attention from the academic site. Though the number of salary surveys of certified IT professionals and IT-related specialists grows over recent years, there have been no academic studies examining IT certification and quite few addressing effects of IT driven changes

in labor market. Krueger (1993) made an attempt to estimate the wage premium to computer use at work using Current Population Survey. He claimed that workers who use computers in their work earn 10 to 15 percent more compare to those who don't use computers. Though he estimated a variety of models trying to control for possible omitted variables, this paper became heavily criticized. Following Krueger's approach, DiNardo and Pischke (1997) found similar wage differential to the use of calculators, telephones, pens and pencils. They showed that Krueger's estimations just reflect the fact that higher paid workers use computers on their job. Cappelli (2000) found that positive wage premium associated in particular with computer use, declines when various controls for human capital are added into the model. Levy and Murnane (1992) and Bresnahan et al. (1999) found that computer use is associated with a higher level of specific skills which are indirectly affect earnings. If wage premium results from specific skills rather than directly from computer use then a better measure of those skills is needed. Our paper analyses certification of IT professionals regarding it as a form of human capital, that is used as a proxy for advanced computer skills rather then a simple fact of computer use.

Certification in IT Industry and the Role of Microsoft

Certification exams and certification requirements are typically designed by a vendor. For example, Microsoft, Cisco, Novell, CompTIA, and others, each have their own certification programs and tests. Some of them are more general, others are very product- or subject oriented. Each vendor typically provides several certificates that may be obtained if one has successfully passed a certain number of tests. For example, to obtain MCP, Microsoft Certified Professional certificate, one needs to pass a single test; while to become an MCSE, Microsoft Certified System Engineer, one is required to pass five core exams (four operating system exams & one design exam) and two elective exams. In many cases, certificates have to be renewed after two to three years. There are also some certifications that considered "vendor-neutral", that is, they have been designed by a group rather than one vendor. The vendor neutral exams cover general topics such as hardware, networks, or web page design rather than specific products.

According to the "Certification and IT Professionals" research conducted by Fairfield Research, Inc. and Certification Magazine (2000) Microsoft was recognized as the industry leader whose certification programs attracted the largest number of certificants.

The Data

The data for this research comes from two sources. The first data set was kindly provided by Microsoft Certified Professional Magazine. The Microsoft sample consists of more then 6,000 Microsoft certified IT professionals. Microsoft questionnaire included a quite detailed list of questions about salary and perks, job characteristics, cost and benefits of certification. The survey also reports a number of demographic characteristics that allow us to extract an appropriate sample from the Current Population Survey at a similar time period to separate the effect of Microsoft certification. Therefore, the other two sets that represent my control groups are taken from the March 2001 Current Population Survey. One set (full CPS sample) includes all full time workers and another (CPS IT sample) represents respondents who reported information technology related occupations.

The Microsoft Sample

The Microsoft data set was collected by Wilson Research Group in 2001 on behalf of Microsoft Certified Professional Magazine. Using every nth name from a Microsoft-supplied list of all Microsoft Certified Professionals in the continental U.S., 33 000 respondents were contacted by email and invited to a password-protected web site to complete the survey. The response rate was about 20 %. Contrasting many other business surveys, which ask a more specific sample (say, the readers of some journal) or invite anyone interested to visit a Web site and enter salary numbers, Microsoft Certified Professional Magazine's data is a random and representative sample of the entire IT workforce. The method by which respondents are selected and the number of respondents polled makes it more statistically accurate and worthwhile a serious economic study. Some data set drawbacks have to be mentioned though. On many questions answers are given in range rather then a plain number, which may cause some noise. Some questions refer to year 2000 and 2001 while others deal with the last certificate that could be obtained one-two years earlier. In most cases this does not cause much trouble, but in some situations this problem pertains.

The CPS Samples

The Full CPS sample is restricted to full time year around workers as they are defined in March 2001 Current Population Survey. In order to estimate log-linear wage equation, the employed sample was also restricted, excluding those employees who reported negative earnings or for whom major variables (income, age, education, female, and time) were incomplete. The IT CPS sample is constructed from the full CPS sample and consists of workers

whose reported occupation is arbitrarily defined as IT occupation. This subset includes the following professions: Computer systems analysts and scientists (CPS code 064), Computer programmers (CPS code 229), Operations and systems researchers and analysts (CPS code 065), Computer science teachers, postsecondary education (CPS code 129), Tool programmers, numerical control (CPS code 233), Computer operators (CPS code 308), Peripheral equipment operators (CPS code 309), Supervisors of the computer equipment operators (CPS code 304).

The Variables' Description

As our research is focused upon the determining the Microsoft certification returns to earnings, the various specifications of the basic wage equation are estimated. The dependent variable in this model is annual income in the logarithmic form. Age, education, gender, and working hours are used as major determinants '. Occupation and firm size variables are included to separate the effect of Microsoft certification. This effect is represented using a dummy variable. In all samples, the education variable (EDU) is calculated by taking the years needed on average to get the reported degree. AGE variable is used as a proxy of the general experience and indicates the reported age of the respondent. FEMALE is a dummy variable that takes the value of 1 for females. The variable MS_IND is a proxy for the Microsoft certification, takes the value of 1 for the respondents from the Microsoft sample and zero otherwise. Variables describing income, time, and those controlling for firms' and occupational characteristics are worth to be explained in more detail.

For the CPS respondents INCOME is calculated as total annual wage and salary plus total own business self-employment earnings. This is done to account for possible additional money streams coming from occasional one-time projects in addition to the main job that are quite common among IT professionals. TIME variable indicates hours usually worked per week.

There are several questions about earnings and hours in the Microsoft survey. Questionnaire was designed so that respondents are supposed to answer one common block of questions and two different blocks, depending on the employment status. One set of questions was designed for self employed, and another set of questions aimed for those working for a company. Some respondents answered both sets of questions. It is likely that most of the people who answer both sets of questions are "contract" workers. They are legally self employed, but have a full time contract with a particular company. Thus, TIME variable in the MS sample is calculated based on the questions "If you are NOT self-employed, what is the average number of hours you work per week for your employer?" and the question for self-employed "What is the average number of hours you work per week?" To avoid double-reporting, whenever the respondent answered both questions, only answer to the former question is recorded. INCOME is calculated on the basis of the question "Please select the range that best describes your base personal income (salary) before taxes in 2000?" Table 1 provides some insights about the data.

Table 1. **Summary Statistics: MS, full CPS, and CPS IT samples**

Data Set	Variables	Mean	Strd. Var.	Min.	Max.
Microsoft sample (7120obs[2])	Income	60.32	24.75	27,500	150,000
	Age	35.31	8.55	16.5	60
	Education	15.17	1.96	10	20
	Female	0.10	0.30	0	1
	Time	42.97	6.83	20	51
Full CPS sample (42 909 obs)	Income	42,761.93	44,045.94	1	511,794
	Age	40.94	11.52	15	90
	Education	13.66	2.46	6	20
	Female	0.42	0.49	0	1
	Time	43.77	7.96	35	99
CPS IT sample (1087 obs)	Income	61,319.16	42,953.06	1	454.915
	Age	38.34	9.97	17	76
	Education	15.18	2.12	6	20
	Female	0.31	0.46	0	1
	Time	43.13	6.43	35	99

[1] The MS data set does not have a race variable, so to make CPS and MS specifications as comparable as possible, we do not include race in the MS regressions. However, the results for CPS data are not affected by the inclusion of a race variable.

[2] The sample size may later vary due to non-response by individuals for other questions of interest. A brief comparison of the demographics of people lost due to non-response will demonstrate that the average age, the distribution of educational levels and the gender breakdown were similar to those remaining in the sample.

The average age of the respondents in the Microsoft sample is slightly above 35, with fewer then 0.10 % being under 18 years, and 0.5 % being above 60 years. An average respondent in the CPS IT sample is 3 years older and an average worker from the full CPS sample is 5 year older. These estimations are quite close, if we keep in mind that 30 workers in the MS sample of the age 60 and above reported their age as 60. Less then 0.5 % individuals in both CPS samples are under 18 years old, while about 2 % in the CPS IT sample and 5 % full CPS sample are above 60 years old.

Fewer then 1 % of Microsoft professionals are people who attained less then high school **education** and 15 % got Master's or Ph.D. diplomas with an average respondent having a college degree. This is comparable to the CPS IT sample: an average respondent holds a college diploma, less than 1 % did not receive a high school degree, and about 18 % of the sample obtained graduate education. In the full CPS sample the average worker either attended college without degree or holds an associate degree, with 7 % respondents having less the high school education, and only 10 % received Master's of Ph.D. diplomas.

Ten percent of Microsoft professionals are **females,** with this proportion being much higher in the control groups: 31 % in the CPS IT and 42 % in the full CPS sample.

On average, in all three samples respondents work the same amount **of time** per week, 43 hours. However, there is a tremendous difference in **income,** defined as a base wage and salary. While we have observe it to be around $60,000 in both the Microsoft and the CPS IT samples, in the full CPS sample earnings are about 40 % lower, on average.

To summarize, MS and CPS IT samples look comparable, whereas the full CPS sample differs significantly form both of them. An IT professional, on average, is younger, more educated, and more likely to be a male, compared to an average worker. Nevertheless, we would like to keep all three samples to illustrate certain effects.

Several words should be said about variables describing the occupation and the firm size. In each survey the **firm size** is recorded by categories with very different range definitions. We employed two approaches to incorporate this variable into regressions: midpoints and dummies. Each of them has its pros and contras. Given the distribution by categories in each sample, it is possible to jointly group firms as small (less then 10 employees), medium (10 to 500), large (500 to 1000) and extra large (above 1000). Unfortunately, Microsoft sample holds zero observations in "small firms" category and a significant number of self-employed workers. This may affect the estimates of the firm size effect on the dependent variable. However, it is not clear whether this may or may not influence the relationship between the IT training variable and income, which we are looking at. On the other hand, using midpoints has its drawback too. In particular, the largest category in the CPS is "1000+ employees" which implies that all "extra large" firms are assigned the value of 1000. In Microsoft sample the upper tail of the firm size variable categories is more detailed, with the largest category being "100000+". Thus, the size of some similar firms in Microsoft sample will be coded as 2500, 15 000, 62 500, and 100 000 employees. As a result, this approach may create some distortion toward the Microsoft data. Since our focus is actually on the effect of the Microsoft training on income, using midpoints can be justified.

Occupational dummies are arbitrary designed using both Microsoft and CPS definitions. Table 2 shows what occupational groups from each survey are included in each category.

Table 2. **Occupational dummies**

Variables	CPS		Mircosoft	
	Occupational Title and Code	N	Occupational Title	N
occ_p	Computer systems analysts and scientists (064)	247	Programming Project Lead (Non-Supervisory)	751
occ_pl	Computer programmers (229)	620	Programmer/Analyst	377
occ_nnl	Operations and systems researchers and analysts (065)	83	Networking Project Lead (Non-Supervisory), Network Engineer	2890
occ_wdba	Tool programmers, numerical control (233)	11	Webmaster/Developer/Producer Database Administrator/Developer	736
occ_t	Computer science teachers (129)	7	Trainer	515
occ_m		0[1]	Management (Supervisor)	828
occ_sup	Supervisors, computer equipment operators (304), Computer operators (308), Peripheral equipment operators (309)	120	Help Desk/User Support	1006
others		41 821		367

Unfortunately, it was impossible to determine from the definitions of occupational variables in MS sample what categories in CPS would relate.

Table 3. **Impact of Microsoft Certification on** Earnings

Sample	Standard Earnings Equation	Earnings Equation with Firm Size Controls		Earnings Equation with Occupation Controls		
		Midpoints	Dummies	No control over firm size	firm size is included at midpoints	firm size is included as dummies
MS + Full CPS	34.21 %	33.69 %	28.49 %	10%	9.75	7.55 %
MS + CPS IT	5.76 %	5.84 %	2.41 %	–	–	–

The Impact of Certification on Income

The Model

The impact of the Microsoft certification on the annual income is measured based on the standard Mincer equation, augmented with a Microsoft certification indicators. Let the general model be specified as

$$\ln W = X\beta + T\gamma + \varepsilon,$$

where W is the annual income, X is a vector of demographics, firm characteristics, and occupational dummies, β is a vector of coefficients, T is a vector of certification dummies with the corresponding vector of coefficients, and ε is an error term that is assumed to possess all the standard properties. At this stage of our analysis the vector T will include one variable indicating whether a person has any Microsoft certificate.

We have estimated OLS model using three types of specifications. Each specification was tested for both groups, for all workers (MS + full CPS) and for the IT professionals only (MS + CPS IT samples pooled together). Basic specification includes only major demographic variables: age, education, female, and time. Firm size calculated at midpoints and in a form of dummies is included into model specifications of the second type. Finally, occupation is added to control for the correlation between the certification and position requirements. Table 3 summarizes all the results. Table 4, Table 5, and Table 6 in the Appendix provide the details of regression equations.

There seems to be a positive and significant premium to getting Microsoft certification. It drops between two samples, from 34.2 % in the full sample to 5.8 % in the IT sample. The estimates with and without firm size are very similar when the variable is included at the midpoints. The effect changes from 34.2 % to 33.7 % in the full sample and from 5.76 % to 5.84 % in the IT sample. When firm size is included as a group of dummies, the coefficients go down slightly, that can be a result of how the categories are determined. However, the order of the estimate magnitude doesn't seem to change much, no matter how the firm size variable is included (midpoints or dummies). Thus, the question is, Why is there such a tremendous difference between two samples? One possible explanation for the larger coefficient in the full sample is that it is just a "proxy" for IT. To test this hypothesis we introduced occupational dummies into our regression. The coefficient immediately drops by approximately threefold. Depending on form of the firm size control it is equal to 10.5 %, 9.7 %, or 7.5 %, which is more similar to the IT sample estimates. Another explanation of the difference in the coefficients is a specific measurement error.

The Measurement Error

When pulling together Microsoft and CPS data sets, we need to recognize that some people from CPS may also keep Microsoft certificates but failed to report this fact. Therefore, instead of T, our certification variable, we observe $Z = (1-C)T$, where C indicates whether a respondent is in the CPS sample. Hence, we deal with an error in the right-hand side binary variable. This type of measurement error constitutes a bias for our γ coefficient (Bollinger, 1993). Unfortunately, at the moment we do not have all the necessary information to estimate the slope, but it is important to mention that problem. We know that whenever we observe training, workers did receive it. On the other hand, we believe that probability of getting people with Microsoft certificates in a CPS sample is the same as in overall population. Knowing the rate of certified IT professional we should be able to derive a consistent estimate of a return to certification from the pooled sample. Since proportion of CPS workers in the full sample and in the IT sample varies, this can the source of the difference in the coefficients.

Concluding remarks

Our paper reveals a positive and significant premium to the Microsoft certification. However, the estimated coefficient is biased due to the measurement error in the binary indicator. If we assume that most CPS IT professionals who keep no Microsoft certificates are highly skilled workers with earnings above average, then they may be not in-

terested in acquiring those credentials. In this case, the expected premium to Microsoft certification will drop even below estimated 2 % - 10 %. If certification is considered by employers as an essential job requirement, and even high-skilled workers receive lower compensation if they are not certified, then the estimated coefficient is downward biased.

APPENDIX

Table 4. **The Standard Earnings Equation**

Independent Variables	MS + Full CPS		MS + CPS IT	
	(I)	(2)	(3)	(4)
age	0.0519 (0.002)	0.0522 (0.002)	0.0511 (0.004)	0.0503 (0.004)
age^2	-0.0005 (0.00002)	-0.0005 (0.00002)	-0.0006 (0.00005)	-0.0006 (0.00005)
edu	0.1188 (0.001)	0.1078 (0.001)	0.0473 (0.002)	0.0472 (0.002)
female	-0.3631 (0.007)	-0.3061 (0.007)	-0.1435 (0.014)	-0.1315 (0.014)
time	0.0088 (0.0004)	0.0102 (0.0004)	0.0105 (0.0007)	0.0105 (0.0007)
ms_ind	–	0.3421 (0.010)	–	0.0576 (0.014)
Intercept	7.4210 (0.042)	7.3975 (0.042)	8.7625 (0.077)	8.7195 (0.078)
η	50029	50029	8207	8207
R^2	21.6%	23.4%	12.5%	12.7%

Notes. Standard errors are given in the parentheses. Independent variable is annual income in the log form.
All variables are significant at 1 %.

Table 5. **Earnings Equation with the Firm Size Controls**

Independent Variables	MS + Full CPS		MS + CPS IT	
	(5)	(6)	(7)	(8)
age	0.0522 (0.002)	0.0497 (0.002)	0.0499 (0.004)	0.0501 (0.004)
age'	-0.00053 (0.00002)	-0.00049 (0.00002)	-0.00059 JO. 00005)	-0.00060 (0.00005)_
edu	0.1080 (0.001)	0.1032 (0.001)	0.0470 (0.002)	0.0465 (0.002)
female	-0.3067 (0.007)	-0.3203 (0.007)	-0.1303 (0.014)	-0.1325 (0.014)
time	0.0102 (0.0004)	0.0122 (0.0004)	0.0101 (0.001)	0.0120 (0.001)
firmsize	6.87E-07 * (3.03E-07)	–	4.22E-07 * (1.70E-07)	–
firm_se	_	0.1284 (0.036)	–	0.1984 (0.021)
firm_s	_	-0.4108 (0.010)	–	-0.3646 (0.044)
firm_m	_	-0.0851 (0.007)_	–	-0.0243 * (0.010)
firmj	_	-0.0325 * (0.014)	–	-0.0439 * (0.018)
ms_ind	0.3369 (0.011)	0.2849 (0.011)	0.0584 (0.014)	0.0241** (0.014)
Intercept	7.3958 (0.042)	7.5162 (0.042)	8.7495 (0.080)	8.7221 (0.079)
n	49706	49706	7884	7884
R^2	23.4%	25.9%	12.5%	14.3%

Notes. Standard errors are given in the parentheses. Independent variable is annual income in the log form.
Firms with more then 1000 employees (firm_xl) are the base category.
All unmarked estimates are significant at 1 %.
* - significant at 5 %.
** - significant at 10%.

Table 6. **Impact of the Microsoft Certification when Controlling Over Occupation and the Firm**

Independent Variables	MS + Full CPS					
	No control over firm size		Firm size is included at midpoints		Firm size is included as dummies	
	(9)		(10)		(П)	
age	0.0519	(0.002)	0.0519	(0.002)	0.0495	(0.002)
age2	-0.00052	(0.00002)	-0.00052	(0.00002)	-0.00049	(0.00002)
edu	0.1046	(0.001)	0.1049	(0.001)	0.1002	(0.001)
female	-0.3007	(0.007)	-0.3012	(0.007)	-0.3150	(0.007)
time	0.0101	(0.000)	0.0101	(0.000)	0.0121	(0.000)
firmsize	–	–	9.83E-07	(3.04E-07)	–	–
firm_se	–	–	–	–	0.1328	(0.037)
firm_s	–	–	–	–	-0.4076	(0.010)
firm_m	–	–	–	–	-0.0849	(0.007)
firmj	–	–	–	–	-0.0326*	(0.014)
occ_p	0.3034	(0.030)	0.2981	(0.031)	0.2779	(0.030)
occ_pl	0.4013	(0.025)	0.3973	(0.026)	0.3714	(0.025)
occ_nnl	0.2616	(0.028)	0.2554	(0.029)	0.2274	(0.028)
occ_t	0.3054	(0.041)	0.3079	(0.042)	0.2591	(0.042)
occ_wdba	0.3042	(0.037)	0.2942	(0.038)	0.2725	(0.037)
occ_m	0.3797	(0.036)	0.3718	(0.037)	0.3505	(0.036)
occ_sup	0.0443***	(0.031)	0.0344***	(0.032)	0.0070***	(0.032)
ms_ind	0.1047	(0.025)	0.0975	(0.026)	0.0755	(0.026)
_cons	7.4347	(0.042)	7.4300	(0.042)	7.5484	(0.042)
η	50029		49706		49706	
R^2	23.97 %		23.94 %		26.37 %	

Notes. Standard errors are given in the parentheses. Independent variable is annual income in the log form.
Firms with more then 1000 employees (firm_xl) are the base category.
All unmarked estimates are significant at 1 %.
* - significant at 5 %.
*** - insignificant.

1. *Krueger A.* How Computers Have Changed the Wage Structure: Evidence from Microdata, 1984-1989, The Quarterly Journal of Economics- 1993.-V. 108.-№ 1.-P. 33-60.
2. *DiNardo J. Pischke J.-S.* The Computer Use Revisited: Have Pencils Changed The Wage Structure Too? // The Quarterly Journal of Economics.- 1997.
3. *Cappelli P., Carter W.* Computers, Work Organization, and Wage Outcomes, NBER Working paper,- 2000- № 7987.
4. *Levy F.* and *Murnane R.* U.S. Earnings Levels and Earnings Inequality: A Review of Recent Trends and Proposed Explanations // Journal of Economic Literature- 1992- V. 30.- № 3- P. 1333-1381.
Bresnahan T., Brynjolfsson E., Hit! L. Information Technology, Workplace Organization and the Demand for Skilled Labor: Firm-Level Evidence. NBER Working Papers- 1999-№ 7136.
Bellinger C. Bounding Mean Regressions: When a Binary Regressor is Mismeasured // Journal of Econometrics.- 1993- V. 72.- P. 387-399.

Г. Вахітова

СЕРТИФІКАЦІЯ ПРОФЕСІОНАЛІВ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ

У статті досліджується норма прибутку сертифікатів фірми «Майкрософт», для чого використовуються дані, зібрані журналом «Microsoft Certified Professional Magazine». Дані щомісячного дослідження домогосподарств (CPS) було використано як контрольну групу. Регресійний аналіз показав, що норма прибутку є позитивною й статистична значимою. Якщо до факторів регресії додати змінні професій, ефект падає з 33 % до 2-10 %, залежно від специфікації моделі, що порівняно з нормою прибутку на інші види людського капіталу. Розглянуто також похибку вимірювання.