Assessing Gender Differences in Software Developers Using the Human Capital Model

ABSTRACT

Previous studies have indicated that discriminatory practices exist in the Information Technology profession. In this paper, we quantify the differences in the current hourly salaries of female software developers with their male counterparts using the human capital model based on economic theory. In addition to the gender factor, the human capital model includes other control variables that may account for the salary differences such as education, experience, and specific skills, such as object-oriented programming and SQL. Our models indicate that gender is still a statistically and practically significant factor in assessing a software developer's salary.

Keywords: Software developers/computer programmers, gender differences, discrimination, human capital model

Assessing Gender Differences in Software Developers Using the Human Capital Model

INTRODUCTION

There is a great deal of interest in regard to gender differences in the Information Technology (IT) profession. Panteli, Stack, and Ramsay (1999) reported that women were under-represented in all areas of the IT industry and cited several studies that indicate the IT culture is a "masculine, engineering type, computing culture". More recently, a special issue of the *Communications of the ACM* (July 2001) focused on issues regarding the global IT workforce. One key issue that emerged was the under-representation of women, minorities, and older workers in the IT workforce (Arnold and Niederman, 2001). Two specific articles in this issue focused mainly on gender issues. Von Hellens and Nielsen (2001) discussed attracting women to the IT profession in Australia. Trauth (2001) focused on two workforce challenges for Ireland relating to gender and socio-economic class.

Truman and Baroudi (1994) examined the extent to which gender differences exist in ranks of senior IT managers, and they found that the mean salary for women IT managers was considerably lower than males even when controlling for job level, age, education, and work experience. They concluded this is "a problem suggestive of discriminatory practices".

In contrast, the neoclassical economic school has indicated that gender inequalities are likely to decline with industrialization or economic growth. Several studies within the neoclassical economic approach have argued that differences between men and women result primarily from human capital differentials (education, skills, expected length of labor-force participation) that are bound to wither away over time (Forsythe, Korzeniewicz, and Durrant, 2000).

Currently, female IT workers still make substantially less than their male counterparts. Recent U.S. Department of Labor (2002) data indicate that currently full-time female computer

programmers make $867 per week (median) compared to their male counterparts who make $975 per week (median). Equivalently, female computer programmers make, on average, only 88.92% of what their male counterparts make or conversely, male computer programmers make, on average, 112.46% of what their female counterparts make. The question addressed in this paper is the extent to which the salary differences between male and female software developers (we prefer using this term rather than computer programmers) can be attributed to human capital differentials. For most professions, the significant human capital factors include work experience (Auster, 1989; England and McCreary, 1987; Hulin and Smith, 1965; Olson and Frieze, 1987) and education (Auster, 1989; England and McCreary, 1987; Forgionne and Peeters, 1982; Weaver, 1978). In addition, specific skills may contribute to the human capital of software developers.

To address this question, we analyze the differences in current salaries between female and male software developers by factoring in the effects of education, experience, and specific skills. We fit the human capital model based on economic theory to provide a quantitative assessment of the salary differences attributed to gender. While the human capital model quantifies the salary differences based on gender, it also controls for the effects of different amounts of technical experience and different levels of education that software developers possess. Further, salary data is adjusted to account for the average number of hours worked per week. In addition, we consider other human capital factors that impact the salaries of software developers. If a set of human capital factors is found that make the gender factor insignificant, this will provide support to the neoclassical economic school's viewpoint that human capital differentials are responsible for salary differences. On the other hand, if the gender factor is still

significant, the model results will provide a quantitative assessment of salary differences for software developers attributed to gender.

In the next section, the relevant literature on gender inequality and discrimination is reviewed. Then, the human capital model, which we employ to assess potential gender discrimination, and its theoretical rationale, the human capital theory, are detailed. Following this, the nature of our survey is discussed briefly and some summary statistics are presented. The human capital model results are then presented and discussed. To provide confirmatory evidence for our human capital models, our sample is divided into female and male subsets then Chow (1960) tests and Oaxaca (1973) decompositions are applied. The paper concludes with a discussion of the results and managerial implications.

GENDER INEQUALITY AND DISCRIMINATION

In the United States of America, Title VII makes it unlawful for employers, labor organizations, and employment agencies to discriminate against employees and applicants on the basis of their race, color, sex, religion, and national origin. Enacted in 1964, Title VII was designed to achieve equality of employment opportunities and to eliminate discriminatory impediments to that equality (Reinsmith, 2002).

Typically, there are three perspectives for studying discrimination: the economic, sociological, and psychological perspectives (Truman and Baroudi, 1994). In this paper, our main focus is on the economic perspective. For completeness, we will briefly review the sociological and psychological perspectives.

The sociological perspective on discrimination states that structural differences in society and the economy may explain gender wage inequalities. Different industries and job types have different expectations and demands on their employees. Daymont and Andrisani (1984) argue

that between one third and two thirds of the gender gap can be accounted for by male/female differences in college majors and occupational preferences. In addition, the psychological perspective can also help explain the reason for gender discrimination. Filer (1983) found the empirical significance in male/female differences in tastes and personalities. Therefore, discrimination by gender not only manifests itself in differences in human capital formation, but also it is evident in distinctive career choices. Different expectations of career paths, tastes and personalities between women and men lead females to prefer certain industries or job functions to others.

Gender discrimination in the workplace can be divided into two types: access and treatment (Levitin, Quinn, and Staines, 1971). Access discrimination occurs when members of a certain gender are not hired into certain jobs because of policies and procedures (written or unwritten) that bar or discourage their recruitment. Treatment discrimination occurs when qualified members of a certain gender receive lower salaries, lower status, or lower positions than comparable members of the opposite sex. In this paper, treatment discrimination will be examined using human capital theory so further discussion of treatment discrimination will be deferred until that section.

In terms of access discrimination, Truman and Baroudi (1994) indicated that there was lower female participation in the IT occupations than in other business occupations, such as Accountants and Auditors, Personnel and Labor Relations Managers, and Financial Managers. They concluded that this was <u>not</u> a result of access discrimination as the percentage of females in IT jobs roughly matched the percentage of females graduating from Computer Science programs.

However, Truman and Baroudi (1994) did conclude that, given the relatively low numbers of women graduating from Computer Science programs, access discrimination was

occurring at some earlier stage. (Recent census data (Statistical Abstract of the United States 2001) indicate that females still constitute only 29.2% of the Computer Systems Analysts and Scientists.) This situation stems from attitudes and computer usage patterns that are traced to the home, media, school and college socialization process (Grundy, 1996; Siann, 1977; Wilson, 1997). Butler (2000) states "that computer access and use by girls and boys in their formidable years is believed to contribute to gender inequalities in the IT profession". There is a tendency for girls to use computers less at home than boys (Nelson and Cooper, 1989; Culley, 1988; Locheed, 1985). Aggressive behavior of boys in the classroom (e.g., grabbing computers in a first-come, first-served situation) further reduces computer use by girls (Stalker, 1983; Sanders, 1985; Elliot, 1990). Software tends to be biased in favor of boys' interests (Lockheed, 1985; Stalker, 1983; Culley, 1988; Forsyth and Lancy, 1989; Wilder, 1985). Boys have a much greater presence and participation in computer programming courses (Locheed, 1985). A possible reason for this may be that there is a lack of women computer teachers serving as role models for girls (Stalker, 1983).

Gender inequality in the IT profession can also be traced to the segregation of labor markets. This segregation can affect labor supply by encouraging or dissuading women in their selection of preferred work. There is a cultural bias that technology is a "male" domain (Stalker, 1983; Sanders, 1985; Thurston, 1990; Lage, 1991; Fredman, 1992). There is also a perception that computer technology is linked to mathematics – a profession that girls feel is identified more with males (Culley, 1988; Eastman and Krendl, 1987; Thurston, 1990). Further, the image of the computing culture as male-gendered work makes the profession unattractive to women (Panteli, Stack, and Ramsay, 1999). The segregation of labor markets also impact the type and status of IT jobs that women are employed in. Von Hellens, Pringle, Nielsen, and Greenhill (2000)

indicate that women in IT jobs are increasingly concentrated in low status job classifications and interface strictly with customers while their male counterparts are employed in higher status and higher paying IT jobs where they interface predominantly with other IT professionals and with managers. This strengthens the social network for males working in IT while segregating or isolating female IT workers.

HUMAN CAPITAL THEORY

Human capital theory is the dominant economic theory of wage determination (Berndt, 1991). Its origins trace back to the 18th century writings of Adam Smith (1937). Schultz (1960, 1961) popularized the idea of "human capital" -- the idea of treating educational spending as an investment. Berndt (1991) states that educated workers are (hopefully) more productive than their less educated counterparts. As a result, they are more likely to command higher wages. This is a straightforward extension of Smith's idea of equalizing differences. This theory also provides an economic explanation as to why a person will forego earnings and incur additional expenses to undertake an education. In addition to formal education, on-the-job training is also important in the accumulation of one's human capital because many job skills are acquired through training sessions, apprenticeships, and similar efforts (Becker, 1962, 1964; Mincer, 1958, 1962, 1974).

Gender treatment discrimination occurs when qualified members of one gender receive lower salaries, lower status, or lower positions than comparable members of the opposite sex (Levitin, Quinn, and Staines, 1971). Treatment discrimination represents a situation in which the treatment of employees is based more on their subgroup membership than on their merit or achievements (Greenhaus, Parasuraman, and Wormley, 1990; Moyes, Williams, and Quigley, 2000).

Assessing Gender Differences in Software Developers Using the Human Capital Model

Economists take the position that some part of the gender gap in earnings is due to average group differences in productivity-linked characteristics (a human capital gap) and some part is due to average group differences in treatment (a discrimination gap). Gaps that can be explained by human capital differences make it easier to assert that labor markets function in a nondiscriminatory manner (Darity and Mason, 1998).

The neoclassical economic school has indicated that gender inequalities are likely to decline with industrialization or economic growth. Several studies within the neoclassical economic approach have argued that differences between men and women result primarily from human capital differentials (education, skills, expected length of labor-force participation) that are bound to wither away over time (Forsythe, Korzeniewicz, and Durrant, 2000). These neoclassical economic models suggest that employer discrimination will be eliminated from the market in the long run (Gerhart, 1990). According to this perspective, the process of economic growth, through the opportunities and constraints created by the expansion of markets, can be expected to undermine the inequalities that result from discriminatory practices (Forsythe, Korzeniewicz, and Durrant, 2000). This is accomplished through competitive advantage realized by firms that hire equally qualified employees from the disadvantaged group at lower salaries in the short run (Gerhart, 1990).

If there are salary differences between female and male software developers, a neoclassical economist might argue that, on average, male software developers are more productive than female software developers so they should be paid more, on average. (However, applying the group average or assessment when performing an assessment of an individual member of that group is "statistical discrimination" (Albeda, Drago, and Shulman, 1997)). A possible explanation for this average greater productive (if it does indeed exist) is that males

have had much greater access to computers and computer courses throughout their lifetimes. Unfortunately, this greater experience may not be accounted for in an experience factor since this factor usually measures only relevant job experience.

HUMAN CAPITAL MODEL

The econometric literature on wage determination based on human capital has for the most part been based on regression models of the following form: the natural logarithm of earnings is a function of a measure of schooling, a measure of experience, possibly other factors, and a random disturbance term. Roy's (1950) research showed that there is a relationship between earnings distributions and the underlying abilities of the employee (such as intelligence, physical strength, etc.). Roy (1950) also showed that if each of the underlying abilities is normally distributed than the logarithm of earnings will be normally distributed.

Mincer (1974) showed that the regression equation for wages is linear in education but quadratic in experience. That is:

$$\log Y_i = \log Y_0 + \beta_1 S_i + \beta_2 X_i + \beta_3 X_i^2 + u_i \qquad (1)$$

where Y_i is the wages, S_i is education, X_i is experience, and u_i is the random disturbance. Because earnings cannot increase indefinitely as experience increases, estimates of β_2 should be positive while estimates of β_3 should be negative.

SURVEY DETAILS AND SUMMARY STATISTICS

We obtained our data from the voluntary web-based survey on salary and skills of IT workers conducted by Dice Incorporated (http://www.dice.com), an on-line placement company. The date was from June 7, 2000 to April 13, 2001. One caveat can be raised regarding the representation of the respondents of this survey: the survey sample was not random since the respondents were self-selecting and voluntary making non-representativeness and the self-

reporting bias a possibility. Further, the on-line nature of the survey may introduce a bias towards younger workers. However, we argue that these two sets of potential biases would introduce little, if any, bias in our comparative analysis towards female or male software developers because it is likely that these biases, if present, would be equally distributed for both subgroups of the sample.

In the on-line survey, a respondent could select from 38 different job titles. To identify software developers, we used 7 of these job titles -- Developer: Applications, Developer: Client/Server, Developer: Database, Developer: Systems, Mainframe Systems Programmer. Software Engineers, and Web Developer/Programmer. Any problematic data was removed from our sample using the rules listed in the Appendix. This resulted in a sample of 5,547 software developers which was used in this analysis.

Females were somewhat under-represented in this sample as only 17.3% of the survey respondents were female compared to recent U.S. Department of Labor (2002) data indicating that 27.2% of computer programmers are female. In addressing this issue, the following explanations are offered. Since the survey was placed on an on-line placement company's web site, the survey respondents were more likely to be actively seeking new employment than typical software developers. Research on gender differences and job searches provide a possible explanation on why the percentage of female survey respondents is low -- women conduct a job search with less intensity than men (Keith and McWilliams, 1999) and women are less likely to separate from an existing job (Kulik, 2000). Therefore, these differences would impact only the relative number of respondents and would have little, if any, impact on our comparative analysis of female and male software developers.

In Table 1, the respondents were categorized by (technical) experience level (6 categories) and gender. The modal and median experience group for male, female, and overall is Level 3 (3 to 5 years experience). For the first 2 experience level categories, females constituted 26.9% and 21.5% of the respondents which is considerably more than their overall percentage of only 17.3%. For the 6x2 contingency table, the chi-square test statistic is 51.6 (p-value < .0001). So the pattern of experience clearly indicates that females have less experience than their male counterparts.

Table 1. Experience and Gender Percentages

Experience Level	Experience (in years)	Number of Respondents	Males	Females
1	less than 1	413 (7.4%)	302 (73.1%) (6.6%)	111 (26.9%) (11.6%)
2	1 or 2	989 (17.8%)	776 (78.5%) (16.9%)	213 (21.5%) (22.3%)
3	3 to 5	2,042 (36.8%)	1,721 (84.3%) (37.5%)	321 (15.7%) (33.5%)
4	6 to 10	1,088 (19.6%)	929 (85.4%) (20.2%)	159 (14.6%) (16.6%)
5	11 to 14	422 (7.6%)	358 (84.8%) (7.8%)	64 (15.2%) (6.7%)
6	15 or more	593 (10.7%)	504 (85.0%) (11.0%)	89 (15.0%) (9.3%)
Overall		5,547	4,590 (82.7%)	957 (17.3%)

Table 2 presents the percentage of males and females for the different education levels. The modal education group for male, female, and overall is College Grad with over half the respondents falling into this category for all three groups. Gender differences are not as pronounced as was the case for experience. 81.2% of females have a college degree (College Grad, Master's Degree, Doctoral Degree, or Professional Degree) compared with only 77.3% of males. For the 8x2 contingency table, the chi-square test statistic is 24.53 (p-value 0.0009). So

the pattern of education indicates that females have different educational backgrounds than their male counterparts but the practical differences may not be substantial enough to make the claim that female software developers are better educated.

In terms of reported yearly salary, males reported making an average of $65,948 per year while their female counterparts made only $59,336 per year. (As expected, the test for mean differences was statistically significant with a p-value < 0.0001). So in our sample, females make 89.98% of what their male counterparts make which is very close to the U.S. Department of Labor (2002) statistic of 88.92% for Computer Programmers.

Table 2. Education and Gender Percentages

Education Level	Number of Respondents	Males	Females
High School	158 (2.8%)	141 (89.2%) (3.1%)	17(10.8%) (1.8%)
Military	29 (0.5%)	29 (100.0%) (0.6%)	0 (0.0%) (0.0%)
Vocation/Tech School	158 (2.8%)	126 (79.7%) (2.7%)	32 (20.3%) (3.3%)
Some College	876 (15.8%)	745 (85.0%) (16.2%)	131 (15.0%) (13.7%)
College Grad	2,893 (52.2%)	2,382 (82.3%) (51.9%)	511 (17.7%) (53.4%)
Master's Degree	1,256 (22.6%)	1,012 (80.6%) (22.0%)	244 (19.4%) (25.5%)
Doctoral Degree	102 (1.8%)	93 (91.2%) (2.0%)	9 (8.8%) (0.9%)
Professional Degree (MD, JD)	75 (1.4%)	62 (82.7%) (1.4%)	13 (17.3%) (1.4%)
Overall	5,547	4,590 (82.7%)	957 (17.3%)

In contrast, the U.S. Department of Labor (2002) data indicate that currently full-time female computer programmers make $867 per week (median) compared to their male counterparts who make $975 per week (median). This converts into female computer

programmers making only $45,004 ($865 x 52) per year and male computer programmers making only $50,700 ($975 x 52) per year. In addressing the issue of why our average yearly salaries are higher, the following argument can be made. Since the survey was placed on an on-line placement company's web site, the survey respondents were more likely to be actively seeking new employment than typical software developers. Given the time frame of the survey (before the demand for software developers dropped off), these software developers were probably better qualified than their non-responding counterparts. Again, we argue that this potential bias would have little impact on our comparative analysis of female and male software developers because it is appears that this bias is equally distributed for male and female software developers.

HUMAN CAPITAL MODEL RESULTS

Model Preliminaries

Mincer (1974) showed that the regression equation for wages is linear in education but quadratic in experience as given in (1). Berndt (1991) suggested that rather than using annual salaries, the hourly salary rate should be employed. Since the respondents also indicated the average number of hours worked per week, we fit the human capital model by taking the annual salary and dividing it by the estimated hours worked per year. The estimated hours worked per year is the number of weeks per year (365 / 7) times the average hours worked per week.

Since the respondents indicated a technical experience level rather experience in years, the experience level was scaled as follows: (1) 0.5 for less than 1 year, (2) 1.5 for 1-2 years, (3) 4.0 for 3-5 years, (4) 8.0 for 6-10 years, (5) 12.5 for 11-14 years, and (6) 17.5 for more than 15 years. The highest education level attained by each respondent was scaled into education years as follows: (1) 12 for High School, (2) 14 for Military, (3) 14 for Vocational/Tech School, (4)

14 for Some College, (5) 16 for College Grad, (6) 18 for Master's Degree, (7) 20 for Doctoral Degree, and (8) 20 for Professional Degree (MD, JD).

Simple Multiplicative Model

Before proceeding with the human capital model results, we present a simple multiplicative model to assess gender differences in software developers that does not take into account any human capital factors.

$$\log Y_i = \log Y_0 + \beta_1 G_i + u_i \tag{2}$$

This model is equivalent to:

$$Y_i = Y_0 e^{\beta_1 G_i + u_i} \tag{3}$$

Fitting this model using our data indicates that males make 107.97% more than females, or conversely, females make only 92.62% of what their male counterparts make. This model is highly significant with a p-value less than 0.0001. This increase from 89.98% to 92.62% can probably be attributed to the number of hours worked per week as males reported working an average of 42.8 hours per week while their female counterparts reported working only an average of 41.5 hours per week. The fact that female software developers work, on average, less hours per week compared to their male counterparts is not that surprising as U.S. labor statistics, over the last 20 years, indicate that full-time female employees consistently work less than their male counterparts (Ehrenberg and Smith, 1997; U.S. Department of Labor, 2002).

Base Human Capital Model

Next, we present overall results for the base human capital model. Equation (1) is modified to account for the gender factor indicator – G_i which is equal to 1 for males and 0 for females. This model is given below in equation (4). Similar to equation (2), this model is multiplicative so $e^{\beta 4}$ gives the percentage (after converting this number to a percentage) that

males make more than their female counterparts. In all the human capital model results presented, the coefficients will be referred to as the base (intercept term), education (β_1), Exp (β_2), Exp^2 (β_3), and Gender (β_4).

$$\log Y_i = \log Y_0 + \beta_1 S_i + \beta_2 X_i + \beta_3 X_i^2 + \beta_4 G_i + u_i \qquad (4)$$

The results given in Table 3 indicate that males make 104.66% ($e^{(Gender)}$) more than females, or conversely, females make only 95.55% ($1/e^{(Gender)}$) of what their male counterparts make. This increase from 92.62% to 95.55% indicates that education and experience are two significant factors that help explain some of the gender differences for software developers. This model is highly significant with a p-value less than 0.0001 and an associated Adjusted R^2 of 0.1984. All coefficients are also highly significant with the highest (least significant) p-value at 0.0009 for the gender coefficient. As expected by the human capital model, experience has a positive coefficient while Exp^2 has a negative coefficient.

Table 3. Base Human Capital Model – Overall Results

Coefficient or Statistic of Interest	Value	t-value	p-value
Base	2.17496	40.70	< 0.0001
Education	0.04695	14.82	< 0.0001
Exp	0.09076	23.89	< 0.0001
Exp^2	-0.00344	-16.78	< 0.0001
Gender	0.04551	3.34	0.0009
Adjusted R^2	0.1984		< 0.0001

Human Capital Model with Specific Skills

The results of a forthcoming study by Dattero and Galup (2004) indicated that choice of programming language differs between genders. For example, males have a greater than expected representation when considering object-oriented languages, such as Java and C++, while females have a greater than expected representation when considering more traditional

programming languages, such as COBOL. In exploring the data set used in this study, software developers with objected-oriented skills had greater salaries than their counterparts without this skill. Hence, the object-oriented programming skill seems to be a very likely human capital factor. This difference may explain some of the differences in salaries since 54.2% of males in our sample have this skill while only 41.3% of females have it.

To balance the advantage of the object-oriented programming skill, programmers with only COBOL skills make substantially less. Similarly, programmers with only Visual Basic skills make substantially less than their object-oriented programming counterparts. If one also included these two factors in a human capital model, fitting the regression would produce negative coefficients for these two factors. Since losing human capital is not a logical consequence of adding an additional ability, these two factors were not included in the model even though the two factors would be significant and slightly increase the Adjusted R^2.

In exploring the data set used in this study, SQL was the one other skill factor that had a significant effect and produced a positive regression coefficient. In our sample, 55.3% of males have this skill and 53.2% of females have it.

Incorporating the SQL and objected-oriented programming skills factor into the human capital model with the gender effect (Equation 4) produces the following human capital model (Equation (5)) where O_i is equal to 1 if the developer indicated knowledge of Java, C++, Smalltalk, or OOP (and 0 otherwise), and Q_i is equal to 1 if the developer indicated knowledge of SQL (and 0 otherwise). Similar to equation (4), this model is multiplicative so $e^{\beta 5}$ gives the human capital increase produced by possessing the object-oriented skill while $e^{\beta 6}$ gives the human capital increase produced by possessing SQL skills. In the human capital model results presented, the two new coefficients will be referred as OOP (β_5) and SQL (β_6).

$$\log Y_i = \log Y_0 + \beta_1 S_i + \beta_2 X_i + \beta_3 X_i^2 + \beta_4 G_i + \beta_5 O_i + \beta_6 Q_i + u_i \qquad (5)$$

The results given in Table 4 indicate that females make only 96.58% ($1/e^{(Gender)}$) of what their male counterparts make. This increase from 95.55% to 96.58% indicates that objected-oriented programming and SQL are two significant human capital skill factors that help explain some of the gender differences for software developers. This model is highly significant with a p-value less than 0.0001 and an associated Adjusted R^2 of 0.2106. All coefficients are also highly significant with the highest (least significant) p-value at 0.0108 for the gender coefficient. The OOP skill produces a 107.71% increase in salary while the SQL skill produces a 103.25% increase in salary.

Table 4. Human Capital with Skills Model – Overall Results

Coefficient or Statistic of Interest	Value	t-value	p-value
Base	2.19670	41.00	< 0.0001
Education	0.04276	13.35	< 0.0001
Exp	0.08972	23.72	< 0.0001
Exp^2	-0.00336	-16.48	< 0.0001
Gender	0.03479	2.55	0.0108
OOP	0.07428	7.11	< 0.0001
SQL	0.03196	3.11	0.0019
Adjusted R^2	0.2106		< 0.0001

CHOW TEST

In economics, an alternative to employing a gender indicator variable in the human capital model is to fit the overall model (with both male and female data) and then fit the model to each of the two disjoint subsets (a model with only male data and another model with only female data). In assessing and testing differences between two or more disjoint subsets (such as male and female) using the same econometric regression model, Chow (1960) is credited with

developing a test of coefficient equality between the two subsets by fitting the regression model to each of the two subsets and the entire or pooled set of data.

The Chow test is conducted as follows. First, fit the model (with k parameters – not counting the intercept term) using the entire set of data and compute the (restricted) residual sum of the squares denoted by RRSS. Second, fit the model using only the first subset of data (of size n_1) and compute the residual sum of the squares denoted by RSS_1. Third, fit the model using only the second subset of data (of size n_2), and compute the residual sum of the squares denoted by RSS_2. Next, compute the unrestricted residual sum of the squares denoted by URSS by summing RSS_1 and RSS_2. The test statistic is distributed according to an F-distribution with (k + 1) and ($n_1 + n_2 - 2k - 2$) degrees of freedom and is computed as follows.

$$F = \frac{(RRSS - URSS)/(k+1)}{URSS/(n_1 + n_2 - 2k - 2)} \tag{6}$$

The results for the base human capital model for the three sets are given in Table 5. The Chow test indicates a significant difference between the male and female models with a p-value equal to .0002. Examining the models more carefully, the base coefficient for males is substantially higher (130%) than for females indicating that males have a substantial initial advantage in salary. Females get a substantially greater increase from education than their male counterparts which somewhat lessens the base impact. Males benefit slightly more from experience than females. With the strong results for the Chow test and the substantially larger base coefficient for males, one could make a strong argument that the difference in wages appears to be based considerably on gender.

Table 5. Base Human Capital Models and Chow Test Results

Coefficient or Statistic of Interest	All	Male	Female
Base	2.21488*	2.29213*	1.76777*
Education	0.04660*	0.04244*	0.07238*
Exp	0.09174*	0.09111*	0.08784*
Exp^2	-0.00348*	-0.00346*	-0.00323*
Adjusted R^2	0.1970	0.1859	0.2454
Model p-value	< 0.0001	< 0.0001	< 0.0001
N	5,547	4,590	957
Residual Sum of Squares	810.62536	677.41909	129.90515

* Coefficient significantly different from 0 at < 0.0001 level
Chow Test statistic = 5.66 with a p-value = 0.0002

The results for the human capital model with the added OOP and SQL skills for the three sets are given in Table 6. Most of the values in Table 6 are similar to those in Table 5. Surprisingly, males get a greater increase (108%) from possessing OOP skills than their female counterparts (105%) while females get a greater increase (105%) from possessing SQL skills than their male counterparts (103%).

Table 6. Human Capital with Skills Models and Chow Test Results

Coefficient or Statistic of Interest	All	Male	Female
Base	2.22806***	2.29799***	1.81804***
Education	0.04235***	0.03864***	0.06650***
Exp	0.09043***	0.09015***	0.08632***
Exp^2	-0.00339***	-0.00339***	-0.00311***
OOP	0.07704***	0.07700***	0.05325**
SQL	0.03227***	0.02758**	0.04711*
Adjusted R^2	0.2062	0.1946	0.2509
Model p-value	< 0.0001	< 0.0001	< 0.0001
N	5,547	4,590	957
Residual Sum of Squares	801.03222	669.89718	128.68867

*** Coefficient significantly different from 0 at 1% level
** Coefficient significantly different from 0 at 5% level
* Coefficient significantly different from 0 at 10% level
Chow Test statistic = 2.83 with a p-value = 0.0095

OAXACA DECOMPOSITION

The Chow test provides only an indicator on whether the gender models are significantly different in structure. The Oaxaca (1973) decomposition breaks up the salary gender differences into two components: explained and unexplained (or residual). The explained portion is due to the differences in human capital while the unexplained portion can be attributed to discrimination. An often unstated or hidden assumption in the presentation of an Oaxaca decomposition is that the human capital model contains the significant human capital factors – if the human capital model is seriously lacking in significant human capital factors, the unexplained portion is not really a measure of discrimination but merely a measure of the poorness of the human capital model.

By the definition of least squares regression, the mean of the outcome (dependent variable) for each group is equal to the sum of the mean of each observed variable (independent variable) times the estimated coefficient of that variable. In the Oaxaca decomposition, the equation of one of the groups is subtracted from the other group producing the "difference equation". If one "accidentally" uses the set of means for males with the female model (or the set of means for females with the male model), this "accidental set of terms" (on the right-hand side of the equation) is precisely what is added to and subtracted from the difference equation. Algebraically rearranging this result produces the explained and unexplained components. Employing this process to the base human capital model (equation (1)) produces the following decomposition process. Equation (7a) is the fitted male human capital model with the corresponding male averages plugged in. Equation (7b) is the fitted female human capital model with the corresponding female averages plugged in. Equation (7c) is the fitted female human capital model with the corresponding male averages "accidentally" plugged in so the set of three

terms on the right-hand side will be added to and subtracted from the difference equation. Equation (8) is the final result with the explained component on the line above the unexplained component.

$$\overline{\log Y}_m = b_{m0} + b_{m1}\overline{S_m} + b_{m2}\overline{X}_m + b_{m3}\overline{(X_m^2)} \tag{7a}$$

$$\overline{\log Y}_f = b_{f0} + b_{f1}\overline{S_f} + b_{f2}\overline{X}_f + b_{f3}\overline{(X_f^2)} \tag{7b}$$

$$\overline{\log Y}_f = b_{f0} + b_{f1}\overline{S_m} + b_{f2}\overline{X}_m + b_{f3}\overline{(X_m^2)} \tag{7c}$$

$$\begin{aligned}&\overline{\log Y}_f - \overline{\log Y}_m = \\ &b_{f1}(\overline{S_f} - \overline{S_m}) + b_{f2}(\overline{X}_f - \overline{X}_m) + b_{f3}(\overline{(X_f^2)} - \overline{(X_m^2)}) + \\ &(b_{f0} - b_{m0}) + (b_{f1} - b_{m1})\overline{S_m} + (b_{f2} - b_{m2})\overline{X}_m + (b_{f3} - b_{m3})\overline{(X_m^2)}\end{aligned} \tag{8}$$

The results in Table 7 show the values needed for the Oaxaca decomposition for the base human capital model. In our sample, females make only 92.50% of what their male counterparts make which is a difference of $2.24 per hour. Females have a slighter higher education average while males have a higher number of average years of technical experience. Since the human capital model is in terms of the log of hourly salary, the left-hand side of the decomposition is the difference between the average of the log of hourly salary for males and the average of the log of hourly salary for males. So 0.07652 is decomposed into explained and unexplained components which are 0.03163 and 0.04494, respectively (the minor differences in the sum of the component parts and the whole is attributed to calculation of the logs and rounding error). Therefore, 58.7% of the salary differences are unexplained by education and experience and could be potentially attributed to gender discrimination.

Table 7. Oaxaca Decomposition for Base Human Capital Model

Coefficient or Statistic of Interest	Male	Female
Average Hourly Salary	29.90006	27.65833
Average of Log(Hourly Salary)	3.32256	3.24604
Base Coefficient	2.29213	1.76777
Education Coefficient	0.04244	0.07238
Average Education	16.06100	16.19018
Exp Coefficient	0.09111	0.08784
Average Exp	6.30196	5.52612
Exp^2 Coefficient	-0.00346	-0.00323
Average Exp^2	65.16362	55.46003

The results in Table 8 show the values needed for the Oaxaca decomposition for the human capital model with OOP and SQL skills. For each skill indicator variable, the average value is simply the proportion possessing this skill. A much greater proportion of males possess the OOP skill but only a slightly higher proportion of males possess the SQL skill. So 0.07652 is decomposed into explained and unexplained components which are 0.04258 and 0.0339494, respectively. Therefore, 44.4% of the salary differences are unexplained by education, experience, OOP skills, and SQL skills and could be potentially attributed to gender discrimination.

If one reviews the earlier human capital model with OOP and SQL skills with the gender indicator (no decomposition) to check for logical consistency with the Oaxaca decomposition, females made only 96.58% compared to their male counterparts. So for the initial 7.5% (100% - 92.5%) difference in hourly salary, 3.4% of this difference can be viewed as "unexplained". This means 45.6% (3.42 / 7.50) of the salary differences could be potentially attributed to gender discrimination which is quite close to the 44.4% figure from the Oaxaca decomposition.

Table 8. Oaxaca Decomposition for Human Capital with Skills Model

Coefficient or Statistic of Interest	Male	Female
Average Hourly Salary	29.90006	27.65833
Average of Log(Hourly Salary)	3.32256	3.24604
Base Coefficient	2.29799	1.81804
Education Coefficient	0.03864	0.06650
Average Education	16.06100	16.19018
Exp Coefficient	0.09015	0.08632
Average Exp	6.30196	5.52612
Exp^2 Coefficient	-0.00339	-0.00311
Average Exp^2	65.16362	55.46003
OOP Coefficient	0.07700	0.05325
Proportion with OOP Skill	0.54205	0.41275
SQL Coefficient	0.02758	0.04711
Proportion with SQL Skill	0.55251	0.53187

DISCUSSION AND CONCLUDING REMARKS

Education and experience are two well-known major factors that impact salary. Similarly, knowledge of object-oriented programming and SQL also contribute to a software developer's human capital. No additional skills were found that had positive coefficients (contributed to one's human capital) and significantly contributed to the human capital model.

In our sample, males reported making an average of $65,948 per year while their female counterparts made only $59,336 per year. So females make 89.98% of what their male counterparts make which is very close to the U.S. Department of Labor (2002) statistic of 88.92% for Computer Programmers. Converting the annual figures in our sample to hourly rates (the respondent indicated the average number of hours worked per week), females make only 92.5% of what their male counterparts make which is a difference of $2.24 per hour. Accounting for human capital factors (education, experience, and specific skills) reduces the gender difference to 96.58%. Despite this difference being substantially less (3.42% instead of

10.02%), females making only 96.58% compared to their male counterparts has practical significance given that this difference translate into a few thousand dollars per year for female software developers. To provide validation of this result, the sample was split into male and female subsets. The results for the human capital models, Chow test, and Oaxaca decomposition provides confirmatory evidence of this result.

Some neoclassical economists may have reservations about our human capital models. If there are salary differences between female and male software developers, these neoclassical economists might argue that, on average, male software developers are simply more productive than female software developers so they should be paid more, on average. Since no explicit productivity measures are included in our human capital models, this factor would be part of the unexplained component of the Oaxaca decomposition.

Some concerns about how well-representative our sample is to the population of software developers can be raised. Throughout the paper, we tried to point out any potential bias or shortcomings in our sample. In final defense of our sample, we feel that our sample is sufficiently good for our analysis but reiterate what Berndt (1991) said: "the practicing econometrician in labor economics is typically forced to make use of data that are considerably less than ideal".

Despite some concerns about our sample and some reservations about our human capital models, we feel that our study provides a good indication of gender/salary differences for software developers. Unfortunately, our best efforts in finding a good human capital model for software developers indicate that gender is still a statistically significant factor in assessing one's human capital.

The recent study by the Information Technology Association of America (ITAA 2003) on diversifying the IT workforce provides some guidance in providing possible remedies for the

apparent discrimination in software developers – "…companies should demonstrate consistent and proactive approaches to attracting, retaining, and promoting qualified women and minority candidates in IT. Without serious leadership from IT employers, many minority applicants may choose other occupational fields." Specifically, the study recommends (1) a stronger commitment from corporate leadership, (2) increased corporate outreach and mentoring, and (3) fostering stronger partnerships between IT companies and colleges and universities.

We concur with the recommendations of the ITAA. We feel it is important to draw more women to the IT profession. Our human capital models indicate that education is an important determinant in salary so female software developers (and aspiring female software developers) should be strongly encouraged to increase their educational background by pursuing degrees in IT programs (such as computer science, computer information systems, etc.). In addition, female software developers should be encouraged to acquire specific higher paying skills such as OOP and SQL skills. Further, it is important to make sure that female software developers are welcomed into the IT community as a number of psychological studies indicate, "through experience, people come to share beliefs about the extent to which tasks are linked to gender" (Vancouver and Ilgen, 1989). Panteli, Stack, and Ramsay (1999) recommend a change in "the attitudes of employers, male colleagues, and managers. Male dominating attitudes and perpetuating stereotypes seem to pre-determine the positions that should be held by women … Indeed, women themselves may come to believe the stereotypes." Finally and probably most importantly, the IT salary structure in companies should be carefully examined to make sure that differences in software developers' salaries can be clearly attributed to human capital factors.

REFERENCES

Albeda, R., Drago, R.W., and Shulman, S. (1997). Unlevel Playing Fields: Understanding Wage Inequality and Disrimination, McGraw-Hill.

Arnold, D. and Niederman, F. (2001). The Global IT Work Force. Communications of the ACM. 44(7), 31-33.

Auster, E. (1989). Task Characteristics as a Bridge between Macro and Micro Research on Salary Inequality between Men and Women. Academy of Management Review. 14(2), 173-193.

Becker, G.S. (1962). Investment in Human Capital: A Theoretical Analysis. Journal of Political Economy. 70(5 Part 2), S9-S49.

Becker, G.S. (1964). Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education. National Bureau of Economic Research.

Berndt, E.R. (1991). The Practice of Econometrics: Classic and Contemporary, Addison-Wesley.

Butler, D. (2000). Gender, girls, and computer technology: What's the status now?. The Clearing House. 73(4), 225-230.

Chow, G.C. (1960). Tests of Equality Between Subsets of Coefficients in Two Linear Regression Models. Econometrica. 28, 591-605.

Culley, L. (1988). Girls, boys and computers. Educational Studies. 14, 3-8.

Darity, D. and Mason, P. (1988). Evidence on Discrimination in Employment: Codes of Color, Codes of Gender (in Symposium: Discrimination in Product, Credit and Labor Markets) . The Journal of Economic Perspectives. 12(2), 63-90.

Dattero, R and Galup, S.D. (2004). Programming Languages and Gender. Communications of the ACM.

Daymont, T.N and Andrisani, P.J. (1984). Job Preferences, College Major, and the Gender Gap in Earnings . Journal of Human Resources. 19(3), 408-428.

Eastman, S. and Krendl, K. (1987). Computers and gender: Differential effects of electronic search on students' achievement and attitudes. Journal of Research and Development in Education. 20, 41-48.

Ehrenberg, R.G. and Smith, R.S. (1997). Modern Labor Economics: Theory and Public Policy (6th edition), Addison-Wesley.

Elliott, J. (1990). Sex equity in computer education: An enrichment program for seventh and eighth grade girls. In Yes, I can, edited by A. Fredman, 25-28.

England, P. and McCreary, L. (1987). Integrating Sociology and Economics to Study Gender and Work, in Women and Work. An Annual Review (Volume 2). Sage Publications, Beverly Hills, CA, 143-172.

Filer, R. (1983). Sexual Differences in Earnings: The Role of Individual Personalities and Tastes. The Journal of Human Resources. 18(1), 82-100.

Forgionne, G.A. and Peeters, V.E. (1982). Differences in Job Motivation and Satisfaction Among Female and Male Managers. Human Relations. 5(2), 101-118.

Forsyth, A. and Lancy, D. (1989). Girls and microcomputers: A hopeful finding regarding software. Computers in the School. 6, 51-59.

Forsythe. N., Korzeniewicz, R., and Durrant, V. (2000). Gender inequalities and economic growth: A longitudinal evaluation. Economic Development and Cultural Change. 48(3), 573-617.

Fredman, A., ed. (1992). Yes I can: Action projects to resolve equity issues in educational computing. International Society for Technology in Education.

Gerhart, B. (1990). Gender Differences In Current And Starting Salaries: The Role of Performance, College Major, and Job Title. Industrial & Labor Relations Review. 43(4), 416-434.

Greenhaus, J., Parasuraman, S., and Wormley, W. (1990). Effects of race on organizational experiences, job performance evaluations, and career outcomes. Academy of Management Journal. 33(1), 64-86.

Grundy, F. (1996). Women and Computers, Intellect Books, Exeter.
Hulin, C.L. and Smith, P.C. A Linear Model of Job Satisfaction. Journal of Applied Psychology. (49:3), 1965, pp. 209-216.

ITAA (2003). Presentation of the Blue Ribbon Panel on IT Diversity, presented at the National IT Workforce Convocation. Arlington, VA.

Keith, K. and McWilliams, A. (1999). The Return to Mobility and Job Search by Gender . Industrial and Labor Relations Review, 52(3), 460-477.

Kulik, L. (2000). A Comparative Analysis of Job Search Intensity, Attitudes Toward Unemployment, and Related Responses. Journal of Occupational and Organizational Psychology. 73, 487-500.

Lage, E. (1991). Boys, girls, and microcomputing. European Journal of Psychology in Education. 6, 29-44.

Levitin, T., Quinn, R.P., and Staines, G.L. (1971). Sex Discrimination Against the American Working Women. American Behavioral Scientist. 15(2), 238-254.

Lockheed, M. (1985). Women, girls, and computers: A first look at the evidence. Sex Roles.13, 115-122.

Mincer, J. (1958). Investment in Human Capital and Personal Income Distribution. Journal of Political Economy. 66(4), 281-302.

Mincer, J. (1962). On-the-Job Training: Costs, Returns, and Some Implications. Journal of Political Economy.70(5 Part 2), S50-S79.

Mincer, J. (1974). Schooling, Experience and Earnings, Columbia University Press for the National Bureau of Economic Research.

Moyes, G., Williams, P., and Quigley, B. (2000). The relation between perceived treatment discrimination and job satisfaction among African-American accounting professionals. Accounting Horizons. 14(1), 21-49.

Nelson, L., and Cooper, I. (1989). Sex role identity, attributional style, and attitudes toward computers. Paper presented at the annual meeting of the Eastern Psychological Association, Boston, MA.

Oaxaca, R. (1973). Male-Female Wage Differentials in Urban Labor Markets. International Economic Review. 14(3), 693-709.

Olson, J.E. and Frieze, I.H. Income Determinants for Women in Business, in Women and Work. An Annual Review (Volume 2).

Panteli, A., Stack, J. and Ramsay H. (1999). Gender and Professional Ethics in the IT Industry. Journal of Business Ethics. 22, 51-61.

Reinsmith, L. (2002). Proving an employer's intent: Disparate treatment discrimination and the stray remarks doctrine after Reeves v. Sanderson Plumbing Products. Vanderbilt Law Review. 55(1), 219-260

Roy, A. (1950). The Distribution of Earnings and of Individual Output. Economic Journal. 60(3), 489-505.

Sanders, J. (1985). Here's how you can help girls take greater advantage of school computers. American School Board Journal. 172, 37-38.

Schultz, T.W. (1960). Capital Formation by Education. Journal of Political Economy. 68(6), 571-583.

Schultz, T.W. (1961). Investment in Human Capital. American Economic Review. 51(1), 1-17.

Siann, G. (1997). We Can, We Don't Want To: Factors Influencing Women's Participation in Computing . In Social Europe: Occupational Segregation of Women and Men in the European Community Supplement, edited by Lander and Adam. 70-76.

Smith, A. (1937). The Wealth of Nations (reprinted edition). Random House.

Stalker, S. (1983). Computers in the classroom: A feminist issue. Paper presented at the National Women's Studies Association Conference. Columbus, Ohio.

Statistical Abstract of the United States (2001). U.S. Census Bureau.

Thurston, L. (1990). Girls, computers, and amber waves of grain: Computer equity programming for rural teachers. Paper presented at the Annual Conference of the National Women's Studies Association. Towson, MD.

Trauth, E. (2001). Mapping Information-Sector Work to the Work Force. Communications of the ACM. 44(7), 74-75.

Truman, G. and Baroudi, J. (1994). Gender Differences in the Information Systems Managerial Ranks: An Assessment of Potential Discriminatory Practices. Management Information Systems Quarterly. 18(2), 129-141.

U.S. Department of Labor (2002). Bureau of Labor Statistics, Employment and Earnings.

Vancouver, J.B. and Ilgen, D.R. (1989). Effects of Interpersonnal Orientation and the Sex-Type of the Task on Choosing to Work Alone or in Groups. Journal of Applied Psychology. 74(6), 927-934.

Von Hellens, L. and Nielsen, S. (2001). Austalian Women in IT. Communications of the ACM. 44(7), 46-52.

Von Hellens, L., Pringle, R., Nielsen, S. and Greenhill, A. (2000). people, Business and IT skills: The Perspective of Women in the IT Industry . Proceedings of the Association for Computing Machinery Special Interest Group on Computer Personnel Research. 152-157.

Weaver, C.N. (1978). Sex Differences in the Determinants of Job Satisfaction. Academy of Management Journal. 21(2), 265-274.

Wilder, C. (1985). Gender and computers: Two surveys of computer related attitudes. Sex Roles. 13, 215-228.

Wilson, E. (1997). Computing, Computer Science and Computer Scientists: How They are Perceived. In Social Europe: Occupational Segregation of Women and Men in the European Community Supplement, edited by Lander and Adam.

Appendix: Data Treatment

We have adopted the following rules in attempt to make the self-selected dataset clean. By doing so, some legitimate observations, in addition to the obviously outliers, may have been removed. But it is our belief that it is better to stay on the safe side.

Items	Exclusion Rules	Rational
Age	Age 1 (18 under), 7 (60-64) and 8 (65 and over)	Lack of representation
Education	Education 1 and Education 10	Education 1 is default value and 10 is Other, which is unknown to us.
Job Title	35: Non-IT: Executive/ Corporate 36: Non-IT: Financial 37: Non-IT: Manufacturing / Operations	Our interest is limited to ITP.
Hours per Week	Category 1 (< 20 hours/week)	This is the default value and if not carefully enough respondents would have unintentionally selected it.
Country	Non-USA countries	Our Intention is to focus on U.S.A.
Age * Exp	(Age 18-24) AND (Experience of 11 years or more)	It is unlikely for young people to acquire this many years of experience
Exp * Yearly Salary	(Technical experience is less than 1 year) AND (Yearly salary is greater than $100K) (Technical experience is 1-2 years) AND (Yearly salary is greater than or equal to $125K)	Unlikely