DOI: 10.21276/sb.2017.3.3.7

Scholars Bulletin **ISSN 2412-9771 (Print)**
(A Multidisciplinary Journal) **ISSN 2412-897X (Online)**
An Official Publication of "Scholars Middle East Publishers",
Dubai, United Arab Emirates
Website: http://scholarsbulletin.com/

Gender Gap in the Heart of Wage Earning Differentials

David Iheke Okorie[1], Dak-Adzaklo Cephas Simon-Peter[2]
[1]Wang Yanan Institutes for Studies in Economics (WISE), Xiamen University, China.
[2]Institute for Financial and Accounting Studies (IFAS), Xiamen University, China

***Corresponding Author:**
David Iheke Okorie
Email: okorie.davidiheke@gmail.com

Abstract: The notion of gender inequality has been a long lasting phenomenon and have been empirically investigated in so many economies by researchers like Blau & Kahn, Jones & Makepeace, etc. in the past. This work sets out to investigate the existence of gender discrimination in the heart of wage earnings. The dataset used is the Current Population Survey of the United State, the analysis showed that the likelihood of earning higher wages depends on the gender of the worker as male workers are likely to end more than their fellow female counterparts are in the same employment category. However, this likelihood depends on the Educational Qualifications of the workers (both male and female). Therefore, this works has empirical shown that there exists gender discrimination in the heart of wage earnings which is not independent of College Degree.
Keywords: gender inequality, gender discrimination, likelihood.

INTRODUCTION

Engaging in a productive activity, which returns some benefit, is not an end but a means to an end. The end has always been to satisfy our wants, literally speaking, it has always possessed the ability (resources) which makes it possible or which gives us the power to provide those things we want. To this end, everybody wants to possess this ability not for its own seek but for what it will help to achieve. This is the aim behind people in every economy working to make some money. It's plausible to explain that at this stage of this work, we consider the category of people working for others and received wages and salaries at as when due and not for the category of people who work for themselves. Wage then is a very critical factor people consider in a bid to engage in a particular work. The importance of wage in one's life cannot be over emphasized but it's important to note that wage is been determined by other factors other than its self-importance. These factors may include the skill level of the worker, level of education of the worker, nature of work, work duration, location of the organization, labour unions, minimum wage laws, etc. therefore, to earn more is to critically consider these other factors that affect wage levels and adjust it in the direction it will result to a higher wage level.In different times, industries, and economies, it has been observed that Gender of the worker is also a very significant and important factor that determines the wage level of a worker, thus gender discrimination. The periodic of National partnership for Women and Families [3] report that there has been a persistent gender based wage gap, which can continuously harmed women, and their families in United States as women are paid 75cents out of every $1 paid to men. This phenomenon is not only observable in the United States but also in the rest of the World. Graddy & Pistaferri [4] reports similar scenario in the United Kingdom (UK) likewise other researchers like [1] and [2].It has always raised concern in the past and presently in most economies of the world and hence economies seek to bridge these gaps using policies and laws as they deem fit. Moreover, this work seeks to establish the existence of these gender gaps using a more recent survey using some parametric and non-parametric methods. This will confirm if the policies and laws meant to close this gap have been effective over time.

RESEARCH QUESTIONS AND TESTABLE HYPOTHESIS

There are cases in households in which females are the main source of income for the families, this could be because of the death of their husbands or unemployment on the part of their husbands or other reasons or factors that will make the females to provide for the family needs. In such situations, it is expected that the woman provide to the extent she could. However, when the woman is consciously paid below the levels of her male counterparts, she can no longer adequately provide for the family as she supposed to and this is not because of her performing below expectations a work but because she is a female. Fortunately, such cases have been discovered and policies and efforts have been made to close the gaps. One may ask, has these gaps been closed given that such wage gender differences are still observed in most

organization [3]. Establishing the existence of these wage differentials with respect to gender is one of the purposes of this work. Secondly, this work will examine factors, which are likely to cause these wage differentials with respect to gender, factors like the Degree of workers. Trying to examine if the differences in earning of male and female depends on the level of degree each possess or is the degree independent of the gender-wage differentials. To do this, the following are the testable hypothesis of this work –

H-01: There are no gender-wage differentials
H-02: gender-wage differentials is independent of college Degrees

To find solutions or answers to this hypothesis, we will adopt the use of appropriate parametric and non-parametric methodologies, which its efficiency depends on the nature of the data set and the underlining distribution.

DATA DESCRIPTIVE STATISTICS

The data used in this work is the Current Population Survey (CPS), conducted by the U.S. Department of Labour, which provides data on labour force characteristics of the population like level of employment, unemployment, earnings, etc. in this survey, approximately sixty-five thousand households randomly selected from an address database are surveyed each month. The variables employed in this work are Average Hourly Earnings (AHE), Age, Bachelor Degree Holder, and Gender. Bachelor is a binary variable which takes one, if the individual has bachelor degree and 0 otherwise, Gender is a binary variable which is 1 if the individual is female and 0 if male. Below is the descriptive statistics of the quantitative variables.

Table-1: Data Descriptive Statistics

Variable	AHE	Age
Minimum	**$2.10**	**25years**
Median	**$14.90**	**30years**
Mean	**$16,83**	**30years**
Maximum	**$61.06**	**34yeasr**

Source: U.S. CPS 2004 survey

From the statistics in Table1, it's like the variable, average hourly earnings is skewed to the right. In the presence of outliers, the median is a better central tendency statistic to report over the mean, which can change in the presence of outliers. It is also is also shown in the table that the age of the individuals that filled the population survey are the young people of the entire population. Therefore, it's assured that these group of persons are healthy, strong, and able to take on any type of job even if it requires manual labour or strength. The rest of the other variables are dummies and requires no descriptive statistics.

INDEPENDENCE TESTS

Just as stated, this work at this stage will use the non-parametric forms of analysis to show the independence and/or nature of dependence among these variables. We will first, try to evaluate the independence of gender and being a *bachelor* degree holder. To do this, we will employ condtional probability distributions, difference in proportion tests, relative risk analysis, odds, and odd ratio tests. Next, we evaluate the independence of average hourly wage earnings and gender using the chi-squared statistics and log ration statistic. We further will test that Age and gender are independent variables using the ordinal independence statistic. Finally, we construct a three-way contingency table to analyze the conditional odd rations (partial and marginal) to validate the existence of homogenous association among the variables; average hourly wage earning, gender and bachelor degree holders. The contingency tables for the above analysis with sample size of 7607 are shown below:

Table-2: Cross classification of Bachelor Degree by Gender

	BACHELOR'S DEGREE	
GENDER	**YES**	**NO**
MALE	1811	2615
FEMALE	1688	1493

Source: U.S. CPS survey.

Table-3: Cross classification of Average Hourly Earnings by Gender

	AVERAGE HOURLY EARNINGS	
GENDER	**HIGH**	**LOW**
MALE	2418	2008
FEMALE	1420	1761

Source: U.S. CPS survey.

Table-4: Cross classification of Gender by Workers Age

	GENDER	
WORKERS AGE	**MALE**	**FEMALE**
$\mathbf{25 \leq Age \leq 27}$	1160	927
$\mathbf{28 \leq Age \leq 30}$	1350	893
$\mathbf{31 \leq Age \leq 34}$	1916	1361

Source: U.S. CPS survey.

Table-5: Cross classification of Average Hourly Earnings by Gender and Bachelor Degree

		AVERAGE HOURLY WAGE	
BACHELOR	**GENDER**	**HIGH**	**LOW**
YES	**MALE**	1331	480
	FEMALE	1074	614
NO	**MALE**	1087	1528
	FEMALE	346	1147

Source: U.S. CPS survey.

Table-6: Partial table of Bachelor and Gender, controlling for AHE = HIGH

	BACHELOR	
GENDER	**YES**	**NO**
MALE	1331	1087
FEMALE	1074	346

Source: table five.

Table-7: Partial table of Bachelor and Gender, controlling for AHE = LOW

	BACHELOR	
GENDER	**YES**	**NO**
MALE	480	1528
FEMALE	614	1147

Source: table 5.

Table-8: Marginal table of Bachelor and Gender, controlling for AHE

	BACHELOR	
GENDER	**YES**	**NO**
MALE	1811	2615
FEMALE	1688	1493

Source: table 5.

Table-9: Partial table of AHE and Gender, controlling for Bachelor = Yes

	AHE	
GENDER	**HIGH**	**LOW**
MALE	1331	480
FEMALE	1074	614

Source: table 5.

Table-10: Partial table of AHE and Gender, controlling for Bachelor = No

	AHE	
GENDER	**HIGH**	**LOW**
MALE	1087	1528
FEMALE	346	1147

Source: table 5.

Table-11: Marginal table of AHE and Gender, controlling for Bachelor

	AHE	
GENDER	**HIGH**	**LOW**
MALE	2418	2008
FEMALE	1420	1761

Source: table 5.

Table-12: Partial table of AHE and Bachelor, controlling for Gender = Male

	AHE	
BACHELOR	**HIGH**	**LOW**
YES	1331	480
NO	1087	1528

Source: table 5.

Table-13: Partial table of AHE and Bachelor, controlling for Gender = Female

	AHE	
BACHELOR	**HIGH**	**LOW**
YES	1074	614
NO	346	1147

Source: table 5.

Table-14: Marginal table of AHE and Bachelor, controlling for Gender

	AHE	
BACHELOR	**HIGH**	**LOW**
YES	2405	1094
NO	1433	2675

Source: table 5.

These are the tables, which we will present their non-parametric analysis in the next chapter. You will notice that table 2 and table 8, just like table 3 and table 11 are the same because marginal table is basically ignoring the existence of the other variable(s). For the test of homogenous association, we will use tables 6 through table 14.

MODEL BUILDING

The aforementioned test analysis on contingency tables only tells how that two or more variables are related but doesn't tell us how. To establish how this variable is related, we will develop suitable and appropriate models to explain the nature of their relationships. To do this, we categorize the average hourly earnings into high and low. High will be any value of average hourly earnings above its median value and otherwise, low. Given this scenario, we assume our random component is Binomial distributed because, for the same values in in the explanatory variable, the logistic regression will treat it as a group which follows Binomial instead of Bernoulli distribution, we will then employ the models appropriate for Binomial distributed response variable models. These models include the Linear Probability Model (LPM), Logistic Models (Logit) and Probit Models. These models will be employed and the best-fit analysis will be done to ascertain which model explains the nature of the relationship between these variables better. Secondly, we will if we use the observed average hourly earnings without categorizing it, we will then assume that the random component is normally distributed and will employ the appropriate methodology for analysis. These models are specified as follows:

$$E(Y_i|x_k) = \beta_0 + \beta_k X_k \quad (1)$$

$$\pi(x_k) = \beta_0 + \beta_k X_k \quad (2)$$

$$log\left(\frac{\pi(x_k)}{1-\pi(x_k)}\right) = \beta_0 + \beta_k X_k \quad (3)$$

$$\varphi^{-1}(\pi(x_k)) = \beta_0 + \beta_k X_k \quad (4)$$

$$where: \beta_k = [\beta_1 \quad \dots \quad \beta_j], \quad X_k = [X_{1i} \quad \dots \quad X_{ji}]^T, \quad and$$
$$\varphi^{-1} = the\ inverse\ CDF\ of\ the\ underlying\ distribution$$

Given that the estimation of Generalized Least Squares (GLM) uses the Maximum Likelihood Estimations (MLE), we will transform these models into their Joint Likelihood Functional (JLF) form as:

$$l(\beta_0, \beta_k) = \prod_{i=1}^{N} \frac{1}{\delta\sqrt{(2\pi)}} e^{-\left(\frac{(Y_i-\beta_0-\beta_k X_k)^2}{2\delta^2}\right)} \quad (5)$$

$$l(\beta_0, \beta_k) = \prod_{i=i}^{N} \binom{n}{k} . (\beta_0 + \beta_k X_k)^k . (1 - \beta_0 - \beta_k X_k)^{n-k} \quad (6)$$

$$l(\beta_0, \beta_k) = \prod_{i=i}^{N} \binom{n}{k} \cdot \left(\frac{e^{\beta_0+\beta_k X_k}}{1+e^{\beta_0+\beta_k X_k}}\right)^k \cdot \left(\frac{1}{1+e^{\beta_0+\beta_k X_k}}\right)^{n-k} \quad (7)$$

$$l(\beta_0, \beta_k) = \prod_{i=i}^{N} \binom{n}{k} \cdot (\varphi[\beta_0 + \beta_k X_k])^k \cdot (\varphi[\beta_0 + \beta_k X_k])^{n-k} \quad (8)$$

$$given\ the\ total\ number\ of\ trials(n)\ and\ \left(Y_i, X_{1i}, \ldots X_{ji}\right)_{i=1}^{N}$$

To find the Maximum Likelihood Estimators of the parameters following the underlining probability distributions of the random component, Cumulative Density Function (CDF) and choice of models structure is to take the First Order Condition (FOC) of equations (5) through equations (8). It is important to mention that the argument in equation (5) is to minimize the objective function because we are minimizing the error, while we maximize the objective functions in the other equations. We will employ the use of analysis of Deviances to examine the model with a saturated model to know the best model in explaining the relationships between the variables. This helps us to compare simple models saturated models to know which is better while testing the hypothesis that the simple model is better against that a more saturated model is better. We will also use the information criterion and log likelihood values to select the best model out of all the models used to estimate the relationship in this work

RESULTS OF TEST OF INDEPENDENCE

Table-15: Independence Result presentation

Table	CPD	DPT	RR	ODDs	ORS
Table 2&8	Male: [0.41, 0.59]	P-value = 0.0000	0.77	Odd1 = 0.69	OR = 0.61
	Female: [0.53, 0.47]			Odd2 = 1.13	
Table 3&11	Male: [0.55, 0.45]	P-value = 0.0000	1.22	Odd1 = 1.2	OR = 1.49
	Female: [0.45, 0.55]			Odd2 = 0.8	
Table 4	$M^2 = 2.838$	P-value = 0.09	M = 1.685	P-value = 0.05	
Table 6	Male: [0.55, 0.45]	P-value = 0.0000	0.73	Odd1 = 1.22	OR = 0.39
	Female: [0.76, 0.24]			Odd2 = 3.10	
Table 7	Male: [0.24, 0.76]	P-value = 0.0000	0.69	Odd1 = 0.31	OR = 0.59
	Female: [0.35, 0.65]			Odd2 = 0.54	
Table 9	Male: [0.73, 0.27]	P-value = 0.0000	1.16	Odd1 = 2.77	OR = 1.59
	Female: [0.64, 0.36]			Odd2 = 1.75	
Table 10	Male: [0.42, 0.58]	P-value = 0.0000	1.79	Odd1 = 0.71	OR = 2.36
	Female: [0.23, 0.77]			Odd2 = 0.30	
Table12	Yes: [0.73, 0.27]	P-value = 0.0000	1.77	Odd1 = 2.77	OR = 3.90
	No: [0.42, 0.58]			Odd2 = 0.71	
Table13	Yes: [0.64, 0.36]	P-value = 0.0000	2.75	Odd1 = 1.75	OR = 5.80
	No: [0.23, 0.77]			Odd2 = 0.30	
Table14	Yes: [0.69, 0.31]	P-value = 0.0000	1.97	Odd1 = 2.20	OR = 4.10
	No: [0.35, 0.65]			Odd2 = 0.54	

Source: computations using R

The table above contains all the non-parametric analysis of the tables listed in chapter three of this work. In table2&8, the conditional probability distribution (column2) of possessing a Bachelor's Degree for Male and Female shows that Females have higher likelihood to have Bachelor's Degree than Male. However, the difference in proportion test (colunm3) shows that these two groups are not the same confirming that Female are more likely to possess Bachelor's Degree than Male. Relative Risk analysis reviews that these two groups of gender are not independent with respect to Bachelor's Degree. The odds for success are also higher for the Female group than Male but are dependent as confirmed by the odd ratio analysis. Considering tables3&11, the male group are more likely to receive higher hourly wage than their female counter parts. Male and Female groups are not independent in view of wage earnings. The Relative Risk result shows that the chances of earning higher wages are higher for the male group. The odds of success for higher wage are higher for the male group relative to the female groups. These two groups are also not independent of Average hourly earnings. Using the ordinal measures of independence for table 4, using mid rank techniques (Wilcoxon Test), the result shows that the male and female works are independent of Age, this suggests that age is not one of the factors that can explain differences in gender with respect to their

earnings.To test for homogenous associations, Table 5 was dismantled into three marginal tables and six partial tables. For these tables, we will focus only on the odd ratio analysis. From the result, there exist no homogenous associations and all these variables are dependent. The odds of getting a bachelor's degree are low for the male category but high for the female category. However, the odds of success for earning higher wages are high for the male category. It is important to note also that the odds of higher earnings are higher with bachelor's degree. All these analysis has shown us that these variables are dependent on each other except for age and average hourly earnings. However, how are they dependent? This is a question the contingency table cannot be able to provide answers to. Therefore, to get the answers to these questions, we will provide a model, which will capture this dependency structure to be able to tell us how these variables are dependent. The result of the model estimations is shown in the next section.

RESULTS OF MODEL ESTIMATIONS

Table 16: Models Estimation

	Dependent variable:			
	Average Hourly Wage Earnings(High & Low)			
	LPM(norm)	LPM(binom)	logistic	probit
	(1)	(2)	(3)	(4)
Female(Yes&No)	-0.187***	-0.179***	-0.886***	-0.532***
	(0.015)	(0.014)	(0.074)	(0.044)
Bachelor(Yes&No)	0.315***	0.319***	1.362***	0.842***
	(0.014)	(0.014)	(0.067)	(0.040)
age	0.020***	0.020***	0.095***	0.058***
	(0.002)	(0.002)	(0.009)	(0.005)
I(female * bachelor)	0.096***	0.085***	0.458***	0.270***
	(0.022)	(0.021)	(0.105)	(0.063)
Constant	-0.188***	-0.164***	-3.177***	-1.927***
	(0.055)	(0.054)	(0.262)	(0.158)
Observations	7,607	7,607	7,607	7,607
Log Likelihood	-4,907.699	-4,678.612	-4,675.960	-4,676.285
Akaike Inf. Crit.	9,825.397	9,367.225	9,361.921	9,362.570

Note: *p<0.1; **p<0.05; ***p<0.01

Source: Computations using R.

This result shows the estimations of the models specified in chapter two section 2.3 called Model Building. All the estimates in each of the models are statistically significant at 1% levels. This implies that they are vital in explaining the variations in probability of success in a higher Average Hourly Earnings in the United States. The first column shows the LPM (Linear Probability Model) result when we assume normality of the random component, holding other variable constant, the likelihood of a female making a higher earning without a Bachelor's certificate is about 19% less than that of the male category. However, the chances of female earning a higher wage increase by about 9.6% if the female has Bachelor's certificate. Once a worker possesses a Bachelor's certificate, he or she will have about 32% chances of earning a higher wage. It is only about 2% the chances of earning a higher wage increase as a worker gets older. When Binomial distribution of the random component is assumed for the response variable, the Maximum Likelihood estimates are approximately the same with the Normal Distribution assumption. The nonlinear estimation of the relationship between gender and average hourly earnings is also statistically significant but their estimates are some worth different. Judging from the Akaike Information Criterion, the Logistic result is relatively better than that of the Probit estimation. The conditional partial log odds between higher average hourly earnings and bachelor is 1.362. The conditional partial odds between average hourly earnings and female is -0.886, this suggests that if you are a female, the conditional odds of success decreases. The odds of success for female will only increase by 0.458 if the female possess a bachelor's degree. Whether the female worker possesses a bachelor's degree or not, the male group always earns higher than their female counter parts. Using the prediction equation of Logistic models, the estimated probabilities for the possible outcomes is shown below.

Table-17: Maximum Likelihood

Conditional Likelihoods	Female	Male
$\hat{\pi} = P(Ahe = 1 \mid Fem = 1, Bac = 0)$	**0.23**	
$\hat{\pi} = P(Ahe = 1 \mid Fem = 0, Bac = 0)$		**0.42**
$\hat{\pi} = P(Ahe = 1 \mid Fem = 1, Bac = 1)$	**0.64**	
$\hat{\pi} = P(Ahe = 1 \mid Fem = 0, Bac = 1)$		**0.73**
source: from the logistic prediction equation		

Setting Age at its mean in all probability calculations, using the logit equation, the probability for higher earnings for the female and male groups on the condition that both groups do not possess a Bachelor's degree are respectively 0.23 and 0.42, however, if we condition on the two groups having a Bachelor's degree, the likelihood of higher earnings are 0.64 and 0.73 for the female and male category respectively. This reveals that given the same conditions, the male group always earn higher than their female counter part. Therefore, the results show evidence against the earlier stated hypotheses of this work, then we reject the null hypotheses and conclude that there exist a significant gender-wage differential and bachelor's degree is not independent of the gender-wage differentials as having a bachelor's degree increases the chances of higher wage earnings for both groups.

MODEL CHECKING RESULTS

Adopting the Deviance approach to model checking, the models used in this work are all adequate, as the hull hypothesis of model adequacy relative to its saturated model was not rejected at 5% size of test, so we conclude that the models are adequate.

More so, the logistic model estimates the best relationship using the Akaike Information Criterion (AIC) while judging with the Log-likelihood, the linear probability model with normal random component is best. For the purpose of systematic problems of Linear Probability Model towards extreme values, we use the logistic regression results for analysis in this study.

SUMMARY AND CONCLUSION

This work is motivated to identify the existence of gender discrimination with respect to wage earnings. Most researchers have in the past confirmed that there exists gender discrimination in wage earnings. Validating that gender discrimination do not exist and wage differential is independent of gender, which are the two hypotheses if this project, we adopted parametric and non-parametric methodologies using the dataset obtained from the United States. From our analysis we rejected the two testable hypotheses and conclude that gender discrimination still exists in wage earnings and this wage differential is not independent of a bachelor's degree

From all our analysis using both parametric and non-parametric methods, even though it shows that the female category is more likely to acquire the Bachelor's Degree, the male group still earns much higher than their female counter parts. When workers are grouped into those with (out) Bachelor's degree, the male counterpart in each of these two groups earns higher than their female counter parts, this fact was confirmed both using the contingency tables and model estimations. This shows a strong evidence of gender gap, in the heart of wage earnings differentials. This however is a gender phenomenon where one is intentional paid lower than her counter parts because she is a female. More so, this implies that females are more likely to be employed by firms which seek to minimize cost because they provide the same services as their male counterpart and yet are paid lower wages.

REFERENCES

1. Blau, F. D., & Kahn, L. M. (1996). Wage structure and gender earnings differentials: an international comparison. *Economica*, S29-S62.
2. Jones, D. R., & Makepeace, G. H. (1996). Equal worth, equal opportunities: pay and promotion in an internal labour market. *The Economic Journal*, 401-409.
3. Sheet, Fact. (December 2015). *Latinas and the Wae Gap*. Washington DC: National Partinarship for women and families.
4. Graddy, K., & Pistaferri, L. (2000). Wage differences by gender: evidence from recently graduated MBAs. *Oxford Bulletin of Economics and Statistics*, *62*(s1), 837-854.