Economy
ISSN: 2313-8181
Vol. 1, No. 1, 15-19, 2014
www.asianonlinejournals.com/index.php/Economy

Median Regression Analysis of Gender-wise Income Gap in Punjab, Pakistan

Muhammad Aslam[1] --- Arslan Saeed[2] --- Saima Altaf[3]

[1,3]Department of Statistics, Bahauddin Zakariya University, Multan, Pakistan
[2]Govt. Science College, Multan, Pakistan

Abstract

This paper primarily examines the impact of gender on the monthly income of the working class in Punjab, Pakistan. The relevant data have been obtained from Pakistan Labour Force Survey (2008-9). A special case of quantile regression i.e. the median regression is used for the desired investigation. In addition to gender, the other covariates are marital status, area of residence, level of education, job type and status etc. As in many other regions and countries, the male workers in Punjab tend to have higher average income and the income tend to increase with increase in level of education. The workers with permanent jobs earn more as compared to temporary job holders.

Keywords: Income gap, Labour force survey, Median income, Poverty, Quantile, Quantile regression.

1. Introduction

Income is the central motivating factor of human activities. According to the classical and modern economists, materialistic life is the game of income earning and spending. According to Robbins (1945) "economics is the study of human behavior adopted between unlimited end and scare means". Every human tries to satisfy his end with his scare resources. The main head of resource is income that is the reward of man's mental and physical effort. Income and income related issues are very important for study to a research scholar. The most relevant income related issues are unemployment, inflation, income inequality and poverty etc.

Income is one of the vast fields of study. So, how much research is made about income or its related issues (development, poverty, inflation, and living standard etc.) is insufficient due to dynamic changes in the world. However, the study of income gap is one of the important areas for the researchers.

Various researchers have analyzed the income gap in different forms. For instance, Falaris (2003) has analyzed the wage differences between males and females at various quantiles of income in Panama. Machado and Mata (2005) have analyzed the wage gap or wage changes over various periods of time in their study.

A number of studies have also been conducted about the income gap in Pakistan. For example, Hyder and Reilly (2005) have analyzed the public and private sector wage gap in Pakistan. Sabir and Aftab (2007) have also given the study about dynamism in the gender wage gap. Some other important studies are given by Ali *et al.* (1999), Ahmad (2000), Shahbaz *et al.* (2007) and Cheema and Sial (2012) etc.

The present article is also about the income related issues focusing the province of Punjab in Pakistan. We primarily aim to discuss the gender-wise gap in income. However, the impact of area (urban or rural), education, and job type on income of the people of Punjab are also being targeted in this study.

It is common to use the ordinary least squares (OLS) method to estimate the values of dependent variable, depending on different covariates. But sometimes the conditional mean of response variable is not desirable when one wants to obtain a good estimate that satisfies the location and shape properties better than the ordinary mean as found in the OLS. Koenker and Bassett (1978) argued that the OLS estimators may be seriously deficient in linear models with non-Gaussian errors. They introduced the quantile regression for this situation.

The traditional regression analysis is focused on mean. The conditional mean models have certain attractive properties under ideal conditions (assumptions). The conditional mean models have some deficiencies i.e. (a) these models are not extendable for non central location (b) the assumptions of conditional mean modeling are not always met in the real world (c) the conditional mean models do not go beyond location. They do not cover all

characteristics (like scale, skewness and other higher order properties) of relationship between response distribution and explanatory variables.

When we study the distribution of income, it is found to be highly skewed and invites the quantile regression to be used. When the distribution is highly skewed then mean is not good representation of location as median. The conditional median regression model is used in modeling the location and shape of distribution of response variable which is specific type of quantile regression. Therefore, in the present study, we aim to use the median regression model. The upcoming section elaborates the structure of such model. We are not the first who are using this approach. Many researchers used the same in their studies about income or wages e.g., see Buchinsky (1997), Tasi and Kuan (2003), Falaris (2003), Machado and Mata (2005) among many others. Similarly, a number of studies about the income modeling in Pakistan can also be found in Hyder and Reilly (2005) and Sabir and Aftab (2007) etc.. However, the present work addresses the income gap focusing of the largest province of Pakistan i.e. Punjab.

2. Material and Methods

Our analysis based on the secondary data taken from "Pakistan Labor Force Survey" (PLFS) 2008-09 conducted by the Federal Bureau of Statistics, Pakistan. Previously, Hyder and Reilly (2005), among many others, have also used the PLFS Data in their research. Such data give detailed and comprehensive characteristics of employed persons of age greater than 10 years.

In our study, the data of earning persons in Punjab consist of 7070 individuals whose monthly incomes are given.

2.1. Variables and Description

For our study, we used the variable as given in Table 1.

Table-1. Variables and Coding

Variable		Notation	Definition
Income		Y	Natural logarithm of the monthly income of a person.
Gender		G	= 1 if a person is male; = 0 if female.
Age		X_1	Age of a person in years.
Area		X_2	= 1 if a person belongs to urban area; = 0, otherwise.
Marital status		X_3	= 1 if a person is married; = 0, otherwise.
Primary	Education level and Training	E_1	= 1 if a person has primary level education but below middle; = 0, otherwise.
Middle		E_2	= 1 if a person has middle level education but below matriculation; = 0, otherwise.
Matric		E_3	= 1 if a person has matric level education but below intermediate; = 0, otherwise.
Intermediate		E_4	= 1 if a person has intermediate level education but below graduation; = 0, otherwise.
Graduate		E_5	= 1 if a person has ordinary graduate level education but below masters; = 0, otherwise.
Profdeg		E_6	= 1 if a person has a professional degree in engineering, medicine, computer and agriculture etc.; = 0, otherwise.
Postgraduate		E_7	= 1 if a person is postgraduate or has higher degree; = 0, otherwise.
Training		E_8	= 1 if a person has received any on/off job training; = 0, otherwise.
Job status	Job nature and Status	J_1	= 1 if a person has a permanent job; = 0, otherwise.
LSOM		J_2	= 1 if a person is legislator, senior official or manager etc.; = 0, otherwise.
Professionals		J_3	= 1 if a person is professional doctor, engineer etc.; = 0, otherwise.
TASP		J_4	= 1 if a person is technician or associate professional etc.; = 0, otherwise.
Clerks		J_5	= 1 if a person is clerk; = 0, otherwise.
SWSMSW		J_6	= 1 if a person is service worker, sale's person etc.; = 0, otherwise.
SAFW		J_7	= 1 if a person is skilled agricultural or fishery worker etc.; = 0, otherwise.
CTW		J_8	= 1 if a person is craft or related trade worker; = 0 otherwise.
PMOA		J_9	= 1 if a person is plant or machine operator or assembler etc.; = 0, otherwise.
EO		J_{10}	= 1 if a person belongs to elementary occupation; = 0, otherwise.

Although Hyder and Reilly (2005) also used the majority of such variables in their studies but we added few variables, displaying the job type and status.

2.2. Median Regression

For the empirical analysis of our study, the relationship between response variable and covariates is established. In statistics, we know that the basic descriptive aspects of any data are location (average) and shape (dispersion). In analysis our concern lies in the both aspects of the distribution of response variable with the connection of covariates effects i.e. how the covariates affect the location and shape of response variable. Thus, we chose to use the quantile regression model. The concept of quantile regression has given by Koenker and Bassett (1978). They have explained the significance of the quantile regression approach when the distribution of response variable is non-Gaussian. In

statistics, for non-Gaussian distribution the suitable descriptive statistic for location is median rather than mean. So the median regression which is a special case of a quantile regression, gives the estimates of median of response variable distribution with the connection of covariates effects. One more advantage of such regression approach is that it is robust technique in handling the extreme values and outliers.

Consider a real valued random variable *Y* characterized by the following distribution function,

$$F(y) = \text{Prob}\,(Y \le y),$$

the τ-th quantile of *Y* is defined as the inverse function

$$Q(\tau) = \inf\,\{y: f(y) \ge \tau\},$$

where $0 < \tau < 1$. In particular, the median is $Q(1/2)$.

The τ-th sample quantile $\hat{\xi}(\tau)$, which is an analogue of $Q(\tau)$, may be formulated as the solution of the optimization problem,

$$\min_{\xi \in R} \sum_{i=1}^{n} \rho_\tau (y_i - \xi),$$

where $\rho_\tau(z) = z(\tau - I(z < 0)), 0 < \tau < 1,$ is usually called the check function.

When covariates *X* are considered, the linear conditional quantile function, $Q(\tau \mid X = x) = x'\beta(\tau)$, can be estimated by solving,

$$\hat{\beta}(\tau) = \arg\min \sum_{i=1}^{n} \rho_\tau (y_i - x'_i \beta)\,, \qquad (1)$$

for any $\tau \in (0,1)$. The quantity $\hat{\beta}(\tau)$ is called the regression quantile. The case $\tau = 0.5$, which minimizes the sum of absolute residual, is usually known as median regression. For more details and use of the median regression, see Koenker and Hallock (2001), Buhai (2004) , Martins and Pereira (2004), Chen and Wei (2005) and Aslam *et al.* (2010) .

Our model of interest is

$$Q_\tau(y_i) = \alpha_{\tau 0} + \alpha_\tau G_i + \sum_{j=1}^{3} \beta_{\tau j} X_{ji} + \sum_{k=1}^{8} \gamma_{\tau k} E_{ki} + \sum_{m=1}^{10} \delta_{\tau m} J_{mi} + \varepsilon_{\tau i}, \qquad (2)$$

where $Q_\tau(y_i)$ is the desired τ-th quantile of the log-income of the *i*th individual, $\varepsilon_{\tau i}$ is the random error with zero mean and constant variance and rest are the respective coefficients and covariates defined in Table 1. However, for $\tau = 0.5$, we have the model of interest, the median regression model.

3. Results and Discussion

There are 7,070 individuals of the Punjab, whose monthly incomes are given in the data (PLFS, 2008- 09). Out of 7,070 persons 4,848 (68.57%) belong to the urban and 2,222(31.43%) belong to the rural areas of Punjab. There are 85.40% males and 14.60% females. It seems that the working class of men is greater than five times of that of women in Punjab. Mean age of the respondents is 33.29 ± 12.43(standard deviation: SD). In the data, there are 76.40% literate and 23.6% illiterate persons and this shows that literacy rate in employed community is high. Only 32.57% of the individuals have permanent job.

The average monthly income is reported to be Rs. 8,293.72 ± 8,405.26 (95% C.I: Rs. 8,097.76, 8,489.67). The median of monthly income is found to be Rs. 6,000. It means than 50% of the working class in Punjab earns just Rs. 6,000 or below per month.

It is noted that the distribution of income is positively skewed the mean is far from median and close to the upper quartile. It is also evident from Fig. 1.

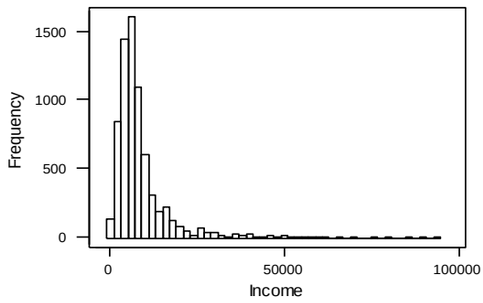

Fig-1. Histogram of Monthly Income

From Table 2, it is reported that the average income of males and females in Punjab are Rs. 8,554 and Rs. 6,770, respectively (using PLFS 2008-9). Thus, the average income gap is Rs. 1,784 which is also statistically significant.

Table-2. Comparison of Average Income of Males and Female

Gender	N	Mean	Std. Deviation	Std. Err.	*t*	*p*-value
Male	6038	8554.09	8413.23	108.27	6.32	0.00
Female	1032	6770.31	8198.33	255.20		

Table 3 presents the median regression estimates which are the chief targets of the present study. All the regression coefficients are found to be statistically significant at 1% or 5% level of significance. Since the logarithmic income is used in Model (2), the coefficients in Table 3 are directly interpretable. They merely tell the

percent change in the median monthly income of an individual in Punjab. It is reported if a person in Punjab is male he can earn 46.77% more income as compared to female for just being male. In other words, if the 50% of the females in Punjab earn monthly income Rs. 10,000 then 50% of their male counterparts will earn Rs. 14,677. However, when we compare our result with that given in Hyder and Reilly (2005), this figure is 21.28% for the entire country (using LFS 2001-02). Thus, the income gap between males and females almost double in Punjab when compared with entire Pakistan. Being married can increase 9.53% of income and it is evident due to change in the responsibilities after getting married. Moreover, if we compare the urban and rural residents of Punjab, the gap in income is found to be 9.87%. The urban workers tend to earn more. However, it should be noted that in PLFS (2008-9), the area means the area of residence not the work area so we cannot assess the true income gap between the rural and urban workers. It may possible that a rural resident works in some urban area and vice versa.

When we focus on the education level of the workers in Punjab, we note expectedly that with the increase in the education, there is increase in the income. Professional degree holders earn at highest rate as compared to the others. According to Hyder and Reilly (2005) , the change in the median income is 6.83% if a person has primary level of education in Pakistan. This figure is almost double i.e. 11.35% in Punjab. It shows that there are fair available sources of income for low educated persons in Punjab as when compared with entire Pakistan.

If we focus on the job status and nature in Punjab, we report that after having a permanent job, the median monthly income can be increased as much as 41.17%. The median income of is legislators, senior officials or managers is at the highest level in Punjab.

Now we elaborate the results of the estimated median regression model with the help of hypothetical information of a working individual of Punjab. Suppose, if we consider a graduate married male (E_5 =1, G =1, X_3 =1) worker of age 30 (X_1 = 30) who lives in an urban area (X_2 = 1) and is a clerk (J_5 = 1) on permanent (J_1 = 1) basis. The median monthly income of the persons in Punjab having such characteristics can be computed to be Rs. 11,762. In other words, 50% of the male workers with the above stated characteristics will have monthly income more than Rs. 11,762 and 50% less than this amount. If such person is a female then the median income is Rs. 7,368. Thus, there is a gap of Rs. 4,394 between the males and females of the same cadre and status as stated above.

4. Conclusion

A data set of 7,070 individuals of the Punjab, whose monthly incomes are given in PLFS (2008-09) is considered. In addition to income, several other variables are included in the study whose impacts are being measured on the monthly income. These covariates are gender, age, marital status, education level and job status etc. It is found that the female labour force participation is very low (14.60%) in Punjab. More than 75% of the working class is literate 68.57% belongs to the urban areas. Majority of the employers do not have permanent jobs. The average monthly income is reported to be Rs. 8,293.72. Moreover, 50% of the employees in Punjab earn less than Rs. 6,000 per month according to the information available in PLFS (2008-9). Similarly, from the distribution of income, it is depicted that a huge working class has low income. The average monthly income of males and females are reported to be Rs. 8,554 and Rs. 6,770, respectively (using PLFS 2008-9). The males tend to earn Rs. 1,784 more income, on average, as compared to their female counterparts in Punjab. The median regression analysis has the following main findings:

a) A male can increase 46.77% in the median monthly income as compared to working female in Punjab; b) The factor of marriage increases 9.53% in the monthly income; c) The people with urban origin can earn 9.87% more; d) Permanent job is prime factor in the increase of income.

Table-3. The Median Regression Estimates

Variable	**Coef.**	**Std. Err.**	***T***	***p*-value**	**[95% Conf. Interval]**	
Constant	7.1111	0.03821	186.10	0.0000	7.0362	7.1860
Age	0.0097	0.0007	13.89	0.0000	0.0083	0.0111
Gender	0.4677	0.0207	22.60	0.0000	0.4272	0.5083
Marital Status	0.0953	0.0185	5.15	0.0000	0.0590	0.1315
Area	0.0987	0.0146	6.76	0.0000	0.0701	0.1274
Primary	0.1135	0.0488	2.33	0.0296	0.0179	0.2092
Middle	0.1758	0.0525	3.35	0.0029	0.0729	0.2787
Matriculation	0.2768	0.0521	5.31	0.0000	0.1747	0.3789
Intermediate	0.4510	0.0553	8.15	0.0000	0.3426	0.5593
Prof. Degree	1.0322	0.0682	15.13	0.0000	0.8985	1.1659
Graduation	0.5280	0.0316	16.71	0.0000	0.4661	0.5899
Post Graduation	0.8989	0.0364	24.69	0.0000	0.8275	0.9702
Training	0.0634	0.0206	3.08	0.0055	0.0230	0.1037
Job Status	0.4117	0.0165	24.95	0.0000	0.3794	0.4440
LSOM	0.7308	0.0369	19.80	0.0000	0.6584	0.8031
Professional	0.4715	0.0376	12.54	0.0000	0.3978	0.5452
TASP	0.2965	0.0322	9.21	0.0000	0.2334	0.3596
Clerk	0.3685	0.0372	9.91	0.0000	0.2956	0.4414
SWSMSW	0.3087	0.0287	10.75	0.0000	0.2524	0.3649
SAFW	0.2601	0.0541	4.81	0.0001	0.1540	0.3661
CTW	0.3977	0.0238	16.71	0.0000	0.3510	0.4443
PMOA	0.4673	0.0305	15.32	0.0000	0.4075	0.5270
EO	0.2705	0.0238	11.36	0.0000	0.2238	0.3171

References

Ahmad, M., 2000. Estimation of distribution of income in Pakistan, using micro data. The Pakistan Development Review, 39(4): 807-824.

Ali, S.S., S. Tahir and G.M. Arif, 1999. Dynamics of growth, poverty, and inequality in Pakistan. The Pakistan Development Review, 38(4): 837-858.

Aslam, M., A. Saeed, G.R. Pasha and S. Altaf, 2010. Median regression analysis of body mass index of adults in Pakistan. Pakistan Journal of Nutrition, 9(6): 611-615.

Buchinsky, M., 1997. Changes in the U.S. wage structure 1963-1987: Application of quantile regression. Econometrica, 62(2): 405-458.

Buhai, S., 2004. Quantile regression: Overview and selected applications. Rotterdam: Tinbergen Institute and Erasmus University.

Cheema, A.R. and M.H. Sial, 2012. Poverty, income inequality, and growth in Pakistan: A pooled regression analysis. The Lahore Journal of Economics, 17(2): 137–157.

Chen, C. and Y. Wei, 2005. Computational issues on quantile regression. Sankhya, 67(2): 399-417.

Falaris, E.M., 2003. A quantile regression analysis of wages in Panama. Department of Economics, University of Delaware. (Mimeographed).

Hyder, A. and B. Reilly, 2005. The public and private sector pay gap in Pakistan: A quantile regression analysis. The Pakistan Development Review, 44(3): 271–306.

Koenker, R. and G. Bassett, 1978. Regression quantiles. Econometrica, 46(1): 33-50.

Koenker, R. and K.F. Hallock, 2001. Quantile regression. Journal of Economic Perspectives, 15(4): 43-56.

Machado, J.A.F. and M. Mata, 2005. Counterfactual decomposition of changes in wage distribution using quantile regression. Journal of Applied Econometrics, 20(4): 445-465.

Martins, P.S. and P.T. Pereira, 2004. Does education reduce wage inequality? Quantile regression evidence from 16 countries. Labour Economics, 11(3): 355-371.

Robbins, L., 1945. An essay on the nature and significance of economic science. London: Macmillan and Co. Ltd.

Sabir, M. and Z. Aftab, 2007. Dynamism in the gender wage gap: Evidence from Pakistan. The Pakistan Development Review, 46(4): 865-882.

Shahbaz, M., N. Aamir and M.S. Butt, 2007. Rural-Urban income inequality under financial development and trade openness in Pakistan: The econometric evidence. The Pakistan Development Review, 46(4): 657- 672.

Tasi, S.L. and C.M. Kuan, 2003. Assortative mating and income inequality in Taiwan: A quantile regression approach. Conference of Research Committee on Social Stratification and Mobility of the International Sociological Association. University of Tokyo, Japan.