
Walmart - Human Resource

Introduction: -

The data analysis based on the effect of salary on employment longevity, used the dataset collected for different groups of employees of Walmart at one point in time.

Hypothesis: Increase in salary = Increase in longevity,

Alternative Hypothesis: Increase in salary \neq Increase in longevity

From the analysis it shows that effect of salary does affect the employment longevity in a positive way, for every unit increase in salary there is an increase in longevity of 0.156 unit

As there are other factors that affects the longevity and correlates with salary that tend to create a bias estimate due to endogeneity, did include variables to the base model to improve the estimates values. Outlined the fact that the included variables do not influence the salary in a major way but do have an impact in the longevity. Highlighted the possibility for omitted variable bias and listed the possible omitted variables-(1) Average employee satisfaction survey rate, (2) Total benefits package. Including with suggestion for instrumental variable-(IV) Merit based compensation policy- as it is not feasible to include all the controls to the model due to unavailability of data or as it is hard to measure.

As still there is a bias in the estimate due to the unavailability of control variable, proxy and instrumental variables- based on the omitted variable mentioned above the predicted value will be greater than the true value, as both the omitted variables are positively correlated with AvgSalary and positively impact the AvgLongevity.

Data: -

The dataset is collected for 400 groups of employees from different region of Walmart at one point in time,

Data Type: - Cross sectional data

Dependent Variable: - AvgLongevity

Explanatory Variable: -AvgSalary

Number of observations: - 400

For reference, the summary statistic of the data is listed in ([Appendix A](#))

Analysis: -

Building the base model based on the expected hypothesis: the salary to have a positive effect on employee longevity

As a process to determine the effect of salary on employee longevity, estimated the model ([Appendix B-Model\(1\)](#)) with a simple regression-AvgLongevity on AvgSalary, AvgSalary is statistically significant ($p < 0.5$) and got the coefficient of (0.149) for AvgSalary means for every unit of increase in AvgSalary then AvgLongevity increases by 0.149 unit.

To add evidence to the model used the quadratic functional form ([Appendix C](#)), The regression line in the scatter plot reflects the evidence, that as the AvgSalary increases the AvgLongevity increases along. It shows that salary have a positive effect in longevity.

But along with AvgSalary there are other factors in the dataset that influence the employee longevity that creates a bias estimate caused by endogeneity. As an attempt to improve the estimate expanded the model ([Appendix B-Model\(2\)](#)) to include individual related control variables like AvgAge, AvgEducation for the group, which changed the AvgSalary coefficient to (0.150). Both the control variable AvgAge and AvgEducation are statistically significant ($p < 0.5$) and negatively correlated with AvgSalary, which is reasonable as highly educated people are tending to face more challenges which leads to career development and the same way as age increase people get more exposure and more experience wanting them to look out for more opportunity to enhance their career, both reduces the stay within the company.

In addition, ran the regression to estimate the model adding the environmental and surrounding influence control variables like CompetingLocalFirms, LocalUnemployment, AvgHouseholdSize and MaleProportion that correlates with the treatment AvgSalary and that affects the outcome AvgLongevity to the expanded model along with AvgEducation and AvgAge ([Appendix B-Model\(3\)](#)). By adding this additional control variables changed the coefficient of AvgSalary to (0.156), where all the variables are statistically significant ($p < 0.5$) except a variable MaleProportion, MaleProportion and CompetingLocalFirms are negatively correlated with AvgSalary, and it implies that compared to female the male employees are more tend to switch jobs, as in average women tend to stay more in known environment as a practice to balance work and family, and more CompetingLocalFirms means more opportunity which becomes easier for employees to switch without concerning on shifting the whole house hold. The other two variables Local Unemployment and AvgHouseholdSize positively correlated with AvgSalary, which means increase in unit increase the longevity of employee in units, When there is an increase in the rate of unemployment people avert from taking risk, similarly with AvgHouseholdSize, in average people with growing or bigger families stay away from risk factors as switching job or challenging more than an extend for the benefit of their family which supports the factor to increase in longevity as the family grows.

Comparing the Beta1, the coefficient of AvgSalary from all the three model – (0.1488) from model(1), (0.150) from model(2) and (0.156) from model(3) the included variables do not influence the salary in a major way but do have an impact in the longevity.

Still the endogeneity exists due to omitted variable bias, some of the omitted variables that support the existence of endogeneity are (1) Average employee satisfaction rate- as satisfied employee will contribute his work with an involvement leads to good performance and higher salary longer survival. In contrast if the employee is not satisfied with the employer or management or coworker despite other factors they will find a way to move out. Another omitted variable can be (2) Total benefits package- example Health insurance, based on the package when the salary goes up there will be added benefits which makes the employee to stay longer within the company.

Better way to get the precise estimate with this model is by introducing instrumental variable that correlates with salary but does not affect the employee longevity- Merit based compensation policy can be used as instrument for salary as salary for each year is determined by the performance assessment and the rate based on it Good performance leads to higher salary leading to better longevity. As mostly good performance is also an indication of job satisfaction, better work environment and adaptable surroundings.

Conclusion: -

From the above analysis using the dataset I want to conclude that when the AvgSalary increases by 1 unit, the AvgLongevity increases by (0.156) units. But still after adding the available controls from the dataset the endogeneity still exists to have a bias estimate. One suggested option is to overcome the bias is by gathering the data for instrumental variable- Merit based compensation policy and run the regression using two stage least square(2SLS).

As still there is a bias in the estimate due to unavailability of control variable, proxy and instrumental variables- based on the omitted variable mentioned in the analysis the predicted value will be greater than the true value.as both the omitted variables are positively correlating with AvgSalary and positively impact the AvgLongevity.

Appendix

Appendix A

The summary statistic of the dataset,

Variable	Obs	Mean	Std. Dev.	Min	Max
AvgLongevity	400	4.514111	2.368723	.5505339	10
AvgSalary	400	43.07492	6.524539	25.77125	58.61459
AvgHouseho~e	400	3.302942	.2941328	2.345307	4.180555
AvgEducati~s	400	11.93929	.9665447	9.750813	14.69448
AvgAge	400	35.42803	3.916715	23.24201	45.39655

Appendix B

Regression output

	(1)	(2)	(3)
	AvgLongevity	AvgLongevity	AvgLongevity
AvgSalary	0.149 (0.0166)	0.150 (0.0117)	0.156 (0.00726)
AvgEducati~s		-0.563 (0.0790)	-0.581 (0.0495)
AvgAge		-0.366 (0.0195)	-0.351 (0.0120)
CompetingL~s			-0.384 (0.0335)
Proportion~e			-0.670 (0.644)
AvgHouseho~e			0.472 (0.161)
LocalUnemp~t			1.048 (0.0475)
_cons	-1.897 (0.723)	17.75 (1.271)	10.42 (1.020)
N	400	400	400

Standard errors in parentheses

Appendix C

The graphical representation of the positive effect of salary on longevity based on hypothesis using the quadratic functional form

	(1)
	AvgLongevity
AvgSalary	-0.139 (0.177)
AvgSalary2	0.00337 (0.00206)
_cons	4.128 (3.748)
N	400

Standard errors in parentheses


