Using Negative Binomial Regression Model to minimize over dispersion: An application onwoman's labor force data

Reem Ismail Elsybaey*

CAPMAS, CAIRO, EGYPT – **reemismail_2008@yahoo.com**

Prof Dr: Amany Moussa**
Dr: Yasmin Ibrahim***

ISSR, CAIRO, EGYPT – **yasminority@yahoo.com**

Abstract

The participation of Egyptian women in the labor force is an important issue that occupied an important place recently. In spite of all efforts that were made in this framework, it still appears that there are many challenges that impact negatively on the ability of Egyptian women to actively participate in economic life. This paper used the female's labor force data of Upper Egypt in years 2011 and 2013. Also it used the plots of residuals versus the mean to determine if the varianceis too large ornot. The plots show that in both years 2011 and 2013 there is over-dispersion in the data and the variance is too large. In this case the Negative binomial regression can be used to minimize over-dispersed data that's when the variance is greater than the mean. TheNegative binomial regression model can be used as a generalization of Poisson regression because it has the same mean formation as Poisson regression and it has an extra parameter to detect the over-dispersion. The results show that using the Negative binomial regression helps in minimizing both over dispersion and the value of the dispersion parameter.

Keywords: Upper Egypt, Poisson regression, dispersion parameter.

1. Introduction

There are many conferences referred to the need to promote the social status of women and strengthen their role in the political, economic, and social development. The most important conference is the International Conference on Population and Development which concerned with Women's issues, held in Cairo in 1994. It called for women's equal right with men in all areas and the elimination of discrimination against them.

The participation of Egyptian women in the labor force is an important issue that occupied an important place recently. In spite of all efforts that were made in this framework, it still appears that there are many challenges that impact negatively on the ability of Egyptian women to actively participate in economic life. The most important of these constraints is the low participation of women in the labor force, high unemployment rate among femalescompared to males where the rate among women are more than three times higher than that among men.

*Reem Elsybaey: The Author is a Master degree Student in ISSR and Statistician at (CAPMAS) National Statistics Office in Egypt.

** Amany Moussa: Professor Dr in ISSR (Institute of Statistical Studies and Research) – Egypt, Author's Supervisor.

***Yasmin Ibrahim: Dr in ISSR (Institute of Statistical Studies and Research) – Egypt, Author's Supervisor.

there is over-dispersion in the

assumption in Poisson regression that at

number of cases has variance equal to the mean is

Table (1): Comparison between the results of Goodness of fit for Poisson Regression and Negative Binomial in years 2011 and 2013:

Goodness of Fit (2011)	Poisson Regression			Negative Binomial		
	Value	df	Value/df	Value	df	Value/df
Deviance	2971.306	142	20.925	168.784	141	1.197
Pearson Chi-Square	3076.850	142	21.668	125.296	141	.889
Log Likelihood	-1833.021			-644.754		
Akaike's Information Criterion (AIC)	3674.042			1299.508		
Bayesian Information Criterion (BIC)	3685.976			1314.426		
Goodness of Fit (2013)	**Poisson Regression**			**Negative Binomial**		
	Value	**df**	**Value/df**	**Value**	**df**	**Value/df**
Deviance	1777.360	143	12.429	160.670	142	1.131
Pearson Chi-Square	1790.216	143	12.519	134.594	142	.948
Log Likelihood	-1231.674			-599.834		
Akaike's Information Criterion (AIC)	2471.348			1209.669		
Bayesian Information Criterion (BIC)	2483.310			1224.621		

-We tested over dispersion in the Negative Binomial regression model using the ratio of the sum of Pearson chi square over the number of degrees of freedom (chi (2)/df) or the (deviance / df). The results according to Poisson Regression Model in 2011 and 2013 was (chi (2)/df) = (21.668, 12.519) respectively. However, the results of Negative Binomial Regression Model decreased to be (1.2, 1.1).

-According to Negative Binomial Regression Model the over-dispersion decreased more than the Poisson Regression Model.

Table (2): Comparison between the results of Negative Binomial regression Model in years 2011 and 2013:

Exp. Variables	2011				2013			
	B	Std. Error	Sig.	Exp(B)	B	Std. Error	Sig.	Exp(B)
Intercept	3.066	.3213	0.000	21.446	2.633	.2478	0.000	13.909
Ur_Ru = 0 or 1	-.524	.0750	0.000	.592	-.229	.0961	0.000	.796
Number of Educated women (Edu)	.007	.0033	.046	1.007	.003	.0027	.238	1.003
Number of Married women (Mar)	.011	.0024	0.000	1.011	.009	.0021	0.000	1.009
Negative Binomial	0.755	.0908	—	—	0.516	.0640	—	—

In 2011:

Log (Work) = 3.066 - 0.524 (Ur_Ru) + 0.007 (Edu) + 0.011 (Mar).

This implies that:

Work = Exp [3.066 - 0.524 (Ur_Ru) + 0.007 (Edu) + 0.011 (Mar)].

= Exp (3. 066) * Exp (- 0.524 Ur_Ru) * Exp (0.007 Edu)

* Exp (0.011 Mar).

In 2013:

Log (Work) = 2.633 - 0.229 (Ur_Ru) + 0.003 (Edu) + 0.009 (Mar).

This implies that:

Work = Exp [2.633 - 0.229 (Ur_Ru) + 0.003 (Edu) + 0.009 (Mar)].

= Exp (2.633) * Exp (-0.229 Ur_Ru) * Exp (0.003 Edu)

* Exp (0.009 Mar).

Parameter Estimates:

- From table (2), the intercept is the Negative Binomial Regression estimate when all variables in the model are evaluated at zero. The log of the expected count for the Number of working women is 3.066 units in 2011 and in 2013 it was 2.633 units.
- According to the results of Negative Binomial Regression Model the number of educated females has an insignificant effect on the number of working

women in Upper Egypt. The coefficient for the Number of educated women in years 2011 and 2013 was (0.007, 0.003) units respectively. This means that, for a one unit increase in the number of educated females the expected log count for the number of working females is expected to change by (0.007 , 0.003) units respectively.

- Table (2) also refers to the important significant effect for the marital status, where the coefficient for the Number of married females in 2011 and 2013 was (0.011, 0.009) units. This means that, for a one unit increase in the number of Married females the expected log count for the number of working females is expected to change by 0.011 in year 2011 and will change by 0.009 in 2013.
- The indicator variable Ur_Ru is the estimated Negative Binomial regression coefficient comparing Rural to Urban. The coefficient for this dummy variable residence in the two years 2011 and 2013 is negative and statistically significant indicating that rural regions have fewer working females than urban regions.
- From table (2) we can notice that female's place of residence or female's region (Ur_Ru) has a significant effect on the number of working females. Also in both years 2011 and 2013 the expected log count of working females who are in rural areas decreased by about 41% and 21% respectively compared to the expected log counts of working females who are in urban areas.
- Thus, we would expect the urban areas to have more number of working females than their rural counterparts.
- Also there is an estimate of the dispersion parameter, which is the value of Negative binomial in table (2). In Poisson Regression Model this value is constrained to zero. In this case, the parameter with 95% confidence interval does not include zero, furthermore the negative binomial model form is more adequate than the Poisson Regression Model.
- The dispersion parameter in years 2011 and 2013 was equal to (0.755, 0.516) respectively. The Negative Binomial Regression Model was able to handle and decrease over dispersion in the count data for both years 2011 and 2013. However, the over-dispersion still exists because the value of the dispersion parameter is greater than zero.
- Also the Generalized Poisson Regression Model (GPR) can be used to handle over-dispersion where the GPR is a generalization of the standard Poisson regression (PR) model.

Table (3): Choosing the best Model

Model	Significant Variables	AIC (2011)	AIC (2013)
Poisson Regression	**Ur_Ru Edu Mar**	**3674.042**	**2471.348**
Negative Binomial	**Ur_Ru Mar**	**1299.508**	**1209.669**

- For Akaike's Information Criterion (AIC) in table (3) the smaller the AIC, the better the model is.
- The smallest AIC value is a negative binomial regression model. Then the best model for the number of working females is obtained from the negative binomial regression model. So the Negative Binomial is more adequate and appropriate in the case of over-dispersion than Poisson Regression Model.

5. Conclusions

- According to the results of Negative Binomial Regression Model the number of educated females has an insignificant effect on the number of working women in Upper Egypt.
- The coefficient for the dummy variable residence in the two years 2011 and 2013 is negative and statistically significant indicating that rural regions have fewer working females than urban regions.
- We can notice that female's place of residence or female's region (Ur_Ru) has a significant effect on the number of working females. Also in both years 2011 and 2013 the expected log count of working females who are in rural areas decreased by about 41% and 21% respectively compared to the expected log counts of working females who are in urban areas.
- The smallest value of AIC was belonging to the negative binomial regression model. Then the best model for the number of working females is obtained from the negative binomial regression model. So the Negative Binomial is more adequate and appropriate in the case of over-dispersion than Poisson Regression Model.

6. References

1. Bouche, (2009), *Application of detecting and taking over-dispersion into account in Poisson regression model,* University of Poitiers.
2. Elhadad, (2009), *Low participation of women in the formal labor market and wage discrimination,* Social Research Center, AUC.
3. "*Labor Force survey*", (2011, 2013), Central Agency for Public Mobilization and Statistics.
4. Lawrence L.Kupper, Keith E.Muller et al., (1998), *Applied Regression Analysis and other Multivariable methods,* 3rd Edition, Thomson.
5. Melliana, Setyorini,Eko,Rosi and Purhadi, (2013),*The* Comparison of Generalized Poisson Regression and Negative Binomial Regression Methods in Overcoming Over dispersion, *IJSTR vol.2.*
6. Sakr and Shehata, (2009), *Woman Empowerment,* Social Research Center, AUC.